Salmon on the run: Practicing scale in the study of wild Alaska salmon

Sarah Inman

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

David Ribes, Chair

Charlotte Lee

Daniel Schindler

Program Authorized to Offer Degree

Department of Human Centered Design and Engineering

University of Washington

Abstract

Salmon on the Run: Practicing scale in the study of wild Alaska salmon

Chair of the Supervisory Committee:

Associate Professor David Ribes

University of Washington

Modern scientific research infrastructure has eclipsed the importance of scaling when understanding ecosystem change. Scale is the lens through which scientists parse complexity. Although scale is central to scientific practice, modern research infrastructure has replaced the importance of scale with a focus on scalability in data. This dissertation engages the topic of scale in an ethnographic study of the State of Alaska's Salmon and People (SASAP) project, a 3-year initiative designed to investigate how data science can aid natural sciences. Through the synthesis of three empirical studies, this thesis proposes a conceptual framework that brings the extant theorizations of scale into conversation with the theorizations of scale in ecology. This study explores scale in three different cases: 1- a data science application for ecological data synthesis; 2- a field program focused on collecting and storing data in the long-term; and 3- a participatory modeling initiative instrumenting the local. As such, this thesis articulates how research practitioners reconcile issues of scale in wild Alaska salmon research and offers general insights about how to define scale. Building on prior work in infrastructure studies, this dissertation also provides methodological contributions for the ethnographic study of contemporary data initiatives. In conclusion, this research offers insights into how scientists define and instrument scale, methodological

contributions for conducting studies of large-scale data initiatives, and a general language for working with scale.

Table of Contents

List of Figures

List of Tables

Acknowledgments

For those who read this manuscript, you should also know the names of the people who helped make this research possible but are not listed as author. I am eternally grateful to my advisor and mentor, David Ribes, who continuously inspired me in his precision in writing. Once he referred to writing as his craft, and I have learned much about that craft over the many years of writing with him. Primarily, I am grateful to David for his holistic approach to mentorship. He was always there for me when I needed guidance and also willing to point out when it needed to be my own journey.

I am grateful to others on my committee — Charlotte Lee, Daniel Schindler, and Nic Weber — for all the support I received during the research and writing of this dissertation. Charlotte would ask me to go back to my research question again and again or to back out and tell her the whole picture. These conversations were valuable for helping crystallize the big picture. Nic stepped in at the last moment and was generous with his time and energy, helping me clarify my writing. Some of my fondest memories in my PhD process occurred during the summers that I worked with Daniel and his students at Lake Nerka. During this time, I learned skills not typically part of the ethnographer's toolkit like how to stream walk, how to drive a boat, and how to chop a salmon head. I really appreciate all the people who I met throughout this research program including the staff, namely, Chris Boatright and Jackie Carter. My work was also supported tremendously by staff in the Human Centered Design & Engineering department, in particular, Elaine Shelley and Allen Lee, who consistently go beyond their job responsibilities.

My research was furthered by ongoing conversations with scholars many of whom are dear friends such as Kiley Sobel, Sarah Fox, Andrew Berry, Charlie Hahn, and Drew Paine – and of course, much of my work was inspired by scholars like Kristin Asdal, Karen Baker, Janet Vertesi, Yanni Loukissas, and Phoebe Sengers. In the final hours of writing, Karen Baker generously offered her time giving me critical feedback that I needed to make my work more relevant.

Much of research is about community, and I am grateful for the community of scholars I worked with through SASAP and during my summers in the Alaska Salmon Program. In particular, I would like thank Janessa Esquible for her tireless and committed energy to every piece of the work we did in the Kuskokwim. Her dedication to adhering to local interests and her work at Orutsaramiut Native Tribal Council helped make my work more attuned to realities on the ground. She also became a dear friend throughout our collaboration. There are many who have been a part of the data ecologies lab to whom I can attribute my early scholarly thinking around these topics, namely, Charlie Hahn, Shana Hirsch, Andrew Hoffman, and Steve Slota. Finally, throughout COVID, I worked closely with Julia Parrish and Ben Haywood on a project with COASST. Being a part of an academic community throughout a time of isolation was invaluable, and Julia has always been a source of inspiration and structure.

I also want to express sincere gratitude to Jeremy Mhire who first believed in me as a scholar and taught me how to follow curiosity. Yianna Vovides also supported my early years of

becoming a scholar during my time at Georgetown. Finally, I am grateful for the support from friends and family, especially: my parents-- Jim and Terry Inman--who have always believed in me and supported my far from linear path, the vashonistas - Shana Hirsch, Anissa Tanweer, and Judy Twedt and the "sister island" resident, Lauren Drakopolous, and Jane Hossman.

Last but not least – I thank Ross Todd for his support. It has been a long and twisted path, but I appreciate his patience and encouragement throughout the last 6 months of writing.

Data in the study of ecosystems

Introduction

Data are central to the production of knowledge. Since the scientific revolution, there has been an increasing professionalization and institutionalization of science. This has been met with critical debates about the role of data in 21st century science (e.g., Crawford, 2014; Kitchin, 2014; Kitchin & Lauriault, 2015). On one side of this debate, there is a call to make data open as a political imperative and as an ethical scientific practice. On the other side, there is a recalcitrance to this imperative to share (Lezaun & Montgomery, 2015). Rather than denigrating one side and praising another, this dissertation takes part in the debate by characterizing how scale plays an integral role in the production of data. This research explores different definitions of scale through three empirical cases specific to ecological research on wild Alaska salmon.

Understanding ecosystem change requires heterogeneous data that covers vast spatial and temporal scales. As problems related to climate change and ecosystem stability become increasingly more complex (e.g., rapid and emergent change, deeper understandings of temporal and spatial complexities, exponential rates of change, and time lags), scientific disciplines turn toward more data-centric, open science practices with an emphasis on producing synthetic, accessible research. Large-scale challenges brought on by climate

change, environmental degradation, and exponentially increasing populations have created a need to "increase the scale of inquiry, from the sample plot, habitat patch and small watershed of traditional ecology, to the landscape, geographic region, continent, ocean and entire earth" (Brown, 1994, p.21). Furthermore, catastrophic consequences of anthropogenic change, such as wildfires, stronger storm surges, and melting glaciers, has led to increasing political pressure to apply big data as a solution. This has led fields such as ecology to problematize their own practices around reproducibility, interoperability[1], and data accessibility (Jones, 2006; Michener & Jones, 2012; Reichman et al., 2011).

Preserving data in archives and databases is not new. However, the optimism for technologically-mediated services marks a moment in which computational power and technological development is rapidly increasing. Furthermore, this is happening across a range of disciplines from medicine to astronomy to ecology to physics. Emerging data practices have led to a new model of contemporary science (Aronova et al., 2010) and has inspired the creation of many organizations like the National Center for Ecological Analysis and Synthesis (NCEAS), a data infrastructure for acquiring, synthesizing, and archiving disparate data using a common language. This study begins with a study of NCEAS through a project called the State of Alaska's Salmon and People (SASAP), an initiative designed to synthesize data about wild Alaska salmon.

Why scale as my object of inquiry

Scale is significant across many scientific disciplines. Due to a deep and varied history, scale is a vast, often ambiguous, concept. This ancient word's polysemous nature is derived

[1] Interoperability is a term from systems engineering which refers to the modularity or flexibility of a product or system to work with other products or systems (Simone et al., 1999).

from its deep roots in Proto-Germanic language. The version of the word as we know it in science comes from Latin. *Scala* meaning 'ladder' presents a linear view of scale and the inherent relations. In this definition, it is clear how the idea of ratios became central to scale as well as the predominantly linear shape that scale and scalability take. Scale is commonly used to refer to measuring devices, ratios, dimensions, and levels of complex systems.

Given the myriad consequential ways that scale impacts different scientific domains, the American Association for the Advancement of Science (1989) named scale one of the four interdisciplinary themes that cut across domains. This thesis engages in this ambiguity in an attempt to provide a framework for attending to scale, particularly in the study of ecosystems and data infrastructure that supports them. Throughout my research, I encounter scale as a major aspect of science that has been largely neglected in the philosophical and historical studies of science. Through my work, I offer a general language for engaging with scale, and I provide scaffolding for conceptually attending to scale in data integration endeavors.

Research questions in this study

Given the critical importance of achieving scale when studying ecosystem dynamics, my central research question is: *how do scientists instrument scale to understand ecological phenomena?* To answer this broad question, I synthesize prior literature on scale in ecology and Science and Technology Studies (STS). I also ask: *How are long-term research infrastructures sustained through time? And, how might a definition of scale help practitioners respond to these challenges?* Or, in other words, I seek to 1- define scale; 2- define how scientists instrument scale, and 3-show how to reconcile scalar issues.

I offer a framework for systematically engaging the different strategies for instrumenting and defining scale. Through the development of this framework, I argue that scale is produced in the interstices of data infrastructure and the phenomena of study. Errors or anomalies are caused by scalar mismatch and happen in the spaces between inner or endogenous scale of the phenomena being studied and the outer or exogeneous scale of the information infrastructure. In practice, this means that better scaffolds are required to conceptualize scale, particularly in large-scale data infrastructure work.

The core motivations of my research are the following: 1- to understand how long-term infrastructure in ecology is sustained through time; 2- to characterize how scientists define and instrument for scale; and 3- to identify how issues of scale challenge the design of robust infrastructures for ecology. The questions I ask do not pertain to one discipline in particular but have drawn from two main bodies of work. First, I review the literature in ecology about the challenges of selecting the appropriate scale for determining data collection protocols. I then draw from infrastructure studies as a lens for uncovering how scientists instrument scale. By bringing together the ways information science and STS as well as ecology has reconciled issues of scale, I contribute a more general language for working with scale.

Contributions

Intellectual contributions

This dissertation is focused on the role that instruments play in the movement from phenomena of research to data and how these instruments play a role in the production of scale. To do this, I provide an empirical and historical account of how data — particular to

salmon ecologies — have been produced, and I explore the dynamics of research, which is reflected in the data work.

My ethnographic account tells a story that brings together open data science, long-term field ecology, quantitative ecology, and Indigenous knowledge. It does not seek to weigh in on which approach is "better" but rather show how each is suffused with cultural perceptions about science. In other words, I tell a story about the ways that knowledge is made and re-made based on different cultures of knowledge production.

This ethnographic account highlights how scientists deal with the issue of scale: how they define and instrument scale, and how scale often complicates scientific research. This in-depth investigation is important to add to discussions about data infrastructure, not only the social and political shaping of data infrastructure, but also the work practices that are involved in the construction and maintenance of such infrastructure.

The major intellectual contribution of this thesis is a definition of scale. Through this work, I shed light on different approaches to scale and offer ethnographic insights into what scientists do when engaged with scalar challenges. Scale is frequently used to refer to everything from ratios to laws to a quality of growth. I do not directly engage with the most popular or colloquial definition of scale, which is a quality of growth, commonly referred to as economies of scale. Instead, I engage with the ecological literature to articulate how scientists in ecology define scale. However, this work has implications that move beyond ecology and can help better articulate the distinctions between scale in data science and scale in natural science.

Practical contributions

In addition to my theoretical contributions, I provide methodological insights into how to study contemporary data initiatives ethnographically. The past decade has brought many calls for inventive approaches to ethnography (Lury & Wakeford, 2012). My approach responds to a need in the data ethnographic space by employing a novel approach technically. To do this, I wrote a Python script to scrape the scientific programmers' GitHub account to analyze their communication traces. I combined this with a novel conceptual move to explore how actors scale the phenomena of study, rather than the management of an initiative. This provides insight into the material issues that plague research scientists and furthers the space of infrastructural inversion (Bowker, 1994), ethnography of scaling (Ribes, 2014), and trace ethnography (Geiger & Ribes, 2010, 2011).

Most importantly, I contribute to both Science and Technology Studies (STS) and Computer Supported Cooperative Work (CSCW) by showing how issues of scale in science are frequently issues of interest to information science. This is the central argument that I make in this dissertation, which I highlight through cases that illustrate the flexibility and recalcitrance of a knowledge infrastructure and the multi-scalar approach to scientific data work. To do this work that spans conceptual and practical implications, I bring together the fields of ecology, STS, and CSCW to better understand environmental research infrastructures.

Ultimately, I conclude with a general language around scale that will help scaffold future data integration endeavors. In taking seriously the concerns of my actors (scientific programmers, ecologists, fisheries biologists and managers), I bring together multiple

perspectives on the topic of data integration and scale and from that produce a framework intended for usage in future projects.

Methods

This study is based on my ethnography of the State of Alaska's Salmon and People (SASAP) project, a partnership with academic researchers, non-profits, Native Alaskans, and the National Center for Ecological Analysis and Synthesis (NCEAS). This ethnographic study eventually led to a broader ethnography of science on wild Alaska salmon research as findings from initial research clarified my questions, and my field sites expanded beyond the immediate sites within the SASAP project. Throughout this ethnographic work, I ask how do scientists instrument scale to understand ecological research phenomena. While this thesis takes on the task of providing a general language around scale, it also focuses on the specifics of science, specifically salmon science, in Alaska. This context is important as there are particular ways that science in Alaska has developed over the last century.

The SASAP project represents a specific approach to understanding change over long-time scales. This first case was the impetus for my being in the field and speaks to the universality in the creation of a database. My initiation into this project was participation in SASAP, a partnership with NCEAS, data professionals, research scientists, natural resource managers, and Indigenous advisors. In this study, scientific programmers encounter errors in data ultimately reconciled by understanding scale.

The project I studied was premised on the view that creating an "ultimate database" might facilitate better access to data for answering questions about some of the most significant and urgent environmental problems. The project goals were two-fold: to break

down barriers to collaboration and to ask questions not previously possible. They argue that the database will facilitate the emergence of new knowledge providing a resource for the broader salmon stakeholder community.

Given the current focus on developing tools that support the integration of data scientific practices in the natural sciences, this research examines user-to-user communication in a GitHub repository used by computer and natural scientists archiving data for a salmon data synthesis project as a point of entry for exploring data collection, preservation, and usage.

This focus on NCEAS led to two additional field sites as I consolidated my research questions. In one site, I worked as a field technician with ecologists in the Alaska Salmon Program (ASP), a field program that has sustained data collected practices since 1946. And, in the other, I looked at how quantitative ecologists and natural resource managers work together to define local scale.

My engagement with both sites was motivated by a need to understand how data are produced in the field. In this research, I explore some of the practices of how salmon are taken from the stream and translated into data. This follows the way scientists attend to issues of scale when producing data and knowledge about salmon and their habitat. This dissertation is not only a history of science, but it offers an ethnographic account of the work practices of scientific knowledge production. This is in the same vein as Bruno Latour's circulating reference (Latour, 1999), in which he defines representation in the sciences as a part of translation chain. The translations he outlines are the ways that objects are able to remain durable and move across different sites. Rather than following what makes an object 'durable', I look at what makes an object change through time.

Theme 1: Multi-scale dynamics in the study of salmon

Pacific salmon have been widely studied. In his critical environmental history of the Pacific Salmon crisis in the Pacific Northwest, Lichatowich asks how salmon declines have been such an *intractable* problem (Lichatowich, 1999). Conversely, I ask what has made salmon—its habitat, its declines, its renewal—*tractable*. While his focus is on the Pacific Northwest, mine is on Alaska; and, there are some major differences, namely, markedly fewer people have moved into Alaska, which aligns with his central argument that the introduction of colonial white settlers around 150 years ago brought on a decline in salmon productivity. There is also the more global-scale challenge of climate change, which has rapidly changed salmon habitat in just the couple of decades since his publication. To answer this question of how salmon have been made tractable to scientists, or in other words, have been turned into data and studied, I introduce a debate about the changing role of data in scientific knowledge production.

As change has accelerated, Alaska continues to experience unprecedented heat (Suryan et al., 2021) and subsequently, shifting habitats (Stanford et al., 2005), which make it increasingly important to monitor areas that perhaps were not considered high priority in the past moving beyond the invisible present (Magnuson, 1990). These irruptive climatic and environmental changes have been met with rapid changes in scientific research. Over the last 100 years, there have been not only advances in technological approaches but also increasingly complex and urgent problems. These challenges operate at multiple scales – temporally and spatially but also locally and globally (and regionally). Research in these areas requires not only long-time scales to predict far into the future but also the ability to make findings meaningful at a local level (Bocking, 2004).

Understanding how fine-scale and broad-scale dynamics, or cross-scale interactions (CSI) shape one another is increasingly important for understanding global change (Peters et al., 2008). This growing complexity has led to a focus on the scale at which scientific practice occurs both in terms of the pressure to predict further into the future and in the importance of understanding further into the past. The key focus of ecological science to date is to understand change at these various scales. However, scientists take different approaches to dealing with the challenges of understanding change across scales.

One such approach has been to collect and synthesize large amounts of data while another approach focuses on sustaining a research program in the long-term. Other approaches focus on modeling as a key way to deal with multi-scalar issues. Within all of these approaches, there are debates about the role that data should play in the production of scientific knowledge and how those data can be used to produce legible, relevant, and responsive information to stakeholders.

Theme 2: Salmon as incidental

I grew up purchasing salmon from a Walmart, something farmed, something light orange, kind of bland, a little soft. My first trip to Alaska served up some of the darkest, reddish fish I have ever seen. I did not buy it at a kitschy tourist shop or in a fine dining restaurant. It was served to me by people I just met who welcomed me into their home. It was given to me in a cabin far away in the Alaska wilderness, thawed in the lake. It was outside in an icebox at a store with no sign and one cashier. It was everywhere. It was also in the names of many local establishments; it was on logos and printed on t-shirts. People wore it on their car bumpers, as tattoos on their body. I have stood in streams and almost been knocked down by the thick

stream of salmon swimming upstream frenetically against the force of gravity (figure 1). If there was one thing I learned that summer: *the salmon are not incidental.*

Figure 1. Sockeye salmon swimming to their spawning grounds at Hansen Creek, Alaska, 2017

Salmon have complex life histories (figure 2). They are semelparous, which means they spawn and then die. Salmon are also anadromous, which means that they are born in freshwater but mature in oceans, and then return to freshwater to spawn. These salmon that have fed in marine environments return to their rearing grounds to die in freshwater. Knowledge of this phenomena -- anadromy -- was hard won. As recent as the early 1980s, some argued for a saltwater origin. However, it is now widely accepted that they travel many miles in the ocean only to return to their natal spawning grounds in freshwater. Once they return to freshwater ecosystems, their decaying bodies feed back into the ecosystem.

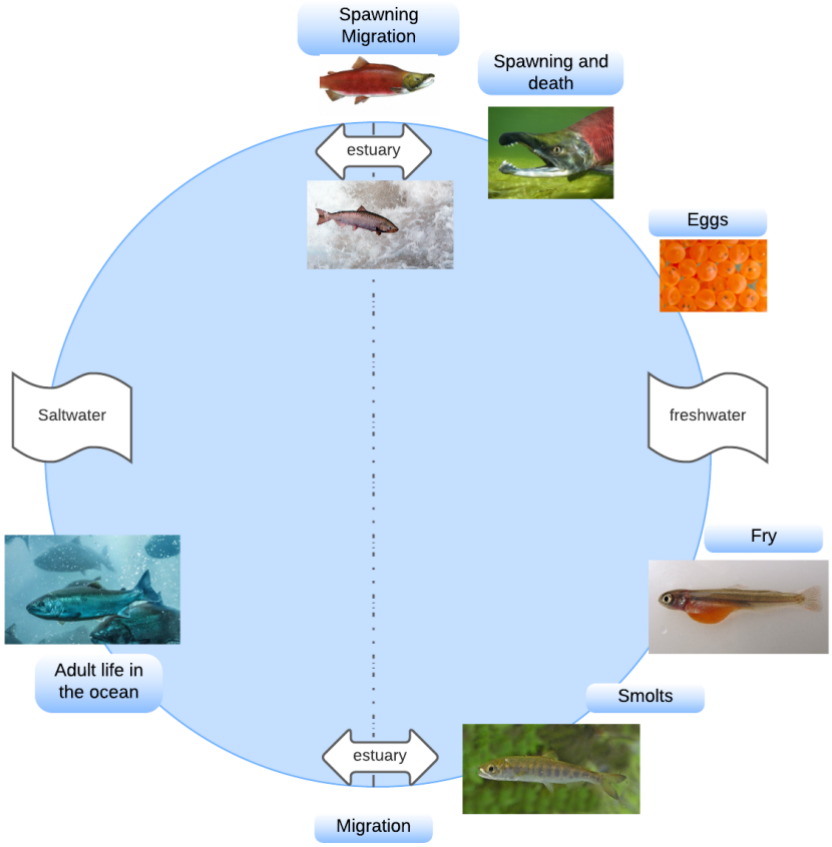


Figure 2. Salmon lifecycle

Not only has the knowledge of their life cycle stages been hard won, but at each stage, research continues to illuminate aspects about those stages in life. These aspects include questions such as what kinds of habitat spawning adults prefer or how time spent in rearing estuaries leads to fitness in the ocean. New theories continue to be proposed. For example, scholars have recently focused on collective movement ecology suggesting that salmon are more collaborative than competitive when migrating homeward (Westley et al., 2018), using

their neighbors as cues to progress homeward[2]. The model of migration that Westley et al. (2018) propose is based on social cues rather than purely individual preference or choice. Using the social model, outputs matched what had been observed for the last 30 years in Hansen Creek (in the Wood River system).

This theory aligns with other fields' changing paradigms about the role of collaboration in realms of nature heretofore considered competitive. For example, Suzanne Simard, a forest ecologist, has shown how mycorrhiza networks provide pathways for trees to communicate with neighboring trees to send information about potential threats. This is a paradigm that emphasizes collaboration over competition for scarce resources, and is just one example of how science is filled with changing paradigms, alternate views on a phenomenon, and instrumentation. Presently, there is no clear way to apply this newer insight about salmon migration to data collection. Fish continue to be counted, sampled genetically, and tagged like they have been for decades. In other words, the instrumentation that facilitates our understanding of salmon remains largely constant while the scientific theory and models change.

"The salmon are incidental" was one of the first things said to me during a call about this project. It referred to how the data science and data work proposed for this project were the main focus. In other words, the salmon could have been anything; what was of interest was developing programmatic and technological solutions to the problem of streamlining and synthesizing ecological data.

[2] As with many research findings, they start with an outlier. Peter Westley's study of salmon straying rates led him to notice that when salmon are abundant, there are fewer strays. He noted that this was counterintuitive to a view that sees salmon as competitive: "if there's competition on the spawning grounds, you would think some fish would go elsewhere. Yet the numbers said the opposite."

Theme 3: Alaska at a crossroads

In the 1990s, native anadromous Pacific salmonids… are at a crossroads. Biologists Willa Nehlsen, Jack E Williams, and James Lichatowich, 1991

There are many points of entry to begin this story. The story of Alaskan settlement could go back to 10000 BCE around the time that the Bering Sea Land Bridge, or Beringia (the Laurentide and Cordilleran ice sheets), connected Siberia and eastern Alaska. Up until about 30 years ago, this theory rested on the presence of an "ice-free corridor" by which people could have walked from Siberia. However, recent studies (Heintzman et al., 2016; Pedersen et al., 2016) have indicated that this supposed corridor would have been blocked between about 30,000 and 11,500 Before Present (BP). Because the Pacific coast was de-glaciated around 14,500 years BP, scholars (e.g., Pedersen et al., 2016) now suggest that initial colonialization came up through a Pacific coastal route.

For the purposes of this research, I trace the history back to a more present era in which these theories are developed and disputed and in which the western world has reshaped the land (and the sea) in an attempt to first harvest its resources, and second to manage those resources. These changes have at times been caused by direct interaction with the land while at other times are outcomes of latent effects of an ever-changing climate. This work looks at how data legibility is often contingent on participant's ontological commitments, which are embedded in infrastructure. A great deal of research has looked at science as extractivist viewing science as part of an imperial history; however, this work offers a more nuanced perspective on how knowledge is co-produced among not only different knowledge systems but also instrumental and infrastructural systems.

Open any book on the history of Alaska and the description will likely begin with an ode to its vast and remote beauty. Characterized as the 'final frontier', Alaska conjures images of

a wildness that most only experience in escape. However, many have disputed this image of a final frontier with Cronon calling it out as an outright myth (Cronon, 1995). Providing texture to the concept of 'wilderness', Cronon (1995) calls out the frontier framing as a particularly American way of demarcating a past that had vanished, a lamentation at a loss of the heroic frontier. Perceptions about Alaska and the appeal of Alaska emphasizes a fierce independence and self-reliance, and yet, the state is deeply reliant on government. Haycox (2020) is perhaps most prominent for making this argument.

There are a number of contradictions or tensions like these that arose during my time in Alaska: a) the legacy of industry (oil and gas, timber, commercial fishing) that put pressure on the very resources they exploit; b) a conflict between the identity of Alaska as a frontier state, a haven for rugged individualists seeking refuge from tyrannical government rule, and yet Alaskan Natives who inhabited the land for centuries have been oppressed in many ways throughout European colonization; c) Climate change is impacting Alaska and northern regions much more rapidly than other places and yet the people who often live in the communities that are experiencing the first effects of climate change are not responsible for the large-scale production of co2 emissions.

My official introduction to the Alaska salmon world was in November 2016 at the Salmon and Society conference. It was here that I met many of the famous people of the Alaska salmon world: Daniel Schindler, Ben Stevens, Sue Mauger, Gale Vick, Courtney Carothers, Peter Westley, Milo Atkinson, Mike Williams, Jim Fall, and others. Conference attendees and presenters expressed a sentiment that Alaska was turning the corner on its long, embedded relationship with industrial development. The view was that the two could coincide: salmon and development - but that environmental protection was becoming more of the norm. Over

and over again, people in the conference focused on how to not be like the "lower 48" with our dams, pollution, rapid population growth, and dwindling salmon runs.

In retrospect, this was a fascinating time to do a study of the infrastructural support of environmental science research as much of it seemed on the brink, not enough to be "good science" but also in need of being socially and politically relevant in an incredibly tense moment politically. Toward the end of 2019, much of the hope that was present in early 2016 had evaporated. Pebble Mine -- a large open pit gold and copper mine that threatens the salmon spawning habitat -- was back on the table. Governor Dunlevy had cut $1.6 billion funds much of it from education. Salmon were dying in the Yukon from overly hot summers. And, warming was at an all-time high.

The conference looked at everything from the design of fish culverts to the impact of urbanization to the technical challenges of identifying cold spots in fish habitat and developing partnerships for a monitoring system to understand rapid change. There was recognition of major issues like Pebble Mine, but also the sense that there are other – perhaps just as important – issues to pay attention to: the "nickel and dime" stuff (Milo Atkinson) like culverts channeling oil on a stream bank. One theme that reoccurred throughout the Center for Salmon and Society meeting and in various conversations about the SASAP project is that Alaska (and its environment) is at a crossroads.

This crossroads is not only the significant change in climate and in the human systems impacted by that changing environment, but also a political move that is less focused on conservation and more attuned to corporate interests. While this event kicked off the beginning of our 3-year project on the State of Alaska's Salmon and People, the Moore foundation funding for salmon research was ending. In the following chapters, I unpack a

couple of the tensions that are common to Alaska, namely, the relationship that industry has with scientific research, the scale of change experienced in a land that has been sparsely populated, and the outsized interest of scientific exploration in the region.

While much of the data used in data integration work comes from the state department, the history of science on salmon in Alaska goes back farther in time than Alaska statehood. One such program that represents early research on salmon is the Alaska Salmon Program.

Background

Big science and international scientific collaboration

Much of the aforementioned change to science has been brought on by advances in computational power. The emergence of the International Biological Program (IBP) in 1964 marks the fruition of an empiricist anxiety in which an increasing emphasis on collecting, archiving, and synthesizing large data sets in order to answer global-scale problems (Aronova et al., 2010; Hampton et al., 2013; Strasser et al., 2011) took precedence to other ways of doing science. IBP was modeled after the more successful International Geophysical Year (IGY), which was a large-scale effort to gather data about the earth in the 50s. IGY, a synoptic data effort (Odishaw, 1957), had two important predecessors — the International Polar Years of 1882-1883 and 1932-1933 (Belanger, 2006). Although IBP was not as successful as its predecessor, it did mark the beginning of Big Science.

Big Science was a term promoted by Alvin Weinberg, then director of the Oak Ridge National Laboratory, to mean that post WWII, academia had become bound with big government and industry, and transformed science from an individual endeavor into a collective enterprise (Weinberg, 1961, 1967). Weinberg famously saw the computer as

"technology's answer to our modern society's demand for more and more data and more information" (Weinberg & Bowers, 1968). While Weinberg was certainly a technological optimist with a focus on the role the computer would play on science, he was also deeply concerned with nuclear energy. Bocking (1995) traces this history locating the Atomic Energy Commission (AEC) in the center for support for ecological research as concerns about environmental impact of test bomb fallout and nuclear reactors grew. The AEC and the subsequent formation of the National Science Foundation in 1950 led to a major boost in post-WWII funding for ecology (Coleman, 2010).

The model of Big Science was centered on data-centrism and field observations, and in the United States, was seen as a way to promote the Big Science model and to move ecology from a "'little science' field of biology into a modern Big Science" (Aronova et al., 2010). However, the ecological community met this initiative with much less acceptance[3]. Despite its shortcomings, it marked the beginning of "big ecology" - the premise that ecology is characterized as diverse individual projects but lacks a culture of data curation and sharing at larger scales and as such, should embrace data-intensive practice to better understand global-scale change. This led to an enthusiasm for synoptic data collection and what Aronova et al. (2010) characterize as Humboldtian science "in a world shaped by the post-atomic age and Cold War sensibilities" (187).

Early documentation on the Long-Term Ecological Research (LTER) networks (e.g., Likens, 1989) highlights how the IBP transformed into LTER after it ended in 1974. Alongside

[3] Most entertainingly defined as a "boondoggle designed to ride the coattail of the IGY" - Lamont Cole Lamont C. Cole to Frank Blair, 6 Mar 1964, IBP Papers, Series 1: USNC/IBP: Ad hoc, Folder Membership: Chairman: S. A. Cain. Survey of Biologists re Interest in International Bio- logical Program (IBP).

these developments was the burgeoning field of ecosystem science. Coleman (2010) writes that Margalef's ideas on nature as a 'cybernetic machine' was a major influence on the field not only due to the scientists who took on this concept but also for getting funding. Reiners (1986) notes that ecosystems lack the clear boundaries that cells and organisms do. As such, ecosystems were comparatively less empirically accessible or tractable.

In the development of the ecosystem or holistic view of ecological function is the idea that nature is a system and should be in a state of stability. As Kwa (1987) writes: "the machine metaphor provides implicitly (and in some cases also explicitly) a structural analogue for an ecosystem" (426). This metaphor is an example of how nature came to be viewed systematically.

Stakes of data openness

Many in the critical data studies space (boyd & Crawford, 2011; Dalton et al., 2016; Levin & Leonelli, 2017) have argued for more attention to how open data projects develop technically and politically. And, Kitchin (2014) has argued that more attention should be paid to "discursive and material moves and their consequences" (p.66). This call for attention to the discursive practices elevate the need to engage with the sociomaterial practices of data work.

To achieve data openness, scientific programmers and data scientists curate scientific records long after data are collected or produced as part of sustaining knowledge infrastructures. These large-scale archival projects typically include a proposed public-facing artifact intended to provide useful information to downstream users. However, challenges arise when trying to commensurate and interoperate data from disparate sources for a

diverse audience. For example, there are disagreements about who the audience is and what types of actions should be privileged over others; when cleaning and archiving data, negotiations emerge around what data seems realistic and what can be flagged as an error. Additionally, some data come with corresponding metadata that has been added after data collection while other data are carefully produced under the governance of data organizations with standards that ensure metadata are already in place.

Data sharing and data openness are by no means a given (Van Noorden, 2021). This is evident in current disagreements between Genbank and GISAID about sharing SARS-CoV-2 genome data. While data organizations focused on the streamlining of heterogeneous data make the politics around data openness and open science seem obvious, there are many reasons why scientists and data producers would be more judicious about with whom to share data. Furthermore, sharing is a matter of substantive concern (Borgman, 2012). Hours of work and resources were poured into the production of these data, and thus, they do not extricate easily. As such, regulatory and institutional arrangements are set up to make data acquisition less painful or at the least, less contestable.

While these regulatory and institutional facets are evident in the negotiations that take place at NCEAS, the value of open data often reigns supreme in discussions about data work. Furthermore, database design typically falls under the purview of data workers who seek to create a certain vision of data integration that is focused on interoperability, reproducibility, and seamlessness. Ecological research, however, has not always been focused on the creation of large-scale data sets or repositories.

Although the importance of data for ecological research is not disputed, the big data movement has ushered in incendiary claims about the potential for data-driven approaches

to revolutionize science (see Jim Gray's vision of data-intensive work as a fourth paradigm of scientific research)[4] as well as make public trust in science by making it transparent. Mazzocchi (2015) argued that big data would elevate inductive reasoning "in the form of technology-based empiricism" as well as lead to future envisioning in which "automated data mining will lead directly to new discoveries."

However, this growing emphasis on and celebration of data-centrism has not come without its critics (e.g., Lindenmayer & Likens, 2013; Mirowski, 2018; Penders et al., 2008). Critiques of this naturalization of data sharing have shown how openness can be a guise for downstream privatization of data. This "imperative to share" ends up being exclusionary to who ultimately has access to that data (Lezaun & Montgomery, 2015).

This movement has sparked fierce criticism from within the ecological domain as well, most notably, the characterization of big data approaches as "fishing trips" in which no questions are asked and results may be erroneous, as such (Lindenmayer & Likens, 2013). They attribute this to the authors of these research projects having limited or no understanding of the datasets they use or the ecosystems in question. They go on to refer to these scientists as "parasitic" and that furthermore, this parasitic scientist is the product of an institution that values competition and pressure to publish which disincentivizes resources spent on collecting new data, alluding to a systemic problem.

Leonelli (2016) conceptualizes this debate about data-centrism to argue that rather than an emergence of a method for doing data-driven research, this period marks a rise data-centrism as an approach to science, a view that values data aspects such as the visualization

[4] This focus on data-intensive work has even been referred to as a new paradigm of science (In Jim Gray's last talk to the Computer Science and Telecommunications Board on January 11, 2007, he described his vision of the fourth paradigm of scientific research.) http://jimgray.azurewebsites.net/jimgraytalks.htm

and integration of data as a "discovery in their own right and not as a mere by-product of efforts to create and test scientific theories" (2). This is in contrast to the theory-centric view, which has been the predominant way of thinking about science within science as well as in the philosophy of science. However, this theory-centric view has been challenged by the data-centric view particularly in sciences that require vast amounts of data (e.g., molecular biology). This is not new (Leonelli & Tempini, 2020). This idea of producing knowledge through synthesis rather than the formulation of new hypotheses is not novel; however, "the institutionalization and marketization of data are making 'data-intensive' approaches more prominent than ever before in the history of science." As is evident in these debates, data openness is not a given value, but creates stark divides about how data and what data should be shared.

A word on data, information, knowledge

In this thesis, I often bandy with the terms data, information, and knowledge. Distinctions between data, information, and knowledge have been widely considered in the field of information science (Zins, 2007). Colloquially, data, information, and knowledge are used interchangeably; however, in the field of information science, they are often conceptualized as part of a continuum with data as the "raw" form of information that can then be turned into information, or a signal, which then leads to knowledge. In Ackoff's (1989) conceptualization, data are symbols or products of observation but are not usable in their "raw" form.

However, STS scholars have argued that "raw data is an oxymoron" and that there is no such a thing as "raw" data, or in other words, data are always produced or can be traced to

an earlier version (Bowker, 2005). All the while, according to Ackoff (1989), data directly stem from observations and are processed into information. This process is vague, but essentially involves making data useful to an imagined stakeholder. This linear notion is unrealistic in practice.

Information is inferred from data and denotes the systems that facilitate this inference. Information provides an explication of what is there while knowledge is about how something happened. Where information is defined as data, knowledge is even harder to define in that it encapsulates tacit and explicit knowledge and is as such more specifically about expertise.

Critiquing a linearity of prominent conceptualizations in which data is filtered into information filtered into knowledge, Baker & Bowker (2007) offer a conceptual definition and propose a data-knowledge grid in which aspects of tacit or embodied knowledge and explicit or more codified knowledge is taken into consideration. This model conceptualizes data movement as a flow in which data are transformed into information which is transformed into domain knowledge represented as movement across the different quadrants.

Whether data moves across a straight line toward knowledge, occupies a gridded tile next to knowledge, or flows freely through life cycles that circuitously make their way to knowledge, one thing is clear: data are useful pieces of information encapsulated in standardized formats (sometimes unruly) and part of a larger ecosystem with which participants in this ecosystem may make claims with these data, sometimes pieces of evidence, sometimes discarded detritus.

Providing further simplification, Baker and Mayernik (2020) propose a two-stream model (figure 3) to highlight some distinctions between data production and knowledge production arguing that there are key differences in data collected for the purposes of knowledge production and a more "recent notion of data production that refers to the management of data for reuse by other communities and larger audiences" (p. 9).

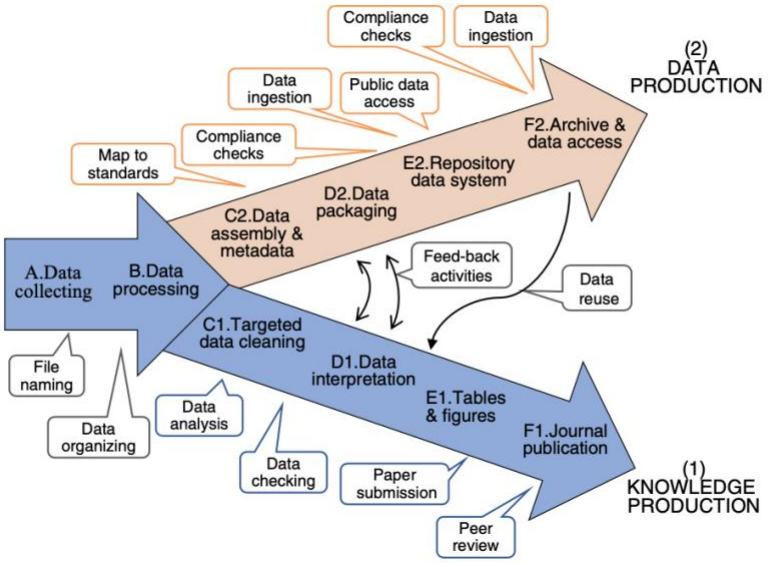


Figure 3. From Baker and Mayernik (2020) distinctions between data production and knowledge production

The production of data is inseparable from the context in which data are generated. Many (e.g., Mayernik, 2016; Millerand et al., 2013; Millerand & Bowker, 2009) have written on the importance of context, often referred to as metadata. However, my study surfaces context as crucial specifically in understanding the achievement of scale. These different ways that data extraction and technological advancements shape knowledge claims are evident through following the data production process. Some of the major restrictions driving questions in scientific knowledge production are the limitations to data availability as well as the

interdisciplinarity of the question. As I argue in later chapters, scale plays a crucial role in the production of data and therefore, of knowledge production.

Structure of the remainder of this dissertation

This dissertation is divided into two main sections. The first section, chapters 2-4, provides an overview of my methodological and theoretical commitments. Chapter 5-7, the second section, discusses the legacy of salmon science, highlighting the three sites in which this research takes place. This section includes my major empirical explorations of three different ways that scale is instrumented.

In my study, I focus on scale as a way of interrogating data integration – both within salmon ecology itself as well as within the data ecosystem. And, what I make is an epistemological point, which is that data production and subsequent curation and usage are a part of the knowledge production ecosystem.

While mostly structured linearly, this thesis is not necessarily chronological. Chapter 2 outlines the conceptual frameworks I use to make sense of scale and introduces the major components of my arguments. Chapter 3 provides a description of my research design and methodological approach, which includes the sensitizing concepts I used to engage with the material. I also make the case for my approach to conducting data ethnographies. Chapter 4 provides an outline of the three different sites of study. Here, I trace a broad view of the history of salmon science in Alaska as well as an historical account of how the data organization, NCEAS, came to focus on synthesis.

To develop these recommendations and scaffolds, my empirical work (chapters 5-7) seeks to characterize three common ways that scientists instrument scale: 1- the creation of an

integrated, open data repository; 2- the sustaining of a long-term field program; and, 3- the translation of locally-produced data into computational models. These three empirical sites can be characterized as three distinct practices that instrument scale: a) synthesis and data cleaning; b) specimen collection and data production; c) modeling and the human as instrument.

The first chapter relies on more novel forms of ethnographic methods and engages with the GitHub communication between scientific programmers and domain experts. This work sheds light on how a data integration initiative encounters issues of scale in their data work.

However, I also engage in more traditional participant-observation. As Latour followed botanists and soil scientists through the Amazon rain forest to look at how empirical evidence is turned into text, he found that in the process of measuring and sampling, "locality, particularity, materiality" is lost while "compatibility, standardization, text, calculation" is gained. In this spirit, I participated as a field technician with the University of Washington's Alaska Salmon Program (ASP) (chapter 6), primarily assisting in collecting long-term program data such as limnological data, adult sockeye spawner counts, occurrence of brown bear hair, the diets of resident fish, and the measurements and quantity of juvenile (or smolts) from tow-netting surveys. This is of particular interest in an era in which data-centric practices for scientific knowledge production are becoming prevalent.

The study of how a field research program has sustained itself highlights the central challenge of scaling for time. Through three instantiations of specimen collection, I show how aspects of the fish can say more about time at fine and deep scales. I argue that these specimens were collected because of alignments between the exogenous and endogenous scale of data and phenomena of study.

Chapter 7 engages with one of the eight working groups from chapter 5 to look at how scientists define local scale when developing models for understanding salmon populations. In the conclusion, I offer a framework for working with scale defining the properties of scale. This chapter engages the issue of local and global scale as I offer a definition of how scientists define 'local' in a specific instance. I show how this is co-constituted by global scale and make the case that asymmetries occur between the local and global which define each other. Finally, I conclude with some proposed properties of scale, methodological implications, and implications for design and for data policy.

Chapter 2: Multi-scalar approaches to sustaining research infrastructure

T*rees grow for hundreds of years, hurricanes may decimate a site every 50 years, and droughts may last for decades; thus, a long-term perspective is needed to understand the ecological response to these slow changes or rare events (Hobbie et al., 2003).*

Why scale?

It is impossible to study scientific practice, particularly as it pertains to data, without encountering scale. Yet discussions of scale within histories of science or STS, when not absent, have been an inconsistent, albeit broad, topic of study. Some of the primary scholarly works that have addressed scale have looked across a wide variety of disciplines and scalar dimensions (e.g., local, global, micro, macro, spatial, and temporal). In one approach to understanding scale, a physicist covers a range of disciplines from business to biology (West, 2017) defining scale as a mathematical ratio that can be applied to any discipline universally. In this depiction, scale is universal, akin to a law. Inversely, an English professor offers almost the opposite: thought experiments that highlight a theory of scale that moves away from the

reductionism of the former (DiCaglio, 2021). In this depiction, scale is contingent and relative.

A philosopher of natural science, Sabina Leonelli (2018), addresses different time scales of data, outlining different temporalities in data. At the same time, information scientist, Karen Baker, researches mesoscale infrastructures for field-based natural science (Baker, 2017). Meanwhile, STS scholar Kim Fortun offers a scalar heuristic for understanding data cultures (Fortun, 2009). In CSCW, Lee and Paine (2015) discuss scale as a dimension in a model of coordinated action that refers to size as "the number of participants involved in the collaboration" (p. 184). These arguments and the scalar dimensions are disparate and have come from a variety of disciplines. Evidently, scale is critical across disciplines demanding a more systematic approach to understanding it with an STS lens, one that does not offer a purely reductionist nor relativist definition of scale.

Size is often a key aspect of scale when referring to scalability as expansion or the ability to scale up. Further, the major scalar dimensions that STS has most consistently engaged with is local and global scale, which came out of early theorizing around infrastructure (Baker & Bowker, 2007; Bowker et al., 2009; Edwards et al., 2009; Edwards, 2010; Ribes & Finholt, 2009).

Despite its critical importance across domains, scale has largely been absent from data integration discussions. This is because technology or innovation literatures typically discuss scale in terms of economies of scale, often referred to as *scalability*. This is not the kind of scale I predominantly engage with in this study, however. I engage ecological scale for two primary reasons: a) the focus on scale of natural phenomena has been largely left out of data discussions; and b) the focus on scale of natural phenomena is a major consideration

for the ecologists, data professionals, and research scientists I study. Scaling for time and space is a practice taken by natural scientists, and scale is considered one the premier conceptual challenges in ecology.

Even scale in the natural sciences is vast. Scale can be both a measuring device or instrument as well as an index or sum of measurements -- at times, a tool for measuring while at other times, a conceptual device. As Schneider (2001) points, scale has a number of common technical definitions: cartographic scale, multi scale analysis, ecological scaling, which refers to "power laws that scale a variable to body size" (Schneider, 2001, p.546). This is the kind of scale that Geoffrey West takes up in his broad sweep of the many definitions of scale. I take up the technical definition for scale to refer to the "extent relative to the grain of a variable indexed by time or space" (Schneider, 2001, p. 546).

Beyond its use in STS and in ecology, scale takes on many meanings. The word itself is ancient with diverse origins. Schneider (2001) provides a brief historical insight into the use of the term identifying two roots to the word: the Old Norse root in *skal* meaning 'bowl' and the Latin root *scala* meaning 'ladder' which leads to scales such as musical scales. Scale can mean everything from an instrument or analytical tool to ratios, measurement tools, or even a descriptive theory and ranges from disciplines such as mathematics, music, biology, chemistry, and social science. Scale can refer to size – macro/micro or large/small or local/global. In the natural sciences, this often appears as fine/coarse and therein follows a discussion of how global entities are reduced to smaller parts. An example of this is in Rader's (2004) discussion of how science has identified different animals as a stand-in for human biology, in particular the mouse as model organism

In this chapter, I theorize scale in order to develop a conceptual framework that can be used to overcome scalar complexity in STS studies of knowledge production. To better understand scale in practice, I ask how scientists instrument scale to produce knowledge about ecological phenomena. I turn to the field of ecology to look at how ecological science has understood and approached issues of scale.

I then turn to the literature on infrastructure to outline how STS and infrastructure studies has written about scale. I identify the domains of scale as organizational, temporal, and spatial and furthermore, that scale issues revolve around resolution and size (or level of complexity) as the primary ways that infrastructure has engaged with scale. Here, I argue that scale has been left out of or looked over in many data integration discussions.

Scale is the organizational, spatial, and temporal levels important to understanding complexity. However, scale is also intimately bound to the extent, or scope, of a study and the subsequent resolution of data for adequately studying a bounded space. In the following section, I overview what has been written in ecological science about these dimensions of scale in two different fields: ecology and information science.

Issues of scale in ecology: Making sense of variability

Variation is the key feature of ecosystems that makes them resilient to change. This complexity, however, presents scientists with decisions to make with respect to scale, particularly when developing an understanding of ecosystem change. Scientists observe, categorize, and study systems at different temporal, spatial, and organizational scales, and while scale is a primary concern for ecological research, challenges remain when operating across these levels of complexity.

Scaling challenges arise when considering how to bound a research site or activity (e.g., What time frame is appropriate for the question at hand; how large of an area should be studied to understand the phenomena; and at what level should a phenomenon be studied?). In answering these questions, scientists determine how fine or coarse data resolution should be as well as how broad the study should extend. These issues are evident in not only research-level theoretical questions but also in the mundane practices of making data ready for usage.

Since its inception, ecology has been a discipline focused on scale. The term 'ecology' was first proposed in the late 1800s, and early on it became an interdisciplinary field investigating both biotic and abiotic phenomena. In understanding ecosystem change, scale was introduced as a premier conceptual challenge. Arthur Tansley coined the term 'ecosystem' and noted that ecosystems evolve over a range of scales "from the universe as a whole down to the atom" (Tansley, 1935, p.299). He reframed the study of nature from groups of individual living things to the study of dynamic interactions between living and nonliving things. As the field became focused on biotic and abiotic communities, questions of scale became more prominent. In a study of the Cedar Bog Lake, Lindeman (1942) noted that the choice to use the lake as the unit of analysis highlighted "that distinctions between the biotic community and dead organisms or inorganic nutrients were artificial" (Kingsland, 2011, p.18). It was here that Lindeman argued that ecosystems should be the ecological unit not just the biotic community. The move to sample the lake as unit of analysis rather than just the biotic community — in other words, to sample for the whole instead of just its constituents — is an example of how scale influences the study of ecosystems versus individual organisms.

In addition to organizational dimensions are spatial and temporal, which are selected as part of a sampling protocol in research studies. Delcourt & Delcourt (1983), for example, are well-cited for providing an understanding of disturbance regimes in the context of space-time domains, and noted the ways in which scale reflected the "sampling intervals required to observe [the phenomenon]." However, Wheatley & Johnson (2009) found that 70% of the observational scales employed in wildlife-habitat research were chosen arbitrarily. This is in part due to disciplinary and institutional expectations that impinge upon scientists as well as logistical, conceptual, and infrastructural challenges of acquiring the appropriate scale of data for the question at hand. This is evident of a growing recognition that data at different observational scales produce different understandings of a phenomenon.

Plainly, scale is the lens through which scientists segment complexity over a range of measurements to make meaningful contributions. It is a decision about what boundaries need to be drawn to best capture the study site. However, it is also often constrained by what is possible (e.g., what advancements in technology have allowed for, what the state of knowledge is, and what kind of resources the scientist has on hand).

Produced through the sampling and measuring of a research phenomenon, scale should be theorized at level of infrastructure that moves beyond data. As instruments are relational to the data they produce, so is scale relational to data and research questions yet is distinct from the datum itself. There is no "natural" scale or perfect scale at which an ecological phenomenon should be studied. Rather, as Levin argues, "observers impose a perceptual bias, a filter through which the system is viewed" (1992, p.1943). In the next section, I go into detail about what aspects are important for ecology when considering which

observational filters to use. There are a number of scales at which understanding can be achieved, organizationally, spatially, and temporally.

Organizational

Levin (1992) argues that scale presents the major conceptual problem in ecology due to variation at spatial, temporal, and organizational scales. Furthermore, this variability is not absolute, but is only meaningful relative to other scales (Levin, 1992). In other words, the organism and the environment are co-constructive of each other due to their interaction. As a point of departure, I outline Levin's (1992) three dimensions of scale — spatial, temporal, and organization.

Ecosystem hierarchies are represented as multiple levels: organisms, populations, communities, ecosystems, biomes, and landscapes. These, however, are heuristics for making multi-scalar interactions visible (Allen & Starr, 2017). The important factor about scale and ecosystem change is what happens in the emergence of interactions. I will evidence this below.

It may seem reasonable that if fish begin to be depleted in lakes and streams then an appropriate response is to stock those lakes and streams with more fish. An approach such as this, however, does not take into consideration unforeseen ecosystem-wide effects. For example, stocking a lake with hatchery-raised fish may cause declines in wild stocks as they begin to mate with wild fish. This effect is considered to be a *direct* effect (e.g., competition with or depletion of the organism of interest); however, there are also important *indirect* effects to consider. For example, changes in pH due to acid rain in the late 1970s led to concern about the effect on fish. While studies in the lab showed that lake trout are not

negatively impacted by minor changes in pH, it has been shown that phytoplankton (e.g., *Mysis*) might be adversely impacted by increased acidity (Nero & Schindler, 2011). This illustrates how food web dynamics can result in downstream, indirect negative consequences on the lake trout. This is an issue of scale at the level of organization because viewed in isolation, the direct effect seems plausible (e.g., looking at interactions at the organizational level of populations). However, this is only a partial understanding that elides the fuller picture of food web dynamics, which take into considered smaller-scale, often invisible, organisms such as phytoplankton.

The emergence of indirect food web dynamics – while important for ecology – are also relevant to information science, and this is a question of how scientific classification systems operate in practice. Classification systems, such as taxonomies, help segment the organization of ecosystems, and are also central to the concerns of information science. Through taxonomies, the differences and similarities of organisms are made visible. At the same time, there is no one direct way to apply these filters. For example, microbial communities might be stable to perturbations if using a taxonomically broad filter, but this is a matter of choice (Allen & Hoekstra, 2015). In this light, taxonomy is a function of information infrastructure, created by scientists to segment the world (Thomer et al., 2018). As such, much of the problems around organization revolve around definitions. Defining a population as a "collection of individuals belonging to the same species" conceals issues such as populations over temporal landscape or the coherence of a population on the landscape. Hence, being a collection of individuals within a particular species group is contingent upon other factors, namely, space and time.

Spatial

This relational quality of scale is evident spatially. For example, species dynamics look different at a local scale look different from dynamics at a regional scale. For some birds, a 4-hectare plot is a local scale. In this small space, one bird species, the Least Flycatcher, negatively influences another species, the American Redstart. Because Least Flycatchers are aggressive, they push the American Redstart out of optimal habitat patches. On a local scale, it looks as if the Least Flycatcher negatively influence the American Redstart; however, when zooming out to look at a regional scale, it is evident that the American Redstart adapts and moves to nearby habitat patches (Wiens, 1989).

As the example above illustrates, knowledge production is highly contingent on the scale at which one investigates the phenomenon. This is compounded by data availability as broad-scale spatial data (e.g., in the form of publicly available satellite imagery or remote sensing data) are easier to acquire than local-scale behavioral data. Furthermore, as I show in following sections, the incentive for using large datasets is higher than the incentive for using small datasets as large datasets promise more statistical significance.

Another complicating spatial dynamic is that impacts of broad-scale change are commonly felt at the local-scale. As such, it is common to study local-scale interactions (e.g. habitat selection by salmon populations) in the study of effects of broad-scale change such as climate change. However, due to the interactions between biotic and abiotic functions and the process of adaptation, it can be challenging to know what spatial extent explains a phenomenon.

Ecology is concerned with understanding the ability of a piece of habitat to support a certain stage of life (life cycle). Stanford et al. (2005) note that in variable habitats, organisms

might display "phenotypic plasticity[5]" to adapt to those variable habitats. However, if an organism remains in a habitat over several generations, it may adapt to those specific "spatial or functional attributes" (123). These are often referred to as ecologically significant units to mean that they are locally-adapted populations. Not being adaptable to dynamism in habitat does not typically bode well for a species or population. The metaphor of a shifting habitat mosaic takes into account the many theories about ecosystems and unifies these theories with a "continuum of 3-dimensional shifting habitat mosaics from headwaters to the ocean" (123). As is evident, the major aspects of scale with respect to spatial scale is how much to localize a site spatially, how local conditions change over time, and how biotic and abiotic factors interact.

Temporal

In addition to organizational and spatial scale, ecologists also contend with temporal scale. There is a temporality in ecosystems and also in data systems. Ecological phenomena occur at geologic timescales, which are not commensurate with the timespans of scientists' careers, of funding cycles, or of instrument functionality that are too short to be used to understand long-term change. Moreover, there is a temporal aspect to sustaining, transforming, and managing data. As such, data do not "speak for themselves" but are made to speak by human and technological actors working together. While research infrastructures are built with dominant temporal logics (Mazmanian et al., 2015) in mind, the issue of temporal scale in ecology challenges the design of seamless infrastructures.

[5] Phenotypic plasticity is a fancy term for adaptation.

Despite limits to collecting data retroactively, there are other issues particular to temporal scale. These are often due to mundane issues such as instrument failure in the field, gaps in research funding, short timespans of researchers collecting data, and a lack of standards across field sites.

One explanation for the challenge of temporal scaling is that ecologists — as people — study phenomena on "anthropocentric scales" (Wiens, 1989). This mismatch between scientists' perception of time and space and ecosystem-wide responses to space and time is perhaps best captured in John Magnuson's conceptual work on the invisible present, which he defines as the time scale at which our responsibilities are most evident. In small time frames (e.g. that of a human life time or a scientists' career), phenomena like acid deposition, deforestation, and invasive species might be evident; however, phenomena as large as climate change require a different time scale. One example Magnuson provides is the ice cover on Lake Mendota in the winter of 1982-1983. Just that single year or even a few decades of data provided few interesting insights; however, with 132 years a general warming trend becomes visible.

This concern with a long-term perspective (Burton & Jackson, 2012) illustrates how phenomena might be hidden in the past and "reside in the invisible present" (Magnuson, 1990). Throughout the literature, it is clear that temporal scaling issues are compounded by latent processes, spatial dimensions, episodic events, mismatch between scientists' and ecosystem lifecycle, as well as emergent properties due to time lags between the causes and effects.

Resolution and extent

Other properties of scale include fine and coarse grain, or resolution, and extent. Grain is the level of resolution, and it encompasses the smallest entities in a study (Allen & Hoekstra, 2015; Allen & Hoekstra, 1991). Grain is relational to the sampling interval selected. Extent, on the other hand, refers to the largest entities that can be seen in data. Small extent could refer to an individual, a household, or a community while large extent includes regional or global-scale processes. It is atypical to combine broad extent (a large-scale area) with fine-scale resolution, which would mean collecting fine-grain data over a larger area. It is more common that broad-scale or broad extent studies have coarse grain data while studies at a smaller scale can offer a finer-grain resolution.

In ecology, local and global can apply to both spatial and temporal scale and are typically used to refer to extent. Extent is the overall area such as a population under study while grain refers to the size of individual units of observation. Statistically, this means that extent is the area defining the population under study. Wiens (1989) notes that extent and grain define the upper and lower limits of resolution of a study. Grain is a component that represents resolution as an average measure while extent sets the boundaries around a study and is a cumulative measure. In other words, 'local' is an attribute of various spatial and temporal scales. In my adaptation of Hoekstra and Allen's (1991) graphic, I illustrate these dynamics of the size of the study and the resolution of the data are related (figure 4).

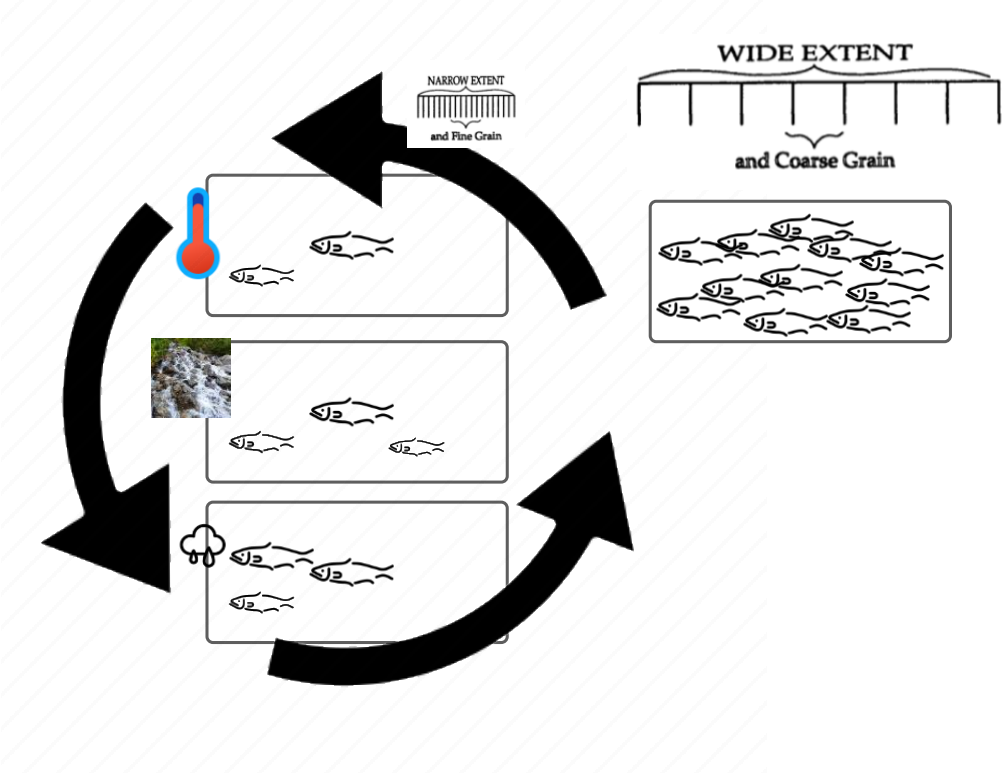


Figure 4. Adapted from Allen and Hoekstra (1991) to illustrate how narrow extent and fine grain reflect local processes while a wider extent and coarse grain can illustrate regional or even global processes

Clearly, the relationship between research objectives (e.g., studying local, regional, or global processes) and the resolution of data selected is critical. However, this matching between scale of research question with the area and resolution of data is rarely considered in data integration initiatives.

As is evident in the literature, different scales of perception can shape the outcome of scientific knowledge production. David Baltimore, Nobel laureate in biology, has argued that biology is an information science. In information science, issues of scale share some similarities as well as differences with issues of scale in ecology, and in the following section, I explore how infrastructure studies has conceptualized scale.

Issues of scale in infrastructure: Making sense of complexity

Infrastructure is a concept that is tightly coupled with the concept of scale (Karasti et al. (2016). Because infrastructure is layered over time, navigating different scales (e.g., time/space, human collectives, or data) is a key challenge (Edwards et al., 2013; Karasti et al., 2016) both for the creation of infrastructure as well as the investigation of it. Misa (1988) highlights the importance of scale in the history of technology, and Edwards (2003) picks up this argument showing that the major methodological issue is the question of scale when he asks: "how do infrastructures look when examined on different scales of force, time, and social organization?" For Edwards, scale is an analytic device intended to change perspective based on scale of analysis. Ultimately, he argues that infrastructure is a bridge across these scales.

Similar to the bridging work of Edwards' infrastructure, Karen Baker (2017) shows how mesoscale infrastructure can provide for flexibility across sites of scientific collaboration and data work. In her argument, the mesoscale is "a transitional point between the data origin where researchers generate data and the larger scope of digital data archives" (p.4). The emphasis on an intermediary layer suggests a traveling back and forth between local and global including multiple potential paths within the micro to macro scales. Likewise, Acker (2015) argues that ethnographically studying infrastructure at the meso-scale offers a view into an in-between area where change occurs.

In many ways, however, the focus on scale in information science and STS has been more centered on technology and scalability than on the scale of ecological systems. Ribes and Finholt (2007, 2009) offered the 'scales of infrastructure' as a broadening of analytic gaze. They argue that infrastructure operates across multiple 'scales of action' which they

categorize as technological, human and organizational, and institutional. In this work, it is evident that the language of scale comes from technology domains rather than ecosystems.

Bringing the mesoscale in conversation with the scales of ecology suggests that moments of breakdown, or scalar mismatches, shed light on the larger infrastructure behind the scenes. In the following section, I outline the important dimensions of scale that the study of infrastructure attends to, namely, that of temporal and local scale.

Size and space

A major emphasis within information sciences and infrastructure studies has been how to manage information at different scales, or how to manage large-scale information ecosystems. Monteiro et al. (2013) use information infrastructure to explain how large-scale information technologies takes place:

> "…interconnections of numerous modules/systems (i.e. multiplicity of purposes, agendas, strategies), of dynamically evolving portfolios of (an ecosystem of) systems and shaped by an installed base of existing systems and practices (thus restricting the scope of design, as traditionally conceived)…stretched across time and space…**shaped and used across many different locales** … over long periods" (p.576).

Engaging the question of "how big, or deep, or old, or widespread does something have to get before it becomes infrastructure?", Edwards et al. (2009) allude to the co-constitutive quality of small or local scale with big or global scale, noting that studies focused on smaller scales, "are relevant to the larger ones as well." This shares a definition with the question of extent in ecological studies — of how broad-scale or small-scale the study should explore.

Offering a variation on this theme of sustainability and scalability, Anna Tsing (2012) engages the issue of scale in her work on commodity chains of Matsutake mushrooms and makes the point that scalability is not something ordinary to nature. Using sugarcane as an

anecdote of nature under control "through the reordering of the social-natural landscape", she shows how Matsutake mushrooms tell a different story about "life in the ruins of scalability" (p.516). Matsutake mushrooms have mostly thwarted efforts to cultivate it due to its requirement of multispecies diversity of the forest. Scalability in a resource-context refers to dissolving or at the very least imitating diversity so that expansion can happen on a large-scale.

Here, scalability means the capitalist drive for expansion. Tsing locates the focus on scalability with the rise of economies of scale. This is similar to Scott's argument in Seeing Like a State: that different forces came together to render once diverse ecosystems into a scalable -- or in other words, expansive -- industry. This 'economies of scale' logic, however, is at odds with ecological diversity. Scalability may mean the making of a diverse ecosystem into an extensible commodity. However, it also requires scalability (of data) to produce knowledge of that diversity in the first place.

Perhaps most clearly, computer models reconcile issues of global and local scale (Edwards, 1999, 2010). While Edwards' object of research is the global climate, he shows how the development of international organizations, computer models, and the emergence of climate change in politics contributed to knowing global climate. Edwards (1999) illustrates how computer models have played a major role in producing global data, highlighting the fuzzy boundaries between models and data when modeling for climate science. However, in the act of making global data, there are many decisions scientists make that define local. Edwards writes: "The best numerical weather prediction (NWP) models today use grid scales below 1 km on a side compared to the 250-500 km grids of most general

circulation models (GCM)." It is not so simple to show how global scale is produced from local scales. This 'scaling up' is an aim in ecological science as well.

Often, the more particular or locally-applicable the results of a model are, the less the model is designed to make theoretical advances (Steger et al., 2021). Specificity is another common theme in considering the concept of local. In his environmental history of ecological research in the Canadian Arctic, Bocking (2013) discusses the tension between local and global. He argues that two concepts from STS are relevant to his study — that of situated science and that of mobile science. Situated science is the idea that scientific knowledge production is unfolding in a particular place while mobile science refers to circulation of knowledge. The essential tension is between locality and circulation.

These themes of local and global (figure 5) – or locality and circulation – are also evident in Latour's (1999) illustration of how science moves from "forest to expedition report." Rather than locating phenomena at the middle ground between things and categories of human understanding, he notes that phenomena are "what circulates all along the reversible chain of transformations" (p. 71). He goes on to argue that instruments help "grasp the practical differences between abstract and concrete." It is, thus, through the practices of its usage and the ways in which those practices are formatted, that the resulting data can be utilized as data.

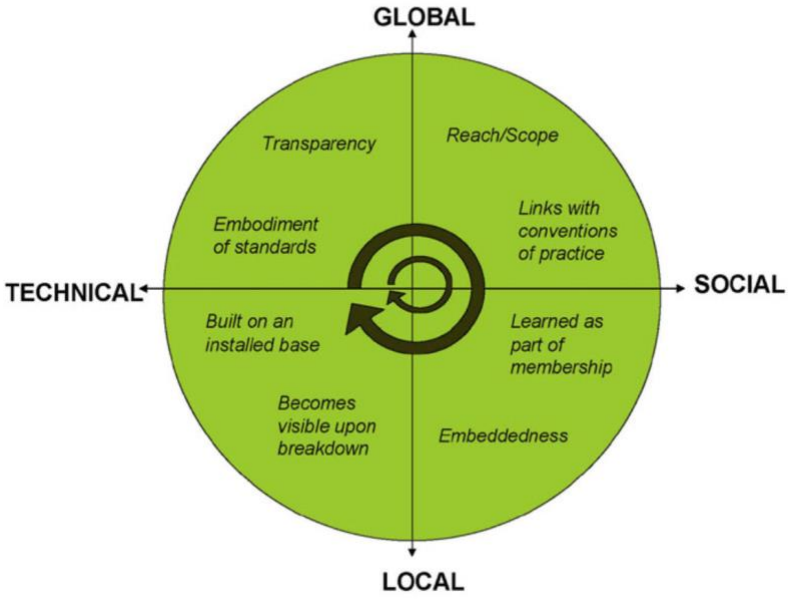


Figure 5. From Bowker et al. 2009 matrix for segmenting movement from local to global and from social to technical.

Temporality

As I showed in the previous section, the issues of scale that the study of infrastructure has contended with has been in part entangled with their attention to technology and the ever-increasing scalability of that technology. The focus has been on size and on tensions between what is particular and what is universal. In addition to a tension between local and global in the study of infrastructure, issues of scale in an information ecology pertain to the spatial, temporal, and organizational aspects (Baker and Bowker, 2007) such as the "challenges of unifying time scales, agreeing on spatial units, and clarifying species lists" (p.129). At the heart of the discussion of temporal scale within STS are two key challenges: that of sustaining infrastructure into the long-term and of understanding complex ecosystems with limited timeframes (e.g., that of a human lifespan).

Much of this engagement with temporal scale stemmed from infrastructural investigations of the Long-Term Ecology Research (LTER) networks (Baker & Bowker, 2007;

Karasti & Baker, 2004; Millerand & Bowker, 2009; Ribes & Finholt, 2009). In these writings, they use natural science as an analogy for how infrastructure can help with the problem of scale noting that a few decades ago, practitioners in ecology were frequently challenged by data loss and the relative short timespans of their careers compared to their phenomena of study. Illustrating this point, Edwards et al. (2013) note that the establishment of infrastructure such as LTER brought the promise of scale to allow ecologists to "look beyond the scale of a field and timeframe of a career" but to allow for the "prospect of studying ecology and climate locally, nationally, globally, and over spans of time that more closely match those of ecological change" (p.20).

Given these points, Karasti argues that there should be a renewed primacy to temporality in the study of infrastructure (Karasti, 2012). This is a scalar dimension that Edwards delves into repeatedly making the point that understanding ecosystems takes a longer timeframe than an individual's scientific career (Edwards et al., 2007). Edwards et al.'s (2007) insight that human time is often incommensurate with ecological time echoes what Magnuson (1990) highlighted in his concept of the "invisible present." Or, in other words, human time and infrastructure time do not always align. This is evident from scholars of infrastructure as well as scholars of ecosystems. Magnuson explains that human time is marked by a few characteristics: "the horizon of death; the salience of extremes; the fading and distortion of memory; the slow, faltering process of learning; and our restless, present-centered, single-focus attention." He goes on to show how geophysical or long-term time scales show infrastructure to be quite ephemeral.

Thus, temporality is also highlighted as a key challenge of developing infrastructure (Ribes & Finholt, 2007a, 2009). As interests or practices change, so too does a knowledge

infrastructure (Ribes & Polk, 2015). However, as Star has argued, there is the legacy of the installed base that must be dealt with. This is based on the idea that there is inertia to change and to decisions that were made when developing an infrastructure. Similar to the concept of the installed base (Star & Ruhleder, 1996) in infrastructure (that there is inertia to what becomes standard) so too do ecosystem dynamics play out. This concept of legacy bears some of this out in the information systems space. The idea is that an object resists, which forces the infrastructure to constantly shift to ensure alignment (Rheinberger, 2000; Ribes, 2015).

Multi-scalar dynamics: Emergence and scalar mismatch

Scalar issues come from the emergence of the aforementioned dimensions, which at the level of ecosystem is frequently complicated by broader influences (Wiens, 1989). The major challenge with reduction (with the making general something local and particular) is that emergent properties are difficult to uncover by studying the smaller parts of that larger whole (Allen & Hoekstra, 2015). This is the essence of one of the three issues of scale that Schneider (2001) provides: 1-issues in ecology are often at much larger scales than what can be studied; 2-related to that issue is that most variables can only be measured in a small space for a small amount of time; 3-patterns at small scales do not necessarily present at larger scales; one cannot simply "scale up". By emphasizing the role of heterogeneity in ecosystems, Schindler et al. (2010) show how current management schemes need to take a more multi-scalar approach.

In short, the main issues of scale can be categorized as follows: a) scale mismatch (the scale of observation selected does not match the scale of relevance to ecosystems); b)

invisible present or time lags; c) inertia and emergence; d) the incompatibility of the life cycles (life cycle of a human with the life cycle of an ecosystem; or envisioned life cycle of data with life cycle of phenomena) and e) issues with defining the correct scale or potentially limited by scale (sometimes we cannot know the processes at which scale we are interested in or studies focus on one scale rather than multi-scales). An additional challenge in issues of scale is outside pressure from the scientific community to present repeatable, statistically significant studies which can often result in selecting a more simplistic scale rather than a better study design.

Additionally, there is often a mismatch between scientists' perception of time and space and ecosystem-wide responses to space and time, perhaps best captured by John Magnuson's conceptual work on the 'invisible present', which he defines as the scale at which our responsibilities are most evident. In short time frames (e.g. that of a human life time or of a scientists' career), phenomena like acid deposition, deforestation, and invasive species might be evident; however, something as large as climate change requires a different scale of perception to understand its characteristics and its impacts on ecosystems[6].

Infrastructure and its inversion as an analytic lens

STS offers infrastructure as a concept for which to study sociotechnical systems. This thesis draws from infrastructure studies literature to theorize the processes and practices that

[6] While my use of ecological scale is to better theorize scale and provide a framework for infrastructural studies of knowledge production, interestingly, the metaphors from computing made their way into the ecology literature as well. Notably, Hutchinson and Odum borrowed from computing history to talk about ecosystems as cybernetic systems with feedback loops. Ecosystem ecology emerged as a systematic way of thinking of the environment, and even drew from computational metaphors. It was around this time that computation entered the scene and would bring about transformational changes in understanding interrelationships and interactions at both small and large scales.

support data management and knowledge production in ecological sciences. In all the different varieties—data, information, knowledge, research—infrastructure encapsulates more than the immediate physical properties of a system. While it may conjure thoughts of railroads, bridges, and highways (e.g., Graham & Marvin, 2002), infrastructure also pertains to the less visible, computational, and social support structures.

Even in early infrastructure studies (Hughes, 1983, 1989; Scott, 1998; Star & Ruhleder, 1996), the focus was on the invisible from electrons to social relations. However, the ubiquity of 'infrastructure' as a conceptual lens has led to a muddled, unclear, and contentious usage of the term. While the roots of the term can be traced to sociology of technology, it has been applied to refer to everything from large technical systems (Edwards et al., 2009; Hughes, 1989) to technical computing architecture to communication technologies (Monteiro et al., 2013) to even broader conceptualizations of entire sociotechnical systems (Lee & Schmidt, 2017).

However, the field took off in the early 2000s with calls for *cyberinfrastructure* - a now generally impotent term as one might ask what infrastructure is not cyberinfrastructure. However, at the time, cyberinfrastructure was signaled as a move to focus on collaborative computing infrastructure. While cyberinfrastructure may be no longer widely used, many of the concepts of infrastructure are still useful as analytic devices.

Star and Ruhleder (1996) provide one of the most cogent accounts of what infrastructure is and how we might study it. Their insight is almost methodological in that they argue infrastructure is always a relational concept. Karasti and Blomberg (2018) pick this up in their definition of infrastructure; however, they distinguish 'infrastructuring' from infrastructure as the empirical study of a phenomenon: "the notion of infrastructuring is

used to denote the open-ended, uncertain and dynamic, qualities of the phenomenon that render their study challenging" (p.236).

Studies of information infrastructure (e.g., Bowker et al., 2009; Bowker & Star, 1999; Cohn, 2019; Edwards, 2010; Ribes & Lee, 2010; Vertesi, 2014) show that infrastructure does not reveal all its internal operations, nor does it grant its users full agency. Visibility is a longstanding issue for technology design, particularly information technology: *how much of its internal operations should design reveal or conceal from its users?* Often, systems are described as usable or transparent, valuing their tidy or simple arrangement, but at other times we value their configurability. Some studies cast the issue negatively, such as the 'black boxing' of machine learning or algorithms. Others laud what seems the very same thing, such as 'infrastructure', which presents a generally positive valence for easy access without having to be mired in technical details.

In early literature on infrastructure, Star and Ruhleder (1996) identify invisibility as a central tenet of infrastructure, and is more aligned with methodological aspects to studying infrastructure. In this thesis, however, I show how scientific programmers invert to build infrastructure. As Star and Ruhleder (1996) have asserted, well-functioning infrastructure tends to fade into a background, embedded in routines and everyday practices. But in moments of breakdown, debate, deliberations, or evaluations such operations may be resurfaced in what Geoffrey Bowker (1994) has called *infrastructural inversion.* In defining infrastructural inversion as an analytic tool, Bowker et al. (2009) argue that infrastructure is relational in its reliance on both "static and dynamic elements" (99). Infrastructural inversion first appears in Bowker's Science on the Run (1994) in which he follows how Schlumberger (an oil field service company) took hold of the marketplace as the company

created their own measurement standards and instruments. In other words, by consolidating knowledge within their reach of expertise, the company inverted the typical scientific management trajectory.

Edwards (2010) engages with infrastructural inversion in his study of how researchers deal with the issues of scaling in climate modeling. In Edwards formulation of infrastructural inversion, it is a less a matter of coordination and more an issue of epistemic concern. Similar to Bowker's assessment that Schlumberger's success was due to inverting the infrastructure of science, Edwards (2010) attributes climate knowledge infrastructure's continuation to its constant infrastructural inversion, which he defines here as "continual self-interrogation, examining and reexamining its own past." There is a major distinction to be made here. In Bowker's study, Schlumberger uses the inversion as a way to justify their measurements while in Edwards' case, the inversions are much more a part of the everyday, routine nature of science. This speaks to one of the core tenants of infrastructure that is that it is visible upon breakdown (Star, 1999).

Bowker and Star (1999) offer some methodological themes for infrastructural inversion: ubiquity, materiality (of standards and classifications), the indeterminacy of the past (e.g., classifications change even while documentation stays the same and translation problems occur when trying to contend with legacy systems), and there are some politics of classifying and standardizing. They put this in the design category as it pertains to the decisions about what will be visible and invisible within a technological system.

Science is supported by knowledge infrastructure, which Edwards et al. (2013) define as "robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds." In other words, knowledge is not

something that is solely within individual actors but is rather a social achievement. Without institutions like the National Oceanic and Atmospheric Administration (NOAA), Alaska Department of Fish & Game (ADF&G), US Fish & Wildlife Service (USFWS), and informal associations like storytelling, communities, and so forth, knowledge would not be possible. Rather than an objective set of findings, knowledge is dynamic, socially and technically produced, and ultimately, a collaborative effort. Part of this dynamism has led the focus on individual expertise to be "replaced by the wisdom of crowds: noisy and endlessly contentious, but also rich, diverse, and multi-skilled" (Edwards et al. 2013) which has resulted in a blurring between knowledge producers and knowledge consumers.

Most of the STS writing on infrastructure of ecological research has come out of engagement with LTER (e.g., Aronova et al., 2010; Bowker et al., 2009; Karasti et al., 2010; Mayernik, 2016; Ribes & Finholt, 2009; Zimmerman, 2008). And Karasti, Baker, & Millerand (2010) show, through an empirical investigation of LTER, how infrastructure studies can be used to understand temporality, or to at least forefront Reddy et al.'s (2006) view that "time matters".

Collaboration in ecology and other field sciences has enabled research on phenomena that are geographically dispersed to be aggregated to say something meaningful about global climate change. However, this level of collaboration across multiple sites presents ecologists and data scientists with several challenges related to scale. In the following chapter, I explore the topic of scale in more detail to outline how it has been theorized in both ecology as well as infrastructure studies.

Temporalities of knowledge production work

Data professionals in general and scientific programmers in particular discuss data life cycles as a way to manage the temporality of a "domain" science. Puig de la Bellacasa (2015) shows how ontologies for understanding phenomena (e.g., soil) are obscured by the predominant timescape, which she defines as a timescape concerned with the intensification of productivity. This dominant focus on productivity obscures other ways that people may know or care about soil.

The temporal metaphor of 'lifecycle' is prevalent in the biological sciences as well as in information and data sciences. In the spirit of making data more accessible, commensurable, interoperable, and reproducible, data and information transformations have been characterized as a life cycle (Higgins, 2012; Strasser et al., 2011; Wallis et al., 2008). To address my research questions, I untangle two distinct concepts: data life cycles and salmon life cycles. Data life cycle models are frequently used in scientific data work to characterize workflows and they have been used in myriad disciplines that work with data (e.g., Fox, 2011; Lenhardt et al., 2014; Michener & Jones, 2012; Paine et al., 2015; Strasser et al., 2011).

Data life cycles: Data categories of time

DataONE's data lifecycle model (Strasser et al., 2011) is the lifecycle model that most frequently comes up in my field sites. It is often used in scientific data work to highlight some common transformations that data undergo. Increasing developments in monitoring technology have brought in a deluge of data with no particular purpose. The DataONE data life cycle model in figure 6 (Michener & Jones, 2012; Strasser et al., 2011) consists of 8 components: 1) *Planning* how the data will be managed and made accessible; 2) *Collecting* observations by hand or with instruments; 3) *Assuring* the quality of data through

inspection; 4) *Describing* the data using appropriate metadata standards; 5) *Preserving* data in a long-term archive; 6) *Discovering* potentially useful data along with its relevant metadata information; 7) *Integrating* data from disparate sources to form a homogeneous set of data ready for analysis; and 8) *Analyzing* data.

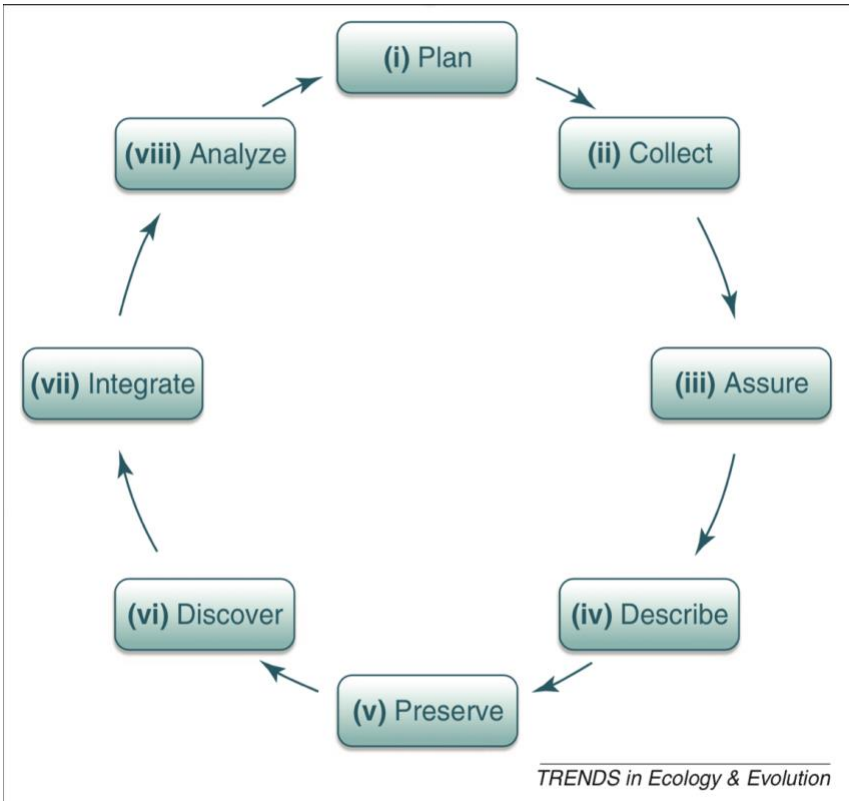


Figure 6. Data Life Cycle model (Strasser et al. 2011; Michener & Jones 2012)

In contrast to this life cycle metaphor, Leonelli & Tempini (2020) propose data journeys as a metaphor for thinking about how data journey across time, space, and social contexts. She proposes data journeys as a theoretical device for interrogating "the conditions for data movement, and the ways in which data mobility and interoperability can be achieved" (p. 2). While not explicitly called out as an alternative to data life cycles, it is clear that 'data journeys' is a way of thinking about the movement of data in a less well-behaved or constrained manner.

Salmon life cycles: "Natural" categories of time

Salmon, however, are both data and phenomena. While these lifecycle models of data and information have a temporality to them that suggest a birth and death rhythm, the rhythm of life and death is rooted in biology. This rhythm is visible in the attempt to produce data to make sense of phenomena observed in nature, that is, natural phenomena. Throughout this work, I show how the lifecycles of salmon shape their movement through data lifecycles.

Salmon have a unique life history: they are semelparous, which means they spawn and then they die. They are also anadromous, meaning they are born in freshwater, but they spend most of their lives in the ocean, returning to their rearing grounds to spawn. Beyond these basic life history dynamics, an aspect of understanding salmon is knowing both their age (a number associated with time units spent away from birth) and their migratory patterns. Helm and Shavit (2017) write about the biological processes such as age (e.g. time elapsed from birth) as well as migration or behaviors as the endogenous perspective, or "organism-centered perspective." Because geophysical properties are highly predictable, "organisms have evolved internal representations of time and space that direct their behavior and prepare them for upcoming conditions" (p. 246).

Even though this perspective situates these behaviors or physiological processes within the organism, the authors note that the behaviors and processes are really a combination of internal and external factors. This suggests some key challenges in drawing boundaries that influence what kinds of scale are associated with ecological functioning and those that are "endemic" to the individual organism.

In other words, there are two essential temporalities to knowledge production work, and they are relational to the object of research. The notion of endogenous (internal representation of scale) and exogenous (the representation of scale outside of an organism) aspects helps capture these distinctions (Helm & Shavit, 2017).

At the intersection of exogenous and endogenous

Using the lens of scientific rhythms (Jackson et al., 2011), I associate organizational, infrastructural, and biographical rhythms with the exogenous or data life cycle perspective while identifying the phenomenal rhythm as equivalent to the endogenous or salmon life cycle perspective. Jackson et al. (2011) note that "phenomenal rhythms are not fully pliable, they push back and circumscribe action, but with technique and technology can be aligned with other registers of time" (p.5).

Probing deeper on the point of data and phenomenal life cycles, Leonelli (2018) distinguishes two types of temporalities implicit in data processing: data time (e.g., the time at which data collection or dissemination or even analysis occurs) and phenomena time (e.g., the time period for which data serve to represent). Her formulation of data time concerns the logistical aspects of data production, for example, spending time in the field collecting data or the cleaning and assuring that data integration initiatives are often focused on, as she defines "the constraints and opportunities posed by the time spent in the production, dissemination, and analysis of data" (743). She builds off of Bogen and Woodward's (1988) study of materiality of making data-to-phenomena inferences to add a relational quality of data production and analysis.

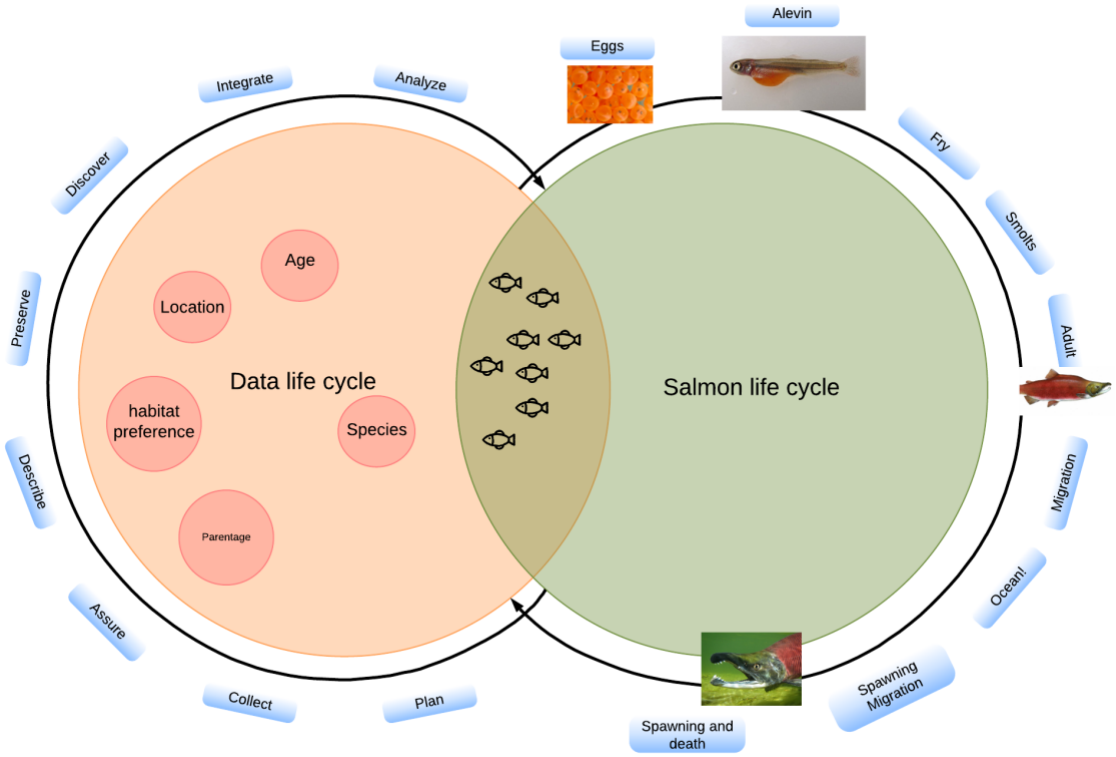


Figure 7. Salmon life cycle and data lifecycle

The idea of inner and outer scales aligns with others' writings on endogeneity and exogeneity (Helm and Shavit, 2017) and data timescale and phenomena timescale (Leonelli, 2017). These binaries are also co-constitutive to each other. The inner time scales of salmon age are important only relative to the datafication and downstream analysis of that data to make sense of age composition, for example. This is difficult to tease out though as the salmon are biotic entities whose intrinsic time-scale shape their migration, reproduction, and behavior. This dynamic illustrates the recursivity of the way data are used to segment phenomena to then produce knowledge about the phenomena.

Drawing from these frameworks, I adapt the terms 'endogenous' and 'exogenous'. I use endogenous to refer to internal representations of space or time (e.g., salmon run upriver to

spawn for reasons related to salmon biology) and the term 'exogenous' to refer to the infrastructural or database representations of space or time (e.g., the date, calendars, streamID). When producing data about a phenomenon, issues can result from the mismatch between these two: internal and external representations. I characterize the ways in which those mismatches occur.

While pointing to the ways that phenomenal time can push back on infrastructural time, Helm and Shavit (2017) do not outline a specific case of the production of "nature". In their concluding thoughts, they make a call for research in the area of instrumentation noting that inventions such as the camera were used to slow down something like a hummingbird's wings to make it legible to the human eye. Due to ethnography's focus on human activity, the instruments that make research possible often go unnoticed. For this reason, I take up this opportunity to further investigate the role of instrumentation in the production of a phenomena, and further, I show how scientists produce scale for our understanding of the natural world.

Conclusion

While Ribes (2014) brought the focus of scale from information science, I bring the concept of scale from ecological science to theorize how a deeper understanding of scale can aid future data integration initiatives. In other words, I highlight the version of scale that is focused on the natural (e.g., ecological or biological) phenomena itself as a primary contribution rather than a scale that has been muddied by a technological or economic focus on scalability. STS emphasizes that nature itself is sociotechnically produced. More clearly

stated, common understandings of nature are the product of mechanisms for segmenting and re-constructing nature as phenomena.

But as scalar challenges highlight, the whole cannot be derived from multiple parts. Its diversity of definitions and applications is perhaps one reason it has been poorly theorized. This research seeks to shed light on the concept and produce a conceptual framework for systematically exploring scale.

In sum, I have provided an overview of the scalar dimensions deemed important to ecology as well as those that make up the study of infrastructure. To make sense of variability, ecologists segment the world spatially, temporally, and organizationally and match the resolution of data to the extent of their field of study. This approach necessarily involves some element of data infrastructure as decisions are made about what kinds of data to collect and over what duration. Similarly, the study of infrastructure has relied on scale as an analytic device for understanding complexity: scaling up or down from an initial data point.

Chapter 3. Methodology: Field devices for the study of scale

"*The salmon are complicated because we study them." -participant in the SASAP project*

Introduction

This chapter has two main sections. The first outlines my approach to data collection. I situate my ethnographic study within contemporary ethnographic work and outline the data collection protocols and analyses that contributed to this thesis. The second section provides an overview of how I studied scale in my research. I define the methodological traditions

from which I draw, and I conclude by offering an adapted version of the ethnography of scaling (Ribes, 2014) to contribute to the study of large-scale scientific initiatives.

In this thesis, I approach instruments as scalar devices, or as tools for providing insight into larger paradigm shifts. In this section, I make the case for approaching scale through instruments, and I conclude with insights into how to employ an ethnography of scaling in such a way to understand the object of research. This is built on the idea that ethnographic accounts can provide deep (or thick descriptions) of a particular place in time. The strength of this method is its ability to refute top-down structures that might shape the researcher's perspective. However, this makes it difficult to say anything on a larger scale. Without reduction, general categories of knowledge cannot be produced. As such, in addition to my traditional ethnographic work, I employed a more novel form of ethnography inspired by trace ethnography (Geiger & Ribes, 2010) and the ethnography of scaling (Ribes, 2014).

By exploring a combination of approaches that look at traces of communication in a current project as well as historical-archival traces from the past, I develop an approach to the study of research infrastructure, particularly when distributed in time and space.

Research Methods

In this dissertation work, I rely on qualitative methods to answer my question of how scientists instrument scale. Although my research was conceived in an environmental data science initiative, data science is not my object of study. Rather, I explore scale — and the data practices that surround the production of scale — as a way of looking more closely at sites of knowledge production. Moreover, data science is one of many fields interested in issues that surround data. As such, I travel into the spaces of data production. In this way,

my field is not strictly a study of work practices in a bounded research center, laboratory, or field site. It is more akin to a multi-sited ethnography (Marcus, 1995) in that distributed work occurred in some instances as well as multi-temporal as this work was also at times distributed across many decades.

The primary data I draw from (table 1) includes, 1- GitHub comments, issues, and commits from the SASAP synthesis project, 2- 300+ hours of ethnographic field data, 3- interviews with 24 managers, scientists, and community stakeholders, and, 4- meeting minutes from the Kuskokwim River Salmon Management Working Group.

Dates	**Site**	**# Pages**
Primary Fieldwork		
2016	NCEAS SASAP	71*
2016, Apr 1-Aug 30	NCEAS working group meeting calls	56*
2017, Feb-Aug 18	NCEAS working group meetings	32
2017-2019	Interviews	610
2017-2018	GitHub data (1000 commits, 691 comments, 185 issues)	NA
Secondary Fieldwork		
2016**	Conference on Salmon and Society	27
2017**	Kuskokwim River Salmon Management Working Group (KRSMWG) Meeting	42
1990-2019	KRSMWG meeting minutes random sample	267
2017**	Kuskokwim Intertribal Fish Commission Meeting	6
2018**	Conference	16
	TOTAL	1127

Table 1. Inventory of data collected

*Includes notes taken by collaborators

**Exact dates not included for the sake of maintaining confidentiality

Participant observation

I conducted fieldwork between August 2016 and 2020, resulting in over 300 hours of observations. This sometimes involved sitting in meetings – virtually or in a room in Santa Barbara. At other times, it included sitting around a dinner table after a day collecting data in the field. And, at others, I interviewed people to ask a series of questions aimed at understanding how they thought about scale at a local level.

The inclusion of ethnographic studies into collaborative work practices is a somewhat new endeavor. Perhaps most renowned was Lucy Suchman's early work at Xerox Palo Alto Research Center (PARC) as she employed novel ethnographic methods to understand the micro-practices of how workers struggled with machines (Suchman, 2007). This kind of research would eventually grow into an entire field called human-computer interaction (HCI).

As a part of the Gordon and Betty Moore Foundation funding that supported my research, the grant specified two research assistant positions for 'data ethnographers' with the goal of understanding how synthesis work occurs. Although ethnography has roots in Anthropology, it has been adapted over the years to be used in the study of expert communities allowing researchers to uncover "how we know what we know" (Knorr-Cetina, 1999). The stated purpose of the data ethnographic portion of the study was to better understand "the challenges, dangers, and benefits of working across institutional cultures with different data management and sharing standards." As I was tasked with understanding

the process of data synthesis work, I was present for many of the virtual and in-person working group meetings.

I was always in a somewhat unusual position in this study. As a first-year Ph.D. student, I was often still figuring out my own methods and objects of research. I was both relying on the ethnographic mode of asking questions and participating where I could while also aware of the power dynamics in the room. How could I study tenured professors in entirely different fields? This is one reason I aligned myself with the more conspicuously data-focused scientists (e.g., those working with biophysical data). It was easier for me to objectify their data and see it as something unique than it was with the social scientists who were also conducting historical and anthropological studies, evaluating participatory governance at the state level (Krupa et al., 2020), and developing indicators of well-being (Donkersloot et al., 2020).

Furthermore, the biophysical data were more tractable to the research organization — NCEAS. Data in the form of spreadsheets with measurement standards and metadata was what the NCEAS data task force had trained for. As such, the GitHub communication platform is more clearly filled with discussions such as these whereas the social science researchers — the anthropologists and historians and even one economist — were less apparent in the data discussions. And as such, this thesis tracks the way that data and data science play a role in natural science.

GitHub traces

As members of the research team, we were granted access to email exchanges, documents, Slack communication, and GitHub communication with the caveat that Slack would not be formally analyzed. Participants in SASAP allowed me to follow their activities

through many email listservs, shared cloud-based docs, and GitHub, a software versioning platform. The primary data source for conducting the trace ethnographic (Geiger and Ribes, 2010) portion of this study was the commit, issue, and comment logs from SASAP's GitHub repository. Trace ethnography is a methodological approach that scaffolds the exploration of technologically-mediated spaces by studying documentary traces such as transaction logs, source code, or version histories (Geiger and Ribes, 2010). As a common versioning software, GitHub provided an avenue for exploring the collaboration traces among the scientific programmers. As such, I relied on the GitHub API to scrape data from NCEAS' enterprise account. With an amalgam of Python code, I was able to access the API and retrieve 1,004 commits, 185 issues, and 730 comments.

Through thematic coding, I identified the areas that drew the most attention from the scientific programmers. Through in vivo coding and memo-writing, I distilled high-level categories of data transformation over time.

Semi-Structured Interviews

In addition to participant-observation and trace ethnographic research, I traveled to the Kuskokwim region to work with the Orutsararmiut Native Council (ONC) to understand the ways that locally-produced data and local observations inform management. I employed a qualitative interview study (Weiss, 1994) using semi-structured interviews with Alaskan Native Tribal in-season managers, ADF&G in-season managers and researchers, and federal managers and researchers as well as community stakeholders in the Kuskokwim region (table 2) about their experience assessing data utility, defined as a participant's perspective

on how useful the data are for understanding salmon ecosystems and how local citizens play a role in the production of that data.

Participant Category	#
Federal manager	5
State manager	7
Tribal manager	5
Community monitor or stakeholder	7
Total	24

Table 2. Interviews conducted with managers, community monitors, and other stakeholders in the Kuskokwim region.

Interviews were audio-recorded, transcribed, and then coded with the Atlas.ti software. Coding was conducted in line with thematic qualitative coding techniques (Saldaña, 2013). Each interview was roughly 60 minutes long resulting in 25 pages of text for a total of 600 pages of text. The initial codebook was developed by taking a first pass through the interviews, highlighting common themes with in vivo codes (Charmaz, 2006). In vivo coding refers to the phrases used by participants, which can often reveal insider knowledge and shared perspectives. In this initial pass through the interviews, I sought to preserve the language used by participants moving on to more thematic coding in later phases.

To gather an understanding of how communities are involved in management and how locally-sourced data is utilized by management, I rely on interpretive inquiry through interviews. Interpretive inquiry emphasizes a phenomenological perspective and refers to the understanding that a researcher's understanding of humans "cannot be separated from their social and cultural world that is always in process" (Morehouse, 2011). Accordingly, because my questions are less focused on measuring effectiveness and more focused on

understanding how particular practices unfold in Kuskokwim communities, this kind of inquiry is appropriate.

I cannot pretend that my account captures everything or even tells the entirety of what data science applied to the natural sciences looks like. It does, however, offer a glimpse at what kinds of discussions and negotiations occur and what aspects about working with data challenge the scientific programmers.

Methodological Traditions

Grounded Theory Ethnography

Grounded theory is a qualitative research method first conceptualized by Glaser & Strauss (1967) as a way of studying data without applying top-down theories to make sense of the world. Glaser and Straus' (1967) strategies develop theory inductively from research data instead of from existing theories (Glaser & Strauss, 1967). More specifically, it is a tradition that looks for action in the data to look at how grand theory or concepts are constructed in the micro-interactions. This also allows the researcher to engage in simultaneous collection and analyses of data, constantly refining and building upon the findings.

Informed by a social constructivist perspective (Bryant, 2002), Charmaz contributes to grounded theory by bringing an interpretive perspective and arguing that not only do we interpret the meanings and actions of participants; they also interpret ours (Charmaz, 2006). Accordingly, this approach is appropriate for studying scientific knowledge production because it allows for the flexibility to encourage the interviewees' perspective to emerge out of the data, rather than prescribing prior assumptions to the interaction with informants.

Practical activities for producing scale

As this work is inspired by an ethnography of scaling approach (Ribes, 2014), I rely on ethnomethodology and actor network theory (ANT) as a way of uncovering the practices of scale. In other words, ethnomethodology and ANT offer a lens for treating scale as the outcome of instrumentation.

Ethnomethodology emphasizes following on-the-ground sensemaking to understand how actors construct reality, categories, and phenomena. As with any field, getting closer to the origin brings forth the tensions and disagreements that lead to the emergence of the thing itself.

In an infamously hard to comprehend text, Harold Garfinkel proposed ethnomethodology as an "investigation of the rational properties" of expressions in a social context (Garfinkel, 1967, p.11). In other words, the assumption of shared meaning is secondary to the contingent and ongoing "practical action" that occurs. If meaning is created in the interaction, from where do norms, stereotypes, and conventions originate?

The world of an ethnomethodologist is a world undergoing constant negotiation and renegotiation. It is 'endogenous' in that it is focused on the everyday as a lived social achievement (Heritage, 1984; Pollner & Emerson, 2001). Stability in the social world (or durability) does not stem from shared meanings, rather it is the "tacit use of the documentary method of interpretation to find the coherence of situations and actions" (Suchman, 2007, p.81) that allows for the constant creation of stability and meaning. Similarly, 'local' or 'particular' does not exist a priori to interactions as it is through the interaction that something is made nuanced. While its roots are in Parsons, Husserl, Pierce, and Schutz,

ethnomethodology did not begin to bloom until Garfinkel broke from the conventional sociologists who he argued did little to explain where social rules and order originate.

Garfinkel saw the world of interactions as achievements of social actors. While Schutz diverged from Husserl's phenomenology, he was influenced by Husserl's focus on the "life-world of mundane experience" (Dourish, 2001). Heidegger, too, was influenced by Husserl. However, he added the perspective of objects or of 'experiences'. In Being and Time, Heidegger shows that objects, or technologies, disappear when they become 'ready-to-hand'. Dourish's interpretation of Heidegger's 'ready-to-hand' is that the "equipment fades into the background" (Dourish 2001, p. 109). This fading into the background presents a problem to the study of technologically-mediated worlds, and this is a problem that makes social scientific studies of computer science and work practices interesting.

This was a central concern to Lucy Suchman in her early work at Xerox Palo Alto Research Center (PARC). To better understand how people and machines work together, Suchman was brought in to study the design of the interface of a photocopier. Here, she draws on her anthropology and ethnomethodology background, defining the interest of ethnomethodologists as an interest in "how it is that the mutual intelligibility and objectivity of the social world is *achieved"* (Suchman, 2007, p.77). She argues that situated action is about the social and material ways that intellectual labor is carried out. Because collaboratively organized artifacts shape cognition and behavior, it is necessary to understand the context in which action is situated as commonplace ideas about the world are "not the precondition for our interaction but its product" (p. 77).

Latour, too, takes ethnomethodology as a starting point in interobjectivity (Latour, 1996), arguing against political philosophers such as Hobbes, against the interactionists[7], and he also pokes a bit of fun at the ethnomethodologists when he says "framed interaction is not local by itself - as if the individual actor, that necessary ingredient for social life with whom one then has to construct the totality, had existed for all time" (p.233). Latour shows how complex interaction predates humanity as is seen in the negotiations of chimpanzees and baboons. For Latour, the social contract is mythical. Since the complexity of social life does not altogether distinguish us from baboons or chimpanzees, the ethnomethodological case does not explain our complicated work. Latour shows that rather than participating in continuous renegotiations, human actors pass abstractions through time and space by means of objects, or non-human actors (e.g., we have clocks that synchronize time without having to negotiate the hour with every meeting participant). This passing along of abstractions is a way that humans globalize from local interactions. The instruments do not disappear; rather, the negotiations behind them disappear.

This is evident perhaps most clearly in computer science where instruments or abstractions are often designed to produced certain behaviors. Failures in system design approaches can be understood by seeing how assumptions written into technology are critical components of human negotiation (Dourish & Button, 1998). Button and Dourish share a lineage with Suchman's situated understanding of human behavior. Dourish summarizes Suchman's critique of the cognitivist model of interaction design showing that

[7] Latour criticizes the symbolic interactionists by showing that symbols have no force of their own so where does the structure come from? He asks, "what do symbols hold on to? If the social is not solid enough to make interactions last – as examples from simian societies show – how could signs do the job? How could the brain alone stabilize that which bodies cannot?" (Latour 1996, p.234).

the organization of action is not a "formulaic outcome of abstract planning" but is rather an "ad hoc accomplishment" (Dourish, 2001, p.121). The paradox of system design is that understanding how a technology organizes a workflow removes the human negotiations that occurred heretofore out of the equation. The paradox of technomethodology is that the focus on the particular, localized interactions makes it a challenge for applying to "inventions of the future" rather it is only possible to see the here and the now on a small scale. This paradox is most clearly captured in the term, sociotechnical, and I will take this up later in this chapter.

Actor-network theory (ANT) emphasizes eschewing a priori categories when studying phenomena (Callon & Latour, 1981). Although I use ANT to emphasize the sociotechnical networks within data synthesis projects, I do not argue that objects or technologies themselves are agentic in the same way that human actors are. I take Star's ecological analytic point that people and objects, such as machines and instruments, are coextensive and "that the voices of those suffering from abuses of technological power are among the most powerful analytically" (Star, 1990, p.267). Star's ecological approach to the study of infrastructure is a critique of ANT for its lack of attention to other participants' interests and for painting too binary a picture of success and failure. For example, in The Pasteurization of France, Latour tells the story of Louis Pasteur and how the model of bacteria came to reign. Though there were alternate models of disease that people believed, in the end, everyone believed in bacteria (Latour, 1984). Rather than acknowledging the many actors that had been translated or subsumed into the story, the Pasteurians became the nodal point in which every other actor gets black-boxed, according to Star.

When engaging with the topic of scale in science, this ecological approach is important in that scale easily slips into the background. In his work on the ethnography of scaling, Ribes (2014) explored how an organization managed itself - looking to the documents they used to make sense of the scale of organization. While GEON, a geosciences research infrastructure, was the organization in question, he did not look at their object of research (e.g., geology). On the contrary, it was an exploration into the information technology, the spreadsheets, the calendars, and other logistical tools that helped actors collaborate. However, in my application of the ethnography of scaling, I analyze traces of communication left in the GitHub repository to uncover how actors manage the scale of their research object (e.g., in this case, salmon). This subtle distinction between subject and object is critical to the future of data ethnographic work.

Ethnography of scale in practice: How to employ ethnographic methods to understand issues of scale in scientific knowledge production

Ethnographic methods have been adapted to meet the demands of studying large-scale, distributed research sites. Large-scale initiatives such as the LTER Network, NCEAS, and GEON, are increasingly requesting ethnographic insights to understand how collaborative work is achieved. This is particularly the case in knowledge infrastructure projects. But ethnographically studying large-scale, distributed enterprises presents challenges to the more established ways of doing ethnography mainly because they are large in scope and ambition, collaborative across national and international locations, and interdisciplinary (Ribes, 2014). In addition to these challenges is the challenge of scaling phenomena in a

rapidly changing world and navigating the tension between what is novel science and what is mundane, needs-to-get-done science.

One way to understand how people make sense of large-scale initiatives is to follow their own "techniques and technologies for knowing and managing large-scale enterprise", or scalar devices (Ribes, 2014). Scalar device (Ribes 2014) is a conceptual tool for uncovering scale - or at least, how the actors of interest scale. Although the ethnography of scaling is presented as a methods paper, Ribes (2014) describes three instantiations of a scalar device as a concept: a) All-Hands Meetings; b) surveys and descriptive statistics; and c) benchmark metrics. The first two concern scaling 'the social', or those activities that involved humans, while the third involves scaling computing capacity illustrating the ways that scalar devices can be used to study sociotechnical enterprises.

While my methodological approach draws heavily from ethnography as it has been used in information science and STS, I also adapt it in a couple of a key ways. Rather than focusing on scaling up as Ribes' (2014) ethnography of scaling focused on scaling ethnography to study larger-scale phenomena and to scale large-scale organizations (e.g., GEON), I adapt his ethnography of scaling to look at how scientists instrument scale in the domain (e.g., the instruments used by scientists to understand their research phenomena, in this case, salmon). As such, this relies on some historical and archival work as well, bringing together contemporary concerns to those archived in the past.

Scientific instruments as scalar devices in the study of scientific knowledge production

Instruments have been widely theorized as playing an intermediary role between phenomena and theory (Baird, 2004; Bowker, 1994; Wise, 1995). Given my earlier point that the crux of the issue of scale in ecology is in the interaction between scales (and, the

emergent properties of scale), Baker's (2017) use of mesoscale is critical to my understanding of what happens in this interstitial space. Baker (2017) uses mesoscale to refer to "a relative scale, fit to purpose for the object and context of interest." In practice, this means that it is not sufficient to only focus on larger volumes of data, data that streams at high velocity, or large-scale repositories. On the contrary, many of the issues with scale occur during instrumenting the field.

STS has a long history of studying science in local sites of knowledge production, particularly in the laboratory. It was the laboratory that offered a field in which the STS scholar could explore knowledge production as a practice like any other form of social achievement (Fujimura & Fortun, 1996; Knorr-Cetina, 1999; Latour & Woolgar, 1979; Traweek, 1988). In the lab, Knorr-Cetina (1999) was able to see how objects are removed from their natural environment to be made available to investigation. As such, the laboratory scientist was able to bring the scale of planetary time into the "time scale of the social" (p.28).

While laboratory studies have offered up sites to produce knowledge about physicists, astronomers, molecular biologists (Knorr-Cetina, 1999; Traweek, 1988), little has looked at the actual instruments of science. In this section, I ask *how does the world go from object to process to collection and back to world? And, how are locally-specific, fine-scale data collections made universal?* Callon & Latour (1981) propose a general theory of translation to illustrate how macro actors are really micro actors built on black boxes[8]. For these actor network

[8] Black boxing is a wildly popular term from STS that refers to technological inner workings that get backgrounded or become invisible.

theorists, scale is a matter of translation and power. Scale is the product of scaling. In other words, there is no such thing as a watershed until you produce a watershed. Or, as someone in my field site commented: "*The salmon are complicated because we study them."*

As such, I began this research by exploring one of the main channels of communication for the data task force team—their Enterprise GitHub repository, a version control software used primarily by software engineers and computer programmers to manage information. In the way that Ribes (2014) traced the use of Gratia, a piece of software for developing benchmark metrics, I, too, have traced the use of GitHub, models, and data visualizations as a way of scaling Wild Alaska salmon and its ecosystem. This approach aligns with others' research on GitHub as a place to better understand collaborative work practices (e.g., Dabbish et al., 2012).

Following the actors' work around issues in data, I categorized their identification and reconciliation of errors and illustrated the major challenges for this particular kind of data organization. I develop strategies for understanding how scientists scale by following their research object. While I do rely on the instruments of science to understand, I was often in need for more participatory methods to acquire the information I needed. Anne Beaulieu (2010) has outlined this challenge in STS scholarship particularly with the growing move to technologically-mediated spaces. She proposes co-presence as an "epistemic strategy" for foregrounding the "relationship between self and other and interaction that achieves presence in a setting" (457). In her work on women's studies, she used a mailing list to participate in a way that established co-presence. In this vein, I sought to establish co-presence among my field sites by engaging with their interests. Specifically, I would write in the public slack channel my own thoughts about the data work practices or questions that

came up. I also volunteered to do a literature review for one of the eight working groups and presented the findings of this literature review back to the group. This led to questions about how they might make their own work more participatory, and ultimately led to my research in the Kuskokwim region exploring how scientists define local scale.

Scientific instruments are scalar devices that help expert communities make sense of phenomena that are larger than themselves. I define instruments broadly, as devices that segment phenomena to produce data. I argue that scientific instruments are a kind of scalar device used to scale down in studying large-scale phenomena (e.g., climate change) as well as to scale up to make sense of fine-scale, often invisible entities (e.g. diel migration of zooplankton). As I show in my research, an instrument not only facilitates collaboration across domains, **making universal what was once local**, but **it also extends across temporal and spatial scale, facilitating the production of new kinds of data**. This extension of scale produces something entirely new that could not have been produced otherwise. There are three different dimensions of instruments that make them important ethnographic tools for this kind of study: they are material; they are often explicitly used by experts; and their function to produce data about phenomena have a recursive quality (table 3).

Material / form	Histories of the production of instruments sheds light on the theoretical concepts and assumptions of the time. By looking at the material ways that phenomena are instrumented, the ethnographer can better understand how change occurs over time.
Expertise / know-how	Instruments play a role in the public and are often developed alongside hobbyist communities, local residents, and other experts. Using instruments as a tool to ask scientists about

	their own embodied expertise is a way of eliciting what is often tacit.
Recursive / networks	A process of mutual instruction occurs in the interaction between instrument and phenomena. These tools of translation help the ethnographer make sense of science as a process.

Table 3. Tricks for using instruments as ethnographic device

Material / form

By employing Garfinkel et al.'s (1981) 'first time through' approach, I look at the moments of negotiation to understand scientific objects before they become *naturalized.* The instrument provides a window into theoretical concepts and assumptions of science at the time. However, it also sheds light on societal values, regulatory nuances, and contested realities. Large investments have been necessary to invent many of our most lauded instruments, which often trace their roots back to industrial development, centralization of state power, and commercialization on a global scale. In Norton Wise's volume of essays on scientific instruments (1997), namely in the form of measurement, Ken Alder shows the trajectory of the meter, or the "cumbersome path by which the French went metric shows that it took decades for new units to become commonplace." In his example, precision is capitalistic, making measurements ever more comparable, centralized, and commensurable.

Wise's volume offers perspectives from state to commercial involvement. The essays illustrate the bureaucratic, centralized state origin of many scientific devices and standards. In one of the most prominent example chapters, we see the term "*today precision must be commonplace*" from a Bausch and Lomb's advertisement in 1944 which referred to the hardware of war they had depicted. However, Wise also argues that "they were associating the urgency and prestige of military power with their own place in modern scientific culture, with 'production line accuracy' of a ten-thousandths of an inch."

As I show in the next chapter on the history of salmon science in Alaska, much of the early development of research was motivated and funded by commercial interests. And instruments used to understand aspects such as water temperature can trace origins to military endeavors. For example, the bathythermograph, a device used to measure water temperature at different depths, traces its history back to WWII submarine warfare, promising to keep submarine tanks out of sonar range of "the enemy".

As the state of knowledge changes, so too does the instrument. In response to Karl Popper's epistemology, or Positivism, David Baird argued that instruments contain epistemic content unlike and alike the content of theories and often sustaining beyond the death of some theories. He notes that "measurement presupposes representation, for measuring something locates it in an ordered space of possible measurement outcomes."

Baird's contribution is that he moves knowledge away from being located within ideas to being located within the artifacts themselves. For philosophy of science which had long-espoused a theory-centric view of data (Leonelli, 2016), particularly in a rapidly changing technological space, this was an important shift away from the major focus in studies of science to look at concepts and theories rather than the tools through which theory was developed, tested, and applied. Where Baird argues that theory is implicit in instruments, Ian Hacking (1983) argues for a separation between theory-realism and entity realism where theories can be true or false and entities are real if they exist, or as he put it: "if you can spray them [electrons], they are real" (p. 22). Essentially, Hacking's move is to turn away from focusing strictly on theory as the critical component to science and to instead look at the objects (representations) and the way that objects are manipulated (interventions). Baird, however, makes a crucial point: that although he segments the different epistemic

parts of an instrument, "the whole is not simply the sum of the parts…the instrument presents an epistemic synthesis, seamlessly joining representation and action to render information" (p.70).

Processes and practices

In addition to being imbued with epistemic content, scientific instruments are contingent upon expertise of use. One aspect of expertise is the inconsistent reliance on tacit knowledge, sometimes occupying a privileged place on stage with science while other times being downgraded to the "basement of the lab." Related to her work on boundary objects, Susan Leigh Star argues that "materials and tools are the detritus of the work, often written out of scientific accounts" (257). Highlighting the level of craft undergirding scientific instrumentation, Star shows how the taxidermists—once a part of hobbyist communities—were absorbed into scientific research endeavors in the early days of natural history. She demonstrates how the tedious and meticulous practices of the hobbyists were taken up by biologists only to then be moved out the sciences once deemed too inconsistent. As science became more modern, it pushed "to erase individual, craft skill from the scientific workplace, to ensure that no idiosyncratic local, tacit, or personal knowledge leaks into the product" (275). She offers two reasons for the dismissal of craft in modern science: a) as biology moved into being a "big science", the craft skill that could not be "industrialized, formalized, or assimilated" was cast as merely decorative. And, b) as representations in biology "became more abstract and quantitative as well as more experimental, the meticulous preservation of specimens for study shifted away from the public display" (275). The move in science from

concrete to abstract is something we will see again in Latour's work in the Boa Vista rain forest.

For Star, the material artifacts of scientific instrumentation are mundane, backgrounded, and menial while the practice points to the historical development and instantiation of a field including the standards, training, and calibration necessary to be *legible* to science. However, the instruments of science also can move into other spaces not necessarily tied to science. Showing the movement of hobbyists in and out of modern science, she illustrates the ways that the move to "sanitize" science in an attempt for more objectivity obscures "the messy face of science."

Related but distinct from embodied expertise in hobbyist communities, Goodwin's Professional Vision shows how expertise shapes the way others see the world. He describes three key discursive practices: coding, highlighting, and the production of graphical representations. Discursive practices refer to the way members of a profession shape events. Through an analysis of the way video was used in a courtroom in the Rodney King case, he illustrates how what was deemed "objective" evidence of police brutality was contested by police lawyers who argued that their – the police officers' – professional vision actually justified the behavior. While this is one way that professionals divide themselves from the public, it is also a tactic for creating distance to say that the profession itself has the type of expertise that allows for vision. In Goodwin's writing, it is not just the instrument that extends human perception, but it is the discursive practices by which experts or professionals can support their expertise. These representational instruments have agency in that they shape the way to see the world, or as Goodwin notes, graphical representations are embodied practices, "sets of perceptual structures, the ability to see what and where to

measure" (Goodwin 1994, p. 615). Vertesi (2012) picks this up to look at how this type of "vision" is embodied interpreting embodiment quite literally to be the actual physical, bodily movements of the technicians moving their bodies like the robots and feeling pain in a possibly psychosomatic way.

The connection to urgency and bureaucracy as well as the material affordances (or accomplishments) of different instruments is apparent in Joanna Radin's work on cryopreservation (2013). Focusing on the technologies of cryopreservation, Radin shows how in the Cold War era, access to cold storage technologies (e.g. dry ice and mechanical refrigeration) renewed requests to salvage blood for its potential usage downstream. In this way, it was both national rhetoric and material capability that facilitated cryopreservation.

These deployments of expertise and embodiment point beyond the materiality of the instrument itself. While national and international scale state logics and precision measurement achieve accuracy and objectivity, expertise and craft undergird the development and deployment of scientific instruments sometimes disrupting the illusion of objectivity. As more terrain is instrumented, claims to objectivity are used to validate findings. However, in resistance to the move to automate, human expertise is used to make claims to what should and should not be instrumented. In this way, there is a privileging of certain kinds of expertise while a devaluing of others.

Networks: Recursive intervention

Thus far, I have argued for viewing scientific instruments as material in that they can shed light on theory and afford certain phenomena physically and conceptually; and, I have argued that practices and expertise are leveraged in myriad ways to make claims to

knowledge. As the third aspect of the instrumental layer of science, I argue that instruments—like infrastructure—are relational. They produce emergent and recurrent interactions downstream of initial usage. See figure 8.

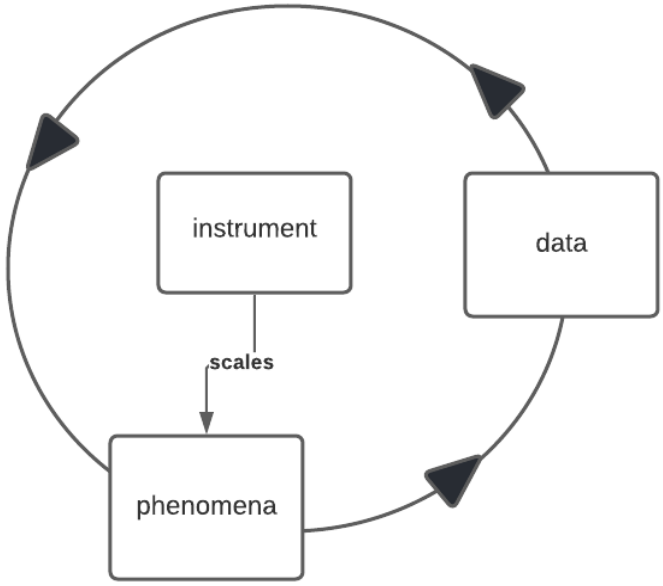


Figure 8. Instruments' role in scaling phenomena to produce data

Instruments play a critical role in producing scale. Primatologists, for example, use tools to create data. They not only study, but also create categories such as structure, rank, order, families, and caste, with their instruments. Shirley Strum, a primatologist, illustrated this in her work. Strum's job was to go out into the field and use forms that were given to her by her advisors. Upon observation of the monkeys, she found the forms not useful as they were constantly shaping her view. Through observation, she discovered that the monkeys engaged in constant negotiation and that the alpha/beta labels were hardly a reflection of reality. On the contrary, the labels were devised to talk about the apes. The only way to produce the idea that chaos is structured is to add a layer to reveal or create this structure.

Hacking evidences the interstices or the interaction which demands a need for a tool. He explores the ways in which scientific knowledge is not understood sequentially but is understood in the interaction between representation and intervention. In other words, science is not a linear succession of progress, but is relational. Resisting the urge to relativism, he notes: "once we have an extensive ability to exploit the putative causal properties of a theoretical entity, we have good reason to conclude that the entity is real, not hypothetical. It is not that we build the device and conclude that electrons are real; we have already so concluded before we built the device." In other words, the device alone does not suffice, but that the interaction between the entity and theory (through the device) is evidence of it being real in the world. As such, scientific knowledge is both relational to the instruments of studying science as well as built into the tool itself.

Rheinberger (2005) (via Bachelard's notion of historical epistemology and recurrence) also notes the temporal trajectory of scientific objects. Taking the concept of science as processual a step further, he argues that the instrument is the intermediary between theory and the world, distinct from technology: "The instrument represents the material existence of a body of knowledge. The phenomenon is provoked as a problem at the knowledge horizon and may itself require new concepts in order to be accommodated. Phenomenon and instrument, object and scientific spirit, concept and method are all joined in the process of mutual instruction" (p. 320). The main point is that in additional to material form of instrumentation and the expertise inherent in the history of instrumentation, there are also relational aspects to the instrument that cause downstream or latent effects.

I have highlighted how scientific instruments help conceptualize movement from phenomena to data. I have shown how they have material histories, which speak to the

values of the time; they are imbued with epistemic content themselves; and, their physical form elide certain phenomena while making others possible.

Conclusion

This chapter summarized my research design and gave an overview of the different strategies I took to understand research infrastructure that is distributed spatially and temporally. This chapter proposed a novel usage of an existing ethnographic approach, which I combined with more traditional ethnographic work.

As I entered into a world of stabilized facts, I fumbled to study these different fields. And, my fieldwork contrasted sharply in the three different sites. At NCEAS in Santa Barbara, California, there were data discussions with which I had already familiarized myself. However, once I arrived to Alaska, the fields changed. I was constantly faced with my own ignorance about the lineage of instrumentation, theories, and standards. My ethnographic experience working as a field technician alongside other researchers highlighted the significance of long-term observations as well as the limits of human observation for saying something meaningfully spatially and temporally.

The three different sites that I engage with represent different relations to scale and data. The first concerns scientific programmers' work to clean and archive data. The second is concerned with ecologists as they instrument temporal scale at different resolutions. And, the third explores how quantitative ecologists and natural resource managers employ the human as an instrument of local. In the next chapter, I outline these three field sites in greater detail to provide more context about how they are connected. The following work highlights some of the data I collected, the things I learned, the relationships I formed, and the fish I counted.

Ch 4 Science on Wild Alaska Salmon

Field legacies: From commercial interests to ecological field site to data repository

"there is no escaping the persistence of the past. … Failing to accept that indebtedness to the past, or to realize how diverse and contradictory that past has been, we will not make much headway toward a deep understanding of our current ideas about nature" -Donald Worster writes in Nature's Economy (1977)

In this chapter, I provide context for the three empirical chapters that follow. To develop a theory of how scientists scale in practice, I explore three sites and three different types of challenges with scale therein. This section introduces those sites and explains why I selected these three for answering my overarching questions.

Scale in sociology, scale in ecology, and scale in information science are vastly different types of scale. While achieving spatial and temporal scale has been a long-standing challenge for ecology, the concept of scale for the management of information is focused on managing the size of an information ecosystem, moving between local particularities and universals. **This is what makes the study of ecoinformatics challenging: the social actors themselves are contending with scale in terms of information systems and scale in terms of ecosystems.**

On the one hand, there are often mundane issues for ecologists to consider such as instrument failure in the field, gaps in research funding, short time cycles of researchers collecting data, and diversity of sampling and collection methods across field sites. On the other hand, there are also profound issues relating to scale that involve balancing priorities, configuring infrastructural elements, arranging governance, and aligning within the political as well as the digital landscape.

To understand scale in practice, I first explore the State of Alaska's Salmon and People (SASAP) project as a specific approach to understanding change over long time scales. Analyzing the scientific programmers' work through the lens of the data lifecycle model highlights the many actions that fall outside of the idealized data lifecycle. Using this lens, I identify field sites, or sites of data collection, as a primary area where understanding the lens by which scientists collected data is crucial to the data integration at hand. While the scientific programmers are far removed from the source of data production, scientists in the two empirical chapters that follow are closer in proximity to the data source.

Theoretically, I connect these with the concept of sedimentary legacy (Hirsch et al., 2021). In their work on the legacy of research infrastructure, Hirsch et al. (2021) show how initial design decisions about data infrastructure have downstream or latent effects on the present-day operation of those infrastructure. Furthermore, they show how these legacies shape how infrastructure can change.

The idea of legacy infrastructure comes out of information systems research from the 1990s (Bisbal et al., 1999; Weiderman et al., 1997) to emphasize the way standards or decisions made in information systems often resist future change. As Star (1999) has argued, there is the legacy of the installed base. This is the idea that there is inertia to decisions that were made when developing an infrastructure. Similar to the concept of the installed base (Star and Ruhleder, 1996) in infrastructure (that there is inertia to what becomes standard) so too do ecosystem dynamics play out.

Legacies evoke a long past making an appearance in ecological research in the 1990s as well. This is used as a way of describing impacts on an ecosystem that last long after the proposed cause or past conditions. This is closely tied with research on human land-use

impacts. Monger et al. (2015) use 'legacy' to refer to multiple scales of effects: soil legacy, geomorphic legacy, and ecological legacy. And more relevant to this study, Lichatowich (1999) discusses the "evolutionary legacy" of salmon:

> *the salmon had to survive a rough trip through evolutionary time. They had to continuously adapt to a landscape with high levels of spatial and temporal diversity. Not only did the landscape change through time, but at any particular time it was also composed of a patchwork of diverse geologic, climatic, and biotic conditions…. The variation in the salmon's life history is an important legacy of their survival and evolution through the geologic and climatic events that marked the history of the Northwest (p.20-2).*

Legacy is used less as a theoretical point and more as a sensitizing concept to reference feedback between past occurrences, duration since the occurrence, and how sensitive a system is to perturbations. Its usage in ecology differs from information systems in that research into ecosystem dynamics uncovers legacy effects of past events; whereas, information science's characterization of legacy is typically of an old computational system being migrated to something new and resisting that migration.

It is also an obstacle to scaling as it prevents extension and often requires an entire overhaul to carry forth. As such, sedimentary legacy (Hirsch et al., 2021) acts as a sensitizing concept to "draw attention to the resources rendered available by a long-term research infrastructure and the impact that these resources have on the ability to inspect emerging objects of investigation." By showing how an infrastructure for ecology (LTER) was constrained by older decisions about data, specimen, and instrumentation, they highlight how an infrastructure struggles against its sedimentary legacy to bring forth change. While this concept is useful for thinking with a bounded infrastructure (e.g., LTER, NSF, NCEAS), I employ the concept of legacy infrastructure to unpack the larger historical

and material changes that can be seen in the present-day operation of my three field sites (figure 9).

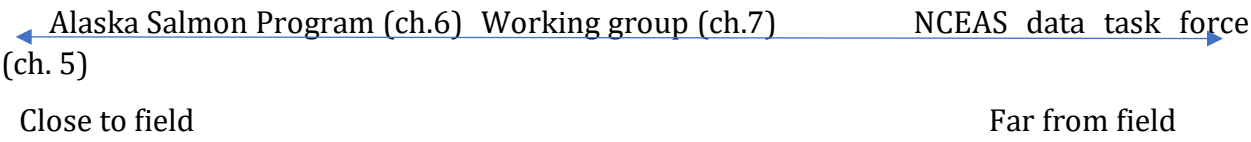


Figure 9. Three different field sites based on proximity to the field

Past endeavors have shaped the field sites that I draw from in this research. By providing a brief history of salmon science in Alaska, I situate this study in a longer history and demonstrate how knowledge has been shaped by specific events. In the next section, I provide a history of the data organization, NCEAS, and discuss my introduction to the data discussions at stake.

Research infrastructures for ecological science and the State of Alaska's Salmon and People (SASAP)

To understand NCEAS' role in Alaska salmon requires exploring the antecedents of its creation as an organization. NCEAS was founded on a principle that attending to data-intensive work such as archiving and synthesizing data could aid ecological science. And today, it represents a modern infrastructural approach to dealing with the aforementioned challenges of scale.

The open science and big data movement has been not only a philosophical debate about inductive versus deductive reasoning, but also open science has been considered a political value for scientific endeavors — particularly those funded by national agencies (Obama 2013). As one journal encourages: "if you have useful data quietly declining at the bottom of a drawer somewhere, I urge you to do the right thing and send it in for publication--who knows what interesting discoveries you might find yourself sharing credit for!" (Lawrence, 2013).

The center's creation story began with the perception among ecologists that important research themes span wide regions and long time periods. This is also a justification for organizations like the Long-Term Ecological Research (LTER) networks, which began in the mid-1980s. Callahan (1984) argues that short-term funding cycles fail to support the long-term research needs and notes that precedents for long-term research had been set with organizations like national parks, wildlife preserves, national laboratories, and the Federal Committee on Ecological Reserves. For example, research on marine fisheries must consider information from wide stretches of the ocean, and studies of long-lived forest communities must span decades. Recognizing that research in such areas cannot be accomplished by a single scientist (nor even a single scientific organization) is part of what has motivated NCEAS' mission.

National Center for Ecological Analysis and Synthesis (NCEAS)

As I have shown, NCEAS has had open science and data synthesis as its motto since they began in 1995. In the 1990s, the National Science Foundation (NSF) established the need for a center whose mission was to foster synthetic research using existing data. In 1994, the NSF

solicited proposals to build such a center. The award (of $12.5 million) was made to the University of California, Santa Barbara under PIs Bill Murdock and Michael Goodchild to establish and operate the center for 5 years. Subsequent grants were awarded to UC Santa Barbara to continue NCEAS in 2001 ($16.6 million) and 2006 ($18.5 million).

The aim to create a data organization that would span decades came out of a critique of the scale of data. Ecologist, Peter Kareiva, found that "half of the field experiments in population dynamics were done on plots a meter or less in diameter." And, in 1991, James Reichmann, program officer at the NSF, asserted that "ecological research problems are inherently multidisciplinary, requiring the efforts of biologists, engineers, social scientists, and policy makers for their solution. Hence, there is a need for sites where a longer-term, multidisciplinary analysis of environmental problems can be undertaken." This view was well underway perhaps most clearly captured in the international biophysical program (IBP). This problem has been addressed before which partially led to NSF's creation of the 18-site Long-Term Ecological Research (LTER) network in the 80s. However, LTER has been criticized for not comparing data across sites (Stone, 1993).

Further, the major impetus for ecological synthesis is the growing awareness that a rapidly changing climate requires a more globally integrated perspective. In other words, there is a growing emphasis that aggregating data across temporal and spatial dimensions will help better comprehend large-scale change.

My time at NCEAS

I came to NCEAS as they hosted a 3-year Gordon and Betty Moore Foundation project called SASAP. In a warm, sunny January in Santa Barbara, the NCEAS group embarks on its 21st year as a center. The new director, Ben Halpern, welcomes a group of interdisciplinary

scientists to begin synthesizing research to understand the state of Alaska salmon. Despite the number of scientists in the room that day and the number of scientists who have come through NCEAS' doors throughout the years, many scientists have concerns about making their data public. These concerns include getting scooped, data manipulation (for nefarious and otherwise innocent reasons), and lack of contextual understanding and expertise to adequately analyze and interpret the data.

Sitting in a room of researchers (figure 10), I participate in discussions about the kinds of data to acquire to answer myriad research questions. In these discussions, researchers across disciplines talk about issues of scale, disagreeing about the temporality of certain disciplinary time scales or discussing the challenge of spatial accuracy in older datasets.

Figure 10. SASAP working group meeting at NCEAS headquarters in Santa Barbara, CA. Photo by Jorge Cornejo.

One of the advisors from the Alaska state department notes that the department is divided into four regions, which each have their own age, sex, length[9] (ASL) database, analyst, and biologist. The ability to query specifics and download raw data varies between regions. He stresses the importance of not requesting an entire "data dump" from the

[9] ASL is a type of data collected about salmon.

department, but that there's "*gotta be some criteria – an initial screening to select the data in a kind of hierarchical manner rather than downloading the entire package."* One of the data scientists disagrees arguing that it is *"easy to take the whole set and then filter through based on your priority. If it's already gathered together in a database then by far, the best way to do this is to just keep it together in data processing. We can write the scripts afterward."* However, the state manager pushes back, arguing that "*without understanding where the data came from, it's completely false information. If it falls into the wrong hands, they will be led the wrong way."*

This represents a common disagreement in the big science and open data space. One view is that errors in data can be resolved with data processing while the other is highly concerned with misinterpretation. An example of the kind of consequences of misinterpretation that are concerning is data about the Pebble Mine.[10] The Pebble Partnership selected shorter time frames within which to study salmon productivity in tributaries that would be impacted by the proposed mine. However, as Schindler noted in his presentation to the Alaska State Legislature House Resources standing committee (April 2019), short-term assessments can misrepresent the long-term viability of the habitat. The 2-3-year time frame selected by the mining company is a poor indictor of potential habitat to produce fish over time. Rather than being about single productive or non-productive tributaries, portfolio diversity stabilizes the

[10] Since 2008, the Pebble Mine, a proposed copper mine, has been a specter haunting the Bristol Bay watershed. As early as 1987, Cominco Alaska Exploration discovered anomalies in the Pebble site and drilled exploration holes in 1988. Reviews of the Draft Environmental Impact Statement (DEIS) in 2019 led many experts to conclude that the data selected by the Pebble Limited Partnership was insufficient to answer questions about long-term damage. The final EIS has been deemed a product of a broken process with a lackluster economic plan of the mine, minimal information on how to handle geochemical risks, and an insufficient mitigation plan for long-term damage.

aggregate (Schindler et al., 2010). In other words, the mosaic of habitat that is constantly shifting through time requires a less linear approach to understanding change.

This debate about the role of data in the sciences is happening as computing and data science are becoming more mainstream. As NCEAS casts the mundane, everyday work of data management as novel, they receive funding to do some of the necessary work of scientific support. This is because organizations focused on best practices for building data infrastructure are forced to cast their work as novel when it mostly pertains to everyday data management. This is due to the mismatched understanding between how much mundane work is required of science and how much funding is allocated to doing this data work. In other words, to segment the world into different scales for adequate scientific investigation, there is not only conceptual work that is required but also technological, routine data work—both the production of data as well as the management of it.

While data managers sometimes fail to highlight the labor of producing data (e.g., walking streams, cutting up fish, collecting scales), domain scientists overlook the labor of getting data ready for usage (e.g., applying standards of storage, interoperating to fit multiple data points together, filling in gaps and reconciling anomalies).

NCEAS is evidence of an increased focus on and funding for the production of synthetic data sets and data integration initiatives. Furthermore, the preferred way to deal with data are still hotly-debated as concerns about data synthesis initiatives or universalist notions of data are focused on the potential for errors to be propagated downstream when there is data misuse, both in the improper archival of such data and in a misinterpretation due to misunderstanding about the context of data. Other concerns include the view that data synthesis initiatives are unjust to labor requirements in that in elevating the role of the data

scientists, the data producers' labor is reduced. And finally, there is a view that synthesis initiatives lack future scalability in that they lead to large amounts of data archived in the present moment; however, these often do not extend into the future making the archived dataset already obsolete.

To put NCEAS in the center on this dissertation would be disingenuous, however. While NCEAS (and SASAP) was my point of departure, the issues that the research in SASAP surfaced were issues much larger than modern approaches to environmental data science. As I introduce the two other sites this thesis focuses on, I show how data are dealt with and how scale is instrumented in longer-term research endeavors.

History of salmon science in Alaska: Alaska Salmon Program

Knowledge is dynamic. A legacy lens illuminates how knowledge is situated in a larger political, technological, and cultural era. This is important to the topic of scale as the discussions I track are often discussions about what standards to apply, how to make data interoperable, and what scale at which a study should be designed. These questions have a direct connection to the legacy of science, particularly in Alaska, as much of the science began due to a commercial interest in harvesting the salmon. In this section, I provide an historical account of salmon science in Alaska to highlight the way commercial interests have underpinned early attempts to understand salmon.

Just over a century ago, little was known scientifically about the life histories of the iconic and now heavily-researched Pacific Salmon. In fact, it was generally accepted that salmon migrate very short distances, if at all. This lack of understanding is evident in early writings

on salmon science as the nation's lead ichthyologist and Stanford University president, David Starr Jordan[11] noted:

> *the descriptive literature of the Pacific salmon is among the very worst extant in science… We fail to find any evidence of [mature fish returning to their spawning grounds] in the case of Pacific coast salmon, and we do not believe it to be true. It seems more probably that the young salmon, hatched in any river, mostly remain in the ocean within a radius of 20, 30, or 50 miles of its mouth" (Jordan, 1887).*

As I will show in this potted history of salmon science in Alaska, greater complexity is not taken for granted in even widespread understanding of salmon life histories.

Catastrophe and commercialization are perhaps the two most important legacies that led to understanding salmon in Alaska. The development of the commercial salmon fishing industry coupled with catastrophic declines in salmon runs are central to much of the research on Alaska salmon.

Throughout the 1700s, many outsiders moved into Alaska in search of gold, otter pelts, and salmon. Much of the "exploration" of Alaska has been driven by a desire for infrastructure such as shipping routes or pipeline routes (Haycox, 2020). And, in the 1800s, many Russians occupied the Bristol Bay region for participation in the maritime fur trade. In search of a northern shipping route between the Atlantic Ocean and the Great South Sea, now known as the Pacific, Captain James Cook named Bristol Bay in 1776. Less than a century later in 1867, the United States Secretary of State, William Steward, purchased Alaska from Russia for $7.2M in a deal infamously referred to as "Seward's Folly" or "Seward's Icebox".

[11] Lulu Miller (2020) has outlined the life of David Starr Jordan in her book, *Why Fish Don't Exist*. He was a renowned scientist who found solace in nature and even broke away from his mentor's doubts about Darwinian evolution. However, he also has been associated with the iniquitous theory of eugenics as well as his involvement in the cover-up of the murder of Jane Stanford, co-founder of Stanford University.

Seward, scorned for his purchase, is now the namesake of a town that is not only the historic start of the Iditarod but also a tourist destination.

Ecological science played a critical role in Alaska since the early 20th century. Much of this ecological science in Alaska has been supported by research infrastructure initially built for industrial development. For example, ecological research on the Toolik Lake Field Station in northern Alaska began after the "haul road" (now Dalton Highway) was put in by the Alyeska Pipeline Service Company[12]. Oil discovery at Prudhoe Bay created not only the rationale for a new biome program in the Arctic but also the physical infrastructure with the construction of the trans-Alaskan pipeline, which ran from Prudhoe Bay to Valdez (Kingsland, 2021). Early research in these regions were facilitated by access to remote areas, and even sometimes provided opportunistic experiments (Chapin & Shaver, 1981).

This is not a story about oil and gas relations, but about the entanglements between the commercial fishing industry, scientists in a research site, and the actual fish themselves. A recurrent theme in this history is the close relationship that the commercial fishing industry has had with scientists seeking to understand salmon biology and salmon ecosystems. While it is tempting to see this as the corrupt role of industry in western science – of which there are many accounts - the Bristol Bay sockeye salmon fishing industry is a different kind of industry. Its difference is in the dependence industries have on the environment around them. Rather than a purely extractive endeavor, the commercial fishing industry – having turned the survival and capture of salmon into a commodity – are heavily invested in ensuring the "resource" is harvested sustainably.

[12] For more detail on this history, see Kingsland's (2021) research on the cold war origins of the LTER site in Alaska.

Introduction to the field: What made Bristol Bay home to one of the largest salmon fisheries?

Shaped by volcanism and glaciation, Bristol Bay is in southwest Alaska (figure 11) and is a convergence of coastal Aleutian, low-Arctic, interior-boreal, and Pacific coastal flora and fauna. Or in other words, southwest Alaska is characterized by biological diversity supported by landscape diversity providing habitat to not only the black bear so synonymous with Alaska wildlife, but also Dall's sheep, Loons, Belugas, and, pertinent to this study, Pacific salmon. The salmon are what have brought me to this place, but why have the salmon come to this place?

The answer to that question is likely habitat. Bristol Bay, in the easternmost corner of the Bering Sea, has been formed from the outflow of large rivers. Situated between the Pacific and North American tectonic plates, its geologic composition along with glaciation following the Pleistocene ice ages has created rivers and streams that provide habitat to a large diversity of species. This habitat diversity has led people to venerate the region for its wildlife, in particular, its globally-renowned sockeye salmon runs. Not only seen as a place with great diversity, the idea that it should be preserved is evident in many of its descriptions as one source describes it as: "a veritable living natural museum that is home to one of the world's last great wild fisheries, the red salmon fishery" (Branson, 2007). In this quote, it is clear that not only does the region represent something no longer in other parts of the world (e.g., the last great wild fishery) but also a 'living natural museum' – a place that is so preserved that others might view it as an artifact of times past.

This place-based focus is important to highlight because it has implications to scale. The number of fish that return to this region have made it the commercial fishing hub that it is

today. As such, this habitat diversity is a component to why research has been conducted on these spawning grounds.

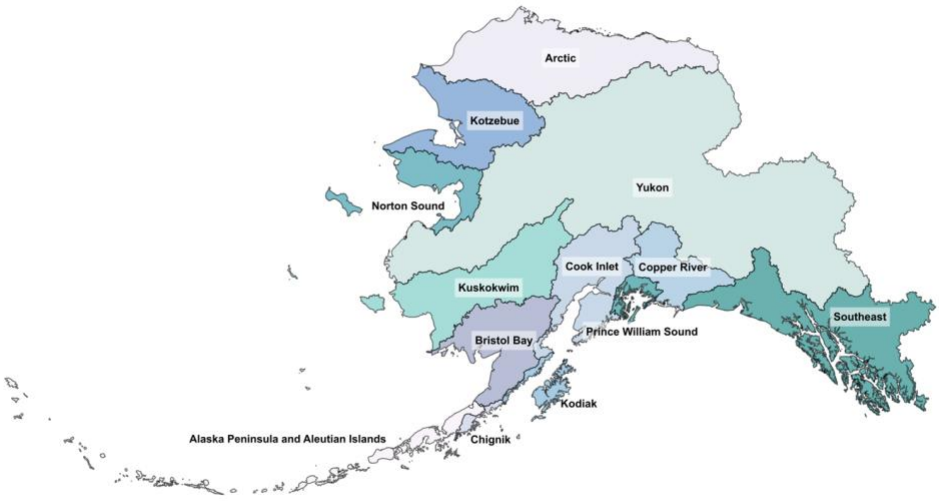


Figure 11. Map of Alaska. The Kuskokwim region is located in the Southwest corner of Alaska right above Bristol Bay. Image by Jared Kibele and Jeanette Clark. 2018. State of Alaska's Salmon and People Regional Boundaries. Knowledge Network for Biocomplexity.

Rural and disconnected from the Alaska road system, it is no surprise that it is considered pristine among environmentalists, sport fishing enthusiasts, scientists, and tourists. However, this area has been fished for centuries first by Indigenous — mostly Yup'ik, Athabascan, and Aleut — people and then by settlers post-colonization[13]. Critiquing the term

[13] Karen Hébert provides a detailed account of how the naming of Indigenous peoples have transformed identity in the region. She notes that while the language map strips away complexity and adds unrealistic boundaries, it does make the case for expanding from a view of Native Alaskans as a monolith (Hébert, 2008).

'final frontier', Cronon--in Arnold's (2009) book about fishing in Alaska--refers to Alaska as a 'resource frontier', a place subject to boom and bust cycles based on external demand for its natural resources: fur, minerals, oil, and fish. As industries took hold in a globalizing Alaska, most destructive to salmon ecosystems were the effects of mining as well as timber runoff and erosion. Mining debates still rage on today, most infamously with the Pebble Mine. However, much opposition has been leveled at this contentious mine, largely due to the interests in salmon that are so characteristic of this region.

These fishing interest groups are a long-standing legacy of the importance of salmon in the region. In 1877, the major commercial industry in Bristol Bay was salmon fishing, and this began with the first salmon cannery opening in Nushagak Bay (VanStone, 1967). In a couple of decades, Alaska fisheries were canning more than any other region on the Pacific coast. By 1920, there were 25 canneries in operation across Bristol Bay (VanStone 1967) and uncoincidentally, declining run sizes. If fisheries were boom/bust cycles, this had been a boom with fishermen averaging about 120-140k pounds of sockeye every year (AHS) and total annual harvests equaling about 15 million pounds.

These incredible harvests are largely due to the lack of regulation in the area up until the White Act of 1924. Prior to 1924, there had been minimal regulation of salmon fishing with the first federal legislation passed in 1889, banning obstructions to salmon streams. In the period between the late 1800s and 1924, there was a patchwork of attempts to regulate salmon fishing.

The White Act came in response to a crisis of low salmon returns, a devastating pandemic that led to a massive loss of human lives, and a growing realization that a better understanding of salmon populations was required. Discussion of this conservation act

ultimately resulted from a feud between Dan Sutherland and Herbert Hoover over how to conserve the salmon populations. Hoover was cast as part of a larger conspiracy with major players like Congress, fisheries biologist Charles Gilbert, and the canners constituting "a powerful oligarchy in salmon management" (Taylor, 2002) while Sutherland saw true conservation as something that focused on social justice and the equitable distribution of salmon; the two – Sutherland and Hoover – represented a binary decision to make between equity on one side or efficiency on the other (Taylor, 2002).

While the White Act went in the direction of efficiency over equity (Taylor, 2002), it was the first time Congress weighed in on regulation at the level of restriction "empowering the Secretary of Commerce to set seasons, restrict gear, and delimit space" (p. 385). This was also the first time that Congress began to listen to science and the concerns about reproduction leading to the infamous 50% goal for escapement to spawning grounds. Prior to this point, little was known about salmon biology. While many of the scientists at the time laughed off long-distance migration, referring to migration as something that only birds do, some began having doubts about this mainstream view. Charles Gilbert[14], David Starr Jordan's student – after being sent to Bristol Bay as part of an Alaska Salmon Commission – began to espouse an "alternative theory" arguing that perhaps each stream had a separate supply of fish, "a run consisting of fish which had been spawned in the stream." During Gilbert's career, he developed better understanding of the local variations among different salmon populations "exploiting the knowledge of regional variations in scale growth

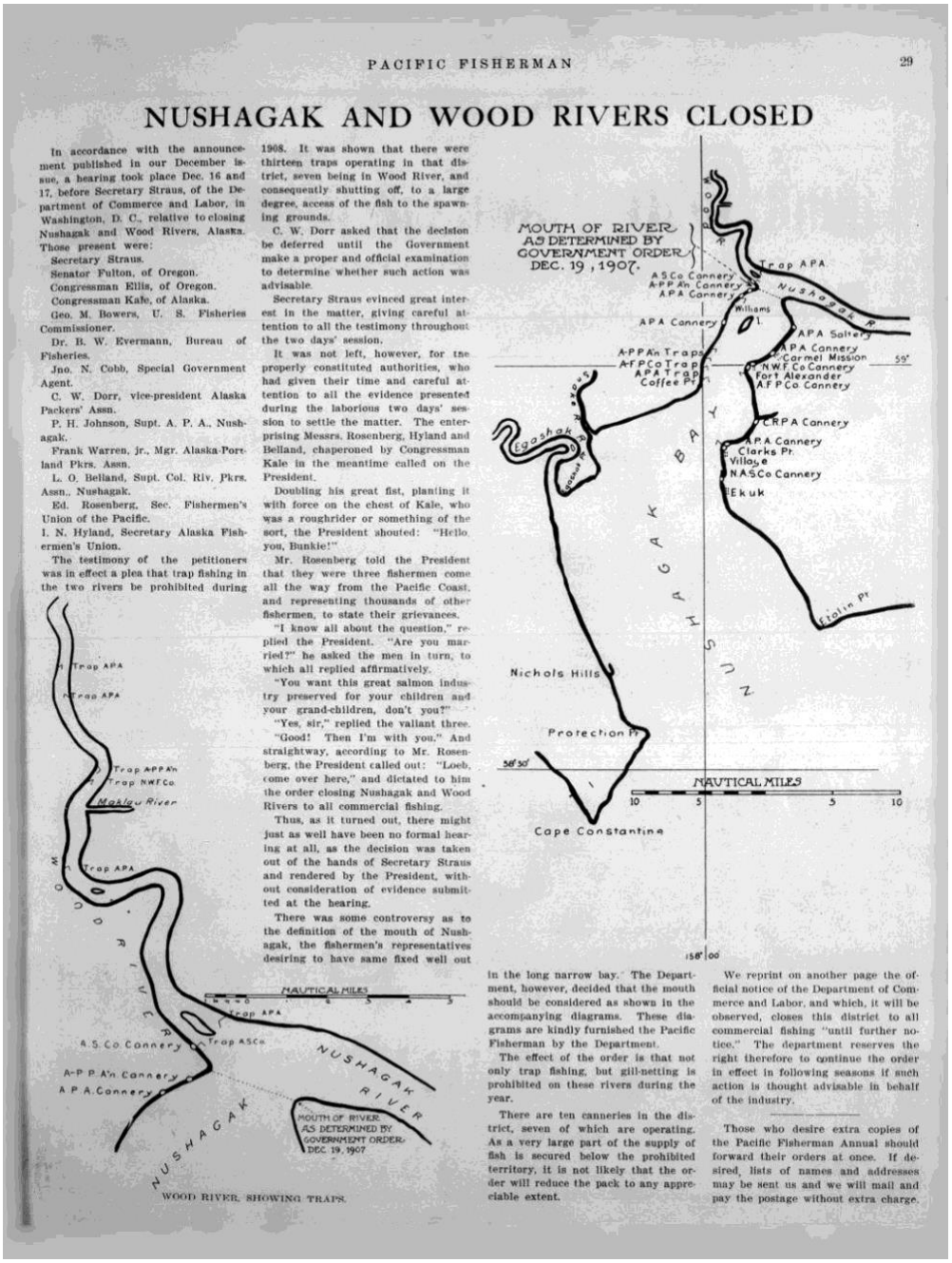
[14] Gilbert would later come to have an important impact on forwarding research on Pacific salmon and eventually would break from his professor.

patterns … to estimate the continent of origin of salmon caught on the high seas" (Pearcy & McKinnell, 2007, p.14). Gilbert and his student, Willis Rich, made important contributions illustrating how salmon home to natal streams, which lead to a fundamental research question around how such homing was accomplished.

Since that time, science has developed a more nuanced understanding of salmon life histories. And, those life histories have played a critical role in Alaska's history. The story of Alaska is one of fish and political will in that the state was founded in part so that Alaskan natives might control their fisheries rather than the federal government managing from afar. This is critical to understanding how scientists instrument scale as much of the context to data synthesis on Wild Alaska salmon contends with data produced by state and federal managers.

Creating a field: How commercial interests and declining salmon runs led to the creation of a long-term ecological field site

After the formation of the Alaska Fisherman's Union in 1907, Roosevelt took action to close fishing in the Wood and Nushagak rivers (figure 12), which according to King (2020) "presented an opportunity for scientists to gather some of the first biological escapement information ever collected." It would, however, be another 40 years before the University of Washington officially began to study the Wood River system.



PACIFIC FISHERMAN 29

NUSHAGAK AND WOOD RIVERS CLOSED

In accordance with the announcement published in our December issue, a hearing took place Dec. 16 and 17, before Secretary Straus, of the Department of Commerce and Labor, in Washington, D. C., relative to closing Nushagak and Wood Rivers, Alaska. Those present were:

Secretary Straus.
Senator Fulton, of Oregon.
Congressman Ellis, of Oregon.
Congressman Kale, of Alaska.
Geo. M. Bowers, U. S. Fisheries Commissioner.
Dr. B. W. Evermann, Bureau of Fisheries.
Jno. N. Cobb, Special Government Agent.
C. W. Dorr, vice-president Alaska Packers' Assn.
P. H. Johnson, Supt. A. P. A., Nushagak.
Frank Warren, jr., Mgr. Alaska-Portland Pkrs. Assn.
L. O. Belland, Supt. Col. Riv. Pkrs. Assn., Nushagak.
Ed. Rosenberg, Sec. Fishermen's Union of the Pacific.
I. N. Hyland, Secretary Alaska Fishermen's Union.

The testimony of the petitioners was in effect a plea that trap fishing in the two rivers be prohibited during 1908. It was shown that there were thirteen traps operating in that district, seven being in Wood River, and consequently shutting off, to a large degree, access of the fish to the spawning grounds.

C. W. Dorr asked that the decision be deferred until the Government make a proper and official examination to determine whether such action was advisable.

Secretary Straus evinced great interest in the matter, giving careful attention to all the testimony throughout the two days' session.

It was not left, however, for the properly constituted authorities, who had given their time and careful attention to all the evidence presented during the laborious two days' session to settle the matter. The enterprising Messrs. Rosenberg, Hyland and Belland, chaperoned by Congressman Kale in the meantime called on the President.

Doubling his great fist, planting it with force on the chest of Kale, who was a roughrider or something of the sort, the President shouted: "Hello, you, Bunkie!"

Mr. Rosenberg told the President that they were three fishermen come all the way from the Pacific Coast, and representing thousands of other fishermen, to state their grievances.

"I know all about the question," replied the President. "Are you married?" he asked the men in turn, to which all replied affirmatively.

"You want this great salmon industry preserved for your children and your grand-children, don't you?"

"Yes, sir," replied the valiant three.

"Good! Then I'm with you." And straightway, according to Mr. Rosenberg, the President called out: "Loeb, come over here," and dictated to him the order closing Nushagak and Wood Rivers to all commercial fishing.

Thus, as it turned out, there might just as well have been no formal hearing at all, as the decision was taken out of the hands of Secretary Straus and rendered by the President, without consideration of evidence submitted at the hearing.

There was some controversy as to the definition of the mouth of Nushagak, the fishermen's representatives desiring to have same fixed well out in the long narrow bay. The Department, however, decided that the mouth should be considered as shown in the accompanying diagrams. These diagrams are kindly furnished the Pacific Fisherman by the Department.

The effect of the order is that not only trap fishing, but gill-netting is prohibited on these rivers during the year.

There are ten canneries in the district, seven of which are operating. As a very large part of the supply of fish is secured below the prohibited territory, it is not likely that the order will reduce the pack to any appreciable extent.

We reprint on another page the official notice of the Department of Commerce and Labor, and which, it will be observed, closes this district to all commercial fishing "until further notice." The department reserves the right therefore to continue the order in effect in following seasons if such action is thought advisable in behalf of the industry.

Those who desire extra copies of the Pacific Fisherman Annual should forward their orders at once. If desired, lists of names and addresses may be sent us and we will mail and pay the postage without extra charge.

WOOD RIVER, SHOWING TRAPS.

Figure 12. An announcement of Nushagak and Wood River Closures (Pacific Fisherman, 1908, January)

In the summer of 1945, William F. Thompson traveled to Bristol Bay at the behest of the Alaska Salmon Industry, Inc. who requested help from the University of Washington to better understand salmon populations. Ultimately, the salmon processors were in hopes of a more

scientific approach to management. After spending the summer of 1945 in Bristol Bay, Thompson wrote to the Chairman of the Bristol Bay Packers:

> *In my review of the Bristol Bay salmon fisheries I have come to some very decided conclusions regarding what has to be done…. I have gone over in detail all available published material on the red salmon of Bristol Bay, and in fact of Alaska. In addition, I have begun to collect data on climatic changes, economic conditions, and regulations. These data will, I believe, be necessary for the interpretation of the history of the fisheries.*

Writing that he believed strongly that a biological basis for regulation should be set, he outlined an attention to long-term consistent data collection protocols that persist today and suggested pathways forward for the proposed research:

> *May I suggest that the spawning ground surveys be continued without change or radical alteration for several years. They will be found, I should guess, a great deal of use. The estimates cannot give any close determination of the numbers escapement the commercial catch, but can be used in a comparative way, if made each year in exactly the same year, over the same ground and at the same time. They will not necessarily forecast the runs, but will enable us to determine from what grounds the runs come each year, and be of use in various ways.*

Some of the standards for data collection techniques used today can be traced back to those early days when the program first began. In spring of 1946, Thompson's proposal for research was approved in agreement with UW President Sieg[15], and shortly thereafter Bud Burgner and Thompson drove to Bellingham to meet with US Fish and Wildlife as well as the Pacific American Fisheries (PAF), the largest cannery in the region at the time, to prepare for their first season in Alaska (Thompson, 1947). As Burgner and Thompson looked at barges

[15] Lee Paul Sieg was president of UW from 1934-1946 and is most known for his unwillingness to participate in Roosevelt's Executive Order, which authorized forced relocation of Japanese American citizens to internment camps. Instead, he found people to take in Japanese American students. He is also the namesake for the building in which I pursued my doctorate.

used to pull seine boats off beaches, they discussed salmon migration with a few of the people with the cannery:

> *the chums ascend the rivers in deeper water than the reds and by fishing the center of the river a much higher percent of chums result. Reds stay near the surface and most fish are gilled high in the net, making depth of net less important.*

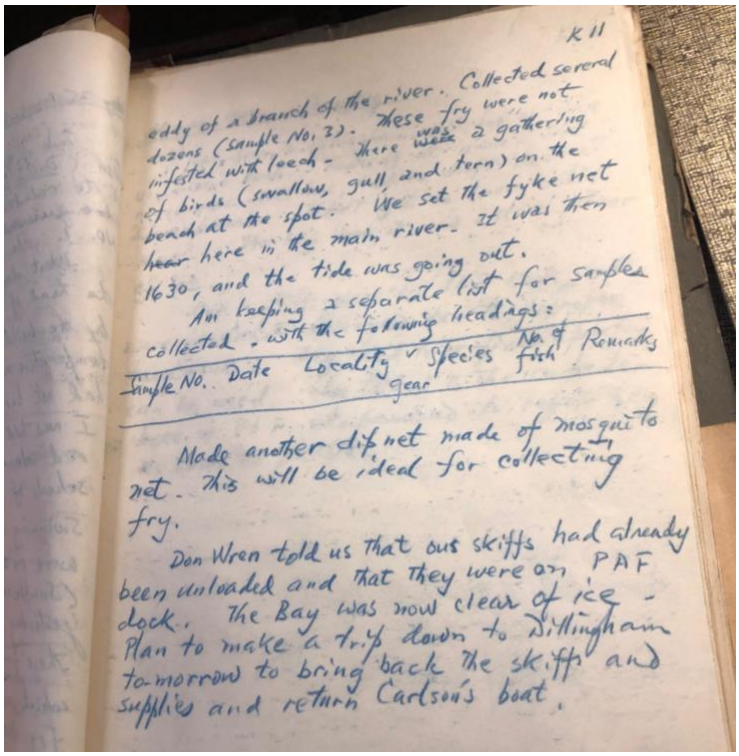
By June 25, 1946, the crew had arrived in Bristol Bay to see the first day of fishing season noting that catches were light [a little over 7000 reds and 600 kings to one scow] and that "fishermen claim this is a late season - cold." And in observing the fishing activity, Burgner jotted down potential questions about those techniques for fishing: "at what point do fishermen cease to set entire net?; how much variability is there between fishermen on this point; how many sets a day are made at different times of the season? How long does it take fishermen to haul and reset net? are scows always in same spot and how far do boats range from scow? Is the length of time between sets uniform? What determines the length of time: fish in net? Tide? Desire to fish in different spot? Nearness to scow? What factors influence shifting of fishing grounds?"[16] Their early notebooks are filled with questions like these, which marked the beginning of scientific questioning about the fish and their habitat.

With passing years, more scientists would come to the field to make observations. Ted Koo, who would eventually write an edited volume on sockeye salmon, made fastidious observations in his field logs.

> *Don Wren flew in with his 'Piper' at 1315. Piper could take one man at a time due to a propeller problem. When she came down again for a second trip, took Dick down to Dillingham — I decided to stay here to make observations… Collected several dozen*

[16] They also observe fishers dealing with oil on the beach. "Beach netters on Clark Beach are not fishing because of oil on the beach left by the APA oil barge. As soon as the grounds clear after a couple of tides it will be safe to fish."

(sample no. 3). These fry were not infested with leech. There was a gathering of birds (swallow, gull, and tern) on the beach at the spot. We set the fyke net here in the main river. It was then 1630, and the tide was going out. Am keeping a separate list for samples collection.


K 11

eddy of a branch of the river. Collected several dozens (Sample No. 3). These fry were not infested with leech - there were a gathering of birds (swallow, gull and tern) on the beach at the spot. We set the fyke net here in the main river. It was then 1630, and the tide was going out.

Am keeping a separate list for samples collected - with the following headings:

Sample No.	Date	Locality	Species	No. of fish	Remarks
			gear		

Made another dip net made of mosquito net. This will be ideal for collecting fry.

Don Wren told us that our skiffs had already been unloaded and that they were on PAF dock. The Bay was now clear of ice - Plan to make a trip down to Dillingham tomorrow to bring back the skiffs and supplies and return Carlson's boat.

Figure 13. Excerpt from field note book from Alaska Salmon Program archives

In later years, Koo would implement more standards to his observations. In 1962, he opens one of his field logs (figure 13) with a memo to staff members:

> *For the convenience of handling and cataloging at the Institute office, all log writing should contain certain information that is necessary for positive identification and classification. Required on each page are the following entries: Name, Project and subproject (and on D.S. include vessel name); Page no.; Date – day, month, year (include day of the week, also); text*
>
> *If you wish to record information which does not specifically belong to any project, enter it under 'general observations – miscellaneous'. Such information may include, for instance, the appearance of a beluga in the lake, the unusually large number of certain type of insects, volcano eruption, forest fire, plane accident, construction work, etc,..*

Much has gone unchanged in the field program. Figure 14 shows a side by side photo of the field site at Camp Nerka – one from my time there and the other from 1949. When looking

back into the archives, the original scientists' field notes are full of everyday observations: about the weather, about the food, about strange insects. Before editing his pivotal book on Red Salmon (1961), Ted Koo traveled to Alaska in 1949 and noted the things he saw. On May 20, 1949 he made observations about weather and temperature as well as comments from locals: *Naknek River had been open for about a month already, so Mrs. Bennett of the PNA station at Naknek told me. She also said that last winter was quite severe, plenty of snow, and temperature went down to 40 below zero.* From the sky, he observed that pools were still covered with ice and Nushagak River was solid with ice. When he returned the following year, he noted on May 12, 1950: *ice was largely gone in most of the pools; only very little remnants of snow left here and there. Kvichak, Nushagak, and Wood Rivers were all open. All these conditions indicated that an earlier spring was here than last year.* Even in the early days, Eicher (1938-1956) argues that the one-man collections of data required a continuity to be able to ask broader questions about salmon.

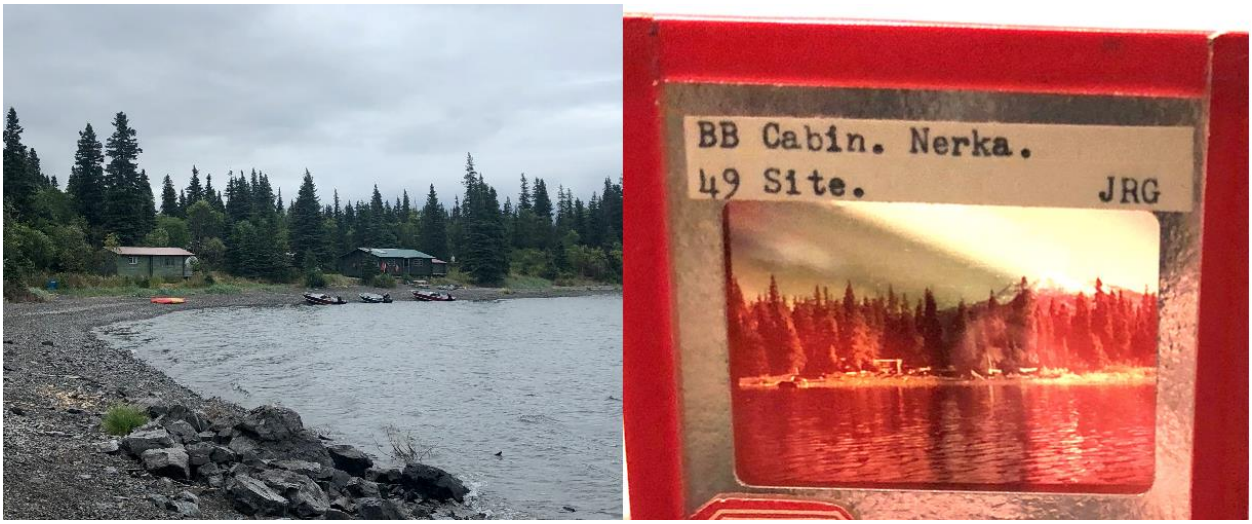


Figure 14. photo of the Nerka field camp, 2021 (left), 1949 (right)

The program continues this tradition to this day; however, their observations have become more standardized over time. One specific project that has occurred over many

decades is a fine-scale study of two streams, A and C streams, which drain into Little Togiak Lake. This project illustrates the novelty of collecting fine-scale data on a long-term basis. By conducting surveys everyday -- calling out each live fish by its corresponding tag, tagging and measuring untagged fish, accounting for the dead fish, measuring rock size in different transects -- the scientists produce fine-scale data that captures seasonal and daily trends in specific locations. In these streams, a team goes out daily – sometimes accompanied by Ray Hilborn and his family – to document each fish with its respective tag number and to gather all untagged fish to tag, measure, and collect genetic samples. Through long-term data collection on these streams, scientists have understood not only the number of spawning adults (Peterson et al. 2015), but also dynamic reproductive strategies (Bentley et al., 2014), evolutionary effects of predation (Lin et al., 2016), the impact of size-selectivity in the gillnet fishery (Kendall and Quinn, 2009), and more.

While counting has been consistently tracked over time, scientists' enumeration of salmon was always about more than knowing the number of fish in the streams. Even as the impetus was to understand the number of fish in the streams, early FRI observations looked at salmon behavior as well. In one of Ole Mathieson's tattered, rite in the rain notebooks from July 27, 1951 (figure 15), he wrote a legend of behavior codes:

1. Migrating upstream with no intention to spawn
2. Migrating perhaps looking for a mate or place to rest
3. Both fish active, female digging / male courting etc
4. Inside the redd circling about
5. Standing in redd, quiet, stranded fish maybe [not legible]
6. Post spawning, swimming around in any direction usually being chased away by other fish

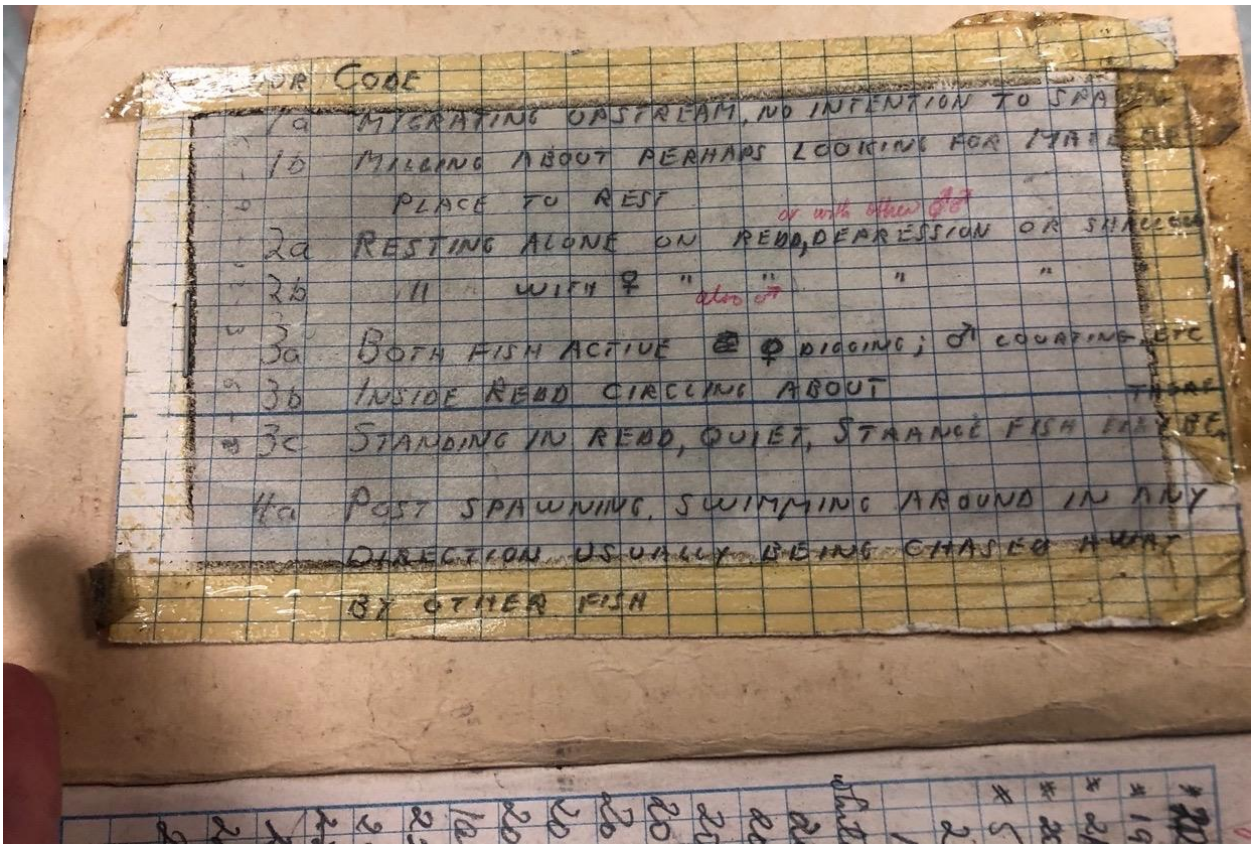


Figure 15. Excerpt from Ole Mathieson's notebook, from the Alaska Salmon Program Archives

This kind of behavioral observation occurs today as well. During summers walking streams in Alaska, scientists would comment on behavior: "the big males are more successful"; "the sneaker males (small in size male salmon referred to as 'jacks') use their small size to their reproductive benefit". Sitting around the dinner table late in the evening after tagging fish, a few of the researchers discuss the optimal spawning areas in streams noting that it could be size of gravel that plays a role in salmon preference for spawning habitat. They argue that the more successful spawners push others' eggs out of the way to make room for their own. One researcher explains that there is a curve that explains optimal success of a stock (Beverton-Holt model), and if the late-spawners are less successful reproductively, they will dig redds (salmon nests) in a place where a previous female has

deposited her eggs. As is clear, scientists have been motivated by understanding the complexity of interactions between fish, a question that ultimately required more computational methods to simulate population dynamics.

As I showed earlier, endeavors to count fish were initially driven by commercial (e.g., processors) interests in exploitation. This close connection between the salmon industry and the scientific presence in the region is evident in the early funding as well as the field notes. In one of Burgner's early field books from 1946, he lists all the canneries in the area. And while they were heading to Alaska to look at potential reasons for salmon declines with respect to biological and ecological factors, they were also paying close attention to the pressures on the salmon fisheries, e.g., the fishermen's technologies and techniques for fishing salmon. The objective of that first research was to "determine the number of fish allowed to spawn in order to assure the perpetuation of the supply and to secure any facts pertinent to management for that purpose" (Thompson, 1962).

What I have shown in this brief history is that the combination of catastrophically low salmon runs and the burgeoning commercial interest in the salmon in the early 1900s led to the processors commissioning the University of Washington to conduct research into the biological reasons for fish declines. The longest continuously maintained dataset is the adult sockeye salmon spawning ground surveys and spawning ground age composition[17]. Developing a basic understanding of populations and interactions, the program has maintained these long-term datasets. However, the desire to know longer-term dynamics led

[17] Spawning adult sockeye are counted by walking through streams and counting individual or groups of 10 on tally clickers. Historically, only counts of salmon were recorded; however, with time, additional information was parsed out such as sex (male, female, jack), status (live, dead), and model of death (senesced, bear kill, etc…).

to newer forms of data and understanding while still calling back to those older sources of data. While the legacy of the program is its support by commercial fishing, it has continued to change the research program and develop new insights into how salmon behave. I will explore this in finer detail in later chapters.

Legacy of commercial fishing in the Kuskokwim

The other field site I explore that is closer in proximity to data collection also has a legacy of commercial fishing. The Arctic-Yukon-Kuskokwim (AYK) region is remote, vast, and vulnerable to climate change as many people depend on the resources of the land and sea at physical, mental, and spiritual levels (National Research Council, 2005). This region has particular data challenges as data about fish are scarcer than in regions more dominated by commercial and sport fishing interests.

I came to this site through the SASAP project as one of the eight working groups (see figure 16) asked questions about how to incorporate local communities in their data collection practices.

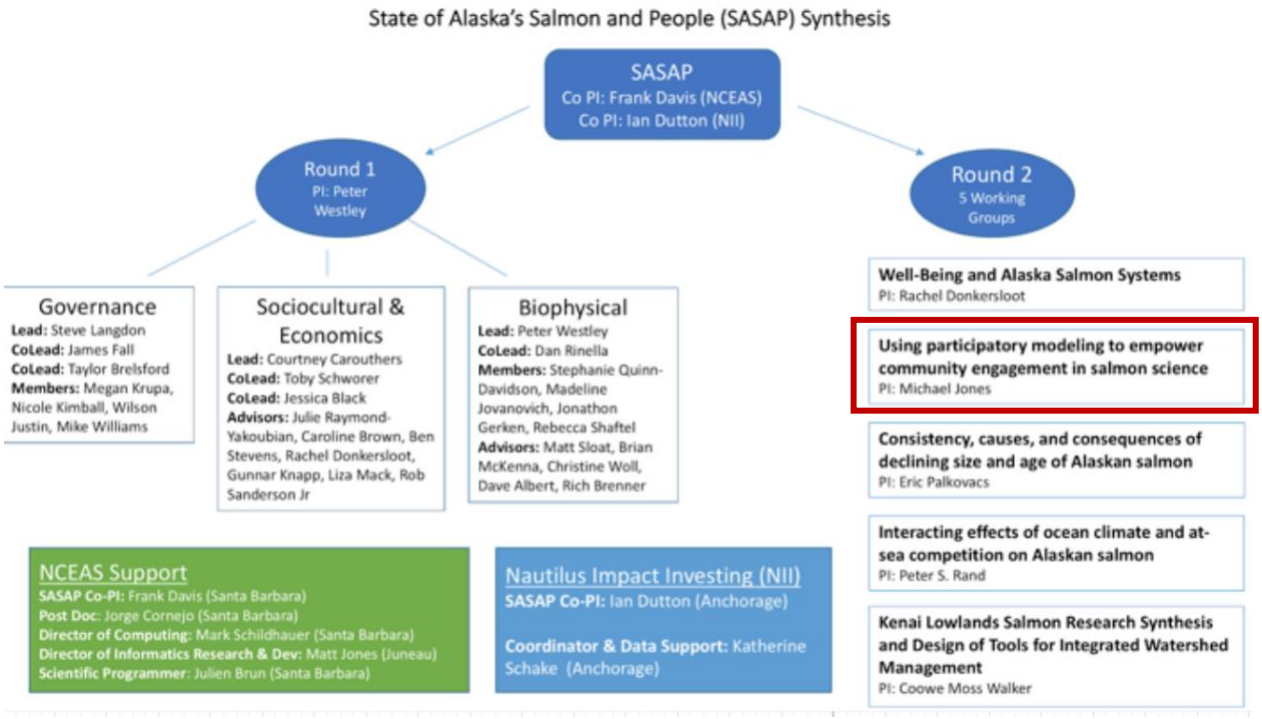


Figure 16. Image from introductory SASAP working group meeting illustrating the eight working groups and how they were organized.

The Kuskokwim is named for its river. Kuskokwim is a loose translation from a Yup'ik word which means a "big slow moving thing." The Kuskokwim River stretches across 702 miles along Southwest Alaska into the Bering Sea. The majority of the 16,000 residents of the Kuskokwim area live in the Kuskokwim River drainage, but this also includes several Bering Sea coastal villages. In the Lower Kuskokwim River, the Alaska Natives are Yup'ik Eskimo, the middle region is Athabaskan, and the Upper Kuskokwim are Upper Kuskokwim Athabaskan. The largest Kuskokwim community is Bethel, which is also a hub for the 56 villages in the Yukon-Kuskokwim Delta. Salmon is a primary source of sustenance in these subsistence communities. However, salmon is much more than food. Within the Pacific Northwest and Alaska, salmon is discussed as a rite of passage. As Nick Kameroff

commented: “my relationship with salmon is my lifestyle” (Kameroff, 2019), a refrain echoed throughout the region.

During my first trip to Bethel, I attended a working group meeting full of research scientists, data task force members, and local stakeholders. They introduced new terms and clarified old ones. I learned that the historical precedent of management in the Kuskokwim region was commercial fishery management. Given this field legacy, many of the models and concepts used in commercial fisheries management are used to manage the largest subsistence fishery in the country. This has particular implications in aligning ways of knowing with ways of being, which I will explore in chapter 7.

Throughout the state of Alaska, most salmon fisheries are managed through designation to Alaska Department of Fish and Game (ADF&G) managers. The commercial legacy of subsistence management is evident in some of the reports from the past. In the preface for the 1999 Annual Management Report (AMR), the Division of Commercial Fisheries (CF) of the ADFG is designated the responsible party for not only commercial fisheries but also subsistence in the Kuskokwim region. Given that commercial fishing has been almost altogether dissolved in the Kuskokwim region, fisheries are managed collectively—by state, federal, and tribal in-season managers.

According to an Annual Management Report (2000), the first commercial sale of salmon in the Kuskokwim happened in 1913. Quotas for commercial catch were not established until 1954, and the first counting tower in the region did not come until after statehood in 1960 at which point the Quinhagak District for commercial salmon fishing was established. By 1999, the Kuskokwim River experienced devastatingly low chinook, chum, and coho salmon returns with very late run timings. At this point, the federal government took

control of managing the fishery, and a year later, the Kuskokwim River was declared an economic disaster area.

The assumption of responsibility by the federal government is determined by policy. Section 802 of ANILCA established a rural priority for subsistence harvesting on federal waters, and the U.S. Congress anticipated the State of Alaska would implement this rural subsistence priority. However, a 1989 Alaska Supreme Court decision found the rural priority in conflict with the common use clause of the Alaska Constitution. Section 804 of ANILCA clarifies that when necessary to restrict subsistence harvests on federal lands in order to protect the continued viability of [fish and wildlife] populations, or to continue such uses, a subsistence priority will be implemented based on: (1) customary and direct dependence upon the populations as the mainstay of livelihood; (2) local residency; and (3) the availability of alternative resources. In cases of ample resource abundance for subsistence uses in the Kuskokwim River region, the federal government defers management to the state of Alaska. However, in recent years of low Chinook salmon abundance, the federal government has assumed management of Chinook salmon harvests on federal waters of the Kuskokwim River. Under a Memorandum of Understanding adopted in 2016, federal managers with USFWS, and under delegated authority from the Federal Subsistence Board, have implemented management measures in cooperative consultation with in-season managers of the Kuskokwim Inter-Tribal Fish Commission. The cooperative agreement also seeks input from ADF&G and an ADF&G advisory group–the Kuskokwim River Salmon Management Working Group, as well as local residents.

The state department still manages most of the salmon fisheries in Alaska; however, not in the Kuskokwim region due to the impact on subsistence users. The fishery was first

federalized in 1999 after a Chinook crash. As one participant noted: "*even if it comes back to what they [the number of salmon] were, I think the amounts needed for subsistence will increase because the population is increasing and so I don't believe that the state will ever manage this fishery again."*

Even as commercial fishing appears to be far out of sight, something that occurs in Bristol Bay and elsewhere, the legacy of commercial fish remains.

The importance of subsistence in this region cannot be overstated. In his ethnographic study of discourse in the lower Kuskokwim, Hensel (1996) notes the significant changes including everything from changes in climate to transportation to fish preservation techniques and the way in which subsistence practice is the thread that runs through the community. Hensel's detailed account of subsistence practices refutes the notion that subsistence is an economic activity, but rather is a cultural practice closely tied with identity. In other words, he shows how identity is sustained through social practices rather than priori ascribed.

In this site, I predominantly engage with tribal, state, and federal in-season managers working with scientists to understand how human activity is considered a data source. While I agree with Hensel's argument that subsistence activity is cultural not economic, I explore subsistence through the lens of management, which is a direct descendent of commercial fishing.

Conclusion

By studying three different sites' approaches to data collection and data synthesis, this research looks at the practices of data integration and explores the ways that expertise is distributed throughout environmental data practices. This approach takes up the

opportunity to fill the gap on studies that focus on micro-level interactions (Ackerman et al., 2013) to better understand how expertise gets shared and consolidated in emerging resources.

I look at various approaches to the question of how scientists instrument scale. While SASAP, a data-centric approach to dealing with issues of the long-term, is focused on data synthesis, there are other approaches to reconciling how to produce long-term understandings of ecosystem function. The Alaska Salmon Program represents one approach to collecting data that captures a long-term perspective. Achieving long-term data – a well-documented challenge for ecology – was one of the main justifications for the initial establishment of the research program. I engage with the Alaska Salmon Program to look at material practices of scientific data work. It is here that I explore the ways in which scientists have used specimens to instrument for temporal scale. And, in chapter 7, I explore the discussion around models and how locals can participate in data collection for management models.

In the following section, I explore how scientific programmers collaborate to clean, integrate, and archive data. My investigation into this work begins with an analysis of GitHub communication traces, which illuminates how challenges particular to the ecological domain challenge data integration work. I explore a data integration initiative to better understand how contemporary data approaches to ecological data management have tackled the challenge of understanding salmon ecosystems. I show how scientific programmers encounter scalar issues in their work to make data interoperable and how their strategies often entail inverting the infrastructure in particular ways.

Ch 5. Inverting scale: Characterizing multi-scalar approaches within environmental data science

Introduction: The State of Alaska's Salmon and People

To answer the question of how scientists instrument scale, I first analyze the data integration undertaken by NCEAS. I examine how individuals involved in the data task force team – a team focused on the cleaning and integration of data for an open science data repository called the Knowledge Network for Biocomplexity (KNB) – make use of information outside of their immediate resources. In this analysis, I characterize how scientific programmers encounter scale as they integrate heterogeneous environmental data into homogeneous data storage.

Modern scientific information ecosystems are not always designed to deal with the complexity of ecosystems. Rather, they are designed to deal with the complexity of information systems. The importance of scaling when understanding ecosystem change is often omitted from open science and big data initiatives. In the following sections, I argue that ecological scale commonly complicates environmental data science endeavors and should therefore be considered as a primary concern in ecoinformatics initiatives.

The point of departure for this study of scientific information ecosystems is the State of Alaska's Salmon and People (SASAP) project, which brought together scientists, Indigenous knowledge experts, fisheries managers, and scientific programmers (or data professionals) to synthesize existing knowledge and build information tools for Alaska's salmon science. SASAP was created in collaboration with NCEAS as an initiative for addressing wild Alaska salmon in its many forms: as industrial commodity, as cultural resource, and as subject of regulation and scientific study. Through the assemblage and synthesis of myriad data on the

one hand, and by bringing together a number of concerned stakeholder communities on the other, the SASAP project aimed to generate new knowledge necessary for restoring healthy populations of salmon to their natural habitat and, as an ancillary goal, for making this information open to the public.

To understand ecosystems on a longer timescale than individual field sites, NCEAS (and other data organizations like it) target the acquisition, annotation, and interoperation of data sets held by multiple institutions. In my study, SASAP represents a contemporary approach to understanding Alaska salmon and people through the synthesis of data held by state and national agencies, companies, and scientific organizations, and supported by the data professionals staffed at NCEAS.

The overarching research question for this thesis is how scientists instrument scale. In this chapter, I explore this question by asking: *How do scientific programmers encounter scale in their work to integrate and archive data? And, subsequently, what strategies do they employ to reconcile scalar issues that occur in data integration*? To explore these questions, I first a) ascertain what complicates data synthesis, b) develop scalar categories for commonly encountered challenges with data curation, and c) summarize strategies scientific programmers take to overcome problems in data integration.

Encountering scale in data integration

In general, data integration processes are challenged by heterogenous data, storage of unstructured data, varied data sharing norms, and lack of context of that data. More specific to my site, ecological data represent phenomena that occur at various spatial-temporal scales. Dealing with this heterogeneity is less of an issue of how to make difference more scalable (e.g., as in the case of large-scale agriculture), but rather, the focus is on how to

accommodate variation in ecosystems through harmonizing data standards (e.g., harmonizing the ways to represent and translate information).

One role for stakeholders in data integration projects is to represent complexity so that downstream users of the data can make informed decisions about how to access, make sense of, and analyze that data. This concept often appears as provenance or metadata. The intended goal of adding metadata is to add nuance to data such that downstream users might easily make sense of it. Scaling the study of ecological systems, however, requires not only specific metadata but involves data that cannot be re-collected. Essentially, if the metadata were not notated in the moment of data collection, context is challenging if not impossible to recreate. There are several reasons for patchy metadata documentation, including lack of long-term data management, lack of standards for data collection, differences across agencies, and inconsistencies in adhering data standards.

The importance of metadata has been widely documented, and while scalar issues extend beyond the presence or absence of metadata, it is one example of how lack of data integration norms can challenge the modern acquisition, archival, and usage of data. As such, determining the scale of study post-data collection is a key challenge for scientific programmers working to integrate and archive data.

Rather than formatting salmon to all be the same or to all be studied the same, the scientific programmers try to harmonize data post-data collection. Putting these findings in context with how other data infrastructures operate, it is evident that the preservation of local is a concern at the forefront. Much like other ecoinformatics initiatives, the data in SASAP are archived with some form of metadata; this metadata is in the ecological metadata language (EML), which is built on an old XML framework. When asked why they had not

updated to a more modern approach such as JSON, the lead explained how many users were already familiar with this format and furthermore, that it was easier to use given the textual nature of the standard. This is just one of many examples of technological lock-in or the inertia of the installed base. This is not to suggest that change does not occur in other ways; however, in an attempt to preserve context and remain useful to the intended users, the standard in place remains.

There are myriad types of data that the scientific programmers encounter and different spatial and temporal dimensions therein. The data that the SASAP team focuses on are data collected in the past across shifting measurement requirements, technological eras, and political regimes; however, the team seeks to bring together these disparate data sets into a single data repository—the Knowledge Network for Biocomplexity (KNB).

This endeavor to "revolutionize the natural sciences" by streamlining data practices at the heart of the SASAP project is the major reason I was involved in the study. I was one of two data ethnographers studying the phenomenon of environmental data science. Given that much of the critical data work occurs in other spaces – in distributed online spaces as well as historically – I turn to more inventive approaches to ethnographic research (Lury & Wakeford, 2012; Marres & Weltevrede, 2012; Ribes, 2019) to gain access to these spaces. This approach involves looking at the logistical instruments of collaboration. In particular, I follow the "issues" in their GitHub account to track down how scientific programmers reconcile errors in data integration work.

Issues is a category of communication in GitHub, largely used for project planning in software development. An issue is a way of tracking bugs and also allows for deeper discussion about how a bug was discovered, how it might be resolved, and what information

is known and can be shared about the particular issue. As such, to uncover how scientific programmers collaborate to resolve problems in data integration, and furthermore, how they utilize resources that are not immediately available, I explore SASAP's GitHub issues. This empirical chapter begins to answer the question of how scientists instrument scale by first investigating the practices and challenges of integrating scientific data.

Following the SASAP GitHub Account

By following the communication, or articulation work (Schmidt & Bannon, 1992), amongst the scientific programmers I sample for both breadth and depth to engage the question of how scientific programmers encounter scale in their data integration efforts. Within the symbolic interactionist perspective, articulation work emphasizes a conceptual understanding of how people coordinate within a situated context (Strauss, 1988). Rather than seeing workplace activities as formalized practices, articulation work emphasizes that activity is situated and local (Star & Gerson, 1987; Suchman, 2007). To better characterize the flow of work, attending to the situated nature of articulation work before it is removed from final representations of the work is tantamount (Star & Gerson, 1986). Or in other words, much is made invisible in the final representation of work. In the infrastructural sense, the articulation process is largely invisible when there are no problems (Strauss, 1988); however, the moment a project presents challenges or goes off track, the articulation becomes visible.

As I seek to understand how scientists scale, the comments and issues in their communication tool are instructive to revealing what is often made invisible in the final data product. These discussions outline the strategies for scaling down in the study of large-scale

phenomena as well as scaling up to make sense of fine-scale entities. By following the issues — which are often about anomalies in the data or possible errors — I outline how the actors struggle with their research object, salmon data, as a way of illustrating how the issues of scale in ecology are tightly coupled with issues of scale in information or data science.

What are issues?

To set the stage for the following discussion, I describe the nature of these data, which pertain to salmon: their habitat, the people who depend on the resource, and the activities centered around their harvest or capture (for commercial, subsistence, or sport purposes). Large amounts of data about harvest have been collected by the state department as the primary entity responsible for management of harvest. Managers are concerned with predicting future fish numbers and setting goals around how much fishing to allow. Because the major organization from which NCEAS developed Memorandums of Understanding (MOUs) and acquired data is the state department, much of the data the scientific programmers are dealing with are data about harvest and abundance.

Of the 185 issues created on GitHub about salmon data integration, the top issues, defined as the issues with the most comments, pertain to escapement data, Age, Sex, Length (ASL) data, hatchery release data, sport fishing licenses, mark-recapture data, and an issue about regions. This is not to suggest an outsized interest in these types of data, but rather, this suggests that these data present many challenges. I outline these issues below and categorize those issues at different layers of research infrastructure relying on an adapted model of Ribes and Polk's (2015) kernel metaphor to highlight three major layers: sites of data collection; instrumentation, data format and databases; and, institutional knowledge.

The purpose of this section is to show how issues in data integration work move beyond the data itself, expanding to the entire infrastructure that supported that data collection.

Issues at sites of data collection: errors, duplicates, typos, often caused by human error

I categorize many of the issues in the scientific programmers' GitHub account as problems that occur at the level of sites of data collection. I define these as a mistake from the field, or a problem that occurs when bringing together different data sources. These often encompass issues that are caused by human error as well.

For example, when merging two different reports from two different agencies -- North Pacific Anadromous Fish Commission (NPAFC) and the Mark, Tab, and Age (MTA) Lab data -- the MTA lab data include much finer-scale spatial information as compared to the NPAFC data. While this pertains to the scale of spatial information, it is more about the data standards with respect to resolution that differ across management organizations.

Spatial scale issues are more generally caused by errors related to the challenge of poorly drawn boundaries *in the field* or inconsistencies with respect to spatial *scale,* e.g., measurement unit. As with most data cleaning processes, a large portion of the errors that scientific programmers have to contend with when making data interoperable are duplicates and typos. These can be relatively mundane, innocuous errors, a typo from some point of data entry.

This mundane quality assurance/data cleaning task is common to many data processing pipelines, not just scientific programming. In some instances, duplicates are just endemic to Alaska salmon and streams. One expert on Alaska salmon comments in a GitHub issue that "there are duplicates of many location names across Alaska (e.g., Moose River, Canyon Creek,

etc..)." This is because the naming convention for the stream is not unique in some cases, and the file may lack actual latitude/longitude data if collected long ago.

This kind of data work is what is often considered the data janitorial labor (Lohr, 2014) of data science (or data wrangling). This is the work that is not particularly flashy but is necessary for getting data cleaned and ready for usage. Most issues begin with data issues, which include issues such as gaps in data, changes or misusage of standards, errors, duplicates, and typos. An error such as triggers an error discovery process, which involves practices such as visualizing data. However, often it leads to a deeper understanding or problem that occurred in data production.

Issues with instrumentation: material formats, challenges turning natural phenomena into data

I refer to instrumentation issues as those that pertain to the material challenges of translating a natural phenomenon into data. These issues could be caused by anything from the way an instrument breaks or is not calibrated properly to mundane issues of different units of measurement. An aspect that can alter the shape of the data is a change in the material medium of the data representation.

One example that illustrates this point is the way data are produced to understand salmon age. To account for and reconstruct the "salmon run" — which refers to the number of fish that return from the ocean to their spawning grounds every year — the ADF&G monitors the number, size, and ages of returning fish with ASL data. A salmon run[18] refers to

[18] Salmon runs are an important step in the life cycle—which varies depending on species—but for Sockeye they tend to mature around age 5, spending 1-2 years in freshwater rearing habitat and 2-3 in the ocean. A run is an annual event, which refers to the time when fish return from the ocean to freshwater spawning grounds. Most of these fish return to the very

a period of time when breeding (and as such, dying) fish return to their spawning grounds. These 'pulses' of fish coincide with human fishing activities as commercial, sport, and subsistence fishers harvest salmon according to the state's temporal allotments. Not only an annual sampling event, it also involves determining the age of the fish that have returned. This helps scientists understand population dynamics and salmon life histories. However, these data have many points of intersection between the fish itself and the instrumentation used to produce data.

There are aspects about the salmon and the salmon ecosystem that impact data formatting and standardization. In response to a scientific programmer's struggle with a particular file, an advisor in the salmon data group notes the ways in which data collection processes and physiological traits of the fish have shaped the actual file structure.

> *The file structure arises from the way these data are collected. To gather age/sex/length samples, scales are taken from fish out of a representative sample caught at an ASL sampling project. Scales from fish taken during a sampling event are placed on "scale cards," along with other information about both the sampling event (such as the date and location) and the fish itself (such as length and sex). For Prince William Sound and Copper River salmon, all Chinook and Coho have a maximum of 10 fish sampled per scale card, whereas sockeye and chum salmon have a maximum of 40 fish sampled per card.*

The domain expert goes on to explain the reasoning for differing file structure, which has to do with aspects particular to the salmon noting that the reason for these differences is

> *Chinook and Coho lose scales easily and therefore a larger portion of the sample scales are 'regenerated' and are not useable. To get around this, up to 4 scales need to be pulled from each Chinook or Coho in the hope of finding a single scale that can be successfully aged.*

---

locations where they were born. Because of this rhythm of salmon returns, human activities with salmon are managed around these runs and scientists and managers have worked to disambiguate between these differences.

This is an example of an issue that begins with understanding varied data formats and ends with a journey into specific aspects about the fish itself – the scales used to age Chinook and Coho are lost more easily.

Instrumentation issues represent the errors that occur when data life cycles and salmon life cycles, or rather the endogenous and the exogeneous, meet. These issues are usually resolved with the acquisition of information about specifics that caused errors or inconsistencies in data. These issues speak to Jackson et al.'s (2011) insight that organizational, biographical, and infrastructural rhythms are distinct from phenomenal rhythms.

Issues with institutional knowledge: Differences across institutions responsible for data curation

Issues I categorized as institutional are issues tied to the distributed expertise across multiple academic disciplines and management agencies. This results in a gap between the general data science knowledge of the scientific programmers and the domain-specific knowledge of salmon biology required to understand the data at hand. In these assessments, the data task force members often have to consult resources outside of the data they request because of missing data, inconsistent or non-existent metadata, or in need of clarification.

While the project acquires most of the data by requesting from partnered institutions (predominately the ADF&G), it often faces myriad obstacles in making this data public due to concerns of data holders. Furthermore, the data workers are often faced with the need to acquire data by mining or scraping it from a public site. For example, much of historic data has to either be digitized from analog formats or found in reports.

However, in the addition of simple metadata about fixing typos, there can also be deeper information about standards employed or the kind of notation. In one example, a scientific programmer adds the following to the metadata file*: Age classes are given in European Notation, where the first number is the number of winters spent in freshwater before going to sea (1 winter in freshwater = age-1.X), and the second number is the number of winters spent at se (3 winters at sea = age-X.3).* This information provides a legend for understanding the numerical notation for aging the fish and the different standards for that notation.

The scientific programmers and scientists in the project discuss standardization as the predominant way of making data commensurable and interoperable in the field. However, before a standard is set in place, there are many negotiations that take place in the classification process. In the following quote, a scientific programmer struggles with how to understand different units of measurement for weighing salmon and asks for some advice on inconsistent data, noting the various units of measurement:

> *Only pink and sockeye salmon have recorded non-zero weights in the dataset. Within those, the weights are hugely inconsistent and the units are generally unclear across the board…it seems like the unit that most projects used/tries to use is kilograms—which makes sense seeing as how length is also in metric. However, I think there are at least 4 different units used at different times/locations/projects/species.*

Data collectors' usage of conflicting standards in different projects and by different researchers has led to the inconsistency highlighted in the above quote.

Some of these differences in standards are more easily explained than others. For example, if data had been collected by NOAA, a federal institution that manages and researches marine environments, the standard for measuring salmon will be different than those used by the ADF&G, a state agency responsible for collecting data in freshwater. This is because as salmon get closer to their spawning grounds, their snouts and jaws develop

and their tails become frayed. As such, the measurement standard for measuring fish in freshwater with larger snouts and frayed tails differs from measurement standards used for marine environments. In freshwater, the dominant standard is the mid-eye to hypural plate (MEHP) standard or mid-eye to fork of tail (MEF) while ocean measurements tend to involve tip of snout to fork of tail (STF) measurements. This is an issue of institutional or organizational data standards, which intersects with biological and ecological aspects of salmon and its ecosystem.

What are the dimensions of scale in the data?

There are three primary attributes of data that are collected: spatial accuracy and coverage, temporal accuracy and coverage, and specific dimensions of the phenomena of study (e.g., salmon). In other words, the critical variables are when or how long data were collected, where they were collected, and what was collected.

In the SASAP case, issues of time scales emerge in discussions about data anomalies. These issues are often due to varied, inconsistent, or uncertain time scales. In other words, there are issues around *how time is measured* while there is also *temporal uncertainty* with respect to what the future will hold or what the past was like. Uncertainty about changes through time is a primary concern for the scientific programmers: both uncertainty about how data are represented from the past and uncertainty about how data might become useful in the future. This is compounded by missing data either caused by an error or that the data was not collected in the past. Ultimately, these are concerns about epistemology, or how knowledge is produced and plans for its downstream usage.

Scale as an epistemological concern is evident throughout much of the data work. These epistemological concerns overlap with infrastructural issues such as the maintenance of data in the long-term or the funding and personnel constraints that impact data production. Below, I categorize the scalar dimensions (spatial, temporal, phenomenon) and how they overlap with the different levels of infrastructure (sites of data collection, instrumentation, and institutional) where issues occur (table 4).

	Spatial	**Temporal**	**Phenomenon**
Data/human error	Level of spatial aggregation	Wrong dates marked	
Instrumentation	Geodatabases or geospatially referenced data	Calendars	Scale cards and material format of the data; standards for aging fish
standards / institutional norms	Stream codes	Annual or daily depending on funds or research needs	Age classes standard notation

Table 4. Table of spatial/temporal/phenomenal qualities of the data issues

As I highlight, the major challenges with scale inherent in data integration can be best categorized as spatial, temporal, and related to the phenomena itself – to the salmon -- in this case. Broadly, the overarching concerns about scale are ensuring certainty (of data accuracy and provenance) through time, achieving an adequate temporal *coverage*, and commensurating standards from the past. These concerns mirror what Ribes and Finholt (2009) uncover as tensions in achieving long-term: 1- contribution over time; 2- alignment of end-goals; and, 3- designing for use. These are distinct from concerns about errors or uncertainties in the data itself. However, I argue that to integrate data, the scientific

programmers adopt an approach to data preservation that centers on alerting potential users to potential errors.

Although many of the issues pertain to spatial and temporal dimensions (e.g., stream IDs, age of the fish, location when counted), the scientific programmers resolve these issues by gathering more context from distributed resources, reconstructing how the error occurred, and flagging data with their potential errors. In other words, programmers are gathering and recording knowledge about potential errors or anomalies rather than just context in the form of metadata.

In all of these issues, the error or anomaly is the catalyst for acquiring a deeper understanding of the context in which the data issue emerged. While categorization is an aspect closely tied with instrumentation, it is clear in these examples that the issue stems from the agency, database, or even data collector responsible for data production. In the following section, I argue that there are three major inversions that the scientific programmers conduct to reconcile scalar issues: 1. Cross-referencing; 2. Locating distributed expertise; 3. Flagging.

Strategies: How scientific programmers invert research infrastructures to reconcile anomalies in data work

To continue developing an understanding of how scientific programmers encounter scale in their data work, I now categorize their actions for reconciling the data anomalies and errors they uncovered. A typical workflow involves a scientific programmer discovering an anomaly (something mundane like a typo or a gap to something more unusual like a seemingly large or small figure). This leads the programmers to consult experts 'in the field'

– those who have a domain knowledge of why such errors occurred. The final action includes flagging a dataset with information about potential errors or anomalies in the data. This is a scalar strategy that emphasizes the preservation of uncertainty over making erasures or additions.

Furthermore, the discussions around issues in the data stretch beyond data cleaning and quality assurance. Using the data life cycle model (table 5) as a sensitizing concept, I highlight how much of the work the scientific programmers take on falls outside of the categories of the life cycle model, which largely include tasks to resolve errors such as visualizing inconsistencies, locating distributed expertise, and consolidating research into a flag. For a full depiction of my codebook, see appendix A.

Data lifecycle model (Strasser et al., 2011)		
Stage	**Description from the literature**	**Example from my data**
plan	"data management planning"	*Is region appropriate? Region may create a lot of redundancy. Perhaps focus on critical regions instead (Bristol Bay; AYK; Cook Inlet; etc)? Regional Governance data can be linked to permit data and stock assessments, but this would be a "case study" approach instead of a broad level view across the state.*
collect	"ecological data are collected and organized in many different ways, including manual recording of observation in the laboratory and field via hand-written data sheets, tape recorders and hand-held computers; automated data collected via laboratory and field instrumentation; satellites and aerial platforms; and, increasingly, sensor networks that are embedded in the environment."	*An issue I ran into is that the count number of permits by age range (for example, 20-30) were all characters/factors* ***because for low values the data collectors used "1 to 3" instead of 1, 2, or 3****. This was dealt with but I was unsure about what to change "1 to 3" to so that the columns could consist solely of integers--I took the average and changed them to 2 for now.*
assure	"QA/QC refers to the mechanisms for preventing errors from entering a data set that are used a priori to ensure high data quality before colletion and to monitor and maintain data quality during and after data collection"	*We are asking for the* ***"raw" data files, so that we can employ consistent QAQC methods across all the data files that we receive.*** *However, we understand that the data collector knows best when it comes to QAQcing data, so if you have cleaned data*

		files that you would like to provide that is great.
describe	"metadata documentation to understand the content, format, and context of a data product"	*Friday, [redacted] noticed that at this point it is difficult for WG members such as ourselves to go from the daily count data set back to the originals, and locate the* ***associated metadata****. As we continue to have multiple people reformat data, it appears having a plan that everyone is on board with is to be pertinent to moving forward (or at least clarification of the what has been happening so far).*
preserve	"deposition of data and metadata in a data center or data repo"	
discover	"sophisticated, user-friendly search tools that enable scientists to search by time and space and also drill down further using faceted search techniques that allow one to filter the results by parameter, sensor employed, author and other properties of the data"	*By the way, is you want to explore the data, the shiny app is now done for Age and Length. Maybe the next step is to add an option to check by sex too?*
integrate	"integrating source data from such studies is labor intensive and time consuming, because it requires understanding methodological differences, transforming data into a common representation, and manually converting and recoding data to compatible semantics before analysis can begin."	*A variety of methods of estimation were attempted prior to 1999. Most recently these have been modeled from a variety of different sources.*
analyze	making analysis reproducible: "scientific workflow systems provide an executable and complete description of analytical procedures that allows scientists to link together processes drawn from multiple different analytical systems."	

While infrastructural inversion has been proposed as a tool for scholars of infrastructure (e.g., Edwards, 2010; Hahn et al., 2018; Parmiggiani et al., 2015), Bowker's (1994) original usage of it was as an action taken by those he studied. Clarke and Fujimura (1992) take on the idea of scientific inversions paying close attention to the ways in which material aspects of scientific research infrastructure enable or constrain researchers. Similar to Bowker's

Table 5. Data lifecycle stages, description, and examples from the data

original formulation of infrastructural inversion, my study of the scientific programmers' work practices could be considered an inversion as I utilized their articulation work to make sense of scale. However, I argue that my approach did not constitution an inversion. On the contrary, it was their work of bringing together multiple, sometimes conflicting, resources, of stitching together distributed pieces of expertise, of creating visualizations that highlighted anomalies in the data, and of flagging a data set for downstream users, that constituted infrastructure inversion. I unpack the three kinds of inversions that the scientific programmers enact: inverting documentation, expertise, and the dataset itself. It is here that I show how these inversions map onto the different layers of an infrastructure.

Inversion	**Examples from the data**
Inverting documentation	
Cross-referencing reports	*Digging deeper into our available location information for ASL data - most of the lat/lons for escapement projects that we currently have are at the end of AWC streamcodes and not at the location of the actual weir. Getting more exact location information for escapement data looks like it will be a difficult task.* ***As a shorter-term solution to get analysis off the ground, I am focusing on populating samples from escapement projects that don't currently have AWC codes with that information.*** *Samples from harvest from cities/villages have lat/lons already, and all commercial catch samples are linked to stat-areas, so we are getting close to having nearly all of the data spatially referenced.*
Identifying redundancies	*I have merged annual sums of escapement data* ***from all four sources (M&V, AYKDBMS, Sportfish)*** *and placed the resulting file in the google drive here. Notably, many of the sources that report on the same project/species/year do not agree - @[redacted] and I are both going to be doing some investigating as to why.*
Inverting expertise	
Consulting expert on historical instrumentation	*In short: inriver-mark recapture estimate minus upriver harvest = escapement estimate. A variety of methods of estimation were attempted prior to 1999. Most recently these have been modeled from a variety of different sources. Please see below from [redacted] (cc'ed), who is the authority on Copper River Chinook escapements. For his master's thesis, [redacted] reconstructed historical (pre-1999) Copper River Chinook escapements. However, he has recently updated these estimates using Bayesian methods that take into account multiple sources of*

	*inference (**mark-recapture, proportions of Chinook in upriver harvests, Gulkana counting tower, black magic**, etc.).*
Brokering or sharing expertise	*The juvenile data out there should be a fraction of that existing for adults. Still, there have been (and continue to be) a few different juvenile projects around the state for which size data have been collected.*
Inverting the final dataset	
Flagging	*I think that is the way to go and instead of remove them, **just flag them as problematic,** in case in the future we know a bit more and we want to include it in the analysis.*

Table 6. Additions to the data lifecycle model: How scientific programmers invert infrastructure to reconcile errors and anomalies in the data

Most of the issues that I highlighted result in a moment of closure in which a flag is added to the data file. This flag encompasses hours of work spent trying to figure out what caused the error in the acquired data. The typical workflow of flagging is as follows:

1. identify the error (e.g. visualize the dataset to find inconsistencies, merge with other files)
2. locate distributed expertise to question about the now potential error (e.g., reach out to subject matter experts who are involved, read through old reports)
3. flag as a mistake if deemed obviously a mistake. Otherwise, preserve a final reason for the flag (e.g., fell outside of reasonable age ranges or appears to be erroneous based on what experts know).

Identifying inconsistencies and redundancies through visualization and cross-referencing, an inversion of existing documentation

One strategy - or inversion - that the scientific programmers take is to merge disparate sources to see where misalignment occurs. This can be through the use of software tools that allow for the merging (e.g., Shiny Apps in R) or it can be through the manual cross-referencing of multiple documents. Errors in the disparate data sets include issues such as

inconsistent standards, missing data, anomalies in instrumentation, changes in data format or otherwise seemingly out of place data. Discovery of these errors is part of the data cleaning process, or Quality Assurance/Quality Control (QA/QC), and in many instances, it involves research into the provenance of data.

Variable temporal coverage is a major issue that the scientific programmers encounter when attempting to create a seamless dataset. It is not uncommon to find inconsistency with respect to time ranges of the data; there are some daily data, some annual (with no daily counts), and "the availability of daily counts differs based on region and time-frame." One of the scientific programmers updates a file to note that escapement data has been acquired from all regions but the "temporal coverage is variable." The duplicates from the data issues are sometimes just deleted; however, duplicates and typos can also be checked systematically. To identify errors such as duplicates or typos, the scientific programmer looks for conflicts in the data:

> *I went through and did a few checks and found some conflicts where there are multiple rows for the same location name but a different region name. Some of these appear to be real – Salmon River in both Kotzebue and Kuskokwim for example, but some look like errors. I attached a file with the duplicates.*

She goes on to note the cross-referencing of other documents:

> *We will need to start implementing a unique identifier for each location. We can either piggyback off the ADFG LocationIDs or use something from the AWC (Anadromous Waters Catalogue).*

Getting rid of duplicates involves restructuring and combining data. Generalizing to a wider region can create duplicates for sample locations on the same dates. Domain experts involved in the study often suggest digging into reports to do some cross-referencing:

> *Commercial harvest data is most consistent source for weight data but it has issues that make it unreliable (at times) – in general it is reported by the processors and at times they just fill-in what everyone accepts to be the average weight without actually weighing anything. The hatcheries monitor weight and fecundity during egg take each year – not sure if that data is archived or if they would share it – I've looked at some PWS hatchery pink salmon fecundity data in the file I sent you ([redacted]) but it is suspect as well…* ***Hopefully while digging through all this data some projects with additional measurements will surface****.*

In the comments, one scientific programmer notes that the available location data is only located at the end of a stream code. A stream code comes from the Anadromous Waters Catalogue (AWC), which is the ADF&G's 'catalog of waters important for the spawning, rearing, or migration of anadromous fishes.' This catalog exists because of AS 16.05.872 protection, which states that for bodies of water to be protected they must be documented as supporting life function of an anadromous fish species. There is a nomination process to consider if a stream should be listed. Because these location data are not at the site of the actual weir, getting the exact location information is difficult. As a short-term solution, they focus on "populating samples from escapement projects that don't currently have AWC codes. Samples from harvest from cities/villages have latitude/longitude information already, and all commercial catch samples are linked to stat-areas, so we are getting close to having nearly all of the data spatially referenced." This approach of bringing together multiple data sets from multiple reports is a common strategy the scientific programmers take to filling in data gaps. It is essentially cross-referencing to find other reports where the missing information might be hiding.

While discovery in the data life cycle model (Michener & Jones, 2012; Strasser et al., 2011) is considered the action that occurs when research scientists are able to access data,

there is a large amount of *error* discovery that occurs among the scientific programmers when cleaning the data. In the salmon project, this was most commonly done through data visualization tools such as Shiny Apps, an open package from RStudio which allows for interactive web visualizations.

Inverting expertise

The data task force team is often operating at the intersection of general and local expertise. While the tenets of data science such as reproducibility, interoperability, and metadata come with themes of generality, much of the work required to make data amenable to interoperable usage demands a highly local understanding of the data before archiving it. In many of the issues, the scientific programmers reach out from Santa Barbara to an expert in Alaska who can answer questions about the data particularities.

In an example about conflicting standards that had been used, the issue is reconciled by reaching out to a subject matter expert who notes that if "*district name or number are included in the ASL data, these could be cross referenced with a GIS database of commercial fishing districts…Alternatively, if it was an escapement sample, then the lat/long can be obtained from the Anadromous Waters Catalogue*". While these comments are not explicitly about the data format itself, they refer to complexities and potential for error within the data ecosystems.

When filling in gaps that occur over multiple years, the scientific programmers hunt down different reports to piece this information together. One of the authorities on these reports comments on the challenge of tracking numbers down before 2001 noting that while the reports provide a good "historical synopsis", the temporal coverage varies. He goes on to note: "*documentations of the history for others might not be as easily found. You would also*

have to be aware of changes in terminology and point goals vs. ranges, lower bounds and thresholds." These data are often hidden in escapement goal reports such as annual management reports or season summaries.

As part of the work of locating expertise, the scientific programmers are frequently piecing together multiple reports found online or asking distributed experts for guidance about how to handle the data. For example, in an issue with ASL data, the scientific programmers begin with visualizing data to identify outliers. A domain expert chimes in to confirm removal of a few large measurements noting that "if they were in cm, a 1000+ cm Chum (salmon) would be a Boone and Crockett contender." One way this issue is resolved is through the establishment of reasonable ranges of length data much in the way that they established reasonable age ranges from a prior example. This essentially builds into the creation of a flag—anything outside of the reasonable range is flagged as potentially suspect.

In some instances, the scientific programmers acquire the data from a data holder who introduces the data with its various known complexities. One example of this is in the hatchery harvest/returns data. A domain expert notes that "the tricky part is separating hatchery vs. wild harvests for individual fishing periods in specific districts, etc…We could attempt to do this, but it might involve a lot of work cleaning up datasets for confidential information." In most situations, the identification of errors leads to a deeper exploration of the underlying causes.

Flagging: inverting seamlessness to highlight potential errors

Flagging is the most commonly used action as it often is the final step in the reconciliation process. Flags are often coupled with multiple types of errors (duplicates, typo, unknown, missing). Though not unprecedented, the scientific programmer rarely applies

transformation algorithms to the data. Instead, they avoid adding potential new errors by adding a 'flag' at certain points to consolidate the research they have done about the data and to denote that the data might be erroneous, duplicated, or inconsistent. The flag is also written in the metadata file appearing alongside information about methods and standards. In the following instance, a scientific programmer added a flag to denote to downstream users that an error may have occurred and included information about the use of standard notation in the data.

While acquiring age, sex, length data, the lead scientific programmer remarks that the ASL age data seem suspect. In concert with one of the domain experts, they specify that age classes are written as total age which is the number of winters in freshwater plus the number of winters at sea, and develop a list of "reasonable ages" for the different species involved:

Fresh Water Age

- Pink: 0
- Chum : 0
- Chinook: 0,1 and 2.
- Coho: 1, 2, 3, and maybe a few 4.
- Sockeye: 1, 2, 3 and 4.

Salt Water Age

- Coho: 0 and 1
- Pink: 0 and 1
- Chum: 0, *maybe some* 2, 3, 4, and 5.
- Chinook: 0, 1, 2, 3, 4, 5, 6, 7 and 8.
- Sockeye: 1, 2, 3, 4, 5, 6, 7 and 8.

They note that anything older than those fresh water and/or salt water ages need to be tracked to their original files. They write up the following script (figure 17) to flag for unreasonable ages.

the flagging script (ASL_merge_flag) has been changed. These two functions were added. Script was run and a new master file was created at the same path as usual

```
    #Fresh Water Age

    #Pink: 0 not flagging for now
    #Chum : 0
    #Chinook: 0,1 and 2.
    #Coho: 1, 2, 3, and maybe a few 4.
    #Sockeye: 1, 2, 3 and 4.

    sock <- c("sockeye")
    chin <- c("chinook")
    chum <- c("chum")
    pink <- c("pink")
    coho <- c("coho")

    sockeye <- filter(aslAYK, Species %in% sock)
    sockeye$Flag[which(sockeye$Fresh.Water.Age > 4)] <- 'FWAge'

    chinook <- filter(aslAYK, Species %in% chin)
    chinook$Flag[which(chinook$Fresh.Water.Age > 2)] <- "FWAge"

    chumdat <- filter(aslAYK, Species %in% chum)
    chumdat$Flag[which(chumdat$Fresh.Water.Age > 0)] <- "FWAge"

    cohodat <- filter(aslAYK, Species %in% coho)
    cohodat$Flag[which(cohodat$Fresh.Water.Age > 4 | cohodat$Fresh.Water.Age < 1)] <- "FWAge"

    pinkdat <- filter(aslAYK, Species %in% pink)

    aslAYK2 <- rbind(sockeye, chinook, chumdat, pinkdat, cohodat)
    return(aslAYK2)
}

AgeFlagging_SW <- function(aslAYK){
    #Salt Water Age
    #Coho: 0 and 1
    #Pink: 0 and 1 not flagging for now
    #Chum: 0, maybe some 2, 3, 4, and 5.
    #Chinook: 0, 1, 2, 3, 4, 5, 6, 7 and 8.
    #Sockeye: 1, 2, 3, 4, 5, 6, 7 and 8.

    sock <- c("sockeye")
    chin <- c("chinook")
    chum <- c("chum")
    pink <- c("pink")
    coho <- c("coho")
```

Figure 17. Flagging script

In the process of defining what data are *erroneous*, the data team flags both those data that are inaccurate but also data that has used a different standard or measurement than what is most commonly used.

This does not engage with the work that actually occurs downstream at the level of the repository, but is strictly about how data is made ready for integration into such a repository. In focusing on moments of closure, it is apparent that there is a temporal aspect to this work

both in terms of engagement with the past but also preparation for the future. Scientific programmers are dealing with both varied time scale in terms of how time is actually captured in the dataset as well as uncertainty about what the future will hold. These are ultimately concerns about epistemology, or how knowledge is produced in different eras. Uncertainty about changes through time is a primary concern for the scientific programmers: both uncertainty about how data are represented from the past and uncertainty about how data might become useful in the future. This is compounded by missing data either caused by an error or simply that the data was not collected in the past. What ends up being archived in the data discussions are concerns with temporal certainty rather than how durable standards are across variation.

Conclusion: Finding scale in inversion

The salmon are incidental was the first quote I scribbled down after an introductory call with one of the Principle Investigators on the project. However, repeatedly the message from stakeholders on the project was that the salmon are anything but incidental. The statement has come to indicate a general expertise as opposed to something local or particular. To the funder interested in furthering environmental data science: the salmon are indeed incidental. Methods could just as easily be applied to oceans, snow leopards, or zooplankton. Applying standards that will make working with environmental data easier—more streamlined, more seamless—is their question of interest, not the specifics of salmon management or salmon science in Alaska.

Although the work of the scientific programmers entails reconciling spatial and temporal aspects of data, their day to day work is more focused on different scales of infrastructure

(e.g., knowing where to locate certain reports and who to contact, adding informational signals for downstream users). Following the technical practices of data work, this chapter ultimately contributes some implications for design of the data infrastructures that house ecological data identifying some challenges as a jumping off point:

1. There is a tension between the data professionals and the research scientists, which is often fostered by different valuation of the work, different funding priorities, recognition of the necessity, and visibility. This has encouraged data organizations to cast relatively routine and mundane daily work as something new. In the SASAP data task force summary, they position themselves as critical of proposals to fund an "ultimate database" because those are often from domain experts rather than experts in information or data science. They argue that synthesis studies often fail to meet expectations given how the requirements of data integration are often underestimated. Their proposal echoes a modern concern, which is that a) databasing and data cleaning work is tedious and labor-intensive and as such, should be adequately funded; and b) data are a kind of politics in the sense that no one *should* exert complete control or ownership over them. This second point is based on an argument that aligns with the position that many advocates of open science and open data make—the idea that more data will lead to better science.
2. Data acquisition draws from pre-existing data and as such, are contending with the initials plans for those data. The data collected for this SASAP project were mostly data collected by Alaska department of fish and game.

3. Most data cleaning work involves a great deal of scientific work of tracking down anomalies and reconciling errors. Much of this work falls under the 'assure' and 'describe' stages; however, many of the errors the scientific programmers encounter can be traced back to the moments of data collection.
4. Scalar issues often occur when there are misalignments between the inner scale of the phenomena and the exterior scale of the data.
5. 'Scaling' is a primarily temporal activity whether it is in the management of the present, in predictions of the future, or in understanding historical change. The most commonly noted issue when working with the data is ensuring adequate and consistent temporal coverage. Attempts to acquire adequate temporal coverage bring up issues around temporal uncertainty, ultimately an issue of provenance, or of knowing from which the data came and what is its intended usage. Temporal uncertainty is one of their primary concerns, that is, both the uncertainty of representation of data from the past and the uncertainty of data that will become useful to the future. In other words, these scientists display a sensitivity to the many ways data have been misunderstood in the past and could be misunderstood in its future usage. This is what Ribes and Finholt (2009) have called the "long now", concerns for potential futures are often considered in daily work.

Despite the generality of the strategies, the scientific programmers are constantly wading through particularities in Alaska salmon data. Field sites—or, sites of data collection—come with their own particular practices, cultural norms around technical expertise, and distributed knowledge not always captured in metadata, but rather shared around a dinner table or held within the data collectors themselves. Even still, database designers and

engineers are central to improving natural resource data management, deploying instruments and standards designed to achieve data interoperation.

These findings show that in the reconstruction of context (e.g., documenting through metadata), the move toward universality requires engagement with local experts. However, what ends up being preserved in the final dataset is the uncertainty that occurred to produce an issue with integrating data, often caused by scalar issues. This approach is one that makes the potential error actionable – in other words, it flags the error for a user downstream to do more with it. This is partially due to the challenge of correcting many errors post-data collection. For example, if data are missing for a particular year due to a funding constraint, those data cannot be collected retroactivity.

The major scalar dimension that this empirical investigation produced was that errors or anomalies occurred when the internal representations of space or time (e.g., salmon run upriver to spawn for reasons related to salmon biology) were mismatched with the external representations of space or time (e.g., the date, calendars, streamIDs). This points to the challenge of capturing scale when translating a research phenomenon to a data point. At a moment when data science is becoming increasingly popular for usage in the natural sciences, it is important to better understand the scalar challenges that scientists have struggled with for decades.

In the following chapters, I trace the wider network of scientists that I met through this synthesis work. As I have shown, major issues of scale in ecological knowledge production occur in the field when producing data. As such, this led me to sharpen my focus on how scientists instrument scale in the field. In the following chapters, I will show how scientists instrument temporal and local scale.

Chapter 6. Salmon specimen: The material production of salmon life cycles

Migrating at night while hunter dreams / the salmon had followed / the scent of this creek home / a pilgrimage repeated faithfully by / his ancestors since the last ice age. – Tom Jay, in Reaching Home: Pacific Salmon, Pacific People.

Introduction

This chapter discusses how scientists instrument time to understand change at different resolutions. Specifically, this chapter outlines how scientists use specimens to produce both fine and coarse temporal scales. This chapter is concerned with how researchers develop an understanding on longer timespans than a human lifespan, and as such, looks at what is required to instrument time (e.g., work across decades, standards for data collection, changes in scientific knowledge).

In this chapter, scientists in a long-term field program produce fine *and* coarse temporal scales with specimens. These were not initially a part of the long-term data collection but have been added throughout the years. I argue that these changes occurred because of alignments between the instrumentation and the phenomena itself. Further, this kind of fitting between instrumentation and research phenomena is characteristic of a long-term research program. This is a distinct difference from predominantly data-centric organization, which has goals that are focused on broad-scale synthesis.

I discuss three different specimens (figure 18) that are used to sample for temporal scale. As such, I trace the moments in which this instrumentation and specimen collection was added to the research program and hypothesize why such change occurred. In this section, I

ask: *How do scientists change long-term, or legacy, infrastructure to meet contemporary concerns? And, how does an ecological field program scale for time?*

Figure 18. Three specimens and the scale produced.

To explore how data are produced and how the infrastructure that supports that data production is sustained through time, I conduct an ethnographic study of a long-term field program in Alaska. Rather than arguing for how data are made useful as evidence, I explore how evolving theories shape the instrumentation in a long-term field program. The series of images and fieldnotes give a view into how data have been shaped through time. Some archival images date back to the beginning of the program while photographs from the field come from my own participant-observation as a volunteer field technician.

Arriving as an ethnographer of science

My research took place during four record-breaking years for salmon runs in Bristol Bay and one record-breaking year on account of the extreme heat and drought. In the summer of 2017, the Wood River system had a 4.5 million escapement and by the time I arrived in August of 2018, there had been a 7+ million escapement. 2018 was a record year with 62.9 million escapement to the entire Bristol Bay region. My last year (2021) broke all records of returning run-size with a whopping 63.2 million fish returning to their spawning grounds in Bristol Bay. 2019 was characterized as one of the hottest summers on record with temperatures rising higher than predicted 50 years from now. The Wood-Tikchiks were hot and dry and many salmon could not make it up to their spawning grounds due to low water levels.

The state-protected area of Bristol Bay is considered exceptional among scientists, particularly fisheries biologists and ecologists. While the pristine quality has imparted the program with the description, "living laboratory", scientists there look to understand aspects about the habitat that lead to salmon abundance, such as stream temperature, gravel size, water temperature, water chemistry, and stream gradient. To develop an understanding of the important aspects of habitat and salmon abundance, they account for the environment with data such as chemical composition of the lakes, the number of juvenile fish present, and what sections of the stream have the majority of fish. This is the essential work of producing this kind of data and research infrastructure – the logistics, justifications, and social dynamics.

In other words, because it is so pervasive, data and its attendant infrastructure are backgrounded, not a thing to be studied but part of the everyday functioning of a field

program. The research support infrastructure has become so backgrounded and a part of everyday life, that it is both essential and mundane (Bowker & Star, 1999; Pollner, 1987). This was evident throughout my time in the field: data were everywhere, loosely tracked in a 'rite in the rain' field note book; people would critique models used for different analyses while walking up the stream; gas tanks were stashed in the bushes to refuel our boat when making a long trip to the upper lakes; the generator hummed every couple of days allowing us to run water, clean dishes, and process otoliths.

While producing long-term data records is still a primary focus of the field work, additional specimen collection has been added along the way. In the following section, I highlight two vignettes about specimen collected for the study of salmon: one vignette about re-instrumentation and one about re-purposing specimen collection to achieve finer-scale understanding. In the first, I tell the story of how genetic research came to fisheries science and how the collection of fin clips as genetic specimen began to be collected in the program. In the second, I show how the usage of otoliths—ear stones that can be used to age fish—were re-purposed to understand salmon at a finer temporal scale. I show how this addition to the field data collection was aided by its fitting to the life history of the salmon itself. Notably, these kinds of data are left out of the SASAP archive suggesting that a data organization focused on synthesis embodies a different kind of temporality than one focused on long-term field inhabitation.

However, this chapter is more focused on *why* an infrastructure changes over time as a means of understanding how scientists scale. Although it is clear that long-term observations can lead to important findings, long-term sampling protocols and archives are not static. Rather, the field program – its data, its students, the research it produces – is constantly

shifting to meet contemporary scientific needs. It is this flexibility that has made the research program central to scientific knowledge production.

My ethnographic account is specifically focused on why and how a research infrastructure changes. My argument is not intended to suggest that there have not been important discoveries based on the long-term data collection. On the contrary, there have been important findings come out of the everyday long-term data collection (e.g., Cline et al., 2017; Schindler et al., 2010) as well.

Re-instrumenting the field: Producing temporally-distinct salmon runs

In this section, I argue that new instrumentation was applied as it became apparent that salmon were conducive to new instrumentation. This is not to suggest that change was accidental but instead to show how flexibility in changing instrumentation and testing new theories is central to a research infrastructure's long-term relevance. Additionally, I show how commercial interest in understanding discrete categories contributed to re-instrumentation in the early development of the field program.

Advancements in technology specific to fisheries genetics made new research possible. However, it also advanced because the object of study itself – the salmon -- was tractable to the research methodology. In other words, life cycles and life histories of salmon aligned with the way molecular biology was evolving as a field.

This exemplifies a case in which the research site was instrumented with a new technology: genetic sequencing. While the theoretical underpinning of population genetics began with Darwin in the 1800s, it arose in fisheries in Alaska in the 1970s as a burgeoning research field. In studying the data journeys or workflows in *human* population genetics, Griesemer (2020) identifies tissue specimen collection as the starting point in which

specimen are then turned into data and put into motion. This workflow is concerned with what happens after data are circulated in various social worlds (communities). In contrast, what I focus on in this section is how specimen collection are deemed necessary in the first place – both in the social as well as technical spaces.

Genetics research was widely contentious with respect to human genetics; however, the application of genetics for understanding non-human pasts was fraught with its own contention[19]. Scholars (e.g., Felsenstein) have shown how theories of evolution such as natural selection and mutation became tractable after mathematical models made them so. While early observers noted morphological differences, the taxonomic distinction solidified this as knowledge. Going back to the 1990s, the Alaska Salmon Program began collecting fin clips more frequently to create a long-term genetic dataset. Much of that early research was around actual methods for understanding salmon genetically (e.g., Seeb et al., 1986) and many of those techniques are continued in the field today as part of the everyday sampling protocols (e.g., figure 19).

[19] Most STS writing on genetics has explored *human* genetics, e.g., the ethical implications of studying genetics and race, genetics and disease etc… For example, Alondra Nelson writes about genealogical testing and how debates arose about whether genetic markers should be used to distinguish human groups on the question of race showing how pragmatists argue that race is a social construct and naturalists argue that differences are real, not constructed. Her argument ultimately gets at a perennial STS thread of epistemology and ontology, exploring scientific objectivity and social constructivist views of a burgeoning field, and how information is made political.

Figure 19. Taking fin clips from live salmon in the field. 2021.

Research techniques and theoretical concepts were closely entangled throughout the history of genetic research. Much of the early fishery genetics research began in Seattle in the 1960s with Fred Utter in the ancestor lab at NOAA's Northwest Fisheries Science Center. This work was informed by concepts developed in the 1930s such as "genes-on-a-string" (Beadle & Tatum, 1941). Grant (2021) tells a story that starts with the most popular population genetic equation of RA Fisher, JSB Haldane, and S Wright (FHW), which came with assumptions of natural selection, the buzzword of evolutionary biology. Countering this mainstream assumption, Kimura (1968) postulated that most evolutionary changes "were neutral to the effects of natural selection." Natural selection was not always focused on survival of the fittest, but sometimes allowed for random occurrences, called gene drift or random drift. This did not signal a dismissal of natural selection, but rather an added layer of complexity to mainstream views that evolution would always favor the most efficient, productive, or otherwise preferable traits.

With new molecular techniques came new research questions and new complexity to grapple with. As genetic research evolved, a method called protein electrophoresis was developed as a way to analyze proteins, mostly enzymes, by extracting live enzymes and identifying the population genetic components through a series of staining cocktails. Enzyme staining was a long process. In fact, most genetic sequencing was tactile and difficult in the early developments. Because this method relied on blood protein, animals were well-suited to this kind of study (as opposed to plants). Furthermore, this method was conducive to the study of salmon population genetics because "populations were isolated by faithful homing to natal spawn sites and because salmon populations were small enough to experience random drift that produced differences between populations" (Grant, 2021). Because salmon had small local populations, it was possible to find out where the population came from with a limited number of markers. This is because they exhibit geographic population structuring (McGlauflin et al., 2011; Miettinen et al., 2021). In contrast, marine fish who are broadcast spawners[20] are much more difficult to detect locally-specific populations. These points reaffirm my main argument which is that aspects endemic to salmon life histories made them conducive to genetic study.

Secondly, salmon were tractable to this kind of study because of their value (Allendorf, P.C.). They were of high commercial interest during the time that genetic research was emerging, and given their life history of leaving freshwater to spend adulthood in the ocean before returning home to spawn, there was concern about international pressure on the wild fisheries. At the time, the major concern was that Japanese fishers would intercept salmon

[20] Broadcast spawning is a reference to a kind of sexual reproduction in which organisms broadcast eggs into their environment for external fertilization. Mushrooms are known as broadcast spawners, but many fish are as well. Salmon, however, are not.

bound for their home in North America. This commercial concern is perhaps why management was so enmeshed with population genetics early on while it was still a nascent field. Today, managers rely on findings that distinguish different salmon populations. The idea of discrete categories of different populations congealed as scientists began to look at early-run vs. late-run salmon as distinct populations (e.g., spring Chinook vs fall Chinook). And Anthropologists have shown how these distinctions of different timing were important to Indigenous harvesters as well (Swezey & Heizer, 1977).

Very recent studies (Narum et al., 2018) have pointed to genetic differentiation between different timed stocks (e.g., spring run Chinook vs. fall run Chinook). Geneticists have also posited theories that salmon species diversity reaches far back in time. For example, Waples et al. (2008) noted that repeated and sudden habitat changes would have impacted Pacific salmon evolution, particularly the creation of diverse subspecies.

A turn toward genetic sequencing has changed the way science is done on these fish and even how the fish are managed. Perhaps most profoundly, the Alaska Department of Fish and Game (ADF&G) instituted a genetic policy dating as far back as 1975, driven by initial concerns about the potential deleterious events hatchery-raised fish might have on wild stocks. Genetic research promised to illuminate those potential effects. In other words, burgeoning genetic research highlighted fine-scale and broad-scale differences between what was once seemingly the same fish. These differences, evident mostly in the data, inspired management agencies such as ADF&G to develop management policies for managing wild fish populations based on this new complexity.

Re-purposing: From age composition to habitat mosaics

In this vignette, I discuss two specimen—otoliths and sediment cores—to show how material samples are used to reconstruct the past, both at fine and coarse scales. Over the years, physical samples of the fish, such as scales, otoliths, and fin clips, have been used to understand salmon more deeply. In the case of genetic research, I showed how a new instrument and subsequent specimen collection was introduced into a long-term field program because of two main alignments: the salmon itself was well-suited to genetic instrumentation, and the need for understanding discrete populations as a growing commercial fishing industry emerged.

In this section, I show how a specimen that had been collected for a specific reason – to age fish – was re-purposed to measure aspects about salmon migration and habitat at deeper timescales. This vignette illustrates the ways that material artifacts are like an archive used to produce understandings that reach beyond the observations of a single human lifespan.

As opposed to a re-instrumentation of the field as occurred with genetic sampling, researchers have also turned to other disciplines to make sense of migratory behavior. While Brown (1994) argues that at larger scales, ecology shares more in common with disciplines like astronomy, geology, climatology and at smaller-scales they align with chemistry, biology, I show that ecologists--while drawing from fields such as geology to answer larger-scale questions--are also drawing from their own fields of limnology, chemistry, and biology to answer these large-scale questions.

Time at fine scale

Although not altogether re-instrumenting salmon in the way that genetic research did, researchers looked at strontium isotope ratios in the otoliths, a specimen that has been

collected as part of salmon research for decades, to trace where salmon journey when they return to freshwater. This approach to developing a scientific understanding of migration is built on the idea that physical specimen of the fish tell their own story about where the fish has been.

Otoliths (figures 20-22) are small calcium carbonate deposits beneath the brain, much like a rock or stone within the fish's ears. These stones accumulate layers (or rings) as the fish grows, incorporating microchemical signatures from the water they live in, which is evident in the rings within the otolith. These otoliths[21] are collected by researchers in the field program. Below is an excerpt from my fieldnotes from summer 2019.

> *Although, my seasoned companions eat kippered snacks on the beach after finishing, the smell of dead salmon is still lingering on every item of clothing months after returning home. After piling all the dead fish on a bank baking in the sun, I slice open the head of each fish right behind their eye. I do it so many times, I forget this oozing slimy carcass was once full of life. Once I can see the brain, I gently use my forceps to feel for two hard pieces nestled in the corners of the brain, often making a scraping sound when my tweezer runs over it. My first time, I do this for an entire creek with one of the other students. By my third year doing this, I'll have gotten pretty quick developing what feels like an ability to sense where exactly that tiny piece of stone sits inside the inner ear of the salmon head, to listen for the scraping sound of my tweezers running over the hard piece of stone.*

Many of these otoliths will be used to determine the age composition from spawning ground samples based on methods that were developed decades prior (Koo, 1962). Because otolith sampling has been a long-term interest of the program, there is a protocol for such sampling.

---

[21] Otoliths are collected during carcass surveys, which occur from July to September when the fish are dying. Salmon start coming back to their spawning grounds in droves. They spawn until they die naturally (senesced) or are predated on. Beaches and streams are littered with carcasses at various stages of rot. Carcass surveys have been conducted since the beginning of the program and involve counting males and females dead, how they have died, and then collecting 110 otoliths for both males and females. That is a total of 220 salmon heads they have to be sliced open and dug into with a pair of forceps.

In the images below, the manual used for aging fish with otoliths shows two examples—the first photo gives guidance about how to age the fish while the second is an example of an otolith that is too opaque to age properly.

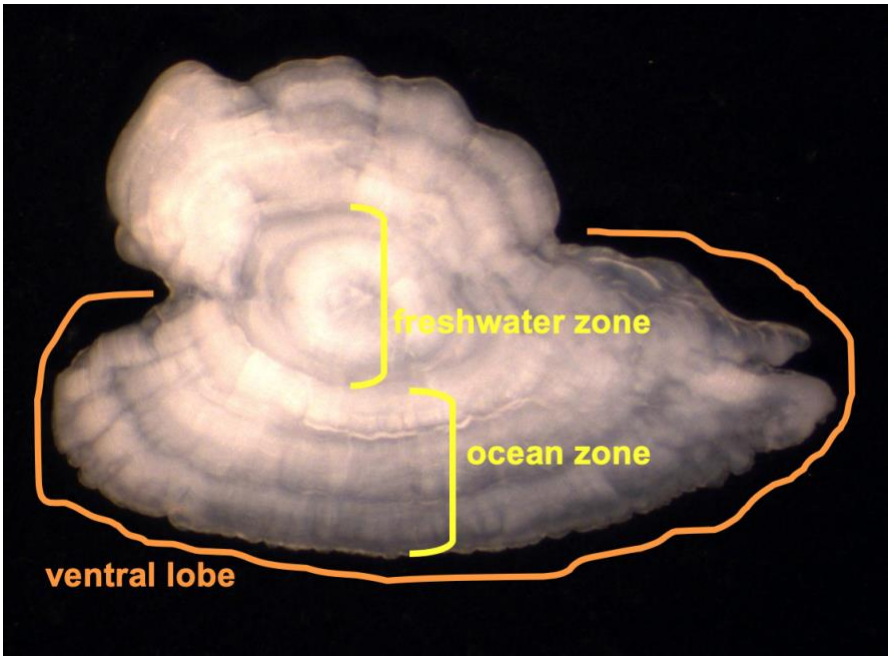


Figure 20. From the Sockeye Otolith Manual (ASP). This image highlights how to age the fish based on years spent in freshwater and years spent in the ocean.

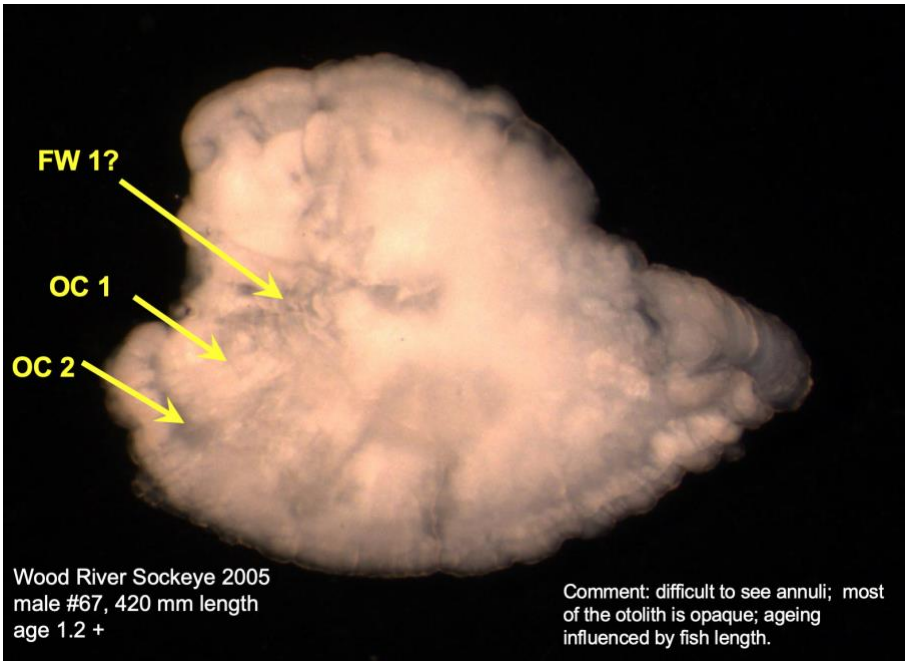


Figure 21. the manual shows an example of an otolith that is too opaque to be useful for aging.

Back at the camp, I process these otoliths by cleaning their encasing membranes off with soapy water and putting them back into their vials to be sent back to Seattle.

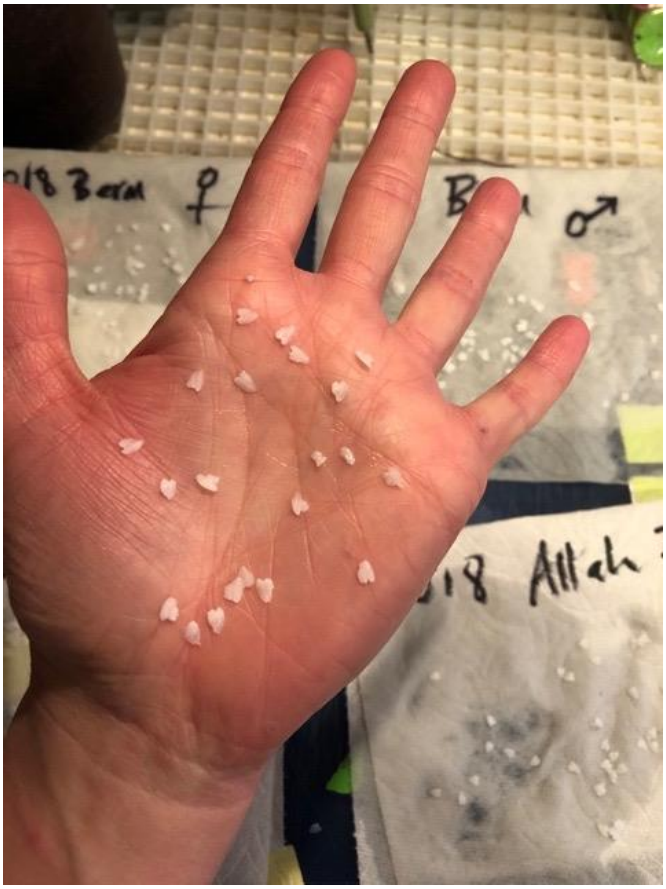


Figure 22. Processing otoliths back at the field camp (hand for scale), summer 2018.

For processing the otoliths, the manual states:

> *When sockeye salmon otoliths are collected on the spawning grounds, they are wiped clean and placed dry in either paper coin envelopes or plastic cryotubes. When otoliths are examined for age determination, they are placed whole (without grinding) in water with the lateral side facing upwards in a shallow, opaque, black dish. Otoliths are examined under reflected light with a dissecting microscope at approximately 32X…Age determination is accomplished by identifying annular marks on the otolith. When examined under reflected light, annular marks appear dark. The annulus is a mark put on the otolith once a year, signifying a period of slow growth, usually considered to represent wintertime growth. Ocean annuli often appear as a depression, or groove in the otolith surface.*

While aging was the dominant reason that otoliths were first collected in the field, they have since been repurposed to develop insights that reach farther back in time. Reviewing the 862 otolith-oriented papers that had been published since the 1998 Otolith Symposium in Bergen, Norway, Campana (2005) shows that about 40% of those papers cover "annual age and growth" while the remaining papers look at otolith microstructure, otolith chemistry, and non-ageing applications. While initially otoliths were used to age the fish, they are now at the center of novel studies (Brennan et al., 2015, 2019) that draw more on geology and chemistry than biology or ecology. Essentially, the fish's migration patterns are tracked by following the chemical signature left behind in these ear stones. These chemical signatures are from strontium isotopes, trace chemicals found in the bedrock. These help to tell a more complex story about life history.

Just as genetic research led to a richer understanding of the local complexities in salmon stocks, this new approach helped illuminate the precision of homing and the ways that heterogeneous spawning habitats "gives rise to locally adapted populations… this biodiversity imparts resiliency to environmental change, lending temporal stability to their overall productivity" (Brennan et al., 2015, p.1). Brennan et al. (2015) argue that while genetic differentiation such as mixed stock analyses can highlight local particularities, those kinds of studies "do not yield migratory information of individuals and are often limited to apportioning harvests to coarse spatial scales" noting the critical need for tools that can explore populations at a finer-scale to develop a more specific time series.

While observations and experiments may have been key tenets of science, the collection of physical specimens offered the phenomena itself as evidence. Moving beyond

human observation, this approach views the otolith as an archive. This metaphor of the archive is not strictly left to the historians of science but has also been used by ecologists (Cohen, 2003), biologists, and geologists. Paleobiology emerged in the late 1970s with the idea that the fossil record might provide insight into the deep history of the earth. Sepkoski (2018) traces the evolution of this highlighting an epistemic break in the early 19th century due to the discovery of stratigraphy. This development focused on the earth as an archive with layers that could be 'read' by fossil traces left behind in the earth's crust theorizing that fossils are "documents preserved in the drawers of its [earth's] cabinet (the strata)" (Sepkoski, 2018, p.57). Reading these fossils in the layers of earth through stratigraphy is analogous to the reading the rings on the otolith.

As the otolith moves from a salmon brain to the laboratory to a data point, it is used to construct a story of how old a fish is or where it has traveled. These little pieces of the fish that were once responsible for navigation and for balance are transformed into an archive that holds information about aspects endemic to the salmon life cycle. Despite its historical usage for aging fish, these otoliths are now used to answer questions much deeper back in time than the original creators of the field site imagined.

Time at coarse scale

In both vignettes (of otolith and of fin clips), a specimen collected from the fish itself offered insight into fine-scale temporal understanding about migratory behavior and stock composition. While these data provided information at a fine-scale, they did little to show how many fish existed over long periods of time. Furthermore, the data used to understand abundance (number of fish at a given time) were data such as catch records or harvest surveys, data from commercial fishing activity. As such, there remained a need to understand

abundance of fish across longer and deeper timescales than data like catch records or survey data could provide. This is because change that impacts ecosystems often occur at longer time scales than commercial fish has had a presence in the region, and as such, data produced through commercial fishing activity does not capture the requisite time scale for the question at hand.

One discipline that offered a solution to the need for a deep time view was paleolimnology, which can be traced back to the late 1950s. Paleolimnology is a subdiscipline of limnology; much like limnology, it is interested in understanding the productivity of lakes; however, rather than measuring lakes in their present form, paleolimnology relies on the geologic record to reconstruct the past (Frey, 1988). Geologic science operates on the order of enormous—somewhat incomprehensible—time scales. Using the metaphor of "earth's archive", Cohen (2003) notes that paleolimnology studies lake deposits because they provide 'historical' data that both "highly resolved in time and of long duration." This orientation toward time and toward earth as archive comes from geology. Charles Lyell was among the first to look at the depositional environments. Lyell was building off of James Hutton, the first to put forth the idea of "deep time". Rather than looking for fossils in the way a paleontologist would, scientists look for indicators or proxies from which they can infer the population dynamics.

Using stable isotopes in lake core sediments, researchers (Finney, 1998; Finney et al., 2000; Gregory-Eaves et al., 2003) developed an understanding of salmon abundance and claimed a deeper-scale temporal understanding of salmon abundance than historical air and ocean temperature records could achieve. In fact, Gregory-Eaves et al. (2003) boast a 2,200-year old reconstruction of salmon-derived nutrients in two lake studies in Alaska. They do

this by comparing two lakes: one (Karluk Lake) identified as a "natural sockeye salmon nursery lake" and one (Frazer Lake) in which the date in which salmon were introduced to the lake is known. This set up a somewhat happenstance experimental condition in which the researchers could study the impact of salmon introduction on the "diatom assemblages and isolate the influence of regional climatic and environmental variability on diatoms like looking that the pre-salmon record." In other words, this provided a baseline by which the scientists could interpret data from the "natural" lake. This natural experiment coupled with longer time scale data produces a scale that observational data heretofore could not rival. Others have researched the Karluk case as a case of deep time (SASAP).

Soutar & Isaacs (1974) refer to these as natural chronographic records, perhaps most commonly known are the rings on a tree. In describing the anaerobic sedimentation at the bottom of the ocean, they write "undisturbed, these threads of information accumulate to form a remarkable sedimentary chronicle combining the rhythmic pulse of the seasons with the vagaries, trends, and inconsistencies of ocean life, chemistry, and currents" (p.257).

In the application of paleolimnological studies in the Alaska Salmon Program, researchers were concerned with using 'sediment chronologies' to provide a synthesis of sockeye salmon abundance (*Oncorhynchus nerka*) using proxies such as diatoms and stable isotopes in lake core sediments (Brennan et al., 2015; Rogers et al., 2013; Schindler et al., 2006). Rogers et al. (2013) argue that centennial-scale variation in salmon abundance is invisible due to the way that modern-day catch records and survey data constrain knowledge temporally. They note that "sediments store information about past ecosystem characteristics, including the abundance of fish. In the case of sockeye salmon, the ratio of stable N isotopes preserved in

lake sediments can be used to quantitatively reconstruct historical abundances" (Rogers et al., 2013, p.1750).

The case of sediment chronologies points to the long-tail of ecological systems. As Magnuson writes about the long-term perspective in ecology, processes that act over decades can be hidden and "reside in the invisible present" (Magnuson, 1990).

Much in the way that genetic research methodologies were suited to salmon life histories, so too does the application of paleolimnological methods. Because salmon return in large numbers and therefore their carcasses contribute significantly to the sediment, the ability to see a record of their presence in the past sediment is high. In other words, due to salmon life histories, their bodies contribute to the sedimentary layer, making the application of geologic methods useful to uncovering deeper understanding of salmon ecology.

Conclusion: The specimen of data archives

The above vignettes highlight an evolution in salmon studies and the way in which emerging methods and instruments aligned with the life histories and life cycle of salmon.

The commercial interest at the heart of salmon science a century ago was the impetus for salmon becoming scientifically tractable in that scientists could rely on observations of people harvesting the fish. In the sections above, I highlighted three examples field instrumentation with specimens to show how aspects of fish biology and life history converge with the scientific methods employed to understand them. More clearly, methods of instrumentation are tightly coupled with the theories for understanding phenomena.

In the first example, I showed how genetic methods were used to understand salmon migration and parentage. This is an example of novel methods being applied to produce

temporally-discrete populations. The second vignette highlighted a transformation in the use of the otolith, a piece of specimen once used to age fish, now used to develop finer-scale understandings of salmon migration over a long history. And finally, I discuss the use of lake sediment to reconstruct the past on a coarse scale, a paleolimnological approach for understanding change at deep time scales.

What I have tried to illustrate with this specific chapter is that no datum exists in isolation or ahistorically (Ribes, 2019) but are always remade contingent on present theories, and sometimes the present tools, at hand. To say to an ecologist or biologist that data have context is moot, because all of the data are situated, in a stream, in an ocean, on a scale card. For the information scientist, it is more a question of how scale has been integrated into the data and how it might be interpreted in future usage.

These vignettes have shown the long history of producing an understanding of the temporal scale of salmon cycles (e.g., migrating timing, population dynamics over time). As technologies for tracking and accounting for salmon become more sophisticated, many of the data collection practices have remained relatively unchanged, particularly in data organizations that laud the benefits of having consistent, long-term data. Rather than a revolution in data collection practices (such as the application of machine learning to videos of salmon, for example), there have been minor additions to data collection practices (e.g., taking a fin clip while tagging fish) and dramatic changes in how data are analyzed relying on more computationally-intensive models to understand complex phenomena. An historical analysis sheds light on the changes in the material infrastructure of salmon science, which comes to bear on the digital and computational infrastructure through which scientists and managers often interact with salmon. As I have shown in the above vignettes,

the shape of data has changed over time with an evolving understanding of how salmon behave and interact with one another as well as the environment around them.

Rather than the shiny programming of data-intensive research, data are collected on the knee of a fish biologist. Nature—the fish, the trees, the river that has been managed for decades—live by their own clock (Helm et al, 2013). This impacts the way they have been quantified and studied. Additionally, time is a relational quality that is difficult to study directly but often studied through something else such as strontium isotopes and other markers of geologic change. In this way, it carries many of the same properties as infrastructure: it is backgrounded, visible upon breakdown, embedded in cultures, and has its own inertia.

The story I have told shows not only the constant negotiation and production of phenomenal time — the rhythm of the salmon themselves — but also shows the places in which infrastructural time and phenomenal time meet. What I have illustrated is that endeavors to track and document all context about all data is insurmountable. However, an historical look at the evolution of data collection can shed light on major paradigm shifts. This research concludes with two questions for future work: How can empirical and historical analyses of existing ecological field sites refine the lifecycle model of data? And, how might an historical approach broaden participation and collaboration in data curation work?

In the next chapter, I participate in one of the SASAP working groups that is uniquely focused on the collection of new data. By understanding how scientists scale for local through a community-based monitoring project, I look at how observations from residents make their way into scientific models. Because the project I participated in involved meeting

with local residents who also participate in the long-term advisory group, the Kuskokwim River Salmon Management Working Group (KRSMWG), I coded the meeting minutes from these sessions to understand what local observations have focused on throughout the last three decades. Taken together, these data outline overlaps and mismatches between local observations and scientific understandings. Through ethnographic accounting of the working group meetings as well as interviews with state, federal, and tribal managers, I examine how scientists use models to scale and how they define local scale in the context of a participatory modeling project.

Chapter 7: Models as an ethnographer's tool for understanding how scientists scale for local

As big data and open science movements are becoming more mainstream, institutions increasingly emphasize public accessibility and co-production as a requirement for science. The emphasis on co-production and public involvement in science rests on the purported benefits of democratizing science (Giddens, 1990; Guston, 2004; Jasanoff, 2004; Wynne et al., 1996). Others focus on publicity of the data as a political imperative (e.g., Nielsen, 2012). Evident in literature on community involvement is a stark divide between the outcomes that local involvement promises and the objectives therein for local involvement. On the one hand, community involvement is proposed as a way to engender trust in science through public participation, thus making science shared and open and making publics more trusting of that science; on the other, it is proposed that local involvement can help fill in data gaps and assist with achieving more data for more accurate science.

Thus far in my field sites, there has been little engagement with this debate. The data task force team at SASAP were heavily focused on interoperability. As such, most of the discussions around data pertained to scalar issues in achieving interoperability, such as how to harmonize different data standards or how to commensurate older data with new. Notably, the issue of public accessibility is largely left out of this data-intensive work. In looking elsewhere in the project documents, I suggest that **this is a subsidiary goal of the SASAP initiative, which prioritizes making data mobile (interoperable) over making data locally accessible**.

In the second empirical chapter, I showed how a long-term research infrastructure is sustained and how it changes over time. These two previous empirical cases tracked what Bowker et al. (2009) have identified as the two main issues associated with large-scale information infrastructure: first, "the idea of a shared infrastructure in the sense of a public good; second, the idea of sustainability, of supporting research over the long term" (98).

In this third and final empirical chapter, I ask: *how do scientific understandings of salmon biology shape, and sometimes break with, local practices of sustainability? And, how do scientists define local scale in this instance?* This is ultimately engaged with the topic of local and global within debates about how to define scale and thus, helps to answer my broader question of how issues of scale in data production challenge scientific knowledge production.

To answer this broad question, I explore local scale by defining what scientists and managers mean by local in a specific case. I start here because local is often assumed—particularly in far north regions—to be Indigenous or traditional knowledge (Wilson et al., 2018); however, it is important to understand what the actors themselves consider to be *local*. As such, I participated in a working group focused on 'participatory modeling' as an

avenue of inclusion of communities in salmon science. Through this participation, I provide an ethnographic vignette of how scientists and managers define local as I follow their discussion around what data they deem "*vulnerable to community-based monitoring*". This was their framing that unfolded during discussions of different models that could be used to better understand salmon abundance and habitat.

I argue that these definitions, and thus operationalizations, of local aligns or misaligns with the knowledge people share in advisory council meetings. Furthermore, the operationalization of local as index of abundance misaligns with on-the-ground salmon harvesting practices in a subsistence context. Local aligns as it pertains to the fine-scale habitat observations or in the broader goals of preserving fish populations. To demonstrate these points, I outline discussions about models that produce salmon as an object of inquiry and management. This research provides an answer to how scientists create an environment conducive to public input, outlining the ways publics have been involved in science.

The chapter is organized as such: 1- I first outline my central argument about how scientists in this particular case define local. 2- I then introduce the concept of modeling and explain its relationship to management in the region. 3- I follow this with a fuller conceptualization of how scientists define local. 4- I then evidence this argument with data from my field research conducted in 2016-2018. Here, I offer a characterization of how scientists define local in the context of these models. I explore how local participation is expressed in the Kuskokwim River Salmon Management Working Group (KRSMWG) calls, which serves as a public forum for federal and state fisheries managers to 1- meet with local users of the salmon resource, 2- review run assessment information, and 3- reach a consensus on how to proceed with management of Kuskokwim River salmon fisheries. In

summary, the puzzle I explore is how can science, which is focused on understanding large-scale change, also attend to issues that occur at the scale of local experience? Furthermore, I look at how local scale is enacted at different temporal scales than the large or universal.

Practicing scale: how local is defined

Local and global are co-constitutive to one another. For every instantiation of local, there is a global counterpart. For example, knowing the number of fish in a transect of a stream contributes to an understanding locally about what habitat salmon prefer. However, this knowledge also connects to global averages of fish returns. In this way, they co-define each other as local is often what comprises the global. However, more than the parts that make up the whole, local is distinct and yet relational to the global. Global scale tends to be a core tenet of science as data build up toward larger, more universal truths. And in fact, even historians of science have taken this production of global as their object of research (Edwards, 2010). To re-state a common theme that this research encounters: the local is often defined in concert with its global counterpart.

A fundamental question for modeling exercises is how to extrapolate from observational data. This dynamic is tied to scale as one proposed way of achieving spatial and temporal coverage is to model out uncertainties so that data might be more accurately scaled up. While not her explicit object of research, scalar issues appear in Oreskes' exploration of how the earth sciences develop technologies, mostly models, to produce temporal predictions (Oreskes, 2003). In modeling discussions, much of the conversations about local data revolves around how it might be scaled up or how it may be validated.

An infrastructural lens illuminates a semiotic relationship between local and global. On the one hand, there is the focus on universality at least with respect to standards and

operation (Faniel & Jacobsen, 2010; Karasti et al., 2010), while on the other there are the particular and local requirements (Edwards et al., 2007). In other words, to achieve the goal of universality, an attention to local practices is crucial (Ribes & Lee, 2010). This tension between local and global is what leads to the development of infrastructure, or as Star and Ruhleder (1996) point out: "an infrastructure occurs when local practices are afforded by a larger-scale technology, which can then be used in a natural, ready-to-hand fashion" (Star & Ruhleder, 1996, p.114). I do not propose a universal definition of local. This is instead an investigation of how scientists define local instantiated here through specific tools and practices.

Participatory modeling: determining what data are 'vulnerable to community-based monitoring'

Early in my introduction to SASAP, one working group in particular noted that its focus was unique in that unlike the other working groups, the team was focused on collecting *new* data. This statement of difference brought up questions about how SASAP and NCEAS could be of service to this team as much of NCEAS' support work was bound up in the acquisition and synthesis of pre-existing datasets. While this group was comprised primarily of quantitative ecologists motivated by value of information (VOI) models, error coefficients, and Bayesian statistics, this group was notably one of the few focused on how science is made available to 'the public[22]' – not science as a product for the public to access but science as a conduit for local residents to participate in data collection.

[22] Much ink has been spilled debating and adding nuance to the term 'publics' and 'the public'. Le Dantec (Le Dantec, 2016) is perhaps one of the most prolific design thinkers in this area of what it means to be public. Drawing on Dewey, he distinguishes his perspective from Habermas' view of 'the public' as public sphere to be more participatory taking Dewey's definition which "sought to define a public not as a single common mass of people, but rather as a specific configuration of individuals bound by a common cause in confronting shared issues." He expands his view of publics to those constituted

While the location of this fieldwork contrasts sharply with my fieldwork at the synthesis center located in sunny, warm Santa Barbara, California, the meetings are largely the same. We sit at a circular table in a room with the curtains drawn so we can see scientists present charts and graphs. We could be anywhere at any time of year. We happen to be in Bethel on the first weekend in May. The ice had just broken up (figure 23) and there was band playing along the river. Later, I learned the significance of 'break up' as one local participant in the project explained that it signals the beginning of a new season of plenty (the end of winter). However, the breaking up of the river has gotten earlier and earlier indicating a warming climate.

Figure 23. Taken from the bank of the Kuskokwim River in Bethel, Alaska May 6, 2017

not only by issues but also attachments and *infrastructuring*. He shows how a public takes shape "over time and across multiple sites through the process of designing technical artifacts, organizational procedures, and social practices" (p.109).

The goal of that first meeting was to look at how community-based monitoring (CBM) can inform management models. The models require information that can be monitored but there are many other facets not in the models that are important to the Kuskokwim (e.g. water temperature, water flows). The CBM project is an example of a common attempt to make models more absolute by incorporating local data. As such, the hallmark of western science has been to overcome uncertainty by applying well-known statistical models to fisheries (Connors et al., 2020; Hamazaki et al., 2012; Staton et al., 2017, 2020). This type of natural resource management involves state and federal managers with a background in biology and statistics who base decisions on the interaction between three key factors: 1) stock-recruitment dynamics of the salmon population under management; 2) expected run size; and 3) the relationship between fishing regulations and harvest. The first factor informs selection of an escapement goal, a target number of adult salmon needed to reproduce each year to provide sustainable future returns. The second factor determines how much harvest can be taken within a given year without risking failure to meet the escapement goal. Finally, the third factor allows managers to specify the necessary conditions to maintain the harvest within target bounds. All factors depend on a retrospective view of harvest and stock responses to previous management actions. As such, the temporal orientation of these discussions was both future-focused (forecasting) as well as firmly planted in the past (reconstruction).

Models provide an ethnographic field device for understanding how science and management scales. As the era of modeling for understanding population dynamics dawned, scientists began to note computational advancements as a major boon to resolving uncertainties. This is because gaps in data could be statistically accounted for without

requiring a complete set of data. My involvement with this working group led to a deeper engagement with modeling as I turned toward models as a scalar device for understanding how scientists instrument scale. In the group meeting that I attended (*participatory modeling to engage citizens in salmon science),* I spent most of my time at the cultural center in Bethel, Alaska listening to scientists describe models.

In these model discussions, they asked "*what type of data might be vulnerable to community monitoring*?" Vulnerability is commonly used in environmental or ecosystem modeling to conceptualize how resilient communities are to various threats. However, the quantitative ecologists' usage of vulnerable to describe data suggests data that are already in existence, vulnerable to the collection by community members. By vulnerable, they meant data that might be easily collected by communities. In this way, these data and their 'vulnerabilities' were both a problem and an opportunity.

There are four models considered in the discussion about which data would be most *vulnerable to community-based monitoring*: 1- stock-recruitment models (a reconstruction of abundance); 2- management strategy evaluation (MSE) (an evaluation of how different strategies perform); 3- in-season management models (an evaluation of seasonal management decisions); and 4- harvest-diversity trade-off models (a longer time scale view of portfolio diversity and harvest). While the modelers set out to define critical uncertainties, the stakeholders were tasked with bringing "practical, local knowledge of monitoring and data collection options" (Jones, 2017). It is through the discussion of the four models that the scientists form an opinion about what role the public can play in the modeling process. All four models have a different temporal orientation spanning from daily (in-season management model) to a longer time scale of 100 years (harvest diversity trade off model).

The time scales of the models are related to their objectives and as such, the data that are deemed useful for those models.

Through workshop meetings, advisory council minutes, and interviews with state, federal, and in-season managers, I began to know Bethel and the Kuskokwim more fully. In the remainder of this chapter, I explore how scientists and managers instrument their field site to understand 'the local', imagined as both opportunity and problem. In other words, the scientists argue that publics can be involved in science (opportunity) as a solution to data that are vulnerable to that kind of local data collection (problem). The problem is that the scientists lack data on certain phenomena and cannot collect those data without incurring exorbitant costs. This presents an opportunity to have local residents involved in the data collection practices.

Theorizing local

There are myriad terms for what I am about to discuss – public engagement with science (PES), community science, community-based monitoring (CBM), participatory science, public participation in scientific research (PPSR), citizen science – to name a few. Much of the literature identifies engagement as a key goal for "democratization" of science. The goals of democratization rest on the idea that on the one hand, science might be made better by making it more transparent and publicly accessible and on the other, that public trust in science might be restored by public engagement. However, many have noted the futility in this move to restore public trust due to underlying institutional norms (Stirling, 2008; Wynne, 2006).

Evidently, the inclusion of local has been long-debated and contentious; nonetheless, it is being taken up by funding institutions as a critical requirement of science. A UNESCO report on traditional knowledge and climate change states: "Indigenous observations and interpretations of meteorological phenomena have guided seasonal and inter-annual activities of local communities for millennia. This knowledge contributes to climate science by offering observations and interpretations at a much finer spatial scale with considerable temporal depth and by highlighting elements that may not be considered by climate scientists" (Nakashima et al., 2012, p.7). More recently, the NSF issued a requirement for researchers conducting arctic research to include a co-production component "when appropriate". This generated a backlash evident in a letter from Kawerak, a non-profit arm of the Bering Strait Native Corporation. It notes:

> *The NNA [NSF's Navigating the New Arctic] has funded projects that claim, among other things, to be collaborative, to do knowledge co- production, to include partnerships with Indigenous communities, and to address questions that will 'help' or 'assist' Arctic residents. Many of these projects (and many more which were not funded) do not and will not fulfill any of those claims.*

Criticizing universalist approaches and calling for more consideration of the local is not novel. Scott (1998) makes lofty arguments against state-centered universality; aligning with Jane Jacobs and Friedrich Hayek, he emphasizes the importance of bottom-up, locally generated solutions that are place-based. Drawing from the Greek *mētis*, which is a "means of comparing the forms of knowledge embedded in local experience with the more general, abstract knowledge deployed by the state and its technical agencies" (p. 311), he emphasizes the practical skills or know-how as preeminent to the universal. It is his view that the practical, local knowledge is supreme in everyday decisions.

Local scale often signals the particular, the nuance, the ever-elusive complexity that universal laws have failed to grasp. Science and Technology Studies (STS), particularly work focused on multi-species or more-than-human relations, has tended to praise the local and the forgotten nuance. Contrariwise, local is considered problematic as the local is hard to act upon as it does not generate theories, objective numbers, or even categories or frameworks from which to operate. As Bowker and Star (1999) point out, categorization is a way of making particular or local legible to the universal; however, it is relational to the universal – not either / or.

The state—while managing many locals—often plays the role of the global as its universalizing force shapes issues. Approaching the Norwegian parliament with a material-semiotic lens, Asdal and Hobæk explore how nature becomes legible or tractable to parliament, and how this enrolls new interests and therefore, publics (Asdal & Hobæk, 2016). Drawing primarily from actor network theorists, Asdal emphasizes the politics of material objects showing how participation is in the legal document-writing, in the literal text of a policy artifact, which promises to result in emergent interested publics. The authority of the state to claim objectivity hinges "on the deployment of such little tools of knowledge as images, graphs, lists, questionnaires, dossiers, tables, and reports" (p. 15-6). Just as inscription devices (Latour, 1987) such as graphs, samples, and lab-rats enabled the movement of nature-objects to scientific papers so too do the documents (public and private), legal proposals, and regulations enable movement in parliament. The enrollment through the use of materials is what I attend to in my study as well: through the surveys, meetings, and models, I show how local is instantiated in specific ways.

As I have shown, these differences are not just between Indigenous knowledge (local) and western knowledge (global). They can be seen *within* the sciences. Field sites come with their own particular practices, embodied expertise, and distributed knowledge. In talking with an ecologist, it is clear that he recognizes himself as an instrument, however, one that has been trained to a level of expertise. In contrast to the view that data can speak for themselves, the ecologist sees the world as uncertain and situated. The crux of two knowledge systems is of particular and general expertise respectively. In the latter, a way of knowing is close to the object of study; its intuition is within the scientist who sees him/herself as instrument; it is embodied. The former, by contrast, is founded upon laws that might be applied to various objects of study. The context of the salmon in particular is less important than the generality of the application.

By outlining how scientific definitions and local practices align and misalign, this research challenges the "naturalization" of the local, or rather the perspective that the local is a given while the global is something that must be created by assembling many locals. On the contrary, what I show is that there are just as many factors that go into validating, interoperating, and producing "local" as there are in the production of its global counterpart.

Three key characteristics of the local

Based on ethnographic work of the modeling discussions as well as interviews with state, federal, and tribal managers, I offer a few key characteristics of local as defined as useful to the science that supports natural resource management. This builds into my larger point about how local and global are co-constitutive. My characterization below is based on

ethnographic data from which I formed primary categories of how scientists formulate what local means.

Local as human

At the level of scientific practice, local is a stand-in for human activity. For the scientists and managers on the project, local is formulated as an index of salmon abundance, an index built with data on human activity. It signifies data about human behavior (particularly in response to the harvest of salmon). Scientists in this case define local as a quality of human activity, not that human activity is relative to something that is not human but is instead an indicator of number of fish in the stream or strength of the salmon run. In other words, **local is the data that can then be translated or scaled up into broader categories.** I evidence this below by inventorying the historical ways that local has been defined. The scientists in the project consider these historical data types not only useful to management but also something that people in the region might participate in.

After defining the goal as centered around synthesizing critical uncertainties, a project lead goes on to say:

> *in these remote regions of Alaska, perhaps the best way to reduce that uncertainty is **to rely on local subsistence users to gather information that would be informative**. Our project is aimed at synthesizing what we understand about the **critical uncertainties** that make it difficult to craft salmon management policies in western Alaska with the capacity of community to gather information about run-timing, quality of the run, the size of the run, the amount of subsistence harvest, and so forth.*

These data around migratory timing of the fish as well as the quantity and quality of fish are all aspects that revolve around the harvesting of the fish, and as such, the practices of local resource users. In shorter-cycled models (e.g., the in-season management model), local can

take the form of daily fishing activities as harvest data are – and have historically been – considered highly useful to management. This includes data from harvest surveys in which percent of need met is reported on. See image below for a visual presented by the in-season management modelers (figure 24). This image depicts the way in which local fishers and community-based monitors produce data such as harvest data, percent of need met, and concerns to then feed into in-season management.

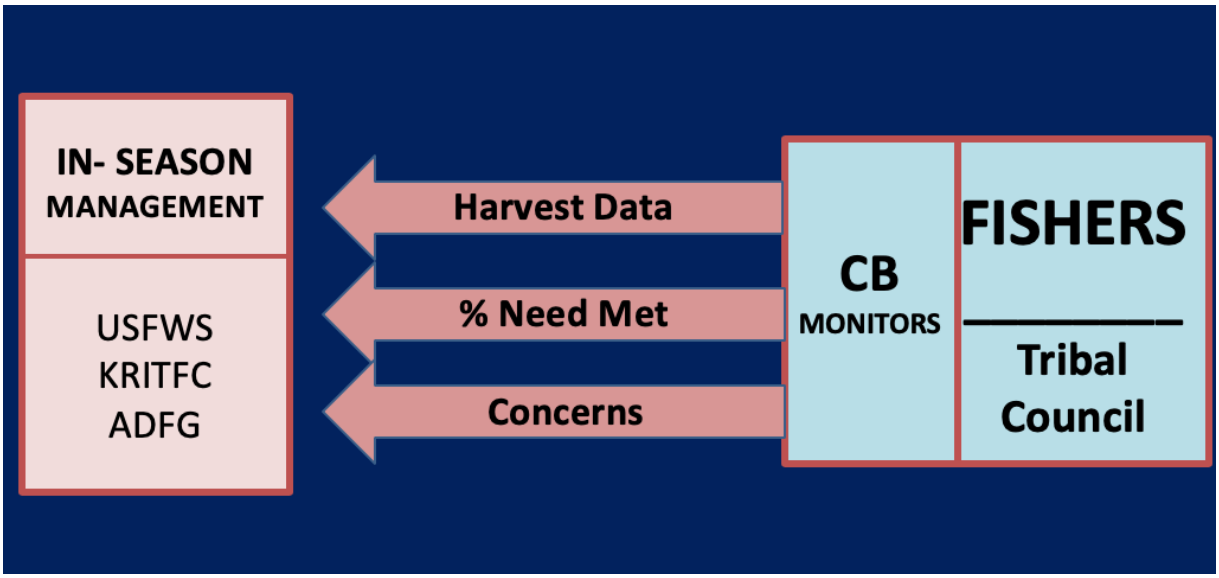


Figure 24. In-season management: Bechtol and Spaeder 2017 - presentation for NCEAS working group meeting

These harvest data are not new. Historically, commercial salmon fisheries from the Kuskokwim River provided an index of the status of returning Chinook salmon; however, directed commercial fishing closures have curtailed data input. CBM programs have filled a gap here with a focus on informing in-season management, supporting equitable harvest opportunities, promoting more inclusive and transparent understanding of management, and involving stakeholders in assessment and management processes. For example, the Orutsararmuit Native Council (ONC) has been collecting qualitative information for 17 years. While this information was considered by managers, it was not directly applied in the years

prior to the Chinook crash. As opposed to the fish ticket data (from commercial fishing), the "qualitative" harvest surveys are much more labor-intensive and requires more time to acquire this data.

Annual management reports show the catch calendars, postcard surveys, and harvest questionnaires that were conducted to translate seasonal subsistence activities into management numbers. While newer studies are positioned as novel in their incorporation of local observations, local resource users and their activities around that resource have served as a proxy or index of salmon abundance for decades. In the image below (figure 25), an annual report from the 1990s articulates an equation for community catch distinguishing between groups that usually fish or usually do not fish.

The average community catch (C_k) was estimated for salmon species from the composite catch per household data using the following formula:

$$C_k = \Sigma^{1}_{i=0} (N_{ki} * C_{ki}) / \Sigma^{1}_{i=0} N_{ki}$$

where
k = community
i = indicates whether the group "usually fishes" (1) or "usually does not fish"(0)
N_{ki} = number of households that "usually fish" or "usually do not fish"
C_{ki} = mean harvest for households that "usually fish" or "usually do not fish"

The total community catch (T_k) was estimated by $T_k = \Sigma^{1}_{i=0} (N_{ki} * C_{ki})$ and its variance (V_k) includes a finite population correction factor:

$$V_k = \Sigma^{1}_{i=0} ((N_{ki}^2)(1-(n_{ki}/ N_{ki}))(\Sigma_{ki}^2/ n_{ki}))$$

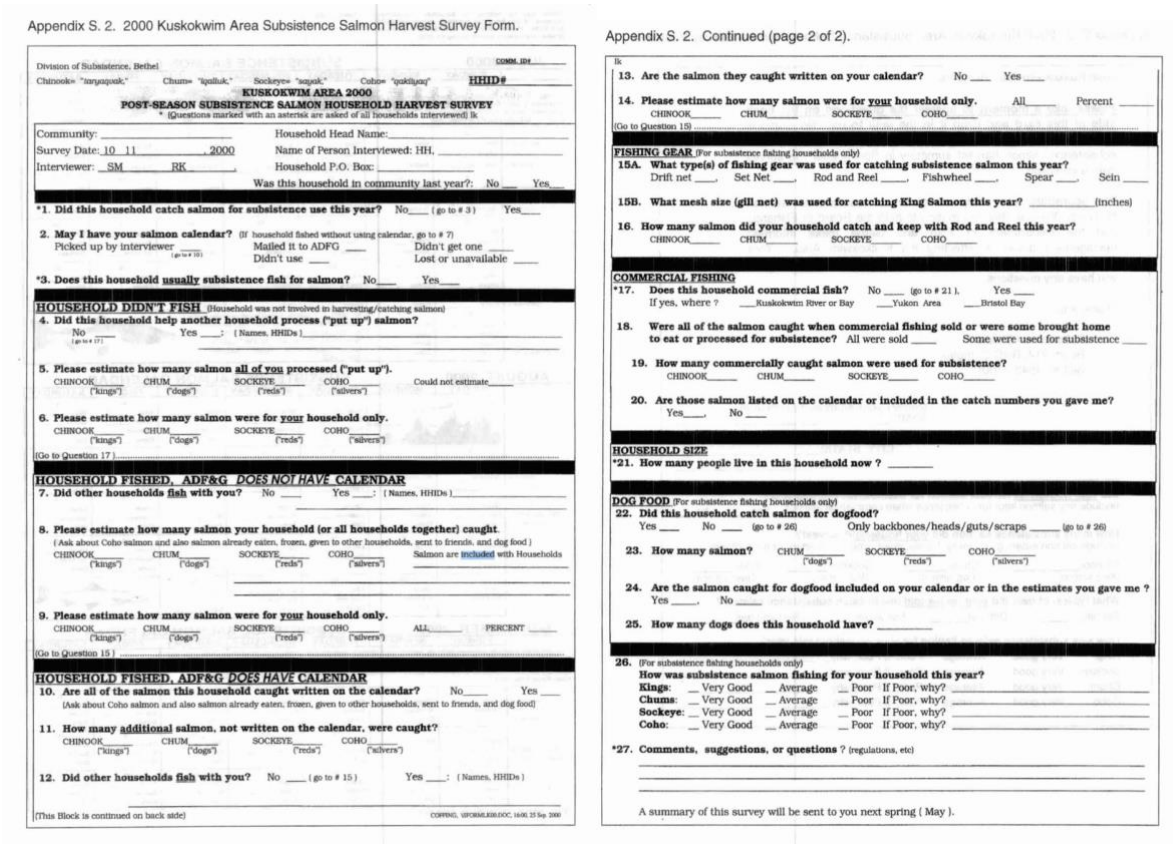
where n_{ki} = number of households for which information is available that "usually fish" or "usually do not fish" and Σ_{ki}^2 = variance for the amount harvested for the "usually fish" or "usually do not fish" households.

If fewer than 30 households or less than 50 percent of all households in a community were contacted, the reported harvest was used for the estimated harvest. Community catch estimates and their variances were summed across communities for region subtotals and across all regions for Kuskokwim Management Area totals.

Figure 25. Community catch equation from Annual Management Report, 1999

In this example equation, the components outlined are community, an indication of likelihood to fish or not fish, the number of households that fish or do not fish, and the mean harvest for households that fish or do not fish. Managers model out household data based on that household's *likelihood to fish*.

A common refrain from managers is that they manage *people* not fish. And as such, much of the ways that they instrument the field is through local fishers in conjunction with modeling. For example, household surveys (figure 26) have been used by managers for decades, which involved management staff traveling house-to-house to communities to interview residents about their fishing effort. While referred to as household surveys, these were harvest surveys used to **track the seasonal subsistence activity formatting it into calendars fitted to be legible to mechanical time.** A 1999 report notes that "information from different sources for a particular species may be different due to the timing of the collection of this information." In other words, timing of migration, species of fish, and the location of the harvester are all data points that matter in determining which fish are targeted and which are not.



Appendix S. 2. 2000 Kuskokwim Area Subsistence Salmon Harvest Survey Form.

Division of Subsistence, Bethel

Chinook= "tanyaquak," Chum= "iqalluk," Sockeye= "aqsak," Coho= "qakliaq" HHID#

KUSKOKWIM AREA 2000

POST-SEASON SUBSISTENCE SALMON HOUSEHOLD HARVEST SURVEY

* (Questions marked with an asterisk are asked of all households interviewed) lk

Community: Household Head Name:

Survey Date: 10 11 2000 Name of Person Interviewed: HH,

Interviewer: SM RK Household P.O. Box:

Was this household in community last year?: No Yes

*1. Did this household catch salmon for subsistence use this year? No (go to # 3) Yes

2. May I have your salmon calendar? (If household fished without using calendar, go to # 7)
Picked up by interviewer Mailed it to ADFG Didn't get one
Didn't use Lost or unavailable

*3. Does this household usually subsistence fish for salmon? No Yes

HOUSEHOLD DIDN'T FISH (Household was not involved in harvesting/catching salmon)

4. Did this household help another household process ("put up") salmon?
No Yes : (Names, HHIDs)

5. Please estimate how many salmon all of you processed ("put up").
CHINOOK ("kings") CHUM ("dogs") SOCKEYE ("reds") COHO ("silvers") Could not estimate

6. Please estimate how many salmon were for your household only.
CHINOOK ("kings") CHUM ("dogs") SOCKEYE ("reds") COHO ("silvers")
(Go to Question 17)

HOUSEHOLD FISHED, ADF&G DOES NOT HAVE CALENDAR

7. Did other households fish with you? No Yes : (Names, HHIDs)

8. Please estimate how many salmon your household (or all households together) caught.
(Ask about Coho salmon and also salmon already eaten, frozen, given to other households, sent to friends, and dog food)
CHINOOK ("kings") CHUM ("dogs") SOCKEYE ("reds") COHO ("silvers") Salmon are included with Households

9. Please estimate how many salmon were for your household only.
CHINOOK ("kings") CHUM ("dogs") SOCKEYE ("reds") COHO ("silvers") ALL PERCENT
(Go to Question 15)

HOUSEHOLD FISHED, ADF&G DOES HAVE CALENDAR

10. Are all of the salmon this household caught written on the calendar? No Yes
(Ask about Coho salmon and also salmon already eaten, frozen, given to other households, sent to friends, and dog food)

11. How many additional salmon, not written on the calendar, were caught?
CHINOOK ("kings") CHUM ("dogs") SOCKEYE ("reds") COHO ("silvers")

12. Did other households fish with you? No (go to # 15) Yes : (Names, HHIDs)

(This Block is continued on back side)

Appendix S. 2. Continued (page 2 of 2).

lk

13. Are the salmon they caught written on your calendar? No Yes

14. Please estimate how many salmon were for your household only. All Percent
CHINOOK CHUM SOCKEYE COHO
(Go to Question 15)

FISHING GEAR (For subsistence fishing households only)

15A. What type(s) of fishing gear was used for catching subsistence salmon this year?
Drift net Set Net Rod and Reel Fishwheel Spear Seine

15B. What mesh size (gill net) was used for catching King Salmon this year? (inches)

16. How many salmon did your household catch and keep with Rod and Reel this year?
CHINOOK CHUM SOCKEYE COHO

COMMERCIAL FISHING

*17. Does this household commercial fish? No (go to # 21), Yes
If yes, where ? Kuskokwim River or Bay Yukon Area Bristol Bay

18. Were all of the salmon caught when commercial fishing sold or were some brought home to eat or processed for subsistence? All were sold Some were used for subsistence

19. How many commercially caught salmon were used for subsistence?
CHINOOK CHUM SOCKEYE COHO

20. Are those salmon listed on the calendar or included in the catch numbers you gave me?
Yes No

HOUSEHOLD SIZE

*21. How many people live in this household now ?

DOG FOOD (For subsistence fishing households only)

22. Did this household catch salmon for dogfood?
Yes No (go to # 26) Only backbones/heads/guts/scraps (go to # 26)

23. How many salmon? CHUM ("dogs") SOCKEYE ("reds") COHO ("silvers")

24. Are the salmon caught for dogfood included on your calendar or in the estimates you gave me ?
Yes No

25. How many dogs does this household have?

26. (For subsistence fishing households only)
How was subsistence salmon fishing for your household this year?
Kings: Very Good Average Poor If Poor, why?
Chums: Very Good Average Poor If Poor, why?
Sockeye: Very Good Average Poor If Poor, why?
Coho: Very Good Average Poor If Poor, why?

*27. Comments, suggestions, or questions ? (regulations, etc)

A summary of this survey will be sent to you next spring (May).

Figure 26. Household harvest survey from Annual Management Report, 1999

One way that this local harvest activity has been used to understand salmon is through catch-per-unit-effort (CPUE), an index of abundance. CPUE has been used as a tool to scale fisheries for over a century (Smith, 1994). It is an indirect measure of the abundance of a species. The decrease of CPUE indicates overexploitation while a stable CPUE translates to sustainability. In an historical account of the concept, Poulsen & Holm (2007) introduce Walter Garstang as one of the first to apply catch-rate to an analysis of British fisheries data sets. This came in response to the challenge of producing calculations that contend with the variations in efficiency of British fishing vessels. Instead, he used the "sailing smack" (a traditional fishing vessel) as a standard unit of effort to measure all other boats against. This stood as a standard account for variability in efficiency of different fishing vessels. Read in this light, CPUE as a concept has always required translations from one unit to another. In

Garstang's (1900) use of it, he used a traditional fishing sailboat as a baseline by which to compare other, more modern vessels (Garstang, 1900).

In the modern usage of CPUE, there is an awareness that management has changed the behavior as one scientist remarks on this process as being "pseudo-realistic" in that an unmanaged stock would not lead to everyone waiting to go out and fish at the same time. Applying the commercial concept of CPUE to subsistence fishing contexts is an example of when dominant logics break from local practices. The asymmetry with respect to catch rates and abundance has been widely noted perhaps most famously by Ray Hilborn who has argued that using catch rates to understand abundance paints an incoherent picture. Hilborn has written extensively about the history of management of fisheries globally, noting that the major changes took place between 1985 and 2010. Hilborn & Walters (1992) note that: "one of the major problems with using commercial catch and effort data to estimate stock distribution and abundance is that the fishermen go where the fish are." The idea is that management is already shaping practices on the ground and therefore, these data represent that shaping rather than a pristine image of how many fish return.

Furthermore, subsistence fishers are constrained by the amount of fish they can process at one time. Some participants discuss this aspect of what the fishery looked like with no restrictions, noting how different it was because people would fish when most convenient or appropriate. One way this has conflicted with traditional practice is that people are "forced into cutting fish when they can instead of when traditionally they would have." Here, a manager notes how management pressures coupled with societal changes have made fish processing more challenging for locals. This theme around the seasonality of fish harvest

and fish processing practices comes up frequently in interviews about how locals have been integrated into scientific management.

In this section, I have argued that local is a stand-in for human activity. It has been calculated based on human activities that involve harvesting fish. It is calculated based on the amount of effort it takes to catch a number of fish, which has come to serve as a quantitative measure of fish abundance. In other words, local is both the human activity that surrounds the harvest of fish as well as the translation of that activity into abundance of fish.

Local is capacity-building for empowerment

Local is also defined in relation to human involvement as a form of empowerment. At a broader scale, the scientists' concern for 'local' pertains to capacity-building or engaging locals to 'engender trust in science'. This is one of the primary goals of the initiative – to empower involvement in salmon science. This is evident even in early descriptions of what the modeling project will focus on. As the proposal states:

> *…better information to inform management models and within-year decision processes, implementation of the monitoring tactics that emerge from our synthesis* ***will engender greater community involvement*** *in management and engagement between stakeholders and decision-makers. It is now well-established that fishery management outcomes result in less conflict when there is meaningful engagement of stakeholders, largely* ***as a result of creating a strong sense of ownership*** *of the management problem by these stakeholders, and* ***engendering trust*** *in the decision process.*

As is clear from the proposal, the goal of involving local residents was also about making management less conflictual by "engendering trust" through the local involvement in data collection. While modern modeling discussions have centered around how local might be scaled up or how it may be validated, local involvement is thought to engender trust in the process by providing avenues of ownership in the process. Imagined as empowering, local is

meant to facilitate involvement in science. Less a point about data, this is more a point about empowerment and capacity-building.

How did a discussion about data vulnerability turn into an opportunity for engendering local trust? This perception of local and local data collection as an opportunity for empowerment is a specifically found in the scientific framing of local involvement. In conversations with local residents and managers, the purpose of local involvement is more about creating a better understanding of the environment – both for managers and local residents. As one participant remarked on the utility of local data:

> *Years ago, when it was the early 2000s, there was a really warm summer and people started complaining about the quality of the fish. Managers went out and found that there were some issues with warm water and a disease that fish get – ichthyophonus. It was affecting the King Salmon, and in fact, significantly enough that they believed it was causing a significant mortality on the fish going up the river. I don't know that they would have ever been aware of that without local people talking about it. (P7)*

In this example, it is the relation between manager and local fisher that has made management more accurate, not just an idea of trust. This is a bi-directional view of how local might be a form of empowerment rather than unidirectional.

One forum that provides insight into these bidirectional forms of knowledge production and local involvement is the Kuskokwim River Salmon Management Working Group (KRSMSWG). In 1988, the Alaska Board of Fisheries formed the KRSMWG as a response to requests from people who wanted a more active role in the management of their salmon. It now serves as a public forum "for managers to meet with local users of the salmon resource to review run assessment information and reach a consensus on how to proceed with management of Kuskokwim River salmon fisheries" (Ward & Horn, 2003, p.1). In the KRSMWG, local observation primarily includes observations about weather, comments on

environmental or river conditions, observations about fishing occurrences (e.g., what is being caught, seen, and what techniques are being employed), and as confirmation or validation that model outputs or predictions are correct.

While scientists considered data types such as the biological measurements of the fish, counts of fish, and harvest surveys as the most useful to management, people also offered observations about fish quality, environmental changes, and fish movement indicating these as major areas of expertise that local residents hold. An example of run-timing and migratory patterns points to a more complex relationship and a bidirectional sharing of knowledge. Managers commented on how local observations about where people were catching fish led to speculation about the run, which coincided with data coming in from the test fishery. Those observations were "*used as subjective support for what folks were starting to interpret the test fishery as saying (that the run was going to be late). So, they had quantitative, numerical evidence of a possible late run and traditional knowledge and local input that made them more confident about the likelihood of a late run*" (P8).

This might be seen as a success story for local observations given the potential biological basis for the migratory pattern. However, research done in the lower portion of the river (Moses et al., 2019) found that fish traveling to the Kisaralik/Kwethluk rivers and passing through the Kuskokuak Slough enter at the same time meaning that they are not indicators of run timing for the aggregate stock. This example demonstrates the mutual shaping that occurs in the exchange of knowledge: On the one hand, the scientific relevance of the observation made it visible to management while the observation data refuted the potential explanation that catching more salmon on one side of the Kuskokuak indicated an early or late run.

As I have shown in this section, exchanges between scientists, managers, and local residents have been happening for decades. In interviews, it was evident that more bi-directional relations between locals and managers not only created what was deemed useful data but also potentially achieved the goals of empowerment.

Local is fine-scale spatially and temporally.

Aside from data about behavior or a form of empowerment, local stands for fine-scale. Local data are defined as more fine-scale than global and, as such, are considered absolute measures rather than relative. Local residents often produce fine-scale data—temporally and spatially—because they are going out to discrete locations at specific times. In discussion with what might be useful to management and what local residents might be able to collect, scientists consider measurements that are more absolute.

After much deliberation, the model that ends up being considered most capable of being aided with local data is the in-season management model. The in-season management model is used to evaluate seasonal management decisions related to spatial stakeholder objectives. This model is particularly important for understanding the intersection between how effort changes over time and the timing of different components of the run. Temporal patterns are modeled on a more seasonal rhythm than larger-scale ecological questions such as "what habitats are best suited for salmon productivity?". This is because the everyday concern with scientific research on Alaska salmon is how to allocate fish to local fishers. The major differentiation is that this model looks at short time-frames to make decisions every season - and often daily; whereas, other models such as those focused on habitat diversity are more focused on long-term sustainability.

Given the commercial legacy of fisheries management, many of the concepts of management stem from commercial fisheries (e.g., catch per unit effort (CPUE), maximum sustainable yield (MSY), surplus, escapement). These methods include taking data from those who fish the salmon and turning it into a number that can stand in for abundance of salmon, and these translations are based on dominant ways the commercial fisheries have been managed. There have been efforts to fit these data to the logic of the models for management. However, change that occurs over longer durations than a management season require models with a different temporal orientation altogether.

One such model is the harvest-diversity trade-off model (figure 27). This model is of interest as it takes a much longer temporal view and seeks to develop alternate visions of management. With the possibility of informing alternative management actions, they ask: 1- what is the trade-off relationship between long-term harvest and the population diversity in the Kuskokwim? And, 2- how do you quantify harvest-population diversity trade-offs and incorporate their consideration into Chinook salmon management? This is the only model that considers longer shifts over time and subsequently, considers how local data might inform a model such as these. In this discussion, the data deemed possible for collection by locals are specimens of the fish. In this way, the specimen provides highly specific data about individual tributaries or sub-stocks. It is a kind of local involvement that is easily made global.

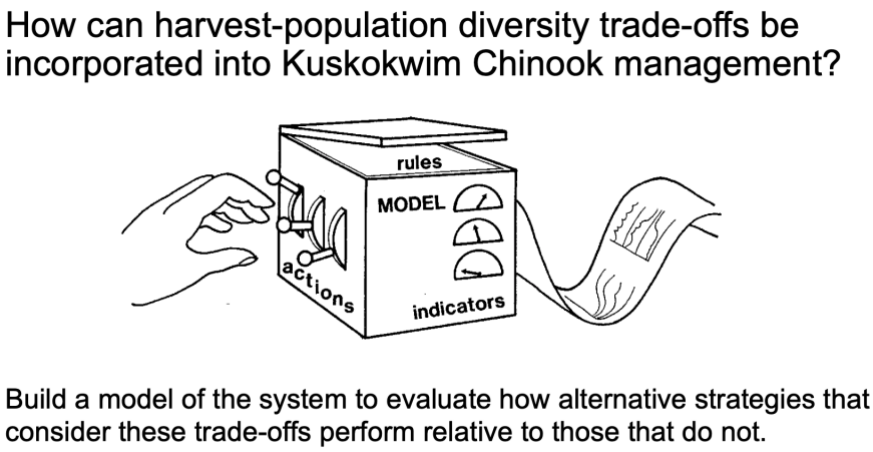


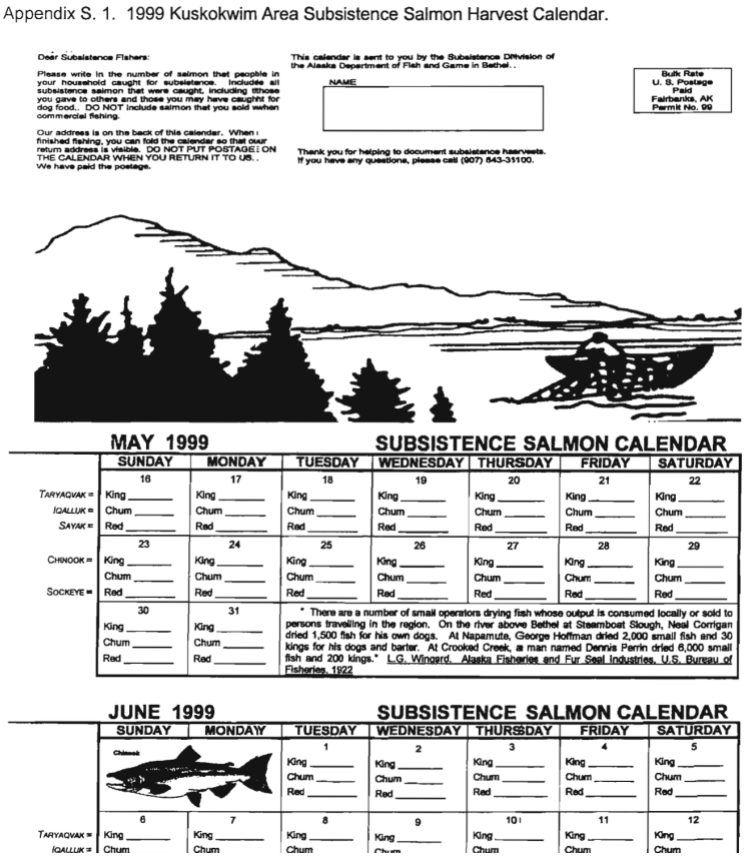
Figure 27. from presentation at working group meeting – Conners et al. (2017) presentation

The data types identified for the model on habitat-diversity are fine-scale specimen collections: 1- tissue samples for genetic stock ID to characterize sub-stock and river section run-timing; 2-scale samples to characterize the age composition[23]. This is because the intention is to have local people go out to collect these opportunistically. This collection would be done as people harvest fish rather than a standardized assessment as might occur in a more traditional field site. As such, the major issue here is that the data would not be standardized in the way that scientific data collection might be.

This way of defining local is fine in scale temporally all the while contributing to models that produce outcomes on longer time scales. Local is seasonal human activity or data collected opportunistically as harvesters of the resource interact with fish. Seasonal activities are formatted in a calendar view making them easy to translate into the models. An annual management report (1999) notes that seasonally, catch calendars (figure 28) are "mailed to all Kuskokwim Area households that had been identified as "usually fish."

[23] They do note that the utility of these data is contingent on the ability of progress in genetic research particularly SNPs and to ensure that representativeness could be achieved.

Location on the river intersects with the kinds of fish caught and subsequent fishing activities. As such, three differently styled calendars were sent out: 1- Lower and Middle regions, communities on the Bering Sea coast, and communities in the Upper Kuskokwim River as far as Stony River; 2- remaining household in the Upper River; 3- households in Quinhagak, Goodnews Bay, and Platinum.



Appendix S. 1. 1999 Kuskokwim Area Subsistence Salmon Harvest Calendar.

Dear Subsistence Fishers:

Please write in the number of salmon that people in your household caught for subsistence. Include all subsistence salmon that were caught, including those you gave to others and those you may have caught for dog food. DO NOT include salmon that you sold when commercial fishing.

Our address is on the back of this calendar. When i finished fishing, you can fold the calendar so that our return address is visible. DO NOT PUT POSTAGE ON THE CALENDAR WHEN YOU RETURN IT TO US. We have paid the postage.

This calendar is sent to you by the Subsistence Division of the Alaska Department of Fish and Game in Bethel.

NAME

Thank you for helping to document subsistence haarvests. If you have any questions, please call (907) 543-31100.

Bulk Rate
U. S. Postage
Paid
Fairbanks, AK
Permit No. 99

MAY 1999 SUBSISTENCE SALMON CALENDAR

	SUNDAY	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY
	16	17	18	19	20	21	22
TARYAQVAK =	King	King	King	King	King	King	King
IQALLUK =	Chum	Chum	Chum	Chum	Chum	Chum	Chum
SAYAK =	Red	Red	Red	Red	Red	Red	Red
	23	24	25	26	27	28	29
CHINOOK =	King	King	King	King	King	King	King
	Chum	Chum	Chum	Chum	Chum	Chum	Chum
SOCKEYE =	Red	Red	Red	Red	Red	Red	Red
	30	31	* There are a number of small operators drying fish whose output is consumed locally or sold to persons traveling in the region. On the river above Bethel at Steamboat Slough, Neal Corrigan dried 1,500 fish for his own dogs. At Napamute, George Hoffman dried 2,000 small fish and 30 kings for his dogs and barter. At Crooked Creek, a man named Dennis Perrin dried 6,000 small fish and 200 kings." L.G. Wingerd. Alaska Fisheries and Fur Seal Industries. U.S. Bureau of Fisheries. 1922				
	King	King					
	Chum	Chum					
	Red	Red					

JUNE 1999 SUBSISTENCE SALMON CALENDAR

	SUNDAY	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY
			1	2	3	4	5
	Chinook		King	King	King	King	King
			Chum	Chum	Chum	Chum	Chum
			Red	Red	Red	Red	Red
	6	7	8	9	10	11	12
TARYAQVAK =	King	King	King	King	King	King	King
IQALLUK =	Chum	Chum	Chum	Chum	Chum	Chum	Chum

Figure 28. Catch calendars, 1999

Levin (1992) makes a similar case with respect to the issues of scale in ecology arguing for universal vs. particular:

> *we trade off the detail or heterogeneity within a group for the gain of predictability; we thereby extract and abstract those fine-scale features that have relevance for the phenomena observed on other scales.*

In Levin's use of 'local', he signals those cases that are unique and unpredictable. The 'localized effects' of a disturbance (e.g., small fires) are "special cases of situations where systems may be viewed as spatiotemporal mosaics, variable and unpredictable on the fine scale, but increasingly predictable on large scales" (1954-5).

I showed how local is fine in scale, and how the collection of fine-scale data has a different temporal orientation. While people are often collecting this data opportunistically alongside everyday activities, the data can be used in models to contribute to much longer time scales.

Discussion

Local is instantiated in the many tools of management. The puzzle I explore is how instruments and theories evolve to meet the challenges of scaling, in this instance, for the local. Science has informed management, and subsequently, management — with respect to natural resources — is necessarily entangled with the industries that harvest those resources. Ostensibly this was the harbinger of management of wild salmon: to establish sustainable ranges that can guide harvest, balancing the commodification of natural resources and the need to allow enough of a species to produce the next generation.

Many of the climate models used today are focused on future visions of climate impacts; however, the predominant temporal orientation for modeling for salmon management is reconstruction. Reconstructing the past to understand the present and ultimately make predictions about the future is the main focus of many of the models used for management. Given this temporal orientation, a common data entity (in the management of Alaska

salmon) is escapement, which is the enumeration of migrating fish as they go upstream. Escapement data is mostly an index of abundance rather than a count of every fish in the system. These "representative" counts are conducted at various sites; counting towers, weirs, aerial surveys, beach seining, gillnet samples, sonar, mark recapture, and catch cards are a few prominent methods of counting. And some of these technologies have been around for centuries[24].

Figure 29 depicts the relationships between different themes that emerged from interviews with local stakeholders. In this network diagram, I explore different co-occurrences in my coded interviews to answer the question of how is local enacted at different temporal scales.

[24] Schiefer (2019) shows how fish weirs are more than a method of data production, but are places where human-fish relations unfold. Archeologist (Losey, 2010) notes that weirs have been known in southwest Alaska since 3800 BC. Once a fishing technique, it is now only used to observe fish rather than harvest them. She notes that "fish weirs too should be understood as a manifestation of governmental ideas about where a salmon belongs and when they can be accessed as a resource." Really, what she shows is the long history of fish weirs

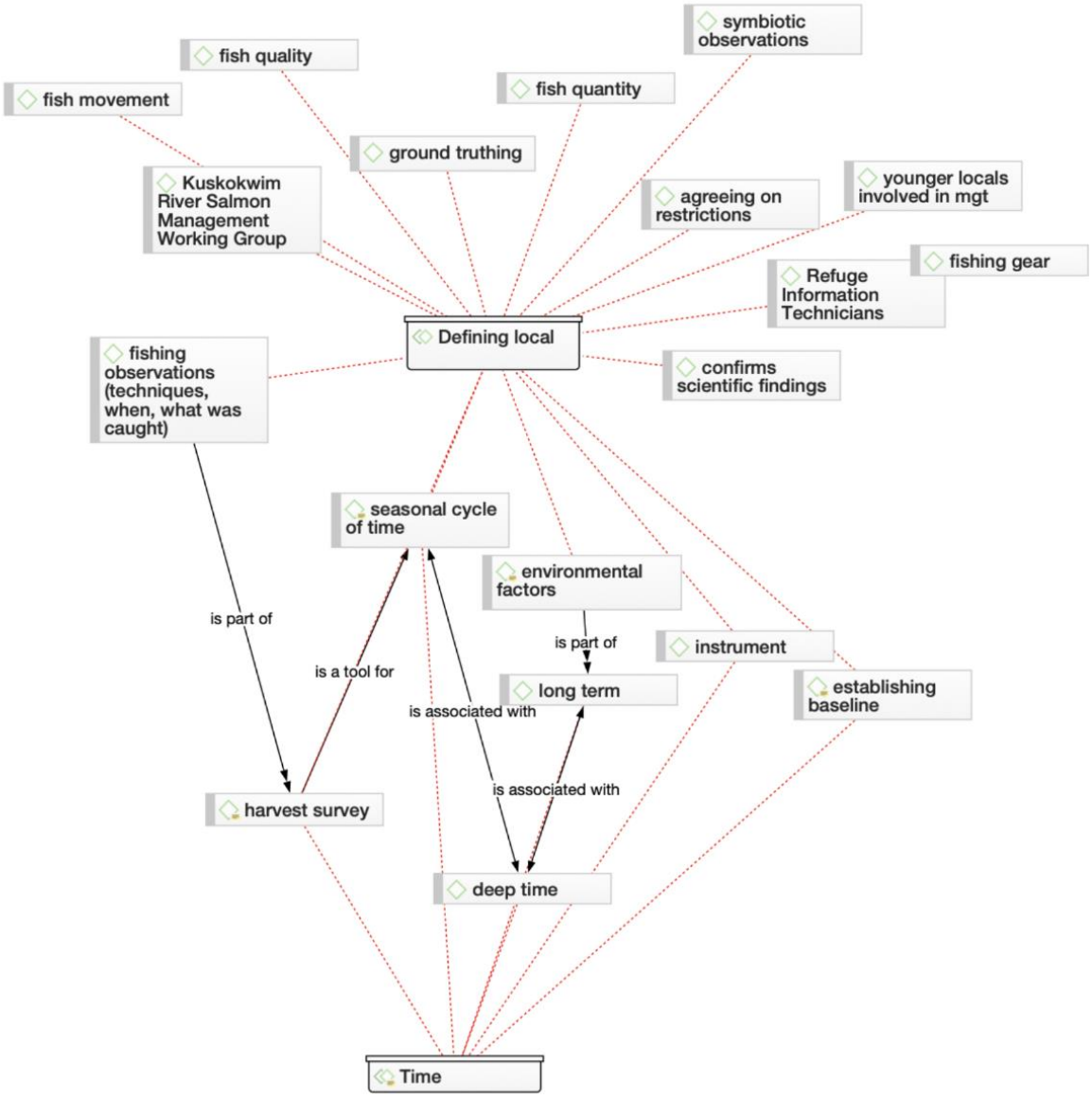


Figure 29. Network map of themes related to defining local participation in data production. I generated this network map with the Atlas.ti software after qualitatively coding interviews and categorizing them thematically.

As Asdal's little tools of democracy create new issues for parliament (Asdal, 2008), so too are these models little tools that translate local to global. Models are a cornerstone in understanding phenomena that span beyond human lifetimes and are one of the most

common ways that scientists deal with issues of scale, e.g., climate models that make predictions far out into the future. And, in the last few decades, computer models have emerged as one of the most important tools for scaling for global climate science (Edwards, 2010; Oreskes, 2003). In his work on the relationship between models and data, Edwards (2010, 1999) outlines the techniques and problems highlighting the problem of scaling as the major challenge in climate modeling. Much of his work outlines how models *produce* global data, which suggests that models are a kind of instrument in their capacity to produce data.

However, local data are also produced, not a given baseline suggesting a linear progression from local to global. Local or particular is a kind of scale. In a systematic review of social-ecological systems (SES) modeling, we argued that models are scalar devices in that they consolidate work across differences into representations that can be legible (Steger et al., 2021). In particular, we looked at aspects of models such as model specificity, intended and achieved model purpose, data types, model extent, and model resolution to determine how scientists use these models to scale for complex systems. In SES modeling, local scale is critical to understanding the context of an environmental issue. It refers to a spectrum of fine to coarse. Often, the more particular or locally-applicable the results are, the less theoretical the study is.

Not only do computer models produce data in the form of predictions, they are also theoretical tools - often being described as a way of explaining data. For example, among the two most popular conceptual models that underpin management models—Ricker Spawner-Recruit model and the Beverton-Holt model—the Ricker model is often the preferred model of state management because it has a more conservative tendency with respect to harvest

numbers. Even if the model does not perfectly fit the data every time, management prefers to underlie escapement projects with a precautionary approach.

Similar to the predictive nature of numerical weather prediction (NWP) models, the salmon forecasting models in Alaska seek to predict by initializing the models with some form of observational data. In understanding Alaska salmon, models produce predictions about salmon abundance. One informant defines a model as a "numerical quantitative explanation for how the data (which are also a number) came to be." Another notes that the in-season management model is a mathematical representation of salmon population and fishery and is used to evaluate seasonal management decisions. Models for the Kuskokwim and Yukon salmon populations promise to consider uncertainty to "provide an objective basis for determining critical information needs to inform salmon management."

The impact of computational advancements cannot be overstated. Some scientists I spoke with noted that the changes from the 1990s allowed people to consider that data on "a much more realistic scale" by providing not just point estimates but also allowing for a view of error possibility. In other words, having confidence in the range of error allowed for greater confidence.

Conclusion: Local in the construction of universality

To understand how scientists define local scale, I offered insights into how local is instrumented with human activity and how local is defined in relation to global scale and to community scale. In this chapter, I offered three qualities of 'local' based on ethnographic field work and archival research in the Kuskokwim region of Alaska. In exploring how scientists and state, federal, and tribal managers as well as native and non-native residents define local, I highlighted the asymmetries in knowledge production and some of the

strategies scientists take to overcome these asymmetries. I showed how a common approach to producing finer-scale data is to rely on residents and subsistence fishers, and I outlined ways that subsequent results can clash with on-the-ground realities.

This chapter showed how local is often instrumented with human activity because the human time scale and spatial coverage is fine in scale. Local stands in for either public, another monolithic word, or complexity and particularity. It refers to the activities of human or social life, or the participatory politics. In this light, co-production has some salience to this point as it refers to the co-production of scientific and policy realms. However, local can also signal nuance, complexity, or particularly. It is the knowing of a location at a fine-scale. This is in part because 'local' people in their everyday activities can provide fine-scale data over a small area and a small amount of time. This is not meant to refute the idea that local people can provide broader scale findings. On the contrary, oral traditions have been cited as a way of providing a deep time perspective, but similar to scientific studies, these insights that speak to broader extents are also coarser in resolution. Lastly, I illustrated the power dynamics in management relations and suggest moving beyond aiming to empower people and instead toward understanding more bidirectional relations between management and local stakeholders.

From this empirical work, meeting minutes as well as interviews with in-season managers, I make the following suggestions:

- Given the number of comments about weather and environmental conditions, managers and management models should consider, and perhaps formalize ways of collecting, other aspects that impact harvest in the long-term. This dovetails with the

goals of the model focused on habitat diversity, and there is already plenty of existing interest on the ground in incorporating this kind of data and information.

- Openers should be timed with optimal drying weather to prevent fish spoilage and lead to more successful salmon preservation. This is because the timing of the fish intersects with environmental changes at a broader scale such as warmer and wetter summers. Management can not only respond to the changes in run sizes of the fish, but also to the local adaptation to a changing climate.
- Knowledge about timing of fish returns develop in a complex exchange of local observations and fishery management model output – in other words, a mutual shaping occurs in which the scientific relevance and potential biological and ecological basis of the local observation is made visible to management. This points to the issue that knowing whether a run is late or just small is often an issue of intuition and expertise rather than exact science.
- Because management is built on a commercial legacy, it uses concepts such as CPUE. However, in a subsistence context, CPUE is largely unrealistic. This is one example in which the local data that fishers provide is sometimes asymmetrical with local realities.
- Salmon return cyclically, such that one species such as Chinook salmon may experience low returns while another species experiences high returns. All salmon, however, are important to subsistence harvests, even though all salmon species do not have the same equivalency in utility or demand. The current management models in use rely on locals primarily for their harvest data. However, should Chinook salmon returns improve and the corresponding management structure change, it is unknown

whether this resource-intensive monitoring program is sustainable and/or needed for this kind of fishery management.

Beyond recommendations for management, this chapter highlighted how local scale is produced, challenging the 'naturalization' of local. Broadly, I showed how one strategy that attempts to reconcile scale issues is to model. The models are designed to scale up from local data input. Better characterizing what factors constrain or enable public involvement sheds light on how to better communicate scientific output and where local observations can be more formally incorporated into management that informs decision-making. This research implores scholars of infrastructure and STS to seriously consider the ways that local is produced in their work that critiques universality.

Chapter 8: Conclusion

Overview of contributions

> *"It matters little what moving body we adopt as our recorder of time. Once we have exteriorised our own duration as motion in space, the rest follows. Thenceforth, time will seem to us like the unwinding of thread, that is, like the journey of the mobile entrusted with computing it. We shall say that we have measured the time of this unwinding and, consequently, that of the universal unwinding as well." -Henri Bergson, Duration and Simultaneity, p.257*

As I illustrate in this dissertation, scale is an integral feature of working with scientific data. My research addresses gaps in STS and information science as I ask how scientists instrument scale when producing knowledge about ecological phenomena and how long-term infrastructures in ecology are sustained through time.

I begin this chapter by reiterating what I traversed throughout this thesis. To draw together my empirical work, I offer a synthetic claim about the properties of scale and

conclude with a theoretical and operational definition of scale. Contrary to philosophical musings, I ground this concept in empirical work about how ecological research infrastructures reconcile scalar issues within data. I close with a few implications for the three domains I engage in this thesis.

Given the current focus on developing tools that support the integration of data science practices in the natural sciences, this research engages the practices and politics of open data by exploring the historical ways that data infrastructures have been created and sustained through time. As an ethnographer, I take part in current data collection, data acquisition, and data cleaning and explore archives to gain a more historical perspective. Thus, this research looks at how knowledge about data is distributed throughout time and space, taking the system as the unit of analysis (Hutchins, 1995). In other words, this research looks at both hardware and software as well as social norms, communication, and research artifacts.

Building on the conceptual lens of knowledge infrastructures (Edwards et al., 2013), I probe the various components that facilitate or constrain scientific research. This thesis takes an expansive view of data infrastructure to explore scale at three different levels of knowledge infrastructure: global, temporal, and local. The main activities in which I explore scale are 1- integrating heterogeneous data; 2- instrumenting temporal scale with specimens in an ecological field program; and 3- instrumenting local scale with human activity in participatory modeling endeavors. I now offer a summary of the findings from the three sites to build my definition of scale.

Scale in scientific practice

As a starting point, I define scale as the lens through which scientists segment complexity. In the preceding chapters, I explored how scientists instrument phenomena at different scales. Although this chapter outlines cross-cutting themes, each individual case answers specific questions. The study of SASAP addresses the ways in which scientific programmers encounter scale in data synthesis and how they develop strategies for contending with data integration issues. Additionally, my ethnography of the Alaska Salmon Program – a long-term research infrastructure - outlines how scientists instrument for time with specimens. This ethnographic work shows how a long-term infrastructure produces durable objects and how it re-instruments itself through time. In the third empirical chapter, I inspect a specific working group from SASAP that is uniquely positioned to produce new data with community members, and engages how scientists instrument the local. This concluding chapter considers some of the cross-cutting themes that this dissertation attends to, such as temporality, locality, and instrumentation.

I find myriad ways that contemporary data infrastructures support scientific research. Some examples of this include the cleaning and assuring of data, providing data support services such as cloud servers (e.g., Aurora), using modern data practices such as Markdown in R, managing in GitHub, and other practices that make data 'scalable'. However, the absence of the field from modern data infrastructure initiatives elides the key to resolving issues of scale post-data collection. Pushing against the popular view that data should be shared as a public mandate, my ethnographic work on sites of data collection – primarily through ethnographic work on the Alaska Salmon Program – highlights that knowledge is embedded

and distributed in the field and is not something that can easily be extricated from the field and into a database.

In the following section, I offer some key properties of scale that my empirical work surfaces. Similar to Star's ethnography of infrastructure (Star, 1999), I provide a general language for thinking with scale.

Properties	**Description**	**Design implications**
Scale is interstitial and infrastructural.	Scale is visible in the interstices of data and phenomena. Endogenous aspects of the phenomena shape the exogenous. Understanding these intrinsic qualities are important to making sense of varied data formats. This is akin to the quality of 'visible upon breakdown' in infrastructure studies (Bowker and Star, 2000).	*Stay close in proximity to the field.* Much of the knowledge about mismatches came from local experts who had first-hand knowledge of the site of data production. (Baker, 2015)
Scale is dynamic.	Scale is not static but is constantly shifting (Stanford et al., 2005). Instrumenting temporal scale is done by relying on the subject matter itself as an archive. Processes acting over decades are hidden and reside in the invisible present (Magnuson, 1990).	*Invest in other narratives of science and consider the practical implications of temporality.* Mismatch between scientists' perception of time / space and ecosystem-wide responses to time / space occludes direct observation of gradual changes.
Scale is reductive as it is also relative.	While scale is relative to the extent or scope of a study, scale is also reductive. It is a way to reduce components of a larger system to produce meaningful insights. Scale in ecology aligns more closely to logics of sustainability (Tsing, 2015) than logics of scalability, less about extensibility and interoperation and more about the correct levels to study phenomena. This refutes the notion of a purely reductionist perspective on scale (e.g., West, 2017) or a purely relativist perspective (e.g., DiCaglio, 2021)	*Question the rhetoric of universality.* Sampling is always partial. Logics of sustainability have more relevance to scale in ecology than logics of economies of scale.

Scale is a material representation.	Scale is a representation produced by an instrument. The material representation of scale appears in the form of standards, data norms, sampling protocols, and visualizations, which facilitate its circulation into other contexts. Much of this work includes ensuring that internal representations of space and time are aligned with external representations. Scale is often represented numerically, but much occurs prior to its placement in a dataset.	*Follow the instrument*. Instruments provide a window into theoretical concepts and assumptions of science at the time and on the societal values, regulatory nuances, and contested realities (Hacking, 1983; Baird, 2004; Wise, 1995; Latour, 1999).

Table 7. Properties of scale

Scale is interstitial and infrastructural

Taking up Helm and Shavit's (2017) call to research instrumentation, I look at the instruments scientists use to make their object of study legible to data infrastructure (e.g., counts of salmon are formatted into excel spreadsheets, which can be easily parsed with scientific models). Leonelli's (2018) data time and phenomena time helps shed some light on the distinctions between endogenous and exogenous that Helm and Shavit propose. In other words, there are both intrinsic (endogenous) and extrinsic (exogeneous) aspects that impinge on the ability to understand a phenomena. What often pushes back against attempts to make data interoperable is the translation of a phenomena (in this case, salmon) to a data point in a database. As such, the scientific programmers take an approach to reconciling their concerns about future unknowns by highlighting that uncertainty. This requires engaging with the local or particular qualities present during data production.

Scale is found in the interstices of endogenous and exogenous. This is because there is a rhythm to the life sciences that is often eclipsed in the data ecosystem. For scientific programmers, scale appears at different lifecycle stages of data and at larger conversations

about the scalability of data. For salmon biologists, it is the life cycle of the fish that is of interest. However, the scientific programmer is frequently reconciling issues produced in the translation of the salmon – their timing and mechanisms of survival – into data points.

As I illustrate in the SASAP case, scale is integral to data science endeavors. One example that illustrates this is in the data formatting for age, which I reference in previous chapters. Due to how some salmon species tend to lose scales more quickly than others, those differences have to be accounted for in the material format of the data (e.g., the scale cards which are turned into data spreadsheets). This is an example of how scientific programmers encounter scale as they work through anomalies in data and construct explanations for differences across data.

All data integration efforts offer a benefit and exact a cost. A benefit might be working with more or more interoperable data. A cost may be some compromise in accuracy, granularity, or context. In the data integration initiative that I was a part of, precautionary decisions were frequently made to avoid losing accuracy or propagating unknown errors downstream. By adding information in a standardized notation to a dataset, scientific programmers engage in a design tradition that makes uncertainty visible to downstream users.

In the areas where misalignment occurs, it is evident that there has been a mismatch between the perceived life cycle of the research phenomenon and the life cycle of data. Given my discussion of the scalar mismatches that transpire in ecology, I show how misalignment between scales often happens at the juncture of these life cycles. In my three empirical chapters, I provide snapshots at different stages of research infrastructure of how mismatches or alignments occur between data life cycles and natural life cycles.

Scale is dynamic

There are different temporal rhythms in scientific collaboration (Jackson et al., 2011), which impact the way scientific phenomena are instrumented. Some collaborations focus on consistent, long-term monitoring through many seasons; others are focused on collecting daily or even annual data; some emphasize deep time, or a temporality that extends beyond human observation. Throughout this work, I attend to questions related to temporal scale, such as, 1- how are different experiences of time integrated into data infrastructure? 2- how does an understanding of temporal scale impact material practices or understanding of salmon? And, 3- what does that mean practically for data systems?

In all three cases, there are many discussions around temporality. These primarily center around achieving time scales that move beyond the view of a human or even of a human organization, but instead speak to a temporality that is closer to the earth. My research highlights the clash of temporalities that emerges between dominant temporal logics (Mazmanian et al., 2015; Puig de la Bellacasa, 2015) and sustainability. In the dominant framework of science for natural resource management, salmon — though "wild" — is discussed as a product (e.g., maximum sustainable yield, stock, surplus, productivity, etc..).

In the chapter on different notions of the local in determining how community-based monitors can collect new data, I speak with Indigenous in-season managers who offer a deep time perspective of the salmon systems they help manage. However, the primary model that scientists consider local community members could collect data for were models that have a daily / seasonal temporal dimension. This suggests that dominant temporal logics are sometimes misalign with local ideas of temporality.

In SASAP, scientific programmers encounter many different orientations toward time. While a major aspect of the data scientific work as it pertains to ecological studies on Alaska salmon is how to conduct studies over a spatially diverse landscape as well as how to coordinate across organizational boundaries, these actions are also occurring over a heterogeneous temporal 'landscape', or timescape. Scientific programmers contend with data that were collected in the past as well as data that might be used in the future. This suggests an awareness of what Bellacasa (2015) points out, which is that studies of environmental change are often obscured by the present timescape. Additionally, it represents an earnest attempt to overcome the tensions between different temporalities: biological, ecological, historical, geological. There are major concerns about ensuring accuracy into the future and determining how accurate data are from the past. Moreover, to achieve data integration, scientific programmers are highly concerned with acquiring adequate or consistent temporal coverage. This is because to house a complete dataset, it is important to account for all years of data. This concern is less front and center for the long-term field program as they have more control over data collection and more knowledge of its creation in the past. The main focus is on collecting consistent data over many decades, and as such, their proximity to the data makes the concern for temporal coverage more achievable.

In the long-term field site, much of the day to day concerns with temporality are the everyday experience of being in the field – deciding when to go collect data, looking at the weather forecast and how far apart streams are; however, zooming out shows how the program has considered time as a critical component in some of the sampling protocols and seasonal data collection norms. The field program's activities are orchestrated by the

behavior of the fish. Because the fish return in August, much of the long-term data collection (such as carcass surveys) coincide with this return. If the fish were to suddenly disappear or start appearing at a different time, sampling protocols would shift. This is because the field program is necessarily more attuned to the research phenomenon that data infrastructures that are more isolated from the field.

Through ethnographic field studies, this research highlights the ways in which ecologists instrument their field sites to understand temporal scale, and the ways in which dominant temporal logics sometimes shape and sometimes break with scientific practice. Scale is the outcome of turning temporality that is endogenous or intrinsic to nature into temporality that is exogenous and compatible with other data types produced for understanding. In other words, shifting the scale is what makes a data point durable over time.

Scale is reductive as it is also relative

Scale is more than a law to be applied to all disciplines. While its Latin root form – *scala* – meaning ladder suggests infinite rungs facilitating upwards growth, scale is more than layers or levels in a hierarchy. Contrariwise, ecological systems show that there are thresholds to growth not captured in the predominant rhetoric of scale. Perhaps the most sobering example of the limits to growth is found in biology with aging that occurs in a human lifespan. This aging ultimately leads to death, and some theorists explain this decay with mathematics. While a mathematical theory brings scale theories closer to law, these do not explain *why* ecological and biological systems fail. The logic of the growth model of scale implies that exponential growth will come with more interoperable, connected systems. In

this logic, there is an upward movement toward a bigger, better future. Importantly, this is not a logic built on sustainability or one with a view of history.

While scale is a relative quality -- as Levin (1992) points out -- scale is also a reductive device. It is a way of reducing components of a larger system into smaller, related components to make sense of the whole. As scale parses and gives a name to nuance, it reduces phenomena to its essence. As Levin (1992) argued, changing the scale of observation can help "move from unpredictable, unrepeatable individual cases to collections of cases whose behavior is regular enough to allow generalizations to be made."

In ecological science, scale contains conceptual challenges, such as how to make sense of phenomena and how scale shapes understandings of complexity. In contrast, information science emphasizes scalability as a way to cleanly interoperate data for future use. To represent complexity, the shifting habitat mosaic (Stanford et al., 2005) is a conceptual device that is useful to understanding how ecosystems respond to emergence. While it offers a theoretical perspective for attending to scale, data issues remain due to the different conceptualizations of scale in information science. In other words, the data solutions are never universal; they are local, contextual, and situated. Thus, what does it mean to create a data infrastructure that is flexible enough to attend to local particularities while universal enough to offer insights cross-scales?

I explore these issues around mobility, scalability, and relationality in my chapter on SASAP, which represents a contemporary data infrastructural approach to the long-term. By tracing the collaboration and communication practices of scientific programmers in the data synthesis project, I uncover how interoperability (mobility or circulation) and accessibility (locality) are achieved. While most of the work conducted was with the aim of achieving

interoperability across diverse datasets, there was a concern with preserving the locality of the data. These moves to reduce while also notating nuance are often at odds with one another.

This relative quality of scale is also evident in the case on modeling in the Kuskokwim. Here, I learn that local is often construed as fine-scale. Furthermore, the human life span – while many have critiqued as having a scalar bias (only seeing human scale) – can produce highly fine-grain data. In this way, the move to produce general insights through quantitative modeling is rich with discussion about the different relationships inherent in different data. Moreover, the relationship that local residents have with the land and the resources makes their involvement in the collection of data useful to the larger scientific and management goals.

A major aspect of what sustains a research infrastructure is its connection to the place in which it operates. This is most evident in the case in the Alaska Salmon Program, a research infrastructure that has been in operation since 1946 and continues to produce scientific research today. In understanding how the infrastructure changes over time, it is clear that data are contingent on present theories, experts, questions, and instrumentation.

As many examples from this study illustrate, the move to make data interoperable is frequently thwarted by scalar challenges that occurred in the field or with decisions that were made previously. It is not useful to science, however, to catalogue every nuance involved in scientific endeavors. In other words, the insight that at each scale, there are different concerns has been known for some time. On the contrary, what is important is that scale is both a phenomenon inherent in non-human systems and also produced

sociotechnically by scientists looking to parse the complexity of the natural world. It cannot be a universal. It is universally contextual and relational.

Scale is a material representation

There are many different ways to scientifically represent an ecosystem. As instruments segment phenomena, scale is produced as a representation of the aspect of the phenomena that is important to the question at hand. This is akin to what Latour calls 'immutable mobiles' (1990), or the smaller pieces of an entity that can be translatable, mobile, and ultimately, scalable.

When considering the whole of research infrastructure, there are material qualities beyond data that facilitate the production of scale. In a book on effective ecological modeling and adaptive monitoring, Lindemayer and Likens (2018) conclude with a call that "little things matter a lot!". They highlight four seemingly small factors that matter for long-term monitoring: field vehicles, continuity of staff, access to a field site, and spending time in the field. As I illustrate in my ethnographic work on the Alaska Salmon Program, the instruments in the field (e.g. Secchi disk, Vandorn, thermometer) may be akin to the laboratory's microscope as these instruments are imbued with epistemic content and have material histories. Furthermore, there are standards that orchestrate data collection, and these standards serve as representations of scale from the past.

As such, I argue that in lieu of the "little things that matter" are: 1) logistical access (field vehicles, getting to remote areas, generators); b) political access (e.g. funding, relevance of research, personnel issues); c) contextual understanding from residing in the field for many seasons; d) institutional knowledge. It is important to separate the instruments from other

infrastructural aspects that aid in the production of scale. This is because the instruments are the tools that produce representations of scale.

Implications for Ecology

Data science, and information science more broadly, render the domain (in the case of this dissertation, salmon science) a service to the advancement of data analytic capability. In other words, the prioritization and preference for data scientific skillsets and advancement renders the domain incidental. But as my engagement with data scientists and field ecologists has shown, the salmon are not incidental – not to the ecologist naturally, but not to the data scientist either. Salmon science and ecology should resist being cast as a handmaiden to the task of data science.

A decade ago, Lindenmayer and Likens (2011) warned that ecology was losing its culture. Even while new tools (such as many that I have written about in this thesis, e.g., genomics, stable isotopes) were pushing ecology forward, cultural changes had led to declines in fields like taxonomy and natural history, they argued. These changes were to make room for an increased emphasis on data. This is evident in claims such as the one Powers and Hampton (2019) make: "ecology has now moved more fully into the 'big science' era", an era focused on broad-scale, macrosystems research.

The moment in which I enter into this debate is the outcome of an institutional momentum that has catalyzed the focus of data science and big data in the natural sciences. As data science casts itself as a novel approach for facilitating research, it also participates in discussions that have occurred in the sciences for decades: *how can we understand change at scales larger than human life spans?* This momentum has partially been set in motion through funding sources but also through training programs and university courses that

emphasize the importance of quantitative skillsets for handling data. This not only makes promises to forward scientific-thinking, but also comes with a promise of job security.

However, my research highlighted the many components that comprise ecological research and showed how working with data is much more than the datum itself. As I evidenced in previous chapters, without an understanding of ecological scale, the scale of data science is constrained (e.g., data scientists cannot produce easily extensible volumes of data when dealing with the heterogeneous and uncertain qualities so common to ecological data). Without understanding the research-based decisions about which spatial, temporal, and organizational level to sample for in a study or what considerations were made when designing a sampling protocol, the goal of expansion in data science is restricted. This suggests that focusing on the variety of data requires different approaches than those that have been used to deal with the volume, velocity, and veracity of data.

By highlighting the ways that scale intersects with the data life cycle model, I argued that the salmon are *not* incidental. What this study has shown is that proximity to the site of data production positions field programs to develop more research-intensive understandings of complexity as well as foundational insights to ecosystem function. The logic around data in data integration initiatives - in contrast - is focused on digitally tractable, scalable data sets that can span long time ranges and spatial scales. In other words, its logic is more align with growth and scalability than it is with complexity and intersecting levels of data.

The action taken by scientific programmers to flag erroneous data marks an opportunity for ecology. One major insight that ecology can bring to bear on data integration initiatives is that the scale at which a research study is designed can impact not only the data that are collected but the insights gleaned. A strategy for being more visible to

data management and data integration initiatives is to speak to an outstanding challenge they currently encounter which is how to properly incorporate context. My suggestion is not to position field ecology as a service to data initiatives. On the contrary, my research has shown that context is not a straightforward 1:1 to metadata. Rather, the implicit or tacit knowledge held by people who invest time in the field is a major gap in data integration work, because this knowledge is not easily replicated through existing data standards or norms.

Baker et al. (2013) addressed technology-oriented vs. science-oriented approaches to data noting that at the TFSE (Therkildsen Field Station at Emiquon) meetings, "stories of the field station and data collection circulated, appearing over time in multiple guises…such stories contributed not only to field station identity and visioning but also to the context within which data were generated and sustained." This points to the social qualities of how data and data infrastructure are sustained in a long-term field program. These social qualities of field programs cannot be understated. Many of the researchers in the field program recounted a love of the field and of field ecology. Many I met in the field as well as in the data integration initiative spoke of the field site as being one of their primary reasons for pursuing science.

This kind of attachment to place (Raymond, 2013) has recently been explored in work that looks at science affinity as a type of attachment to place and how that attachment impacts the development of critical thinking. I propose considering questions such as these for future endeavors: What do people learn in field ecology? Can it be applied elsewhere? How does developing an attachment to place impact scientific research? Rather than being socialized into the science du jour, the insights from this thesis should suggest a need to make foundational field research a field that offers more general skillsets.

Implications for Information Science

To the scientific programmers, this study has shed light on how the domain of ecology defines scale and how this plays a role in data production. A specific insight throughout this thesis is the way that the field work is taken for granted or forgotten in data initiatives. Concerning the question of how to develop data infrastructures for ecology, it is critical to explore how the local specificity of data production is attended to in an infrastructure that tends toward universality.

By offering a view from the field, I showed how scientists instrument temporal scale to produce long-term understandings and contrasted that with the way scale complicates data integration. This ethnography provides a detailed account of the myriad ways that understanding change over long time periods can be achieved and at different scales - fine to coarse.

For the data scientist or information scientist working in the ecoinformatics space this should do a number of things. Namely, it should: a) reveal the different orientation toward scale that ecologists have, b) highlight the challenge of achieving long-term understanding, and c) caution information science to not miss the field but to rather consider funding initiatives that also do the work of collecting data.

Kwa and Rector (2010) argue that until the mid-1990s, field data were held in databases that belonged to principal investigators or individual scientists; however, interdisciplinary collaborations focused on data became gradually focused on "field data as a distinct area of concern." They note that there are two generalizations that can occur: one from published case studies and one from the data assembled for those case studies. They argue that the

latter is what the data revolution invites – a generalization from the data itself. However, this move toward homogeneity is costly: citing Webster and Eriksson (2008, p. 109) that "the reduction of complexity and heterogeneity may lead to black-boxing of information that reduces the changes of biologically relevant inconsistencies."

This study contributed an understanding of the day to day work of data scientists, or scientific programmers, highlighting the many moments in the data lifecycle that —— are faced with challenges from the field. In considering larger discussions about the role that data plays in natural science, I have illustrated the ways that *open data* is not a given virtue nor should more closed field sites be lionized for their lack of data sharing. More critically, I have highlighted the hidden complexity of what may appear seamless or straightforward.

The approach that the field program takes differs from data infrastructure endeavors that focus on the integration of heterogenous datasets to create a consistent trove of data. Rather, the approach of the long-term field program is motivated by facilitating research over time, which emphasizes the ability to change with changing paradigms.

Given the critical importance of scale, how can scale so often be left out of data integration initiatives? Since the crux of the issue of scale in ecology is in the interaction between scales as well as the emergent properties of scale, this study recommends that data science initiatives consider scale more critically in future synthesis endeavors. It is not sufficient to focus on larger volumes of data, data that streams at higher velocities, or large-scale repositories. On the contrary, many of the issues with scale stem from a misunderstanding about the context of data production, which scales up through analysis.

Implications for Science and Technology Studies: Methodological approach for studying scale

For scholars of Science and Technology Studies, the message from this dissertation should be clear: methodological approaches are constantly shifting and to meet the demands of a shifting scientific landscape, we must adapt our methods as well. In this thesis, I showed how uncovering how the actors scale is found not just in the interstices of the domain (a well-known argument) but in the instrument used to scale a research phenomenon.

In all three chapters, my central object of research is the ways in which scientists themselves scale their research phenomena. Rather than looking at how they make sense of the infrastructure or scale that infrastructure as many other studies attempting to offer a theory of scale have done, I look at how scientists work with research phenomena and how they scale that research phenomena. This is tightly coupled with the use of instruments to produce data, but is a component of data production life cycles that is often left out of the equation.

Scientific accounts (usually in the form of a paper) treat phenomena as something waiting to be discovered. The first time through a scientific endeavor, however, it is much more like grasping than discovery (Garfinkel et al., 1981). Once science goes to publication, it becomes "naturalized" or is cast as if it was already there and was a discovery. Drawing on Latour's black box metaphor for the opaqueness of technological systems, Garfinkel et al. (1981) make the point that science is a process that becomes black-boxed in the form of a final publication. In other words, at the stage of the article, science forgets. Once the data are cleaned, annotated, and archived, the work is forgotten and backgrounded; however, a more networked or relational (e.g., infrastructural) view of this discovery process is that there are

factors constantly shaping 'discovery': contextual understanding, local particularities of data collection, negotiations with colleagues as to what 'really happened', and deeper understandings of time.

Similar to this ethnomethodological exploration into the discovery of science, Rheinberger (2005) showcases science as a social process noting that "empirical knowledge is lucid only after the event." Historical judgment is recurrent in that history always appears under the light of the "finality of the present." In this way, he sets up the challenges of scaling temporally noting that the historical epistemologist—unlike the historian—must constantly change in the face of new scientific developments. In other words, an epistemology for assessing scientific thinking "must be as plastic, as mobile, as fluid, and as risky as scientific thinking itself."

Science is contingent in a temporal sense; science "judges its historical past by discarding it. Its structure is in the consciousness of its historical errors." Lynch and Garfinkel as well as Rheinberger open up possibilities to consider methodological implications of instruments in the study of science. Or in other words, how can the instrument serve as a scalar device (Ribes, 2014)?

Related to her work on boundary objects, Susan Leigh Star argues that "materials and tools are the detritus of the work, often written out of scientific accounts" (257). While Star's contributions – on the surface – are more applicable to my explorations into how local publics are created, enrolled, and cycled out of science, she is also notes that the final product of science is only a partial view. As Latour traced the ways in which the savanna was turned into data, he illustrated the network through which science is achieved. As he notes: all the empty forms were set up "*behind* the phenomena, *before* the phenomena manifest

themselves, *in order* for them to be manifested" (p. 49). In other words, what I have tried to show is that science is a series of emergence and recurrence, mediated by the scientific instrument. Furthermore, if we want to understand the ways that scientists reconcile large-scale phenomena, the instruments for segmenting and categorizing the world are a good start.

From a methodological standpoint, this engagement tracks Ribes and Finholt's (2007, 2009) expansive view of scales of infrastructure as stretching across multiple scales of action: technological, institutional, and human/organizational. The technical includes the deployment of durable resources that enable collaboration. The human work includes the work that goes into maintaining infrastructure. And, the institutional involves the provisioning of resources at different governance levels. Read in this light, scales of action may be more akin to 'scaling'. This scaling aligns with different domains (e.g., technical, human, and institutional) and maps onto the categories of instruments that I discussed in my methods chapter: material (technological), processes and practices (human work and the heterogeneous experts involved), and networked (institutional feedbacks).

Methodologically, the study of scale offers many challenges and perhaps is one reason why it has been so under-theorized, particularly in the study of science. This is in part because scale cannot solely be understood through infrastructure alone. Nor can the actors themselves give a complete picture of scale. Because it is tied to perceptual biases, scale is illuminated by exploring the moments of breakage or change within a discipline or initiative. This connection to perception makes ethnography a useful strategy for locating and defining scale in the field.

In my thesis, I find scale through engaging with the data integration team to follow their work around error discovery and reconciliation, but I also find scale in collecting data alongside scientists, by listening to discussions about what kinds of data to incorporate into models, and by reading scientific journals about phenomena I had heretofore been unfamiliar with. I argue that understanding scale requires a deep engagement with the discipline and with the science. Without the science, we as STS scholars or social scientists can say nothing about the field.

All researchers are dealing with anomalies and reconciling what to do about uncertainty; however, when the goal is explicitly for the construction of data infrastructure, the discussion of error is more centered around how to make that uncertainty visible. In chapter 5, I apply a novel approach to qualitative, ethnographic research by acquiring a large data set on the communication traces left behind in a GitHub repository. This provides insight into the research site that just participation in the field site missed. Through this approach, I follow what scientific programmers considered 'issues' by analyzing their discussions on a platform used to coordinate data work. However, I also take a more traditional approach to ethnography by engaging in participant-observation in my field sites.

In sum, I suggest to "follow the instrument" as one might "follow the actor". A scalar device is a conceptual tool defined as an artifact-in-use (revealing the practices and expertise necessary for its usage), instruments that break, reach their limits, or are used in moments of transition (to understand historical gaps and how those gaps are filled), and to attend to scaling as a temporal activity in the way that science is historically contingent. As STS scholars continue to develop the scalar devices concept into an analytical tool, I encourage more explicit engagement with questions of knowledge translation and power.

Recommendations

In this thesis, I theorized scale-in-practice and offered a framework for approaching the issue of scale in data infrastructure projects focused on the natural sciences. In approaching this ethnographic work with the question of how do scientists instrument scale and how might data infrastructure better reconcile these challenges, I provided a synthetic claim about the properties of scale. This has implications for both designers of data infrastructure as well as scientists. In this final section, I summarize a few recommendations based on these aforementioned properties of scale.

- *Stay close in proximity to the field.* Throughout this work, I have wondered how does one bring the nuance of the field site or site of data collection into modern data infrastructure creation? This is a challenge that has typically been resolved through metadata, a standard format for notating the provenance of data. However, the construction of metadata post-data collection has proved challenging, if not impossible. This study recommends new strategies for bringing the field into contemporary data infrastructure. One strategy might be funding existing data production facilities in addition to data synthesis endeavors. Another might involve having advisors or having clear connections to the experts as SASAP did. Other strategies might focus on templating discussions around data to outline fundamental variables important to reconstructing the context of data production. While the scientific programmers were often inverting the infrastructure, an easier approach would have been to have some of these data ready at hand.

- *Invest in other stories of science*. Allen and Hoekstra (2015) suggest a multi-use narrative approach to science; however, they are not clear as to what this narrative approach looks like in practice. It is cast against the reductionist turn at the moment and draws on humanities disciplines to focus more strongly on *narrative* insisting that ecology is a "soft science". My work does not suggest ecology is a "soft science". I suggest, rather, that ecology is suffused with complexity and emergence. This is what makes ecology an interesting data problem: the data do not easily integrate. Focusing on place is one alternative to the common view of scale as a mathematical law and instead, focuses on the place-based knowledge that many of the scientists who study ecosystems have.
- *Make agency commitments longer.* Longer-term connections to the entities responsible for data production is important for building data infrastructure that are attuned to scale. This was evident in some of the work this research uncovered, particularly moments when data was shaped by the government agency responsible for management (e.g., when federal agencies such as NOAA collect length measurements in the marine environment, they use a different standard than those who collect length measurements at salmon spawning grounds). More critically important, building long-term contracts would ensure long-term maintenance of data repositories that materialize from data integration and synthesis projects.
- *Begin projects by defining terms.* This is particularly the case if public accessibility is a goal. The data synthesis project had public accessibility as an ancillary goal to the data integration work; however, much of the conversation about audience was

sidelined until the very end of the project. Bringing these discussions into the foreground would have helped scaffold the kinds of data that hold interest for researcher and public alike.

- *Question the rhetoric of universality when it comes to scale.* Theories that do not take into account ecological insights that scale is relative to other aspects in a study are often myopic and miss the nuance.
- *Consider the practical implications of temporality.* Timescapes of data tend to be linear or chronological. However, as this research has shown, timescapes of data can vary dramatically depending on the question being asked. Having datasets that stretch across long time spans can give the wrong impression and often lead to studies that are more about the size of the data rather than the questions being asked.

As data initiatives continue to grow and focus on the inherent interdisciplinarity of science, organizations will have more or less of a focus on data management. Throughout this work, I propose that data initiatives take an approach that recognizes scale as an integral part of scientific data production. The rhetoric on scale has been inchoate and varied, which has led to ambiguity over the usage of the term. As such, scale has frequently been excluded from data discussions, which has caused downstream effects on data products. This thesis has aimed to provide clarity particular to its use in ecoinformatics, and I define scale as the lens for understanding complexity as that complexity relates to other aspects of a system. Further, scale is a challenging concept to grapple with it as it is often used to refer to everything. Scale is everywhere and as such, nowhere.

REFERENCES

Acker, A. (2015). Toward a Hermeneutics of Data. *IEEE Annals of the History of Computing*, 6.

Ackerman, M. S., Dachtera, J., Pipek, V., & Wulf, V. (2013). Sharing Knowledge and Expertise: The CSCW View of Knowledge Management. *Computer Supported Cooperative Work (CSCW)*, *22*(4–6), 531–573. https://doi.org/10.1007/s10606-013-9192-8

Ackoff, R. L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, *15*, 3–9.

Allen, T. F. H., & Hoekstra, T. W. (2015). *Toward a Unified Ecology*. Columbia University Press.

Allen, T. F. H., & Starr, T. B. (2017). *Hierarchy: Perspectives for Ecological Complexity* (Second). University of Chicago Press.

Allen, T., & Hoekstra, T. (1991). Role of Heterogeneity in Scaling of Ecological Systems Under Analysis. In *Ecological Heterogeneity*. Springer.

American Association for the Advancement of Science. (1989). *Science for all Americans: A project 2061 report on literacy goals in science, mathematics, and technology.*

Arnold, D. (2009). *The Fisherman's Frontier: People and Salmon in Southeast Alaska*. University of Washington Press.

Aronova, E., Baker, K. S., & Oreskes, N. (2010). Big Science and Big Data in Biology: From the International Geophysical Year through the International Biological Program to the Long Term Ecological Research (LTER) Network, 1957–Present. *Historical Studies in the Natural Sciences*, *40*(2), 183–224. https://doi.org/10.1525/hsns.2010.40.2.183

Asdal, K. (2008). On Politics and the Little Tools of Democracy: A Down-to-Earth Approach. *Distinktion: Journal of Social Theory*, *9*(1), 11–26. https://doi.org/10.1080/1600910X.2008.9672953

Asdal, K., & Hobæk, B. (2016). Assembling the Whale: Parliaments in the Politics of Nature. *Science as Culture*, *25*(1), 96–116. https://doi.org/10.1080/09505431.2015.1093744

Baird, D. (2004). *Thing Knowledge: A Philosophy of Scientific Instruments*. University of California Press.

Baker, K. S. (2017). *Data work configurations in the field-based natural sciences: Mesoscale infrastructures, project collectives, and data gateways*.

Baker, K. S., & Bowker, G. C. (2007). Information ecology: Open system environment for data, memories, and knowing. *Journal of Intelligent Information Systems*, *29*(1), 127–144. https://doi.org/10.1007/s10844-006-0035-7

Baker, K. S., & Mayernik, M. S. (2020). Disentangling knowledge production and data production. *Ecosphere*, *11*(7). https://doi.org/10.1002/ecs2.3191

Beadle, G. W., & Tatum, E. L. (1941). Genetic control of biochemical reactions in Neurospora. *Proceedings of the National Academy of Sciences*, *27*, 499–506.

Beaulieu, A. (2010). From co-location to co-presence: Shifts in the use of ethnography for the study of knowledge. *Social Studies of Science*, *40*(3), 453–470.

Belanger, D. O. (2006). *Deep Freeze: The United States, the International Geophysical Year, and the Origins of Antarctica's Age of Science*. University Press of Colorado.

Bisbal, J., Lawless, D., Bing Wu, & Grimson, J. (1999). Legacy information systems: Issues and directions. *IEEE Software*, *16*(5), 103–111. https://doi.org/10.1109/52.795108

Bocking, S. (1995). Ecosystems, ecologists, and the atom: Environmental research at Oak Ridge National Laboratory. *Journal of the History of Biology*, *28*(1), 1–47. https://doi.org/10.1007/BF01061245

Bocking, S. (2004). *Nature's Experts: Science, Politics, and the Environment*. Rutgers University Press.

Bocking, S. (2013). Examining the Environmental History of Arctic Ecological Science. In *New Natures: Joining Environmental History with Science and Technology Studies*. University of Pittsburgh Press. https://doi.org/10.2307/j.ctt5vkgkn

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 20.

Bowker, G. (1994). *Science on the Run: Information Management and Industrial Geophysics at Schlumberger, 1920-1940*. The MIT Press.

Bowker, G. (2005). *Memory practices in the sciences*. MIT Press.

Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2009). Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International Handbook of Internet*

Research (pp. 97–117). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9789-8_5

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. MIT Press.

boyd, danah, & Crawford, kate. (2011). Six Provocations for Big Data. *Information, Communication, & Society*, *15*(5), 662–679.

Branson, J. B. (2007). *The canneries, cabins, and caches of Bristol Bay, Alaska*. U.S. Dept. of the Interior, National Park Service, Lake Clark National Park and Preserve.

Brennan, S. R., Schindler, D. E., Cline, T. J., Walsworth, T. E., Buck, G., & Fernandez, D. P. (2019). Shifting habitat mosaics and fish production across river basins. *Science*, *364*(6442), 783–786. https://doi.org/10.1126/science.aav4313

Brennan, S. R., Zimmerman, C. E., Fernandez, D. P., Cerling, T. E., McPhee, M. V., & Wooller, M. J. (2015). Strontium isotopes delineate fine-scale natal origins and migration histories of Pacific salmon. *Science Advances*, *1*(4), e1400124. https://doi.org/10.1126/sciadv.1400124

Brown, J. H. (1994). Grand Challenges In Scaling Up Environmental Research. In W. K. Michener, J. W. Brunt, & S. G. Stafford (Eds.), *Environmental Information Management and Analysis: Ecosystem to Global Scales*. CRC Press.

Bryant, A. (2002). Re-grounding grounded theory. *Journal of Information Technology Theory and Application*, *4*(1), 25–42.

Burton, M., & Jackson, S. J. (2012). Constancy and Change in Scientific Collaboration: Coherence and Integrity in Long-Term Ecological Data Production. *2012 45th Hawaii International Conference on System Sciences*, 353–362. https://doi.org/10.1109/HICSS.2012.178

Callahan, J. T. (1984). Long-term ecological research. *BioScience*, *34*, 363–367.

Callon, M., & Latour, B. (1981). *Unscrewing the Big Leviathan: How actors macro-structure reality and how sociologists help them to do so*. 17.

Campana, S. E. (2005). Otolith science entering the 21st century. *Marine and Freshwater Research*, *56*(5), 485. https://doi.org/10.1071/MF04147

Chapin, F. S. I., & Shaver, G. (1981). Changes in soil properties and vegetation following disturbance of Alaskan Arctic tundra. *Journal of Applied Ecology*, *18*(2), 605–617.

Charmaz, K. (2006). *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*. Sage Publications.

Cline, T. J., Schindler, D. E., & Hilborn, R. (2017). Fisheries portfolio diversification and turnover buffer Alaskan fishing communities from abrupt resource and market changes. *Nature Communications*, *8*, 14042. https://doi.org/10/f9k6gx

Cohen, A. (2003). *Paleolimnology: The history and evolution of lake systems*. Oxford University Press.

Cohn, M. (2019). Keeping Software Present: Software as a Timely Object for STS Studies of the Digital. In J. Vertesi & D. Ribes (Eds.), *A Field Guide for Science & Technology Studies* (p. 30). Princeton University Press.

Coleman, D. C. (2010). Big Ecology: The Emergence of Ecosystem Science. *University of California Press, Berkeley*.

Connors, B. M., Staton, B., Coggins, L., Walters, C., Jones, M., Gwinn, D., Catalano, M., & Fleischman, S. (2020). Incorporating harvest–population diversity trade-offs into harvest policy analyses of salmon management in large river basins. *Canadian Journal of Fisheries and Aquatic Sciences*, 1–14. https://doi.org/10.1139/cjfas-2019-0282

Crawford, K. (2014). Critiquing Big Data: Politics, Ethics, Epistemology | Special Section Introduction. *International Journal of Communication*, *8*, 1663–1672.

Cronon, W. J. (1995). *The Trouble with Wilderness; or, Getting Back to the Wrong Nature*. 24.

Dabbish, L., Stuart, C., Tsay, J., & Herbsleb, J. (2012). Social coding in GitHub: Transparency and collaboration in an open software repository. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work - CSCW '12*, 1277. https://doi.org/10.1145/2145204.2145396

Dalton, C. M., Taylor, L., & Thatcher (alphabetical), J. (2016). Critical Data Studies: A dialog on data and space. *Big Data & Society*, *3*(1), 205395171664834. https://doi.org/10.1177/2053951716648346

Delcourt, P., & Delcourt, H. (1983). Dynamic Plant Ecology: The spectrum of vegetational change in space and time. *Quaternary Science Review*, *1*, 153–175.

DiCaglio, J. (2021). *Scale Theory: A Nondisciplinary Inquiry*. University of Minnesota Press.

Donkersloot, R., Black, J. C., Carothers, C., Ringer, D., Justin, W., Clay, P. M., Poe, M. R., Gavenus, E. R., Voinot-Baron, W., Stevens, C., Williams, M., Raymond-Yakoubian, J., Christiansen, F., Breslow, S. J., Langdon, S. J., Coleman, J. M., & Clark, S. J. (2020). Assessing the sustainability and equity of Alaska salmon fisheries through a well-being framework. *Ecology and Society*, *25*(2), art18. https://doi.org/10.5751/ES-11549-250218

Dourish, P. (2001). *Where the action is: The foundations of embodied interaction*. The MIT Press.

Dourish, P., & Button, G. (1998). On "Technomethodology": Foundational Relationships Between Ethnomethodology and System Design. *Human-Computer Interaction*, *13*(4), 395–432. https://doi.org/10.1207/s15327051hci1304_2

Edwards, P. (2010). *A vast machine: Computer models, climate data, and the politics of global warming*. MIT Press.

Edwards, P., Bowker, G., Jackson, S., & Williams, R. (2009). Introduction: An Agenda for Infrastructure Studies. *Journal of the Association for Information Systems*, *10*(5), 364–374. https://doi.org/10.17705/1jais.00200

Edwards, P., Jackson, S., Bowker, G., & Knobel, C. (2007). Understanding infrastructure: Dynamics, tensions, and design. *Final Report of the Workshop History and Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures. NSF, Office of Cyberinfrastructure.* http://hdl.handle.net/2027.42/49353.

Edwards, P., Jackson, S., Chalmers, M., Bowker, G. C., Borgman, C. L., Ribes, D., Burton, M., & Calvert, S. (2013). *Knowledge Infrastructures: Intellectual Frameworks and Research Challenges* [Workshop]. University of Michigan School of Information.

Edwards, P. N. (1999). Global climate science, uncertainty and politics: Data-laden models, model-filtered data. *Science as Culture*, *8*(4), 437–472. https://doi.org/10.1080/09505439909526558

Edwards, P. N. (2003). *Infrastructure and Modernity: Force, Time, and Social Organization in the History of Sociotechnical Systems*. 42.

Faniel, I. M., & Jacobsen, T. E. (2010). Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data. *Computer Supported*

Cooperative Work (CSCW), *19*(3–4), 355–375. https://doi.org/10.1007/s10606-010-9117-8

Finney, B. P. (1998). Long-term variability of Alaskan sockeye salmon abundance determined by analysis of sediment cores. *North Pacific Anadromous Fish Council*, *1*, 388–395.

Finney, B. P., Gregory-Eaves, I., Sweetman, J., Douglas, M. S. V., & Smol, J. P. (2000). Impacts of Climatic Change and Fishing on Pacific Salmon Abundance Over the Past 300 Years. *Science*, *290*(5492), 795–799. https://doi.org/10.1126/science.290.5492.795

Fortun, K. (2009). Scaling and Visualizing Multi-sited Ethnography. In *Multi-site ethnography: Theory, praxis, and locality in contemporary research* (p. 14).

Fox, P. (2011). *Data Management Considerations for the Data Life Cycle* (NRC STS Panel).

Frey, DavidG. (1988). What is paleolimnology? *Journal of Paleolimnology*, *1*(1). https://doi.org/10.1007/BF00202189

Fujimura, J., & Fortun, M. (1996). Constructing Knowledge. In *Naked Science: Anthropological Inquiry Into Boundaries, Power, and Knowledge*.

Garfinkel, H. (1967). *Studies in Ethnomethodology*. Prentice-Hall.

Garfinkel, H., Lynch, M., & Livingston, E. (1981). The Work of a Discovering Science Construed with Materials from the Optically Discovered Pulsar. *Philosophy of the Social Sciences*, *11*(2), 131–158. https://doi.org/10.1177/004839318101100202

Garstang, W. (1900). The Impoverishment of the Sea—A Critical Summary of the Experimental and Statistical Evidence Bearing upon the Alleged Depletion of the Trawling Grounds. *Journal of the Marine Biological Association*, *6*, 1–69.

Geiger, R. S., & Ribes, D. (2010). The work of sustaining order in wikipedia: The banning of a vandal. *Computer Supported Cooperative Work (CSCW)*, 10.

Geiger, R. S., & Ribes, D. (2011). Trace Ethnography: Following Coordination through Documentary Practices. *2011 44th Hawaii International Conference on System Sciences*, 1–10. https://doi.org/10.1109/HICSS.2011.455

Giddens, A. (1990). *The consequences of modernity* (Repr). Polity Pr.

Glaser, B. G., & Strauss, A. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research.* Aldine De Gruyter.

Graham, S., & Marvin, S. (2002). *Splintering Urbanism: Networked infrastructures, technological mobilities and the urban condition*. Routledge.

Gregory-Eaves, I., Smol, J. P., Douglas, M. S. V., & Finney, B. P. (2003). Diatoms and sockeye salmon (Oncorhynchus nerka) population dynamics: Reconstructions of salmon-derived nutrients over the past 2,200 years in two lakes from Kodiak Island, Alaska. *Journal of Paleolimnology*, *30*, 35–53.

Griesemer, J. (2020). A Data Journey Through Dataset-Centric Population Genomics. In S. Leonelli & N. Tempini (Eds.), *Data Journeys in the Sciences* (pp. 145–167). Springer International Publishing. https://doi.org/10.1007/978-3-030-37177-7_8

Guston, D. H. (2004). Forget Politicizing Science. Let's Democratize Science! *Perspectives*, 5.

Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge University Press.

Hahn, C., Hoffman, A. S., Slota, S. C., Inman, S., & Ribes, D. (2018). *Entangled Inversions: Actor/Analyst Symmetry in the Ethnography of Infrastructure*. 16.

Hamazaki, T., Evenson, M., Fleischman, S. J., & Schaberg, K. L. (2012). Spawner-Recruit Analysis and Escapement Goal Recommendation for Chinook salmon in the Kuskokwim River Drainage. *Alaska Department of Fish and Game, Fishery Manuscript Series*, *12*(8), 68.

Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., Duke, C. S., & Porter, J. H. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment*, *11*(3), 156–162. https://doi.org/10.1890/120103

Haycox, S. (2020). *Alaska: An American Colony*. University of Washington Press.

Hébert, K. E. (2008). *Wild Dreams: Refashioning Production in Bristol Bay, Alaska*. University of Michigan.

Heintzman, P. D., Froese, D., Ives, J. W., Soares, A. E. R., Zazula, G. D., Letts, B., Andrews, T. D., Driver, J. C., Hall, E., Hare, P. G., Jass, C. N., MacKay, G., Southon, J. R., Stiller, M., Woywitka, R., Suchard, M. A., & Shapiro, B. (2016). Bison phylogeography constrains dispersal and viability of the Ice Free Corridor in western Canada. *Proceedings of the National Academy of Sciences*, *113*(29), 8057–8063. https://doi.org/10.1073/pnas.1601077113

Helm, B., & Shavit, A. (2017). Dissecting and Reconstructing Time and Space for Replicable Biological Research. In W. J. KRESS, A. SHAVIT, & A. M. ELLISON (Eds.), *Stepping in the Same River Twice* (pp. 233–249). Yale University Press; JSTOR. http://www.jstor.org/stable/j.ctt1n2tvtj.21

Hensel, C. (1996). *Telling Our Selves: Ethnicity and Discourse in Southwest Alaska*. Oxford University Press.

Heritage, J. (1984). *Garfinkeland Ethnomethodology*. Polity Press.

Higgins, S. (2012). The lifecycle of data management. In G. Pryor (Ed.), *Managing Research Data* (pp. 17–45). Facet Publishing.

Hilborn, R., & Walters, C. (1992). Observing fish populations. In *Quantitative Fisheries Stock Assessment.* Springer.

Hobbie, J. E., Carpenter, S. R., Grimm, N. B., Gosz, J. R., & Seastedt, T. R. (2003). The US Long Term Ecological Research Program. *BioScience*, *53*(1,), 21–32.

Hughes, T. (1983). *Networks of power: Electrification in Western society, 1880-1930*. Johns Hopkins University Press.

Hughes, T. (1989). The evolution of large technological systems. In *The social construction of technological systems*. MIT Press.

Hutchins, E. (1995). *Cognition in the wild*. MIT Press.

Jackson, S. J., Ribes, D., Buyuktur, A., & Bowker, G. C. (2011). Collaborative rhythm: Temporal dissonance and alignment in collaborative scientific work. *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work - CSCW '11*, 245. https://doi.org/10.1145/1958824.1958861

Jasanoff, S. (2004). *States of Knowledge: The Co-production of Science and Social Order*. Routledge.

Jones, M. (2006). The new bioinformatics: Integrating ecological data from the gene to the biosphere. *Annu. Rev. Ecol. Syst.*, *37*, 519–544.

Jones, M. L. (2017). *Using participatory modeling to empower community engagement in salmon science* (Submission to the SASAP Project, p. 10).

Jordan, D. S. (1887). *The fisheries of the Pacific coast.* (Sec.11, Part 16; The Fisheries and Fishery Industries of the United States, by G. B. Goode).

Karasti, H. (2012). Long-term temporality in STS research on infrastructural technologies. *Yearbook of the Institute for Advanced Studies on Science, Technology and Society (IAS-STS)*, 75–89.

Karasti, H., & Baker, K. S. (2004). *Infrastructuring for the long-term: Ecological information management*. The 37th Annual Hawaii International Conference on System Sciences.

Karasti, H., Baker, K. S., & Millerand, F. (2010). Infrastructure Time: Long-term Matters in Collaborative Development. *Computer Supported Cooperative Work (CSCW)*, *19*(3–4), 377–415. https://doi.org/10.1007/s10606-010-9113-z

Karasti, H., Millerand, F., Hine, C. M., & Bowker, G. (2016). Knowledge infrastructures: Part I. *Science & Technology Studies*, *29*(1), 2–12.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, *217*, 624–626.

King, B. (2020). Bristol Bay Sockeye 1919: The Salmon Collapse and Fishing Regulations after World War I. *ONCORHYNCHUS: Newsletter of the Alaska Chapter, American Fisheries Society*.

Kingsland, S. (2021). Cold War Origins of Long-Term Ecological Research in Alaska. In *The Challenges of Long Term Ecological Research: A Historical Analysis*. Springer Nature.

Kingsland, S. E. (2011). The Role of Place in the History of Ecology. *The Ecology of Place: Contributions of Place-Based Resarch to Ecological Understanding*, 25.

Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. SAGE Publications Ltd.

Kitchin, R., & Lauriault, T. P. (2015). Small data in the era of big data. *GeoJournal*, *80*(4), 463–475. https://doi.org/10.1007/s10708-014-9601-7

Knorr-Cetina, K. (1999). *Epistemic Cultures: How the Sciences Make Knowledge*. Harvard University Press.

Koo, T. (1962). *Studies of Alaska red salmon.* University of Washington Press.

Krupa, M. B., McCarthy Cunfer, M., & Clark, S. J. (2020). Who's Winning the Public Process? How to Use Public Documents to Assess the Equity, Efficiency, and Effectiveness of Stakeholder Engagement. *Society & Natural Resources*, *33*(5), 612–633. https://doi.org/10.1080/08941920.2019.1665763

Kwa, C. (1987). Representations of Nature Mediating Between Ecology and Science Policy: The Case of the International Biological Programme. *Social Studies of Science*, *17*, 413–442.

Latour, B. (1984). *The Pasteurization of France*. Harvard University Press.

Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers Through Society.* Harvard University Press.

Latour, B. (1996). On Interobjectivity. *Mind, Culture, and Activity*, *3*(4), 228–245. https://doi.org/10.1207/s15327884mca0304_2

Latour, B. (1999). Circulating Reference. In *Pandora's Hope: Essays on the Reality of Science Studies*. Harvard University Press.

Latour, B., & Woolgar, S. (1979). *Laboratory Life: The Social Construction of Scientific Facts*. Sage.

Lawrence, R. (2013, March 14). Data: Why openness and sharing are important [Blog]. *F1000 Blognetwork*. https://blog.f1000.com/2013/03/14/data-why-openness-and-sharing-are-important/

Le Dantec, C. (2016). *Designing Publics*. MIT Press.

Lee, C. P., & Paine, D. (2015). From The Matrix to a Model of Coordinated Action (MoCA): A Conceptual Framework of and for CSCW. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 179–194. https://doi.org/10.1145/2675133.2675161

Lee, C., & Schmidt, K. (2017). A Bridge too Far?: Critical Remarks on the Concept of "Infrastructure" in CSCW and IS. In *Socio-Informatics: A Practice-based Perspective on the Design and Use of IT Artifacts* (pp. 177–218).

Lenhardt, W. C., Ahalt, S., Blanton, B., Christopherson, L., & Idaszak, R. (2014). Data Management Lifecycle and Software Lifecycle Management in the Context of Conducting Science. *Journal of Open Research Software*, *2*. https://doi.org/10.5334/jors.ax

Leonelli, S. (2016). *Data-Centric Biology: A Philosophical Study*. University of Chicago Press.

Leonelli, S. (2018). The Time of Data: Timescales of Data Use in the Life Sciences. *Philosophy of Science*, *85*(5), 741–754. https://doi.org/10.1086/699699

Leonelli, S., & Tempini, N. (Eds.). (2020). *Data Journeys in the Sciences*. Springer International Publishing. https://doi.org/10.1007/978-3-030-37177-7

Levin, N., & Leonelli, S. (2017). How Does One "Open" Science? Questions of Value in Biological Research. *Science, Technology, & Human Values*, *42*(2), 280–305. https://doi.org/10.1177/0162243916672071

Levin, S. (1992). The Problem of Pattern and Scale in Ecology. *Ecology*, *73*(6), 1943–1967.

Lezaun, J., & Montgomery, C. M. (2015). The Pharmaceutical Commons: Sharing and Exclusion in Global Health Drug Development. *Science, Technology, & Human Values*, *40*(1), 3–29. https://doi.org/10.1177/0162243914542349

Lichatowich, J. A. (1999). *Salmon Without Rivers: A History Of The Pacific Salmon Crisis*. Island Press.

Likens, G. (Ed.). (1989). *Long Term Studies in Ecology: Approaches and Alternatives*. Springer-Verlag.

Lindenmayer, D., & Likens, G. E. (2013). Benchmarking Open Access Science Against Good Science. *Bulletin of the Ecological Society of America*, *94*(4), 338–340. https://doi.org/10.1890/0012-9623-94.4.338

Lohr, S. (2014). For Big-Data Scientists, "Janitor Work" is Key Hurdle to Insights. *The New York Times*.

Losey, R. (2010). Animism as a Means of Exploring Archaeological Fishing Structures on Willapa Bay, Washington, USA. *Cambridge Archaeological Journal*, *20*, 17–32.

Lury, C., & Wakeford, N. (Eds.). (2012). *Inventive methods: The happening of the social*. Routledge.

Magnuson, J. J. (1990). Long-Term Ecological Research and the Invisible Present. *BioScience*, *40*(7), 495–501. https://doi.org/10.2307/1311317

Marcus, G. (1995). Ethnography in/of the World System: The Emergence of Multi-Sited Ethnography. *Annual Review of Anthropology*, *24*, 95–117.

Marres, N., & Weltevrede, E. (2012). Scraping the social: Issues in real-time social research. *Journal of Cultural Economy*.

Mayernik, M. S. (2016). Research data and metadata curation as institutional issues. *Journal of the Association for Information Science and Technology*, *67*(4), 973–993. https://doi.org/10.1002/asi.23425

Mazmanian, M., Erickson, I., & Harmon, E. (2015). Circumscribed Time and Porous Time: Logics as a Way of Studying Temporality. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15*, 1453–1464. https://doi.org/10.1145/2675133.2675231

Mazzocchi, F. (2015). Could Big Data be the end of theory in science?: A few remarks on the epistemology of data-driven science. *EMBO Reports*, *16*(10), 1250–1255. https://doi.org/10.15252/embr.201541001

McGlauflin, M. T., Schindler, D. E., Seeb, L. W., Smith, C. T., Habicht, C., & Seeb, J. E. (2011). Spawning Habitat and Geography Influence Population Structure and Juvenile Migration Timing of Sockeye Salmon in the Wood River Lakes, Alaska. *Transactions of the American Fisheries Society*, *140*(3), 763–782. https://doi.org/10.1080/00028487.2011.584495

Michener, W. K., & Jones, M. B. (2012). Ecoinformatics: Supporting ecology as a data-intensive science. *Trends in Ecology & Evolution*, *27*(2), 85–93. https://doi.org/10.1016/j.tree.2011.11.016

Miettinen, A., Palm, S., Dannewitz, J., Lind, E., Primmer, C. R., Romakkaniemi, A., Östergren, J., & Pritchard, V. L. (2021). A large wild salmon stock shows genetic and life history differentiation within, but not between, rivers. *Conservation Genetics*, *22*(1), 35–51.

Miller, L. (2020). *Why Fish Don't Exist: A Story of Loss, Love, and the Hidden Order of Life*. Simon & Schuster.

Millerand, F., & Bowker, G. (2009). Metadata standards: Trajectories and enactment in the life of an ontology. In *Standards and their stories. How quantifying, classifying, and formalizing practices shape everyday life* (pp. 149–165). Cornell University Press.

Millerand, F., Ribes, D., Baker, K. S., & Bowker, G. C. (2013). Making an issue out of a standard: Storytelling practices in a scientific community. *Science, Technology & Human Values*, *38*(1), 7–43.

Mirowski, P. (2018). The future(s) of open science. *Social Studies of Science*, *48*(2), 171–203. https://doi.org/10.1177/0306312718772086

Misa, T. J. (1988). How Machines Make History, and how Historians (And Others) Help Them to Do So. *Science, Technology, & Human Values*, *13*(3–4), 308–331. https://doi.org/10.1177/016224398801303-410

Monteiro, E., Pollock, N., Hanseth, O., & Williams, R. (2013). From Artefacts to Infrastructures. *Computer Supported Cooperative Work*, *22*(4–6), 575–607. https://doi.org/10.1007/s10606-012-9167-1

Morehouse, R. (2011). *Beginning Interpretive Inquiry: A Step-By-Step Approach to Research and Evaluation*. Routledge. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/washington/detail.action?docID=958418.

Moses, A. P., Staton, B. A., & Smith, N. J. (2019). Investigation of Migratory Timing and Rates of Chinook Salmon Bound for the Kwethluk and Kisaralik Rivers Using Radio Telemetry. *Journal of Fish and Wildlife Management*, *10*(2), 38. https://doi.org/10.3996/082018- 48 JFWM-074

Nakashima, D. J., Galloway McLean, K., Thulstrup, H. D., Castillo Ramos, A., & Rubis, J. T. (2012). *Weathering uncertainity: Traditional knowledge for climate change assessment and adaptation*. UNESCO ; UNU-IAS.

Narum, S. R., Di Genova, A., Micheletti, S. J., & Maass, A. (2018). Genomic variation underlying complex life-history traits revealed by genome sequencing in Chinook salmon. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1883), 20180935. https://doi.org/10.1098/rspb.2018.0935

National Research Council. (2005). *Developing a research and restoration plan for Arctic-Yukon-Kuskokwim (western Alaska) salmon.* National Academies Press.

Nero, R. W., & Schindler, D. W. (2011). Decline of Mysis relicta During the Acidification of Lake 223. *Canadian Journal of Fisheries and Aquatic Sciences*, *40*(11). https://doi.org/10.1139/f83-221

Nielsen, M. (2012). *Reinventing Discovery: The New Era of Networked Science*. Princeton University Press.

Oreskes, N. (2003). The changing role of prediction in the earth sciences. In *History and Philosophy of Science for African Undergraduates* (pp. 358–368). Hope Publications.

Paine, D., Sy, E., Piell, R., & Lee, C. P. (2015). *Examining Data Processing Work as Part of the Scientific Data Lifecycle: Comparing Practices Across Four Scientific Research Groups*. 12.

Parmiggiani, E., Monteiro, E., & Hepsø, V. (2015). The Digital Coral: Infrastructuring Environmental Monitoring. *Computer Supported Cooperative Work (CSCW)*, *24*(5), 423–460. https://doi.org/10.1007/s10606-015-9233-6

Pearcy, W., & McKinnell, S. (2007). The Ocean Ecology of Salmon in the Northeast Pacific Ocean: An abridged history. *American Fisheries Society Symposium*, *57*, 7–30.

Pedersen, M. W., Ruter, A., Schweger, C., Friebe, H., Staff, R. A., Kjeldsen, K. K., Mendoza, M. L. Z., Beaudoin, A. B., Zutter, C., Larsen, N. K., Potter, B. A., Nielsen, R., Rainville, R. A., Orlando, L., Meltzer, D. J., Kjær, K. H., & Willerslev, E. (2016). Postglacial viability and colonization in North America's ice-free corridor. *Nature*, *537*(7618), 45–49. https://doi.org/10.1038/nature19085

Penders, B., Horstman, K., & Vos, R. (2008). Walking the Line between Lab and Computation: The "Moist" Zone. *BioScience*, *58*(8), 747–755. https://doi.org/10.1641/B580811

Peters, D. P., Groffman, P. M., Nadelhoffer, K. J., Grimm, N. B., Collins, S. L., Michener, W. K., & Huston, M. A. (2008). Living in an increasingly connected world: A framework for continental-scale environmental science. *Frontiers in Ecology and the Environment*, *6*(5), 229–237. https://doi.org/10.1890/070098

Pollner, M. (1987). *Mundane Reason: Reality in Everyday and Sociological Discourse*. Cambridge University Press.

Pollner, M., & Emerson, R. (2001). Ethnomethodology and Ethnography. In *Handbook of Ethnography*. SAGE Publications Ltd. https://doi.org/10.4135/9781848608337

Poulsen, R. T., & Holm, P. (2007). What Can Fisheries Historians Learn from Marine Science? The Concept of Catch per Unit Effort (CPUE). *International Journal of Maritime History*, *19*(2), 89–112. https://doi.org/10.1177/084387140701900205

Puig de la Bellacasa, M. (2015). Making time for soil: Technoscientific futurity and the pace of care. *Social Studies of Science*, *45*(5), 691–716. https://doi.org/10.1177/0306312715599851

Rader, K. (2004). *Making Mice: Standardizing Animals for American Biomedical Research, 1900-1955*. Princeton University Press.

Reichman, O. J., Jones, M. B., & Schildhauer, M. P. (2011). Challenges and Opportunities of Open Data in Ecology. *Science, New Series*, *331*(6018), 703–705.

Reiners, W. (1986). Complementary models for ecosystems. *The American Naturalist*, *127*, 59–73.

Rheinberger, H. J. (2000). Cytoplasmic particles: The trajectory of a scientific object. In L. Daston (Ed.), *Biographies of Scientific Objects* (pp. 270–294). University of Chicago Press.

Rheinberger, H.-J. (2005). Gaston Bachelard and the Notion of "Phenomenotechnique." *Perspectives on Science*, *13*(3), 313–328. https://doi.org/10.1162/106361405774288026

Ribes, D. (2014). Ethnography of scaling, or, how to a fit a national research infrastructure in the room. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '14*, 158–170. https://doi.org/10.1145/2531602.2531624

Ribes, D. (2019). *Materiality Methodology, and Some Tricks of the Trade in the Study of Data and Specimens*. 18.

Ribes, D., & Finholt, T. (2007a). *Planning Infrastructure for the long-term: Learning from cases in the natural sciences*. the Third International Conference on e-Social Science, Ann Arbor, MI.

Ribes, D., & Finholt, T. (2007b). *Tensions across the scales: Planning infrastructure for the long-term*. international ACM Conference on Supporting group work.

Ribes, D., & Finholt, T. (2009). The Long Now of Technology Infrastructure: Articulating Tensions in Development. *Journal of the Association for Information Systems*, *10*(5), 375–398. https://doi.org/10.17705/1jais.00199

Ribes, D., & Lee, C. P. (2010). Sociotechnical Studies of Cyberinfrastructure and e-Research: Current Themes and Future Trajectories. *Computer Supported Cooperative Work (CSCW)*, *19*(3–4), 231–244. https://doi.org/10.1007/s10606-010-9120-0

Ribes, D., & Polk, J. B. (2015). Organizing for ontological change: The kernel of an AIDS research infrastructure. *Social Studies of Science*, *45*(2), 214–241. https://doi.org/10.1177/0306312714558136

Rogers, L. A., Schindler, D. E., Lisi, P. J., Holtgrieve, G. W., Leavitt, P. R., Bunting, L., Finney, B. P., Selbie, D. T., Chen, G., Gregory-Eaves, I., Lisac, M. J., & Walsh, P. B. (2013). Centennial-scale fluctuations and regional complexity characterize Pacific salmon

population dynamics over the past five centuries. *Proceedings of the National Academy of Sciences*, *110*(5), 1750–1755. https://doi.org/10.1073/pnas.1212858110

Saldaña, J. (2013). *The Coding Manual for Qualitative Researchers* (2nd ed.). SAGE.

Schiefer, P. E. (2019). *Cultivating Salmon. Human-Fish Relations in Bethel, Alaska*. University of Aberdeen.

Schindler, D. E., Hilborn, R., Chasco, B., Boatright, C. P., Quinn, T. P., Rogers, L. A., & Webster, M. S. (2010). Population diversity and the portfolio effect in an exploited species. *Nature*, *465*(7298), 609–612. https://doi.org/10.1038/nature09060

Schindler, D. E., Leavitt, P. R., Johnson, S. P., & Brock, C. S. (2006). *A 500-year context for the recent surge in sockeye salmon (Oncorhynchus nerka) abundance in the Alagnak River, Alaska*. *63*, 7.

Schmidt, K., & Bannon, L. (1992). Constructing CSCW: The First Quarter Century. *Computer Supported Cooperative Work (CSCW)*, *22*(4–6), 345–372. https://doi.org/10.1007/s10606-013-9193-7

Schneider, D. C. (2001). The Rise of the Concept of Scale in Ecology. *BioScience*, *51*(7), 545. https://doi.org/10.1641/0006-3568(2001)051[0545:TROTCO]2.0.CO;2

Scott, J. (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press.

Seeb, J. E., Seeb, L. W., & Utter, F. M. (1986). Use of Genetic Marks to Assess Stock Dynamics and Management Programs for Chum Salmon. *Transactions of the American Fisheries Society*, *115*, 448–454.

Sepkoski, D. (2018). Data in Time. *Historical Studies in the Natural Sciences*, *48*(5), 581–593. https://doi.org/10.1525/hsns.2018.48.5.581

Simone, C., Mark, G., & Giubbilei, D. (1999). Interoperability as a means of articulation work. *ACM SIGSOFT Software Engineering Notes*, *24*(2), 39–48. https://doi.org/10.1145/295666.295671

Smith, T. (1994). *Scaling Fisheries: The Science of Measuring the Effects of Fish, 1855-1955*.

Soutar, A., & Isaacs, J. D. (1974). Abundance of Pelagic Fish During the 19th and 20th Centuries as Recording in Anaerobic Sediment Off the Californias. *Fishery Bulletin*, *72*(2), 17.

Stanford, J. A., Lorang, M. S., & Hauer, F. R. (2005). The shifting habitat mosaic of river ecosystems. *SIL Proceedings, 1922-2010*, *29*(1), 123–136. https://doi.org/10.1080/03680770.2005.11901979

Star, S. L. (1990). Power, Technology and the Phenomenology of Conventions: On being Allergic to Onions. *The Sociological Review*, *38*(1_suppl), 26–56. https://doi.org/10.1111/j.1467-954X.1990.tb03347.x

Star, S. L. (1999). The Ethnography of Infrastructure. *American Behavioral Scientist*, *43*(3), 377–391.

Star, S. L., & Gerson, E. M. (1987). The Management and Dynamics of Anomalies in Scientific Work. *The Sociological Quarterly*, *28*(2), 147–169. https://doi.org/10.1111/j.1533-8525.1987.tb00288.x

Star, S. L., & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research*, *7*(1), 111–134. https://doi.org/10.1287/isre.7.1.111

Staton, B. A., Catalano, M. J., Connors, B. M., Coggins, L. G., Jones, M. L., Walters, C. J., Fleischman, S. J., & Gwinn, D. C. (2020). Evaluation of methods for spawner-recruit analysis in mixed-stock Pacific salmon fisheries. *Canadian Journal of Fisheries and Aquatic Sciences*, cjfas-2019-0281. https://doi.org/10.1139/cjfas-2019-0281

Staton, B. A., Catalano, M. J., & Fleischman, S. J. (2017). From sequential to integrated Bayesian analyses: Exploring the continuum with a Pacific salmon spawner-recruit model. *Fisheries Research*, *186*, 237–247. https://doi.org/10.1016/j.fishres.2016.09.001

Steger, C., Hirsch, S., Cosgrove, C., Inman, S., Nost, E., Shinbrot, X., Thorn, J. P. R., Brown, D. G., Grêt-Regamey, A., Müller, B., Reid, R. S., Tucker, C., Weibel, B., & Klein, J. A. (2021). Linking model design and application for transdisciplinary approaches in social-ecological systems. *Global Environmental Change*, *66*, 102201. https://doi.org/10.1016/j.gloenvcha.2020.102201

Stewart Grant, W. (2021). My life with the Red Queen in fishery genetics. *ICES Journal of Marine Science*, *78*(7), 2351–2358. https://doi.org/10.1093/icesjms/fsab112

Stirling, A. (2008). "Opening up" and "closing down": Power, participation, and pluralism in the social appraisal of technology. *Science, Technology & Human Values*, *33*, 262–294.

Stone, R. (1993). Long-Term NSF Network Urged to Broaden Scope. *Science*, *262*(5132), 334–335. https://doi.org/10.1126/science.262.5132.334

Strasser, C., Robert, C., William, M., Amber, B., & Rebecca, K. (2011). DataONE promoting data stewardship through best practices. *In Proceedings of the Environmental Information Management Conference*, 126–131.

Strauss, A. (1988). The Articulation of Project Work: An Organizational Process. *The Sociological Quarterly*, *29*(2), 17.

Suchman, L. (2007). *Human-Machine Reconfigurations: Plans and Situated Actions* (2nd ed.). Cambridge University Press.

Suryan, R. M., Arimitsu, M. L., Coletti, H. A., Hopcroft, R. R., Lindeberg, M. R., Barbeaux, S. J., Batten, S. D., Burt, W. J., Bishop, M. A., Bodkin, J. L., Brenner, R., Campbell, R. W., Cushing, D. A., Danielson, S. L., Dorn, M. W., Drummond, B., Esler, D., Gelatt, T., Hanselman, D. H., … Zador, S. G. (2021). Ecosystem response persists after a prolonged marine heatwave. *Scientific Reports*, *11*(1), 6235. https://doi.org/10.1038/s41598-021-83818-5

Swezey, S. L., & Heizer, R. F. (1977). Ritual Management of Salmonid Fish Resources in California. *The Journal of California Anthropology*, *4*(1), 25.

Tansley, A. (1935). The use and abuse of vegetational concepts and terms. *Ecology*, *16*, 284–307.

Taylor, J. E. (2002). "Well-Thinking Men and Women": The Battle for the White Act and the Meaning of Conservation in the 1920s. *Pacific Historical Review*, *71*(3), 357–387. https://doi.org/10.1525/phr.2002.71.3.357

Thomer, A. K., Twidale, M. B., & Yoder, M. J. (2018). Transforming Taxonomic Interfaces: "Arm's Length" Cooperative Work and the Maintenance of a Long-lived Classification System. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1–23. https://doi.org/10.1145/3274442

Traweek, S. (1988). *Beamtimes and Lifetimes: The World of High Energy Physicists*. Harvard University Press.

Tsing, A. L. (2012). ON NONSCALABILITY: The Living World Is Not Amenable to Precision-Nested Scales. *Common Knowledge*, *18*(3), 505–524. https://doi.org/10.1215/0961754X-1630424

Van Noorden, R. (2021). SCIENTISTS CALL FOR OPEN SHARING OF PANDEMIC GENOME DATA. *Nature*, *590*, 2.

VanStone, J. (1967). *Eskimos of the Nushagak River: An ethnographic history*. University of Washington Press.

Vertesi, J. (2014). Seamful Spaces: Heterogeneous Infrastructures in Interaction. *Science, Technology, & Human Values*, *39*(2), 264–284. https://doi.org/10.1177/0162243913516012

Wallis, J. C., Borgman, C. L., Mayernik, M. S., & Pepe, A. (2008). Moving Archival Practices Upstream: An Exploration of the Life Cycle of Ecological Sensing Data in Collaborative Field Research. *International Journal of Digital Curation*, *3*(1), 114–126. https://doi.org/10.2218/ijdc.v3i1.46

Waples, R. S., Pess, G. R., & Beechie, T. (2008). Evolutionary history of Pacific salmon in dynamic environments: Evolutionary history of Pacific salmon in dynamic environments. *Evolutionary Applications*, *1*(2), 189–206. https://doi.org/10.1111/j.1752-4571.2008.00023.x

Ward, T. C., & Horn, N. (2003). *Kuskokwim River Salmon Management Working Group Support*. Alaska Department of Fish & Game, Division of Commercial Fisheries.

Weiderman, N. H., Bergey, J. K., Smith, D. B., & Tilley, S. R. (1997). *Approaches to Legacy System Evolution.:* Defense Technical Information Center. https://doi.org/10.21236/ADA336213

Weinberg, A. (1961). Impact of Large-Scale Science on the United States. *Science*, *134*, 161–164.

Weinberg, A. (1967). *Reflections on Big Science*. Oxford Pergamon Press.

Weinberg, A. M., & Bowers, R. (1968). Reflections of Big Science. *Physics Today*, *21*(4), 95–96. https://doi.org/10.1063/1.3034938

Weiss, R. (1994). *Learning from Strangers: The Art and Method of Qualitative Interview Studies*. The Free Press.

West, G. (2017). *Scale: The Universal Laws of Growth, Innovation, Sustainability, and the Pace of Life in Organisms, Cities, Economies, and Companies*. Penguin Press.

Westley, P. A. H., Berdahl, A. M., Torney, C. J., & Biro, D. (2018). Collective movement in ecology: From emerging technologies to conservation and management.

Philosophical Transactions of the Royal Society B: Biological Sciences, *373*(1746), 20170004. https://doi.org/10.1098/rstb.2017.0004

Wheatley, M., & Johnson, C. (2009). Factors limiting our understanding of ecological scale. *Ecological Complexity*, *6*, 150–159.

Wiens, J. A. (1989). Spatial Scaling in Ecology. *Functional Ecology*, *3*(4), 385. https://doi.org/10.2307/2389612

Wilson, N. J., Mutter, E., Inkster, J., & Satterfield, T. (2018). Community-Based Monitoring as the practice of Indigenous governance: A case study of Indigenous-led water quality monitoring in the Yukon River Basin. *Journal of Environmental Management*, *210*, 290–298. https://doi.org/10.1016/j.jenvman.2018.01.020

Wise, N. (1995). *The values of precision*. Princeton University Press.

Wynne, B. (2006). Public engagement as a means of restoring public trust in science: Hitting the notes, but missing the music? *Community Genetics*, *9*, 211–220.

Wynne, B., Lash, S., & Szerszynski, B. (1996). *May the sheep safely graze?*

Zimmerman, A. S. (2008). New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data. *Science, Technology, & Human Values*, *33*(5), 631–652.

Zins, C. (2007). Conceptual approaches for defining data, information, and knowledge. *Journal of the American Society for Information Science and Technology*, *58*(4), 479–493. https://doi.org/10.1002/asi.20508

Appendix A: Codebook for GitHub data analysis

Category	Code	Description
Content	Ontology	Changes in what is captured & represented as data across time
	Epistemology	Changes in instrumentation that have generated the data
	Instrumentation	what instruments, forms of calibration, and practices of collection have led to particular data points?
Form	Categories	columns/rows; parsing data into concrete groups
	Standards	Standards changing
	Residual states	when things do not fit into category

	Representation & Media	material medium of the data (papers/disks/drives)
Socio-organizational	Regulatory	Regulatory changes
	Institutional	Institutional capacity or as limiting factor
	Accessibility	Who provides accessibility
Data Values	Reproducibility	Reproducibility
	interoperability	interoperability
	bad data	
	good data	
Errors	data entry; missing; unknown; duplicates	
	Commensuration	how data have been combined with other datasets
	Equivalences	Translations
	Erasures	Removing data
	Additions	Additions
	Flag	Flagging as potentially problematic
	reassign/rename	reassigning data (sort of like translations but just renaming)
data lifecycle model (Strasser et al. 2011)		
	plan	"data management planning"
	collect	"ecological data are collected and organized in many different ways, including manual recording of observatiosn in the laboratory and field via hand-written data sheets, tape recorders and hand-held computers; automated data collected via laboratory and field instrumentation; satellites and aerial platforms; and, increasingly, sensor networks that are embedded in the environment."
	assure	"QA/QC refers to the mechanisms for preventing errors from entering a data set that are used a priori to ensure high data quality before colletion and to monitor and maintain data quality during and after data collection"
	describe	"metadata documentation to understand the content, format, and context of a data product"
	preserve	"deposition of data and metadata in a data center or data repo"
	discover	"sophisticated, user-friendly search tools that enable scientists to search by time and space and also drill down further using faceted search techniques that allow one to filter the results by parameter, sensor employed, author and other properties of the data"
	integrate	"integrating source data from such studies is labor intensive and time consuming, because it requires understanding methodological differences, transforming data into a common

		representation, and manually converting and recoding data to compatible semantics before analysis can begin."
	analyze	making analysis reproducible: "scientific workflow systems provide an executable and complete description of analytical procedures that allows scientists to link togheter processes drawn from multiple different analytical systems."
Additions to the model		
Error discovery; identifying inconsistency		Often connects to larger themes of epistemology and standards
	anomaly	discovering errors or anomalies
	identifying inconsistency	identifying inconsistencies or redundancies in data
	uncertainty	expressing uncertainty/sometimes seeking explanations for unknowns
inverting documentation		
	digging into reports	going back in documents and past archives to understand changes
	consult	consulting expert
	provide	providing expertise
inverting the final dataset		
commensuration	missing	filling in missing data; identifying missing data
	flag	flagging data for removal or further attention
Accessing data		micro-level interactions illuminate macro-level institutional issues
	receiving/requesting	receiving or requesting data
	brokering	brokering relationships with other experts
actions taken in commits	knit	knit
	merge	merge
	reformat	reformat
	connecting	connecting to other working groups

Appendix B: Interview script

Participatory Modeling Interview Script

For a PI/researcher/modeler

1. How would you describe your relationship with salmon?
2. What has been your role in the working group?
3. What does “Participatory Modeling” mean?

4. Your working group is a bit of an "outlier" among the other SASAP working groups, why is that? How did this group get started? What was the inspiration behind it?
5. What questions is your working group trying to answer, and why are they important?
6. How is synthesis playing a role in your working group's work? What role(s) do data play?
7. How is this group integrating science with community-based knowledge about salmon systems? What has that process been like?
8. Why is community-based monitoring seen as important for salmon management in Alaska?
9. What kinds of salmon data have communities been collecting, and how are managers using those data?
10. What are some of the data needs that have emerged from your work?
11. How do the information needs of managers align with or differ from those of salmon-dependent communities?
12. How do you think your working group could help salmon science and management?
13. How has your perspective on salmon science and management changed as a result of your involvement in SASAP?
14. What big lesson or takeaway have you learned as a result of your involvement in SASAP?

For community member

1. How would you describe your relationship with salmon?
2. What has been your role in the working group?
3. Do you think community-based monitoring is valuable for salmon management in Alaska; why/why not?
4. What types of salmon knowledge and data are important to your community, and how do you use that information?
5. Are there any challenges that inhibit the success of a community-based monitoring program?
6. From your perspective, what elements are important to have a successful community-based monitoring program?
7. How do you think your working group could help salmon science and management?
8. How has your perspective on salmon science and management changed as a result of your involvement in this project?
9. What big lesson or takeaway have you learned as a result of your involvement in SASAP?

Appendix C: Network diagram for interview codes

Curriculum Vitae

Sarah Inman

PhD, Human Centered Design & Engineering **2022**

University of Washington, Seattle, WA

Committee: David Ribes (Chair), Charlotte Lee, Daniel Schindler, Nic Weber (GSR)

Salmon on the Run: Scaling Practices for Wild Alaska Salmon Research

M.A., Communication, Culture, & Technology **2013**

Georgetown University, Washington, D.C.

Fractured Consent: Public Participation in Environmental Complexity

B.A., Political Philosophy **2011**

Louisiana Tech University, Ruston, LA

PEER REVIEWED PUBLICATIONS

Hirsch, S., Ribes, D., & **Inman, S.** (2021). Sedimentary Legacy and the Disturbing Recurrence of the Human in the Long Term Ecological Research. Social Studies of Science.

Inman, S., Esquible, J., Jones, M., Bechtol, B., & O'Connor, B. (2021). Opportunities for Community-Based Monitoring to Provide Management-Relevant Data for Data-Poor Salmon Fisheries. Ecology and Society. Special issue.

Steger, C., Hirsch, S., Cosgrove, C., **Inman, S.,** Nost, E., Shinbrot, X., Thorn, J., Agrawal, A., Brown, D., Grêt-Regamey, A., Müller, B., Nolin, A., Reid, R., Tucker, C., Weibel, B., & Klein, J. 2021. Linking Model Design and Application for Transdisciplinary Approaches in Social-Ecological Systems. Global Environmental Change.

Inman, S. & Ribes, D. 2019. "Beautiful Seams": Strategic Revelations and Concealments. Proceedings of the Human Computer Interaction (CHI) conference, Glasgow, Scotland. **Best Paper Honorable Mention Award.**

Hahn, C., Hoffman, A. S., Slota, S. C., **Inman, S.**, & Ribes, D. 2019. Entangled Inversions: Actor/Analyst Symmetry in the Ethnography of Infrastructure. Interaction Design and Architecture(s). IxD&A. Special issue on "Inquiring the way we inquire."

Inman, S. & Ribes, D. 2018. "Data Streams, Data Seams: Towards a Seamful Representation of Data Interoperability." Design Research Society.

Vovides, Y. & **Inman, S**. 2016. "Using Learning Analytics to Support Reflective Sensemaking of Ill-structured Ethical Problems" in Future Internet.

Kruger, D., **Inman, S**., Ding, Z., Kang, Y., Kuna, P., Liu, Y., Lu, X., Oro, S., & Wang, Y. 2015. "Improving Teacher Effectiveness: Designing Better Assessment Tools in Learning Management Systems" in Future Internet, 7(4), 484-499.

Vovides, Y. & **Inman, S**. 2013. Storytelling: Discourse analysis for understanding collective perceptions of medical education. International Association for Development of the Information Society (IADIS) eLearning Conference.

Vovides, Y. & **Inman, S**. 2016. "Enabling Meaningful Certificates from Massive Open Online Courses (MOOCs): A Data-Driven Curriculum e-map Design Model" in Open Learning and Formal Credentialing in Higher Education: Curriculum Models and Institutional Policies.

CONFERENCE WORKSHOPS AND PROCEEDINGS

Inman, S. 2020. Salmonscapes: A case of temporal scaling in the Kuskokwim. Society for Social Studies of Science Conference.

Inman, S. 2019. "How data are (or are not) tractable to management: A case study in the Kuskokwim," Matanuska-Susitna Salmon Symposium, November 2019, Palmer, Alaska.

Inman, S. 2018. "Agencies in the Database: The Role of Flagging," EASST Conference, July 2018, Lancaster, UK.

Inman, S. 2018. "Revealing spaces for community engagement", In Workshop: "Untold Stories" at the ACM CHI Conference, April 2018, Montreal, Canada

Inman, S. 2017. "Data Upstream", Society for Social Studies of Science Conference, 2017 in Boston, MA.

Inman, S. 2017. "Multi-disciplinary collaboration for Wild Alaskan Salmon" at the International Congress of Arctic Social Science, June 2017 in Umea, Sweden.

Inman, S. 2015. "Metacognition in the 21st Century" at the 21st Annual Online Learning Consortium 2015, Orlando, FL.

Inman, S. 2015. "Metacognition in a Technologically-Enhanced Environment" at EDUCAUSE 2015, Indianapolis, IN.

Inman, S. 2015. "Virtual Learning Environment and Informal Learning" at NJEDGE, 2015 Princeton, NJ.

Inman, S. 2013. "A model for reflective sensemaking" at the Learning Analytics and Knowledge Conference, 2013 in Leuven, Belgium.

INVITED TALKS

Inman, S. How data are (or are not) tractable to management: A case study in the Kuskokwim. Invited talk for the UW Alaska Salmon Program's Salmon Symposium. Seattle, WA. December 2019.

Inman, S. Seamful Representations of Data Science. Invited talk for the Human Centered Design & Engineering Theory 101 course, Seattle, WA. February 2018.

Inman, S. Visualization Design Workshop. Invited to lead 2-day workshop in Jaime Snyder's Visualization Design Course. Seattle, WA. January 2018.

Inman, S. Data Upstream. Invited talk at Participatory Workshop for the State of Alaska's Salmon and People project meeting in Kenai, Alaska. September 2017.

Inman, S. Organizational Accidents and Dialogic Interaction. Invited talk at Stevens Institute of Technology, Hoboken, NJ. September 2016.

GRANTS AND AWARDS

2021, **Barclay Simpson Scholars in Public Fellowship, Simpson Center for the Humanities grant, $6000**

2020, Selected participant for the **Association of Polar Early Career Scientists (APECS)** Nunataryuk T-MOSAiC School on "Arctic Coastal Adaptation: Capacity building and knowledge exchange across borders", Abisko Scientific Research Station, Abisko, Sweden

2019, Selected as **Storytelling Fellow** at the University of Washington Research Commons Library

Featured in **NCEAS Portrait Series** on "What it Means to be a Data Ethnographer"

Travel Award, Social Studies of Science

CHI Best Paper Honorable Mention Award, "Beautiful Seams: Strategic Revelation and Concealment", CHI'19

Finalist for 4Culture's Rain Garden x Public Art Grant, $40,000

Travel Award, Graduate School of University of Washington funding to present at international research conference

CRA-W Grad Cohort for Women Workshop, Chicago, IL

2018, Travel Award, Graduate School of University of Washington, $500 travel funding to present at international research conference

Travel Award, National Snow and Ice Data Center, Polar Data Summit, Boulder, CO

2017, Travel Award, Graduate School of University of Washington, $500 travel funding to present at research conference

2016, Research Grant, Data Task Force for Better Synthesis Studies (PI: David Ribes), The Gordon and Betty Moore Foundation, Full tuition + stipend for 3 years ($200,000)

2006-2010, Louisiana TOPS Scholar, Louisiana Tech University, Outstanding Student Scholarship, $8000