NORTH PACIFIC ANADROMOUS FISH COMMISSION

TECHNICAL REPORT 14

Report of the Proceedings for the IYS Workshop
First International Year of the Salmon Data Laboratory (ISDL) Workshop

Technical Editors: Scott Akenhead, Nathan Bendriem, and Jeongseok Park

Vancouver, Canada, 2019
TECHNICAL REPORT 14

Report of the Proceedings for the IYS Workshop

First International Year of the Salmon Data Laboratory (ISDL) Workshop

Technical Editors: Scott Akenhead, Nathan Bendriem, and Jeongseok Park

Vancouver, Canada, 2019
Report of the Proceedings for the IYS Workshop

First International Salmon Data Laboratory (ISDL) Workshop

Vancouver, BC, Canada, January 25, 2019

Location: Pacific Salmon Commission Boardroom, 600-1155 Robson Street, Vancouver, BC, Canada
Sponsors: Fisheries and Oceans Canada (DFO), North Pacific Anadromous Fish Commission
Conveners: Jim Irvine (DFO) and Mark Saunders (NPAFC IYS)

Organizing Committee:
Scott Akenhead (Workshop Convenor)
James Irvine (Canadian Lead, International Year of the Salmon)
Mark Saunders (IYS North Pacific, Director, International Year of the Salmon)

NPAFC Secretariat

Vladimir Radchenko, Executive Director
Jeongseok Park, Deputy Director
Jennifer Chang, Administrative Officer
Mariia Artiushkina, Administrative Assistant

ISSN 1029-5917
North Pacific Anadromous Fish Commission

TECHNICAL REPORT 14

Report of the Proceedings for the IYS Workshop
First International Year of the Salmon Data Laboratory (ISDL) Workshop

Vancouver, BC, Canada, January 25, 2019

Technical Editors: Scott Akenhead, Nathan Bendriem, and Jeongseok Park

Published by: North Pacific Anadromous Fish Commission
Tel: (604) 775-5550
Suite 502, 889 West Pender Street
Fax: (604) 775-5577
Vancouver, BC, V6C 3B2
E-mail: secretariat@npafc.org
Canada
Website: https://npafc.org
Summary

A mixture of leading salmon ecologists and information technologists met to discuss the modernization of data processing in the context of salmon research and management. An important conclusion is that, while salmon data processing problems may appear unique, that is not true: the problems of data assembly, cleaning, standardization, integration, and reproducible analyses are universal, and solutions from leading edge of information technology are applicable. Experiments to apply these new technologies to salmon datasets and analyses—including the International Salmon Data Laboratory (ISDL)—were discussed as next steps. Participants expressed interest in communicating what they had learned about the efficacy of labeled graph databases to other salmon research fora and to granting agencies.

Given that salmon ecologists must try to improve the resilience of salmon to unmitigated anthropic climate change, a paradigm shift is required: from monitoring, passive conservation, optimal harvesting, and enhancement (spawning channels and hatcheries) to predicting how salmon will respond to habitat changes that exceed the range of historical observations. To be useful, that new prediction ability must be translated into effective actions; salmon fisheries managers and salmon habitat managers (e.g., river flows) need to be fully informed and able to react quickly to surprising events.

The International Year of the Salmon (IYS) has presented this as a challenge for the world to respond to. ISDL is one response, with initial support from IYS and by Neo4j Inc. (Graphs For Good). The workshop asked what information flow the salmon ecologists need, and how that can be successfully delivered.

Reliable prediction ability requires mechanistic models as opposed to correlations. Developing and applying those models requires:

- A new perspective on what drives salmon population dynamics.
  - This involves comparing the success and failure of many salmon populations, with many life-history strategies in many different habitat situations, across all salmon species within the northern hemisphere.
  - This is a goal of the IYS Salmon Status and Trends (IYS-SST) initiative, as described in an immediately preceding workshop. Nine “designated survivors” from the IYS-SST workshop attended this ISDL workshop (plus IYS and NPAFC organizers).
- Assembling and integrating an unprecedented breadth and depth of information about salmon and the habitats they encounter.
  - This workshop proposed a specific recent technology, the neo4j graph database and associated libraries and tools, as the basis for a breakthrough in data mobilization.
  - This capacity was explained by Jeff Morris, with practical elucidation and a cogent demo by John Song.
- Modernization of salmon data processing from field collections to integration and analysis.
  - Immediate mobilization of all types and formats of field data with the GeoOptix platform was described by Matt Denniston and Keith Steele.
  - Scott Akenhead pointed to the development of technologies that are capable of modernizing estimates of spawner abundance by resistivity counters and estimates of age-1 smolt survival by increasingly tiny acoustic tags, just two examples of many possible.
  - How DFO Pacific Region manages hundreds of data sources, many databases, and many data customers was reviewed by Bruce Patten, highlighted by a SWOT analysis of the migration from 1990’s technology and practices to new tools and practices.
- Mobilizing data for improved decision support.
  - Sue Grant described DFO’s new State Of Salmon (SOS) program that she is leading.
• Effective knowledge transfer from scientists to policy makers.
  o Kelly Chapman described how the requirement for extensive personal engagement contrasts sharply with
    the implicit assumption, in all preceding presentations, that “decision support products” were sufficient—an
    illuminating moment.

A suite of ideas for data experiments by a salmon data laboratory were discussed. The ecologists asked for
immediate action to standardize (by a ‘semantic web’ of ecological structures, functions, and methods) and
integrate the datasets that will be tabled through the IYS-SST program. Several ecologists offered to carry the
result of that experiment forward to international workshops in May (NPAFC, Portland), June (ICES,
Copenhagen) and September (ICES, Bergen) of this year—an encouraging result that emphasized the potential
value of the ISDL initiative. The suggestions for projects from this workshop and from preceding IYS workshops
were combined as a strategic plan for ISDL, which outlined the specific goals, quantifiable objectives, projects,
and experiments that were discussed. Resources will be required for the implementation of the strategic plan
Effecting that plan requires resourcing.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of Contents</td>
<td>iii</td>
</tr>
<tr>
<td>BACKGROUND</td>
<td>1</td>
</tr>
<tr>
<td>WORKSHOP GOALSs</td>
<td>1</td>
</tr>
<tr>
<td>Workshop goals</td>
<td>1</td>
</tr>
<tr>
<td>WORKSHOP REPORT</td>
<td>2</td>
</tr>
<tr>
<td>The Revolution Begins with Salmon</td>
<td>2</td>
</tr>
<tr>
<td>Dynamic Data Management in the Cloud</td>
<td>4</td>
</tr>
<tr>
<td>The Graph Platform</td>
<td>5</td>
</tr>
<tr>
<td>Data Management: Building an Open, Centeralized Standard for Data Publication</td>
<td>7</td>
</tr>
<tr>
<td>An Overview of Salmon Data Management at DFO</td>
<td>9</td>
</tr>
<tr>
<td>State of the Salmon Program (SOS)</td>
<td>12</td>
</tr>
<tr>
<td>On Becoming Effective Ecologists</td>
<td>15</td>
</tr>
<tr>
<td>NEXT STEPS</td>
<td>21</td>
</tr>
<tr>
<td>IYS International Gulf of Alsaks Expedition 2019</td>
<td>21</td>
</tr>
<tr>
<td>Salmon Data Service</td>
<td>21</td>
</tr>
<tr>
<td>Strategy</td>
<td>21</td>
</tr>
<tr>
<td>NCES Southern Alaska Salmon and People</td>
<td>22</td>
</tr>
<tr>
<td>Coordinated Assessments Project</td>
<td>22</td>
</tr>
<tr>
<td>Immediate Next Step</td>
<td>23</td>
</tr>
<tr>
<td>ISDL STRATEGIC PLAN (A Work in Progress)</td>
<td>24</td>
</tr>
<tr>
<td>Goals</td>
<td>24</td>
</tr>
<tr>
<td>Quantifiable Objectives</td>
<td>24</td>
</tr>
<tr>
<td>Strategies</td>
<td>24</td>
</tr>
<tr>
<td>Projects</td>
<td>25</td>
</tr>
<tr>
<td>FOR MORE INFORMATION</td>
<td>28</td>
</tr>
<tr>
<td>Appendix 1. Attendees</td>
<td>29</td>
</tr>
<tr>
<td>Appendix 2. Presentations</td>
<td>31</td>
</tr>
</tbody>
</table>
BACKGROUND

This workshop builds upon two preceding workshops, also sponsored by the North Pacific Anadromous Fish Commission via International Year of the Salmon (IYS, see https://yearofthesalmon.org/events/) and Fisheries and Oceans Canada:


To set the background for this workshop, Mark Saunders presented an overview of IYS. He identified the International Salmon Data Laboratory (ISDL) as a project within IYS to address IYS Theme #5 Information Systems. ISDL thereby underpins the data assembly, integration, and analyses required to address Themes #1 Status of Salmon and #2 Salmon in a Changing World.

Dr. Kelly Chapman reviewed the immediately preceding IYS Salmon Status and Trends (IYS-SST) workshop wherein salmon ecologists discussed “legacy datasets” for Pink, Sockeye, Chinook, Chum, and Atlantic Salmon. The focus was on decadal patterns in abundance, survival, and biotraits (measurements from individual fish) at three scales: ocean basin, regional (e.g., Northwest Atlantic), and population. The next steps for the IYS-SST are linked to ISDL:

- Assemble more “legacy datasets” for salmon from the North Pacific and North Atlantic, and
- Link events and trends in salmon abundance to habitat changes (drivers) with an emphasis on predictions, i.e. hypotheses about mechanisms to explain and extrapolate correlations.

WORKSHOP GOALS

WHAT salmon ecologists need to predict the responses of salmon around the world to anthropic climate change will involve the assembly, standardization, and integration of new and historical datasets, complete with metadata about practices and precision, plus myriad habitat indicators. Modern analyses of that integrated data will build an understanding of the mechanisms underlying salmon dynamics. Acceptance and implementation of these new tools and practices will require automation and user-friendly interfaces.

HOW those needs can be met with modern technology for data processing—capture, management, integration, analysis, and presentation—was the focus of this workshop. The intention was that more complete, up-to-date, and effectively communicated ecological knowledge about salmon will lead to better decisions about salmon habitats and fisheries, thereby improving the resilience of salmon populations in a rapidly changing world. However, the technical results alone may be insufficient.

Workshop Goals

- Consider examples provided by salmon ecologists of critical issues, historical datasets, proposed analyses, desired products, emerging practices, and future requirements. Determine what these imply for data processing.
- Specify the next generation of technology, tools, and practices for collecting, integrating, analyzing, and communicating salmon information. This design will be valuable throughout the environmental sciences.
- Identify means to lower barriers to implementing these new practices and striking new collaborations throughout the international salmon community.
- Decide on the strategy, plan, and budget required to effect this transformation.
- Identify experiments that will lead to the adoption of new practices based on examples from the data laboratory.
The Revolution Begins with Salmon

Scott Akenhead

Pacific Biological Station, 3190 Hammond Bay Road, Nanaimo, BC V9T 6N7, Canada

Keywords: climate change, graph database, data pipeline, automation, collaboration, models.

Because anadromous salmon integrate changes in both terrestrial and marine habitats in their population dynamics, they are uniquely suited to be “the canary in the (global) coal mine” and warn humans about the accumulated deterioration of ecosystems from industrialization and the projected deterioration from the impacts of unmitigated and irreversible anthropic climate change—climate chaos. A paradigm shift is required: from short-term thinking about optimizing salmon harvests and protecting threatened populations, to long-term thinking about predicting the responses of salmon populations to habitat changes.

The resulting new requirements for global data integration and analysis exceed the capabilities of technologies presently available to, and useable by, salmon ecologists. ISDL will identify how those needs can be met with new technology for data capture and processing, with an emphasis on data integration that will allow the fitting broader and deeper models that address larger spatial scales and generate long-term predictions. For these new technologies to be effective, three considerations must be met:

a) implementation will require thoughtful experiments by this data laboratory, resulting in “irresistible examples” that attract widespread acceptance;

b) The new tools must be easy to use by non-programmers. Workflow diagrams, automation, GUIs, and interactive data visualization are among the strategies to lower barriers to implementation; and

c) Diverse new decision-support products will be required for fully informed and rapid decisions. The resilience of salmon depends to a large extent on decisions by humans regarding harvests, habitats, and hatcheries.

Salmon ecologists at Pacific Biological Station grumble about the state of the salmon data, such that writing a paper requires years of “dumpster diving for confetti.” IYS is proposing that we deliver transformational advances data processing for salmon (Fig. 1). But how? Looking ahead, I can see three opportunities to support advances in salmon science, management, and awareness:

Fig. 1. Conceptual workflow for capture of field data with quality control, then subsequent data integration for further quality control in the context of previous and related data.
1. Salmon Net

A combination of knowledge management and collaboration support to answer: who is doing what activities, where, for whom, with whose money, how, why. This must be personalized to be rewarding and must support “informed conversations” that are cross-linked to source information. All of the proceedings, workgroups, datasets, models, reports, maps, documents, and events are inter-connected. A working prototype was developed using the neo4j database, through the combination of ideas and links for a labeled property graph model. Funding is required for a pilot study prior to widespread application.

2. Salmon Data Service

If the reference to a dataset is one node in Salmon Net, then ideally that data has been standardized and integrated to allow advanced analyses and new decision support products. Proper analysis requires that the data be available as the original observations with metadata about validation, provenance, practices, precision, and caveats. These analyses leverage the value of integrated data and can be automated (workflows) for routine use by non-programmers.

To avoid obsolescence, the integrated data must be updated via automated "data pipelines.” Lacking that automation, maintaining integrated data has been prohibitively expensive. Because the schema for a graph database (neo4j) is mutable rather than fragile, capable of evolving as we learn, we can start to build the integrated database—adding new nodes, links, and new fields in old nodes and links—without knowing the final design. Formal standard processes can catch-up and sync later.

3. Data Laboratory

Saving the world is a tall order. Providing an example of how the world might be saved is easier. Thus, “the revolution begins with salmon.” ISDL is a data laboratory for data experiments to show what can be done.
Dynamic Data Management in the Cloud

Matt Deniston and Keith Steele

Sitka Technology Group, 920 SW Sixth Avenue, Suite 111, Portland, OR, 97204, USA

Keywords: data management, API, automation, field data, habitat monitoring, Umatilla Tribe

After over a decade of building custom data management systems for large fish and wildlife agencies, many of them focused on restoring Pacific Northwest salmon runs, we concluded, “There’s got to be a better way!” We started by defining our design goals: make it easy to use, self-service, secure and reliable, scalable, and extensible. Eighteen months later, our 30-person technology firm is happy to share we did find a better way and managed to deliver on those design goals.

The result is a dynamic, API-based data management system in the cloud called GeoOptix (geooptix.com). Resource managers can customize and centrally control their programs, projects, methods, sites, measurements and metrics. Field technicians can more quickly and cost-effectively collect, and quality check their data. Analysts can integrate and automate a generation of metrics, running of models, and conduct other post-collection analytics. GeoOptix frees up staff time to work on higher order information tasks. Given its infinitely flexible structure for organizing and storing data, it is ideally suited to serve as a central hub for data from multiple programs with datasets spanning multiple years and geographies.

For example, the Confederated Tribes of the Umatilla Indian Reservation in northeastern Oregon uses GeoOptix to collect and centrally control data from many of its habitat and species monitoring programs. Kaylyn Costi, a Fish Biologist from the Umatilla Tribe said, “With GeoOptix, I can set up my survey plans in the morning and have the crew out collecting data in the field that afternoon.” Sitka also worked with Kaylyn and her team to tap into the extensibility of the platform—we showed them how to write a few lines of code against GeoOptix’s API that “listens” for new field measurement data and as soon as it arrives, runs a job that automatically generates metrics. Kaylyn’s managers use these real-time metrics to track the status and trend of the resources they are responsible for managing. To learn more about how the Umatillas are using GeoOptix, see http://sitkatech.com/documents/50/GeoOptixCaseStudy_CTUIR.pdf.

Fig. 1. GeoOptix platform components.
Knowledge Graph

Neo4j (neo4j.com) provides a new way to manage data that is distinctly different from dealing with tables in a relational database management system like Oracle, MySQL, or Access. Neo4j stores data as a network nodes and relationships, equivalent to nouns linked to other nouns by verbs. Formally, neo4j is a “labeled property graph” with important new features. Nodes can be labeled as different types: Person, Place, Organization, Specimen, Document. Every node can hold a trove of information, including named fields, space and time coordinates, and URLs to files and web pages. Similarly, relationships (links, verbs) can be labeled as specific types, such as WorksFor, CollectedBy, IsAuthorOf.

Organizing and structuring data converts it to information. In neo4j, data is organized similarly to how people think: everything (node) is richly connected to other things; things have context; context is everything. Critically, in neo4j, links can be as richly described as nodes are. For instance, a link that is labeled WorksFor might have fields named jobTitle and startedWork. Knowing how things are linked, exactly, and in full context, moves you from information to knowledge. Thus, data about salmon can be transformed into a salmon knowledge graph, organized by the way that you think about salmon: growth is connected to survival is connected to life stage is connected to habitat. Each data point can be connected to the observation method used and related precision. This is a valuable set-up for statistics and simulations.

Use

Neo4j provides tools to:

- visualize and explore your data, to illustrate the rich connections in the data;
- store, reveal, and query nodes and relationships;
- traverse and analyze at any level of depth, in real-time;
- help facilitate technology and data integration; and
- reveal new relationships between data.

There is a natural evolution when building a knowledge graph:

- 1st graph: Connect similar objects into networks, e.g., the social network between persons. Start to do some path finding within the network.
- 2nd graph: Connect dissimilar objects: Who is doing what, where, with whom? Documents to datasets and authors.
- 3rd graph: Layers from different databases; leads to new discoveries and innovation; e.g., habitat indicators, standardized glossaries for parts and methods, libraries of digital maps and documents.
- 4th graph: Advanced analyses, forecasting, automation, AI, IOT, blockchain contracts.

What differentiates neo4j?

1. Performance: millions of “hops” per second; works in memory; billions on nodes.
2. Agility: schema can evolve easily, without breaking anything; contrasts to RDMBS.
3. Developer Productivity: easy-to-learn Cypher language for graph queries; huge libraries of functions.
4. Hardware Efficiency: requires about 10X less hardware and 10X less CPU than competing.

Analytics

For the past three years, Neo4j has focused on graph analytics. One milestone from this work is a new book "Graph Algorithms" by Neo4j data scientists Amy Hodler and Mark Needham. These algorithms are applied to reveal the structure of large and complicated datasets. Such analyses will discover the people who are a “bridge” between otherwise separated groups of people and will discover clusters of nodes based on connectivity (different from clusters based on measurements).

Perhaps “over the horizon” for salmon ecologists is the exciting development of artificial intelligence (AI) that is leveraged by Neo4j capabilities. This takes pattern recognition, and automated responses to those patterns, to a whole new level. As a cursory introduction, there are four pillars of graph-enhanced AI:

1. A knowledge graph is the context for decisions suggested by AI;
2. Connected feature extraction provides context for assessing the accuracy of AI results;
3. Graph accelerated AI provides a new efficiency for AI algorithms; and
4. AI provides credibility for results by explaining the connections it detects.

More knowledge, more discovery

The value of a knowledge graph may expand disproportionately as the size of the graph increases. New connections, new patterns, and new ideas will emerge from visualization and exploration. Broader and deeper analyses are possible. More people will find the information, tools, and results to be valuable.

An example of expanding value was the increased ability of the International Commission of Investigative Journalists to detect offshore money laundering by combining a huge first leak, the Panama Papers (11.5 M documents) with a subsequent big leak, the Paradise Papers (13.4 M tax accounts). More data → more connections → more evidence → more convictions.

We can only speculate about how “multiplication of value from addition of data” might apply to the integration of multiple salmon datasets.
Data Management: Building an Open, Centralized Standard for Data Publication

John Song

Systum Inc., 71 Stevenson Street, Suite 400, San Francisco, CA, 94105, USA

Keywords: data management, database schema, data visualization, data governance, API

Problems surrounding data management and publication are universally experienced by industries, institutions, and entities that rely on the free exchange of information to support their activities. The problems can be attributed to disparate data sources, an ocean of data formats, and information that is neither standardized nor flows freely between individual parties seeking to find universal truths or accessible patterns that extract value from the information.

Salmon data, as shared by researchers and fisheries at the ISDL workshop, are particularly afflicted by this problem. They cite multiple data sources, multiple data formats, and a lack of standard practice for data collection and data definition to support their research activities and to publish their findings to a larger community of policy makers and activities.

The core issue to the data problem lies in the lack of a standard schema for consolidating the information extracted or contributed by researchers across multiple geographies, industries, and institutions. Data measurements and the meta data describing the conditions under which the measurements were made are scattered in different information repositories and with differing definitions of the sampling conditions. Thus, the utility and value of the information gathered is greatly compromised.

A standard or "God" schema is a proposed solution to this disparate data problem. The God schema will provide standard data definitions that will allow universal contribution from different institutions and industries without compromising data integrity. Data submitted to the "God" schema will follow a standard format and will be validated and governed to ensure data quality before it is published to the wider community. Data access, validation, and documentation will follow current best practices for data management in the cloud including data governance, API access, and JSON data exchange.

The God schema will be accessible through a Salmon Data Service that will provide interfaces for data discovery, access, and contribution. The service will also provide documentation and information retrieval for all resources under its management. By enabling the management of God schema-based data through the Salmon Data Service, the data will be more widely accessible and will build the foundation for the creation of ancillary tools to support Salmon research activities. It will enable researchers to speak a common language when discussing data and it will enforce a common standard for publishing data to the international community. Having a standard understanding of data will make it more accessible; it will result in dialogue and activities by contributors outside the salmon research community, which will initiate a change in attitudes toward global conservation.

The Problem

Problems surrounding data management and publication are universally experienced by industries, institutions, and entities that rely on the free exchange of information to support their activities. The problems can be attributed to disparate data sources, an ocean of data formats, and information that is neither standardized nor flows freely between individual parties seeking to find universal truths or accessible patterns that extract value from the information.

Salmon data, as shared by researchers and fisheries at the preceding IYS-SST workshop, are particularly afflicted by this problem. Ecologists cite multiple data sources, multiple data formats, and a lack of standard practices for data collection and data definition to support their research activities and to publish their findings to a larger community of policy makers, agencies, and activities.

The core issue to the data problem lies in the lack of a standard schema for consolidating the information extracted or contributed by researchers across multiple geographies, industries, and institutions. Data measurements and the meta data describing the conditions under which the measurements were made are
scattered in different information repositories and with differing definitions of the sampling conditions. Thus, the utility and value of the information gathered is greatly compromised.

The Solution

A standardized God schema for a graph database is a proposed solution to this disparate data problem. The God schema will provide standard data definitions that will allow universal contribution from different institutions and industries without compromising data integrity. Data submitted to the God schema will follow a standard format and will be validated and governed to ensure data quality before it is published to the wider community. Data access, validation, and documentation will follow current best practices for data management in the cloud including data governance, API access, and JSON data exchange.

The God schema will be accessible through a Salmon Data Service that will provide interfaces for data discovery, access, and contribution. The service will also provide documentation and information retrieval for all resources under its management. By enabling the management of God-schema-based data through the Salmon Data Service, the data will be widely accessible and will build the foundation for the creation of ancillary tools to support salmon research activities. It will enable researchers to speak a common language when discussing data and it will enforce a common standard for publishing data to the international community. Having a standard understanding of data will make it more accessible; it will result in dialogue and activities by contributors outside the salmon research community, which will initiate a change in attitudes toward global conservation.

Other Remarks

• One issue in the “data problem space” is overcoming preconceived notions about data, data standards, and data management. One such preconception is the need to completely design a database before being able to use it, which is wrong when dealing with a graph database.
• Data becomes information when you can describe it. The implications of that include meaningful (standardized) names for parts and practices, and something about the organization (format) and accessibility (medium, security, tools) of that data.
• A graph database schema is the skeleton of the database, a visualization of the types (labels) of nodes and links in that graph. Like an RDF schema but not like a relational database schema.
• Application programming interface (API): functions that allow creating applications that will use the database, including “featuring” that data through an operating system, web pages, and other services. Data submission, validation, and publication done through an API. For instance, an API that is the query interface from statistics models written in the data science language R to the Salmon Data Service (directly or indirectly using the language Cypher) would facilitate applying those models to salmon data.
• Salmon ecologists need to be able to visualize their data—interactive data visualization—and thereby see the context of any data in relation to all sources of salmon-related data.
• John Song demonstrated the NuSEDS database (DFO Pacific Region spawning salmon annual abundances, provided by Bruce Patten of this workshop) imported to neo4j, then joined to the international salmon catches database from NPAFC\(^1\), then linked to the D3 library, finally interacting, live, with an example of data visualization. An auspicious event.
• With little information sharing, often times people are doing the same thing and it is not efficient to have it be repeated by each company or software dealing with the same data.
• Enrich the data and transmit.
• Three main points:
  o This problem is not unique;
  o You need common knowledge, a God schema; and
  o You want to reach a wider audience, so the program should be accessible.

An Overview of Salmon Data Management at DFO

Bruce Patten

Pacific Biological Station, 3190 Hammond Bay Road, Nanaimo, BC V9T 6N7, Canada

Keywords: data management, data architecture, vision

Data Types / Topics

Science Branch and Fisheries Management Branch in DFO Pacific Region requires that a wide variety of data types must be managed. The “big three” for us are:

- **Fishery Data**: catch, effort, biological traits—anything that pertains to the operation of the fisheries. Most significant is catch and effort data: how many fish were caught and how much work was involved to catch them (e.g., the number of boats fishing).

- **Escapement Data**: in-river abundance, biological traits—escapement data traditionally refers to the information about adult salmon that have returned to their birth river after some period at sea.

- **Mark and Recapture Data**: tag releases, tag recoveries—the information from our major stock identification program that we coordinate with US agencies. This program is how we determine the origins of salmon that were caught at sea.

We also manage extensive and diverse datasets for:

- **Laboratory analytics**: age of fish and shellfish, genetics and genomics assays (voluminous)

- **Research data**: investigative surveys that produce unique datasets

- **Biological traits**: data at an individual fish level (e.g., length, sex, fecundity)

Data Inflow—Who and How

Our data program receives data from a variety of people and agencies in a wide variety of processes and formats. Perhaps the highest priority is assigned to data from the fishers who are required to report information to the Department regarding their fishing activities. We have data sharing arrangements with US state and federal agencies.

Data Program Activities and Deliverables

Operations are multi-faceted and include management of the data acquisition program, quality control, standard reports, providing time-series for analyses among other data requests, and assisting analysts via data interpretation expertise. The data stewardship team focuses on making sure that information is received as expected and is quality reviewed prior to use by analysts. We do produce a variety of standard reports, but the highest-value use is models for fish populations (e.g., stock assessment). This often requires that analysts connect a data analysis tool directly to a database view or must ingest some large data export.

Government data are considered “open by default”. This means that anyone can request a copy of the data. However, particularly in the case of fishery data, there are privacy concerns that can restrict the level of detail that is provided to a requestor. The data stewards often need to work with a data requestor to prepare a data extract that is appropriate.
Data Uses

Much of the salmon data that we acquire is for operational use in fisheries management. That is, to operate or manage fisheries, often with tight turnaround times between data acquisition and data provision. Salmon stock identification, an important data use, involves determination of the stream of origin for individual or groups of salmon, e.g., fish from the same river system, and “conservation units.” Fisheries data and stock ID are the basis for stock abundance estimates and are used to determine the population size of a particular salmon stock. The scale of stocks can vary from an individual stream to an entire drainage like the Fraser River system (0.22 × 10^6 km², ~¼ BC, ~½ California).

Stock abundance is useful in the context of stock status determination. This refers to the combined concepts of the trajectory for the stock (stable, increasing, or decreasing) and the relative abundance compared to a specific benchmark, such as the number of fish required to support the production of future generations. An important aspect of stock status is quality, which refers to the genetic diversity of the fish and their ability to produce successful offspring. Status is often categorized using a stoplight scheme of green for good, yellow for caution, and red for danger status.

A key function for DFO Science is to generate formal scientific advice, usually for DFO Fisheries Management. This involves a formalized process of analysis, documentation, peer-review, and publication. The data are also used by DFO staff to explore interesting questions and to produce new scientific knowledge.

Drivers for Change

No other time in my career have I experienced some much change and potential for change as I am experiencing right now:

- **Aging Infrastructure**: the infrastructure investments of the late 90s are largely unchanged but now are rusting out. For instance, since NuSEDs went live in 2000, it has not evolved and is consequently outdated.
- **Big Picture Issues**: we are faced with pressures from global challenges like climate change. Designing appropriate responses was a motivation for this workshop.
- **Fisheries Act Renewal**: The Auditor General released a not-so-complementary report on fish stocks a couple years ago. This led to Fisheries Act revisions that will likely have a specific focus on science. That means there will be legislated implications for how we do our work, legislation backing up the need to conduct science in a modern way, a new situation for us.
- **Whole of Government Initiatives**: beyond just the environmental sphere, there is lots of pressure from technical change. For instance, a recent report to the Clerk of the Privy Council (the most senior civil servant in the Canadian government) outlined “A Data Strategy Roadmap” for the federal public service. If followed, it will result in significant changes to the way we organize and share data.
- **Open Government**: there are international aspects to the design of our program, and how that design will change. In June 2013, Canada endorsed the G8 Charter on Open Data, meaning “…maximize the release of government information and data of business value to support transparency, accountability, citizen engagement, and socio-economic benefits through reuse…”
- **Consolidation and Centralization**: we seek efficiencies from fewer but larger data management solutions.
- **Security Concerns**: with more data, more diverse data, and more portals to access data being provided, protecting confidential information such as the identity of fishers and the location of individual catches, has become increasingly important and increasingly demanding.
**SWOT Analysis**

The following table summarizes some initial thoughts about strength, weakness, opportunities, and threats to our data program.

<table>
<thead>
<tr>
<th>Strength: data, people</th>
<th>Weakness: infrastructure, people</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Long time-series of data.</td>
<td>• Lack of agility: cannot be easily modernized.</td>
</tr>
<tr>
<td>• Consistency of time-series.</td>
<td>• Based on old technology (and old expertise?).</td>
</tr>
<tr>
<td>• Stability of data governance.</td>
<td>• Variety of data designs: data developed in isolation between different agencies.</td>
</tr>
<tr>
<td>• Credibility of institution.</td>
<td>• Poor governance: mostly bottom-up.</td>
</tr>
<tr>
<td>• A team with real-world experience.</td>
<td>• Lack of metadata: precision, methods.</td>
</tr>
<tr>
<td></td>
<td>• Lack of data consolidation.</td>
</tr>
<tr>
<td></td>
<td>• Lack of formal and recent training.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Opportunity</th>
<th>Threat</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Interest in exploring new approaches.</td>
<td>• Staff departures without succession plan or corporate knowledge capture.</td>
</tr>
<tr>
<td>• Renewal projects.</td>
<td>• Explosion of data volume. Large binary objects: imagery, audio.</td>
</tr>
<tr>
<td>• National and international collaboration.</td>
<td>• Loss of data management autonomy.</td>
</tr>
<tr>
<td>• Improved recognition of the importance of data and information to support sound decisions.</td>
<td>• loss of data management features.</td>
</tr>
<tr>
<td>• Interest in data standardization for sharing.</td>
<td>• increased response times for issue resolution.</td>
</tr>
</tbody>
</table>

**Vision for the Future**

Given these significant pressures to change, and the current state of our data systems (not just for salmon), I anticipate a lot of system renewal over the next few years. With this will likely come pressure to consolidate into fewer but bigger systems. If you wanted to reduce the number of data systems to less than 50, for example, what is that process?

The need for more open access, interoperability, and sharing of data will lead to better and broader standardization of the data architecture. New systems to meet present requirements must also be applicable in the future. We need to plan for the data management systems that will be required 25 years from now.
State of the Salmon Program (SOS)

Sue Grant and Bronwyn MacDonald

Fisheries and Oceans Canada, Annacis Parkway, Unit 3, Delta, BC, V3M 6A2, Canada

Keywords: State of the Salmon, Canadian salmon, salmon status, salmon trends, ecosystem trends, data treatment, data visualization, productivity

State of the Salmon (SOS) Program Goal:
To track and understand Pacific salmon ecosystem trends. This includes tracking and integrating:
   a) population abundance and productivity (survival) trends; and
   b) “biotraits” including size, fecundity, age, etc. trends
   c) observations in salmon freshwater and marine ecosystems

The SOS Program includes three pillars:
1. Integration: foster integration across DFO and non-DFO technical experts related to salmon and their ecosystems; this includes integrating expertise on a global scale from organizations such as the Canada-U.S. Pacific Salmon Treaty process, the North Pacific Anadromous Fish Commission, and the International Year of the Salmon.
2. Data visualization: through innovative data visualization approaches, develop tools to assist with track and understanding salmon status, salmon trends, with linkages to environmental conditions.
3. Communication: through presentations, reports, emails, SOS meetings etc. communicate SOS activities and results

A key input for this State of the Salmon Program is salmon data, particularly related to abundance and productivity. These data are produced by individual stock assessment groups within DFO, supported in a number of cases by First Nations, and other organizations. There are two steps related to these data, that include data collection & processing, and data treatment.

Data Collection & Processing
Fraser Sockeye is the first group of salmon data we are including in our data visualization analytical tools. This is a data rich group of salmon populations, where data has been previously consolidated, and is annually updated. Abundance and recruitment data alone have many steps for annual processing. It takes years of work to deliver another year of recruitment data to update each time-series. The data processing steps (information flow) involves:

- field enumeration: catches (multiple fisheries) and escapements (multiple spawning sites);
- data entry;
- error checking of data inputs;
- escapement estimation and testing of assumptions;
- DNA sampling and analyses for stock identification in mixed catches (mixture analysis);
- linking data sets: age, catches, escapements, populations, and stocks (CUs);
- gap filling to deal with missing data;
- reconciling with hatchery data and other linked databases; and
- combining all the preceding data into formats that will feed into analytical code.

There can be significant lags to receiving these data, and typically the only data that is readily available and timely is the data required to manage high profile salmon fisheries.

Data Treatment
1) Examples of analytical approaches to estimate productivity trends:
13

a) recruits produced-per-adult spawner;
b) Ricker model residuals: largely represents density-independent salmon productivity;
c) Kalman filter models: a model form that deals with non-stationary temporal patterns in salmon productivity; and
d) further methods yet to be discovered and/or applied.

2) Examples of removing signal from noise
   a) loge transformation: intended to provide a normal (Gaussian) distribution for ratio estimators such as recruits/spawners;
b) smoothed running average: removes some of the inter-annual variability in salmon productivity and observation error;
c) further methods yet to be discovered and/or applied.

Data Visualization

Design Process

One pillar of the SOS program is an integrative analytical data visualization tool, which will consolidate and present salmon trends and status information for technical salmon-expert end-users. This will enable users to explore these data, and answer their salmon-related questions. A key to our approach was to gather user input in advance of the design process. To that end we conducted questionnaires with 30+ DFO Pacific Region staff who were involved with salmon, to determine the questions about salmon populations that they needed answered on a regular basis. We also conducted a structured user-testing process throughout the development phase, to fine tune and optimize the tool.

Our objective is to allow for effective exploration and communication of salmon data, and to maximize the utility and adoption of this tool upon delivery. To advance this work we collaborate with academic leaders in the field of data visualization, cognitive psychology, advanced coding, and salmon populations.

Where We Are Now

Currently, we are developing a synoptic status module for Fraser River populations. This is in response to several high-priority, status-related initiatives: Fisheries Act Rebuilding Plans, COSEWIC Recovery, and Wild Salmon Policy Red Responses. Our approach emphasizes collaboration; we are working with Carrie Holt (DFO), Gottfried Pestal (SOLV Consulting), Michael Barrus (Consultant), Brigitte Dorner (Driftwood Cove Consulting), and Randall Peterman (SFU emeritus).

The technologies we are using to deliver a user-friendly, interactive interface for analysis, visualization, and communication, are in packages (software libraries) available through the R language, and delivered as interactive graphics on the Internet via Shiny (also an R package).

Next Steps

Development plans include
• automate analyses of Fraser Sockeye salmon data: keeping the assessment of status up-to-date as the data available are updated annually;
• workshops to test the tool to refine the design;
• expand with further modules: in particular a mapping module is planned;
• add data sets for other groups of salmon and species, starting with Southern BC Chinook: the tools and processes developed for Fraser River sockeye CUs will be widely reapplied to other species and other areas;

We are working from a four-year workplan that defines the required projects, deliverables, budgets, and benefits. We have defined the workgroups and projects for some immediate experiments that will provide the success stories and proof-of-concept deliveries required for further support and wide adoption of our approaches.
How can ISDL help us

There are many parallels between the ISDL goals and the SOS goals; in general, we see the same problems and the same solutions. The main differences are in scope (hemispheric vs Pacific Canada) and immediacy—SOS is currently required to deliver results from the available tools, ISDL is future-oriented and unconstrained by operational necessities. We appreciate the value of accessing a much larger community of people involved in data processing and visualization. That community will present many opportunities to enhance our current work in the SOS program. Like ISDL, the SOS program is about pushing the boundaries of what can be done. In addition to a general interest in what will be possible, we are specifically and immediately interested in:

(a) ideas about signal processing in the many time-series we are working with;
(b) ideas on data visualization that will leverage the value of the data, tools, and analyses we are assembling; and
(c) ways to make comparisons—between the patterns in different CUs, regions, and species—to extract knowledge about the status of salmon, and what we should expect in the near future.
On Becoming Effective Ecologists
Kelly Chapman

Vancouver Island University, Powell River Campus, 100-7085 Nootka Street, Powell River, BC V8A, 3C6, Canada

Keywords: decision, knowledge transfer, cognitive dissonance, wicked problems, communication, strategy, dialogue, complexity

In his remarks to opening this workshop, Scott Akenhead spoke of a growing need for wise and nimble decisions by salmon policy makers. He suggested that better transfer of knowledge transfer from ecologists to decision makers would lead to “better informed decisions.” Scott proposed that the title for my talk be “Effective Communication Between Ecologists and Decision-Makers.” I declined for reasons that will emerge from the following discussion, which is adapted from Chapman (2016).

Finding ways of continuously generating, sharing and applying knowledge from a diversity of disciplines and viewpoints is a cornerstone of adaptive management (Holling 1978; Walters 1986). It is also understood as central to building the adaptive capacity (i.e., resilience) (Walker et al. 2002) of governance systems such that they are able to detect and respond to environmental change (Berkes et al. 2003; Olsson et al. 2007) and ecological crises (Folke et al. 2003). Knowledge transfer refers to the spectrum strategies used to bridge the gap between research, policy and management (Roux et al. 2006). The main goal of knowledge transfer is to create useful information for decision-makers (McNie 2007), and the unobstructed flow of knowledge between managers and researchers (Roux et al. 2006). This is representative of a view widely held by many scientists: if one’s research and modelling results are salient, credible, and legitimate—as well as packaged and delivered to target audiences in a timely, relevant, and compelling way—they will lead to better informed decisions (McNie 2007, citing others therein).

![Adaptive management cycle outlined by CMP’s Open Standards Project](source: Conservation Measures Partnership 2007).

This traditional view of knowledge transfer is based on the following string of assumptions:
(a) Humans process information like computers, and make rational decisions based on their knowledge of reality;
(b) When given complete information about the utility of different options, humans will always choose the “best” one; and
(c) The more information we have about a system/problem, the better choices we’ll make.

But are humans rational? Is decision-making a conscious or unconscious process? In reality, 98% of what the brain does is non-conscious and without conscious guidance (Beratan 2007, citing others therein). People don’t actually know what makes them behave or think as they do (e.g., Nesbitt and Wilson 1977), and it appears there is no conscious process of evaluation before responding (Nisbett and Wilson 1977). Many of our common beliefs about how people make decisions emerge from Rational Choice Theory, which began to gather influence in economics and other social sciences in the 1950s. The theory attempts to explain and predict socio-economic patterns and phenomena arising from multiple individual choices, as individuals attempt to maximize their benefits and minimise their costs (Hausman 2008). Rational choice became an “article of faith” among mainstream economists until the 1990s (Hodgson 2012). However, since then, the assumptions and predictive capacity of rational choice models have been widely disputed (including by six Nobel prize winners for Economics), particularly the notion that people consistently act in ways that maximise their rewards (Hodgson 2012). Behavioural economist Daniel Read points out that people:

... ignore important decision factors, put undue weight on some factors relative to others, plan to do the right thing but fail to follow through with those plans, they are more sure about their decisions or beliefs than they should be, they trust others more than they should, and they even fail to do simple calculations that could solve important problems. (cited in Foley and Griffiths, 2011, p. 21)

Clearly, we are not as rational as we would like to think. Applying knowledge transfer strategies premised on the assumption that people primarily act and make decisions by rationally processing information is more than just an exercise in futility—it can also have seriously counterproductive consequences. When a person is given information that goes against their beliefs and values, they will often experience an intolerable psychological discomfort termed cognitive dissonance (as per Festinger 1957). To eliminate this discomfort, the mind applies one of two strategies: (a) altering its beliefs and values to conform to the new information, or (b) resisting the new information.

Most knowledge transfer efforts assume that people will opt for the former strategy. However, research shows this is often not the case (Sides and Citrin 2007), and that, perversely, presenting facts or correcting misinformation often causes people to become hostile and entrench their existing behaviours, as per cognitive backlash (Nyhan & Reifler 2010). Indeed, some studies show that the more uninformed (Kuklinski and Quirk 2000), misinformed (“fake news!”) or ideological (Nyhan and Reifler 2010) people are, the more likely they are to adopt the latter position. So, although cognitive dissonance can lead some people to change their beliefs and behaviours, it can also backfire, causing others to reject the new information, shoot the messenger and further entrench their positions.

A simplified summation of the climate change debate illustrates how knowledge transfer efforts that fail to shift a learner’s beliefs and values can create a vicious circle, thereby escalating errors being committed (as per Watzlawick et al. 2011). Many scientists believe that presenting a persuasive argument based on rational facts should influence the beliefs, decisions and behaviours of others. From this assumption, scientists concerned with the lack of action around climate change present a barrage of facts to the public and decision-makers. However, the scientists do not realise that information that conflicts with people’s existing beliefs has the polarising effect of causing some people to change their beliefs, while causing others to entrench their existing positions and discredit the messenger. Because these scientists have not recognised their underlying unsubstantiated assumption that facts change people’s beliefs and behaviours, they fail to see that presenting climate facts helped create climate denial and scepticism in the first place. In absence of this insight, scientists then respond to mounting climate scepticism by presenting more contradictory facts, thereby escalating their error by intensifying a strategy that inadvertently increases rejection of climate science among those they are trying to convince.

The above is an example of how environmental problems that appear to defy resolution are often framed such that their assumptions preclude effective solution. It also shows how these problems are often unintentionally reinforced (and sometimes even created) by the very strategies applied to resolve them. Such
seemingly impossible or difficult to solve problems such as climate change, rampant extinctions, and habitat degradation, are often referred to as *wicked problems* (as per Rittel and Weber 1973). Wicked problems are characterized as (a) being difficult to define; (b) having no right or wrong solutions, just better or worse as subjectively defined by stakeholders; (c) and having numerous subjective and competing causes. In addition, their implemented ‘solutions’ have significant consequences (meaning there is no opportunity for trial and error learning) and may in fact exacerbate the problem they are meant to solve.

How then do people really make decisions? In the traditional behaviourist view, learning is believed to occur automatically via information processing *within* individuals (Huber 1991). Scientists and other specialists, each working within their individual silo organizations and fields, collect and process information about the world and transfer it (in the form of papers, reports, briefings, etc.) to policy and decision-makers, who then process the information from various sources to make rational decisions that maximize public good. Hence, the objective of knowledge transfer and adaptive management is to transfer the information stored in scientists’ heads, into that of decision-makers.

This contrasts with a *complexity-based* understanding of learning and organizing, which views human organisations (i.e., social systems) as self-organising, complex systems, and knowledge as continuously emerging via the active process of relating *between* individuals² (Stacey 1996a, 2001; Stacey et al. 2000). Stacey (1996a) discusses human organisations in relation to Kauffman’s (1993) NK models of complex self-organising systems. When connectivity, diversity and rates of information flow between their members and/or outside groups are low, organisations are relatively stable and unchanging. These organisations require little energy or information flow to sustain them. However, they are not really learning or adapting because existing patterns of connection have locked-in or ossified³. In order for an ossified organisation to become receptive to change, the connectivity, diversity and rates of information flow between people need to increase. This begins to destabilise the organization, causing it to spontaneously become more responsive to its environment (i.e., it is learning). In doing so, novel patterns of organisational behaviour (i.e., new basins of attraction) may begin to emerge.

A complexity-based understanding of learning and organizations proposes that it is *relationships* and patterns of interaction—not information processing—that directly drive most decisions and system behaviour. In this context, knowledge is not an object or thing that can be transferred or exchanged (i.e., a product of learning). Rather, the ‘knowledge’ of a system is better understood as a function of its *internal and external connectivity* (i.e., relationships between its members and other systems, which confer the system with *plasticity*, in terms of its range of possible behaviours—or *creative capacity*, in terms of the range of future realities⁴ its members may co-create through their interactions) than as the amount of information held inside the heads of individual system members. In sum, decision-making is a process of coordinated interaction between people rather than individual information processing. Hence, *conversation⁵* is the currency of change in human systems (Shaw 2002). Repeated, reciprocal interaction (conversation) drives patterns of behavior in groups of people. It can destabilize old patterns of relating, thereby creating opportunities for change and novelty to emerge (albeit in unpredictable and not always desirable ways).

Supplying people with credible scientific information has limited impact when people seldom make decisions rationally or consciously apply logic to their thinking. When decisions are made that incorporate scientific perspectives, it may have less to do with the credibility or legitimacy of the supplied information and more to do with scientists entering into repeated personal interactions and conversations with decision-makers. By adhering to the assumption that information, rather than relationships, leads to ‘better’ decisions, knowledge transfer efforts become unnecessarily convoluted. Excessive time and resources are often spent on improving

---

² This is also the case for social constructivist perspectives, whereby mind and learning are likewise understood as emerging from ongoing social interactions between individuals (e.g., Ertmer and Newby 1993, Gergen 1997; Shotter 1994).
³ As the term is applied by Seel (2006).
⁴ I.e. the system’s *phase space*.
⁵ Because conversation is integrated into people’s day-to-day interactions, it directly shapes their behaviour in the here and now, and thus the reality they are co-creating. Hence, conversation has a more immediate influence on a system’s patterns of interaction than do abstract discussions carried out in isolation of ordinary practice and requiring formal pathways of planning and implementation (Shaw 2002).
information when they would often be better spent on improving relationships. As Lomas’ (2007, p.130) points out: “human interaction [is] the engine that drives research into practice.”

Hence, ecologists are likely to have greater influence on the world (as a continuously emerging complex system) via establishing relationships and actively connecting with other components of the social system, than they are by passively supplying information to others. As such, ecologists may wish to shift their attention towards improving their own creative capacity. This means improving the impact of their local participation, and therefore helping create a reality that is more sustainable. This requires that ecologists reframe their role of objective observer into that of active participator, i.e., building relationships, changing our conversations, and expanding our day-to-day personal interactions with others. Some strategies for scientists and research programs wishing to improve their impact (see Chapman 2016 for detail) include:

- Increasing the number, diversity, and intensity of personal interactions;
- Engaging in everyday conversation;
- Seeking opportunity in crisis;
- Participating in collaborative planning;
- Fostering creative dialogue focused on possible futures (less focus on discussion and problem analysis based on what went wrong in the past);
- Using narrative to convey a compelling vision of the possible;
- Using modeling to catalyze dialogue and extend imagination;
- Enlisting knowledge brokers or boundary organizations to facilitate interactions with other groups and individuals;
- Changing metrics and incentives to reflect the scope and quality of interpersonal interactions, and
- Conducting qualitative research to deepen understanding of social phenomena.

In conclusion, wicked problems are often the result of how they are formulated. Consider the crucially different formulations of the problem that we have gathered to address:

A. How can we make communication between ecologists and decision-makers more effective so that our decision-makers are better informed?
B. How can ecologists more strongly influence the patterns of reality emerging from the complex systems in which they are embedded?

By examining our deeply held assumptions about how humans make decisions we can see that the traditional problem statement (A) is best reformulated into problem statement (B). This entails leaving the rarefied world of scientific abstraction, and actively participating in the messy and unpredictable process of co-creating reality, with all the tension, cooperation, and uncertainty that it entails.

REFERENCES


---

6 Pielke (2007) suggests that scientists can become actively engaged in policy debates by taking on the roles of ‘issue advocate’ (taking a definite stand on an issue and preferred policies) or ‘honest broker’ (providing a wide range of policy options without advocating for any one of them), neither of which focus on sharing imagined futures. Pielke argues, however, that advocacy by scientists politicizes and damages the credibility of science. This stance is disputed by some critics (e.g., Howe 2007) who suggest scientific advocates are needed to balance the views of powerful interests and prejudices in policy debates.

7 For years, scientists have battled the Rosenthal Effect—the impact that the opinions, outlooks, theoretical and practical biases of researchers have been shown to have on performance of their subjects, be they rats or humans (Watzlawick et al. 2011). Scientists typically view observer bias as negative and are always looking for ways of eliminating it. However, as quantum and chaos theory have shown us, this is a utopian ideal that can never be achieved. If environmental scientists reframed their understanding of observer bias from that of a problem to that of a potential advantage, they could potentially capitalise on the subjective influence research has on a system’s behaviour.


NEXT STEPS:

IYS International Gulf of Alaska Expedition 2019

All of the data are to be shared (among researchers, possibly not public). Could this be an ISDL experiment in data integration? This would facilitate analyses of this and subsequent expeditions.

- In regard to research cruise in February–March: 21 researchers, going to lead to patchy data: one period, one area. Must be linked to pre-established, rich datasets.
- What are links between data nodes: scientists, date/time, cruise location/station, cruise, individual salmon?
- Overlay with environmental conditions data
- A key benefit is that it engages five nations in collaboration on salmon ecology.
- The 21 participating scientists should be the people who design the schema for a knowledge graph for these expeditions.

Salmon Data Service

- The architecture that DFO’s data is sitting in does not matter: if the metadata is there, we can bring it online from an API.
- Intend to have the God Schema agreed to [and operating] by the end of 2022
- Get ship and existing data online, then connections between the data can begin.
- Engage with PICES and ICES to start linking oceanographic data to salmon data.
- What actions must be initiated to work towards those goals?
  - IYS has to build trust with partners, including government, in order to implement a successful strategy
- We need to define a target for the data: where will the data go after collection? That is what the God schema is for: it enables data products that will support action and policy
- Issues with undocumented uncertainty and methods for easily available data: NuSEDs and NPAFC.
  - Appropriate metadata is required for IYS-SST legacy datasets. These have priority for IYS research.
  - Should reconstruct information about data points as best we can (e.g., category of relative precision).
- How do we want to capture new data moving forward, using the new technologies that are laid out in front of us?

Strategy

- Let researchers who are not in this room know and they can volunteer and participate and decide the common goals and projects.
- We identified possible research question in last two days of workshops. This requires identifying and involving the data holders.
  - These question (with work groups) could be data experiments in applying neo4j
  - End result is a workshop at the end of 2019 to see if we can use neo4j on a larger scale.
- What is the question, who is it for, what is the methods, what is the outcome, can it influence decisions?
- Everybody’s project team can investigate this type of “design thinking” following the examples that Jeff Morris (Neo4j Inc.) and John Song (System Inc.) have laid out for us.
- We wish to maintain participation of the 25 people here. How can that proceed?
- “Innovation lab”—we will need a lengthy and structured workshop to learn these new methods
  - Timeline: approximately six to eight months
• Process:
  o An example of “Here’s what we need to prove”
  o Two or three datasets with linkages
  o Build a prototype
  o Socialize it
  o Spawn subsequent projects out of it
• The group would rather incorporate information from ICES (North Atlantic)
  o Move that forward as an ICES working group
  o Make the resulting knowledge graph accessible through an “innovation lab” workshop
  o Not so worried about DFO data
• Start a salmon knowledge graph with what we can access and work with, right now instead of going to a relational database.
  o Always a way to connect one data source to another
  o Taking any datasets and merging them is a good exercise
  o More dataset can be added continuously
  o Possible because neo4j graph database is agile, dynamic, and modifiable
• IYS-SST “Legacy datasets” needs to be alive, and continuously updated as new data is collected.
• Worthy investment: to better relate old datasets to today’s datasets.
• A graph can have more than one schema.
  o You can have sub graphs and master graphs
• We can assemble detailed data on individual populations; status: it’s alive but not moving much.

NCES Southern Alaska Salmon And People (SASAP)

• This project has natural sciences and social sciences brought together to 2016, should be a dataset to look into.
  o Still collecting the data and also updating datasets, although that was not part of the NCEAS model.
  o Usefulness diminishing with each year without updates.
• How were Alaska Department of Fish and Game datasets transferred?
• Is there a better way to share data for these types of projects?
• Linking social scientists and natural scientists is an important goal.

Coordinated assessments project

• Group called Pacific Northwest Aquatic Long-Term Partnerships
  o Coordinated across different agencies
• Annual process for Fraser Sockeye combines scientists working on different life-history stages.
• Previous pattern was people often working in isolation, i.e. freshwater and marine scientists with little dialogue among each other.
• Provides a structured process for collaboration and sharing of information.
• How can we take that into the next step?
  o New data management designs, e.g. group by cohort/ year-class.
• Fraser River Sockeye have a high profile in Canada, poor returns led to Cohen Commission
• Fishers and fisheries managers are unsatisfied because ecosystem is “becoming unusual” and both returns and predictability are low.
• Participants need to collaborate more than ever.
• Synergy leads to connections and better qualitative predictions
• We had no way to unify data in a modeling context until we were shown today’s presentations.
• We also need to create tools to foster dialogue and to integrate that team effectively.
Immediate next steps:

- How should we communicate? Email chain? Discussion group? Connect through Scott Akenhead (scott@s4s.com)?
- Engage graph data visualize companies.
- Act immediately upon Gerald Chaput’s proposal re ICES database.
- Get more researchers involved and build momentum, keep this going.
- Start on standardized glossary: a resource description framework (RDF) of ideas re ecosystem parts and processes, human practices, etc.
- Move quickly. The ICES annual conference in about 6 months, now is the time to obtain buy-in for new projects.
- Mark Saunders and Scott Akenhead to connect with Matt Jones (NCEAS/SASAP) to mutual benefit.
- Establish that we have working tools/technology to make connections between Atlantic and Pacific salmon and habitats, then identify gaps in funding that prevent proceeding.
  o Needs a report by May for IYS workshop in Portland
  o Needs a proposal ready for ICES workshop in 3 months
  o A place to introduce these ideas in one of the workshop themes
- Take a data set and start involving salmon ecologists and key people, so that by the time of NPAFC workshop, we have a model that to demonstrate at NPAFC and NASCO
  o Start by identifying datasets now.
  o For demonstration purposes, what products are required to wow the audience?
  o When we go to senior managers, may not be the same ‘wow’ factor as the audience
- Gulf of Alaska Expedition has a lot of influence in Pacific Region now, a lot of politicians are interested.
  o That should be one of the datasets to demonstrate the integration that that we are talking about.
- Identify high level stakeholders and assigning who should be talking to them
- Identify research questions and take them to innovation lab in a couple of months
- Salmon activities catalogue (Neo4J) for knowledge management and collaboration support needs to be demonstrated and used.
  o Supports controlled sharing of guarded datasets, e.g., released (visible) only to acknowledged co-authors.
ISDL STRATEGIC PLAN (A WORK IN PROGRESS)

The following is organized as Goals, Quantifiable Objectives (to measure progress toward goals), Strategies (to advance specific objectives specifically and specific goals generally), Projects (actual work to effect one or more strategies), and Experiments (to demonstrate feasibility or working prototypes that would be used in one or more projects). This strategic plan is a web rather than a strict hierarchy, which could be potentially built using DB tools and displayed as a graph.

**Goals**

The following goals are written as qualitative statements about a desired future. These interim goals might be improved by an exercise to reconcile competing scenarios.

1. Information flow has important new efficiencies from new data processing technologies:
   a. data collected and delivered: new data capture, QA, transmission, database management, and QC has replaced old practices;
   b. data rescued: raw data isolated on PC spreadsheets eliminated, all “proprietary” data has sharing deals;
   c. standardized and cleaned: parts, functions, and practices related to salmon have been formally described and named, all incoherent names and formats have been replaced;
   d. integrated: multiple data sources are easily combined for wide and deep analyses; and
   e. new analyses and visualization: the context for salmon data is readily available.

2. Decisions and policies are faster and smarter:
   a. new and more effective decision support products are routinely produced; and
   b. improved practices for communication and knowledge transfer have been widely adopted.

3. Implementation of new technologies and practices was swift and widespread:
   a. new tools were easy to use and reapply: enabled by GUIs, workflows, and automation; and
   b. new tools are widely used by salmon ecologists; the result of valuable examples (see “experiments”) and access to user support services.

**Quantifiable Objectives**

These are numerical (including binary) metrics of success, of progress toward goals and intended outcomes. Targets are objectives with dates and assume resources. Individual technical indicators can be combined into progress indices.

1. Fraction of legacy datasets identified IYS-SST for integration that have been standardized and integrated.
2. Count of salmon ecologists and related database managers who are (a) trained in, and (b) using, various new tools and practices.
3. Count of workflows (automation, pipelines) implemented and in use.
4. Count of new data visualization and decision support products developed and available to ecologists.
5. Count of nodes and links in Salmon Graph. [Salmon Lake?] [involves radical assumptions and goals]
6. Dollars raised for ISDL, including equivalent for “support in kind”: labour, participation, waived licenses.

**Strategies**

How, generally, will goals be attained?

1. Capture field data in the cloud: G1a; G3.
2. Integration via neo4j graph database: G1b, d.
3. Workflow for automation: G1a, b, d, e; G3a.
4. Cogent examples, easily reapplied: G1d, e; G2a, b; G3b

**Projects**

A. Funding and Related Support

The success of ISDL will require focus, acceptance, and support from program directors and managers in order for this revolution to advance out of the hobbyist stage. Consolidating and transforming data into a central repository is a resource intensive activity on its own. Where and how these resources are allocated was not made
clear at the workshop. Further work in terms of extracting value from the consolidated data will occasionally require significant effort from hobbyists, along with the proper allocation of resources for the practice to be sustained.

In short, revolutions need resources, which is another way of saying that funding is required to move this project forward in any discernible timeline. All available personnel should discuss and plan how to fund this activity.

1. Prepare cost projections.
2. Build a strategy to chase funding.
   a. Identify sources
3. Prepare a business plan and work plans suitable for funding agencies.
   Action: Mark Saunders and Scott Akenhead. Lay out the long-term goals and objectives of ISDL project.
   Audience includes philanthropic foundations and corporations.
   Action: John Song and Scott Akenhead. Draft and circulate a work plan for the first year, including deliverables and budget.

B. Salmon Schema

Continuous development of the salmon database scheme for neo4j. This provides an integrated salmon database for all relevant IYS initiatives. Initial node types and content, and initial link types and content, are used to implement and initial graph database. John Song demonstrated important initial progress at implementing the Darwin Core schema as a foundation for this schema during ISDL workshop, and describe “the God schema” as an outcome as opposed to a barrier.

4. Document the neo4j instance build by John Song. NuSEDS and NPAFC catch were converted to neo4j graphs and linked, producing > 1M nodes. This was a convincing example.
5. Continue building the Salmon Schema by meeting needs while ingesting more datasets. This exercise will inform areas where the schema needs updates, reconsideration, and expansion beyond the Core Darwin schema.
6. Recast Darwin Core as a graph with many of the Darwin Core fields represented as nodes linked to core nodes with the original record names.

C. Salmon Ideas (Standardized Glossary, RDF Graph, Ontology)

The naming of parts and methods varies within and among salmon data sources. Names of parts are typically column names in spreadsheets, and names of methods would appear in text describing the data in that spreadsheet (metadata). Those names represent ideas, and those ideas are related (nested, cross-linked).

Discovering how things are related by ideas is valuable. This project could, for instance, create standard names to identify and organize the methods employed for estimating spawner abundance; ditto the names for ecosystem functions. A datum about a “salmon” needs a link to one of 15 extant species of *Oncorhynchus*, each properly described and referenced. Linking things to ideas, and ideas to ideas, is an advance on the concept of hyperlinked tags.

Whether or not an Ideas Graph for ISDL should conform to international standards for ontologies to avoid lugubrious conventions and stimulate innovation, is open to debate. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5345125/

7. Examine mobilizing the following glossaries as a start.
   a. https://www.monitoringresources.org/Resources/Glossary/Index
c. http://rs.tdwg.org/dwc/terms/—An experiment has been started with this.

8. Develop a tool to mobilize, expand, organize (link), and describe the Ideas in the salmon glossary. See: https://neo4j.com/blog/neo4j-rdf-graph-database-reasoning-engine/
Action: (not assigned). Build a tool for RDF glossaries that leverages neo4j and related tools and libraries.
Action: (not assigned). Begin assembling a glossary from existing sources and for ideas (names) that are immediately required; perhaps in Google Sheets, preferably in neo4j.

D. Graph Analysis
- AI and ML examples, based on data in the neo4j instance (above). A Neo4j intern to demonstrate the AI and Machine Learning insights that can be derived from the dataset. This will exhibit the differentiating factor that Neo4j prides itself on; it is a database built for analytics. See A. Hodler and M. Needham. 2019. Graph Algorithms. Specific real-world examples:

Action: (not assigned).

E. Visualization
Interactive visualizations that applied the D3 library to data in neo4j were demonstrated by John Song at the ISDL workshop. Sue Grant referenced web-based interactive graphics via R and Shiny, see also Pacific Salmon Explorer.
- Visualization examples
  d. Further application of D3 library to visualization of NuSEDS+NPAFC graph from John Song.
Action: John Song and Scott Akenhead. E.10.a

F. APIs
- Build API requirements for data publication, access, validation, and governance. [DETAILS REQUIRED]
Action: (not assigned)

G. Workflows and Automation
- Prototype workflow examples that will allow data consumers to spend more time using the data rather than prepping the data.
- Review data transformation pipelines that will help data providers publish data to the God Schema.
Action: (not assigned)

H. Training
Graph database training workshops,
- The first is proposed for 2019-03 to allow a progress report to ICES and IYS meetings 2019-05.
Action: Jeff Morris, Scott Akenhead, Matt Deniston, Keith Steele. Organize an Innovation Lab workshop, modeled after Neo4j approach, with research questions posed in ISDL and IYS-SST workshops.
Action: (not assigned). Link with Sue on integrating collaborative process being used in DFO State of Salmon program.
I. Implementation, Institutionalization, Operationalization

Ability to deliver. Includes in-house permission and support for new data processing projects, installation or access to servers and software, adoption of reproducible research, e.g., sharing documented code and sharing (perhaps as deals) standardized datasets. Noted: training for implementation overlaps training for engagement.

• Build an institutional example.
Action: (not assigned)

J. Events

Opportunities to engage more people with more new tools and practices, new datasets, and new problems.

• Relevance. Applies to all events. Ask high-level [sic] people to identify important salmon questions they need answered, before presenting ISDL and IYS-SST to them.
Action: (not assigned)

• ICES salmon workshop, May 2019.
Action: Gerald Chaput. Communicate with Jeff Morris and Scott Akenhead re presentation.

• National Centre for Ecological Analysis and Synthesis.
Action: Mark Saunders, Jim Irvine. Begin a conversation with Matt Jones (NCEAS) about outcomes of ISDL meeting indicating an opportunity to add value to SASAP workshops.

Action: Scott Akenhead, Mark Saunders, John Song, others. Develop ISDL presentation and demos.
Action: Mark Saunders. Will NCEAS support travel for some people to Portland?

• NASCO workshop in Tromsø, Norway, June 2019.

• ICES Science Conference, Sweden, September 2019.
Action: Scott Akenhead, Catherine Michielsens, others. Present ISDL and available experiments to gain wider engagement internationally and beyond salmon.

K. Experiments

Designed to produce “irresistible examples” that provoke widespread acceptance of the new tools and practices involved. Intention is proof-of-concept demonstrations and “working prototypes” that apply examples of data, as opposed to complete solutions implemented with extensive data.

IYS Cruise Data. Connect with the research scientists from the Gulf of Alaska cruise when they return in March and coordinate with them on linking the data gathered into the ISDL. People in the North Pacific are very excited about the cruise. The faster we can gather insights from this, the more we will be able to capitalize on the attention.
Action: Mark Saunders to investigate and coordinate.

Flagship dataset. Determine one data set to start presenting examples of ISDL concepts and potential to other groups. Goals: (a) funding, and (b) start involving key people (resource managers, etc.) in the development process. Leverage the neo4j salmon graph started by John Song via cloud-based neo4j database server provided by Jeff Morris. That graph links catch data from NPAFC website and spawner abundance and metadata from DFO NuSEDS (OpenGovernment.ca)

Action: (not assigned)
**Discovery.** Develop interesting ways to interrogating the Salmon Graph. Intention is “wow the audience.”

Need to establish the audience and cater this presentation to them. What will we say to them?

**Action:** (not assigned)

**Insight.** Demonstrate and example of data integration and visualization (“some synthesis”) that provides, insight, progress, or answers to one or more important questions that ecologists have not been able to address. The idea is to assemble a small number of legacy data sets that have been rated for accuracy and precision into “the graph database world.”

e. Assemble and integrate (the term “synthesize” was used incorrectly in the workshop) the available data from the IYS-SST workshops to illustrate trends and themes [sic] between basins. Show how salmon in the North Pacific and North Atlantic both exhibit the 1990s event, a drop-in salmon marine survival and similar surrounding trends.

f. Show how 1990s event was correlated with synchronous basin scale habitat changes: warm intrusions from the south in Northeast Pacific and cold intrusions from the north in Northwest Atlantic.

**Action:** (not assigned)

**Quality Control.** Salmon ecologists emphasized that data integration—John Song’s example was impressive but worrying—must be preceded by processing to exclude or flag data that is likely to be misleading or misinterpreted. Appropriately filtered data (i.e., with appropriate metadata) that showed connections between data holders and other nodes, as can be done with neo4j, “would be a great product that could form the skeleton of an International Salmon Database that could be built upon over the next several years.” Quality control is an issue with datasets (NuSeds) and thereby limits the value of Pacific Salmon Explorer, suggesting collaboration. An example of a database where the data have been categorized by quality is the Pacific Salmon recruits/spawner data on Open Government, where a simple algorithm was used to categorize data quality data based on the various types of input data. That database has not been updated since it was developed.

**Action:** (not assigned)

**FOR MORE INFORMATION**

Scott Akenhead Scott@S4S.com 1.250.210.4410

See also

https://salmondatalab.slack.com/
https://github.com/int-salmon-data-lab
Presentation of ISDL at GraphConnect 2018, New York
## APPENDIX 1. ATTENDEES

<table>
<thead>
<tr>
<th>#</th>
<th>Person</th>
<th>Interest</th>
<th>Contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mark Saunders, Director, IYS – North Pacific Region, Vancouver, BC</td>
<td>International Year of the Salmon</td>
<td><a href="mailto:msaunders@yearofthesalmon.org">msaunders@yearofthesalmon.org</a> 1.778.928.3474</td>
</tr>
<tr>
<td>2</td>
<td>Jeff Morris, Head, Product Marketing, Neo4j, San Mateo, CA</td>
<td>Adoption of neo4j graph database approaches and technologies</td>
<td><a href="mailto:jeff@neo4j.com">jeff@neo4j.com</a> <a href="http://www.neo4j.com">www.neo4j.com</a> 1.415.260.5325</td>
</tr>
<tr>
<td>3</td>
<td>John Song, CTO, Systum Inc., Los Angeles, CA</td>
<td>Data systems architecture and user interfaces.</td>
<td><a href="mailto:jsong222@gmail.com">jsong222@gmail.com</a> systum.com</td>
</tr>
<tr>
<td>4</td>
<td>Matt Deniston, CEO, Sitka Technologies Group, Portland, OR</td>
<td>Technology and services support for environmental programs.</td>
<td><a href="mailto:Matt@sitkatech.com">Matt@sitkatech.com</a> sitkatech.com</td>
</tr>
<tr>
<td>5</td>
<td>Keith Steele, Systems Architect, Sitka Technologies Group, Portland OR</td>
<td>Data systems architecture and user interfaces for environmental data.</td>
<td><a href="mailto:Keith@sitkatech.com">Keith@sitkatech.com</a> sitkatech.com</td>
</tr>
<tr>
<td>6</td>
<td>Kelly Chapman, Vancouver Island University, Powell River, BC (missing in the photo)</td>
<td>Knowledge transfer; wicked problems.</td>
<td><a href="mailto:Kelly.Chapman@gmail.com">Kelly.Chapman@gmail.com</a></td>
</tr>
<tr>
<td>7</td>
<td>Scott Akenhead, Convenor, ISDL, Pacific Biological Station, Nanaimo, BC</td>
<td>Data integration to enable advanced statistical models.</td>
<td><a href="mailto:Scott@s4s.com">Scott@s4s.com</a></td>
</tr>
<tr>
<td>8</td>
<td>Bruce Patten, Fishery &amp; Assessment Data, DFO, Pacific Biological Station, Nanaimo, BC</td>
<td>Fisheries and ecology data processing in DFO</td>
<td><a href="mailto:Bruce.Patten@dfo-mpo.gc.ca">Bruce.Patten@dfo-mpo.gc.ca</a> DFO Pacific</td>
</tr>
<tr>
<td>9</td>
<td>Jim Irvine, Research Scientist, DFO, Pacific Biological Station, Nanaimo, BC</td>
<td>IYS Salmon Status and Trends workshops. LINK</td>
<td><a href="mailto:James.Irvine@dfo-mpo.gc.ca">James.Irvine@dfo-mpo.gc.ca</a> DFO Pacific</td>
</tr>
<tr>
<td>10</td>
<td>Kim Hyatt, Research Scientist, DFO, Pacific Biological Station, Nanaimo, BC</td>
<td>Extrapolating from intensively studied salmon populations to many salmon populations. LINK</td>
<td><a href="mailto:Kim.Hyatt@dfo-mpo.gc.ca">Kim.Hyatt@dfo-mpo.gc.ca</a> DFO Pacific</td>
</tr>
<tr>
<td>11</td>
<td>Gérald Chaput, Research Scientist, DFO, Gulf Fisheries Centre, Moncton, NB</td>
<td>Canadian Atlantic Salmon. LINK</td>
<td><a href="mailto:Gerald.Chaput@dfo-mpo.gc.ca">Gerald.Chaput@dfo-mpo.gc.ca</a> DFO Gulf Region</td>
</tr>
<tr>
<td>12</td>
<td>Sue Grant, Program Head, State of the Salmon Program, DFO, Vancouver, BC</td>
<td>New tools for interactive analysis and visualization of salmon data</td>
<td><a href="mailto:Sue.Grant@dfo-mpo.gc.ca">Sue.Grant@dfo-mpo.gc.ca</a> DFO Pacific</td>
</tr>
<tr>
<td>13</td>
<td>Catherine Michielsens, Pacific Salmon Commission, Vancouver, BC</td>
<td>Advanced statistical models for salmon population dynamics and fisheries forecasting</td>
<td><a href="mailto:Michielsens@psc.org">Michielsens@psc.org</a></td>
</tr>
<tr>
<td>14</td>
<td>Andrew Munro, Fisheries Scientist, Alaska Dept, Fish &amp; Game, Anchorage, AK</td>
<td>Southern Alaska Salmon and People: an experiment in salmon data assembly alaskasalmonandpeople.org</td>
<td><a href="mailto:Andrew.Munro@Alaska.gov">Andrew.Munro@Alaska.gov</a> 1.907.267.2260</td>
</tr>
<tr>
<td>15</td>
<td>Nathan Bendriem, North Pacific Anadromous Fish Commission, Vancouver, BC</td>
<td>International Year of the Salmon IYS</td>
<td><a href="mailto:NBendriem@yearofthesalmon.org">NBendriem@yearofthesalmon.org</a></td>
</tr>
<tr>
<td>16</td>
<td>Stephanie Taylor, North Pacific Anadromous Fish Commission, Vancouver, BC</td>
<td>International Year of the Salmon IYS</td>
<td><a href="mailto:STaylor@npafc.org">STaylor@npafc.org</a></td>
</tr>
<tr>
<td>17</td>
<td>Bruce Baxter, Fishery &amp; Assessment Data, DFO, Pacific Biological Station, Nanaimo, BC</td>
<td>Fisheries databases in DFO</td>
<td><a href="mailto:Bruce.Baxter@dfo-mpo.gc.ca">Bruce.Baxter@dfo-mpo.gc.ca</a> DFO Pacific</td>
</tr>
<tr>
<td>18</td>
<td>Shlee Hamilton, Fishery &amp; Assessment Data, DFO, Pacific Biological Station, Nanaimo, BC</td>
<td>Fisheries databases in DFO</td>
<td><a href="mailto:Shlee.Hamilton@dfo-mpo.gc.ca">Shlee.Hamilton@dfo-mpo.gc.ca</a> DFO Pacific</td>
</tr>
<tr>
<td>#</td>
<td>Person</td>
<td>Interest</td>
<td>Contact</td>
</tr>
<tr>
<td>----</td>
<td>---------------------------------------------</td>
<td>-------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>19</td>
<td>Steve Schut</td>
<td>Fisheries databases in DFO</td>
<td><a href="mailto:Steve.Schut@dfo-mpo.gc.ca">Steve.Schut@dfo-mpo.gc.ca</a></td>
</tr>
<tr>
<td></td>
<td>Fishery &amp; Assessment Data, DFO</td>
<td></td>
<td>DFO Pacific</td>
</tr>
<tr>
<td></td>
<td>Pacific Biological Station, Nanaimo, BC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Jason Parsley</td>
<td>Fisheries databases in DFO</td>
<td><a href="mailto:Jason.Parsley@dfo-mpo.gc.ca">Jason.Parsley@dfo-mpo.gc.ca</a></td>
</tr>
<tr>
<td></td>
<td>Fishery &amp; Assessment Data, DFO</td>
<td></td>
<td>DFO Pacific</td>
</tr>
<tr>
<td></td>
<td>Pacific Biological Station, Nanaimo, BC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### APPENDIX 2. PRESENTATIONS

<table>
<thead>
<tr>
<th>Information Step</th>
<th>Presenter</th>
<th>PowerPoint</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Introduction</strong></td>
<td>Mark Saunders</td>
<td>IYS Objectives, Participation, Highlights, and Timeline.</td>
</tr>
<tr>
<td></td>
<td>Kelly Chapman</td>
<td>A Précis of the Preceding IYS Workshop: <em>Salmon Status and Trends</em></td>
</tr>
<tr>
<td></td>
<td>Scott Akenhead</td>
<td>“A Vision for ISDL”</td>
</tr>
<tr>
<td>Step 1. Data Capture</td>
<td>Matt Deniston and Keith Steele</td>
<td>“Dynamic Data Management in The Cloud”</td>
</tr>
<tr>
<td>Step 2. Data Assembly</td>
<td>Bruce Patten</td>
<td>“An Overview of Salmon Data Processing in DFO Pacific Region.”</td>
</tr>
<tr>
<td></td>
<td>Jeff Morris</td>
<td>“Neo4j Uniquely Enables Data Integration and Modern Analyses.”</td>
</tr>
<tr>
<td>Step 4. Analysis</td>
<td>Sue Grant</td>
<td>“An Introduction to DFO’s State of the Salmon (SOS) Program.”</td>
</tr>
<tr>
<td>Step 5. Communication</td>
<td>Kelly Chapman</td>
<td>“Effective Communication between Ecologists and Decision-Makers.”</td>
</tr>
</tbody>
</table>