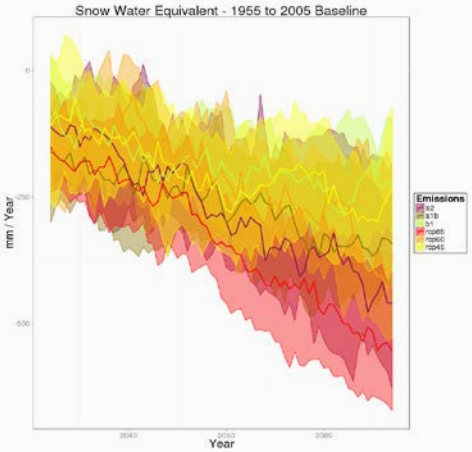
LANDSCAPE
CONSERVATION
COOPERATIVES



Landscape Conservation Cooperative Integrated Data Management Network

February 2015

Summary

The Landscape Conservation Cooperative (LCC) Integrated Data Management Network project (IDMN) gathered representatives from over 20 organizations to develop a framework for data management across all Cooperatives. During the two-year project, the group built not only a framework for data management, but also worked with partner projects to implement the core components LCCs identified as technical and support gaps.

The project outputs included:

- a survey of all LCC coordinators on their priority data management needs;
- a project tracking model that allows individual LCCs to either use a set of existing national systems or to tie their system to this national effort;
- a suite of data visualization options along with working examples of each of these options;
- a robust set of software tools that LCCs can use for data lifecycle management, metadata development, project organization, product distribution and visualization, and data analysis;
- an assessment of outreach options that allow LCCs to connect with partners on these capabilities;
- and, generated user training documentation along with a set of recommendations on training approaches.

The products from the IDMN project are currently in everyday use by all LCCs and many partner landscape conservation projects. Enterprise applications working with the IDMN project are cooperating at a level not seen in the past, identifying connection points and avoiding duplication of efforts. Project members provide constant support to LCC project staff, coordinators, and principal investigators alike. Finally, the next generation of tools being developed by partner organizations all have LCC needs and user training as priorities, thanks to outreach by IDMN team members.

About This Document

The Landscape Conservation Cooperative (LCC) Network Office funded the Integrated Data Management Network (IDMN) project to develop an information management framework that all LCCs could use. This IDMN report is the result of that effort, synthesizing the work of six focus groups who addressed:

- the current state of the LCC information management landscape (Survey and Synthesis);
- project tracking options and recommendations (Project Tracking);
- different ways to present data developed by the LCCs (Visualization);
- different data management software tools and software development best practices available to the LCCs (Tools);
- outreach options to communicate these findings to LCC partners (Outreach);
- and, the best ways to support users working within this framework (User Training).

This project report contains the collective work of the IDMN focus groups. Each group used their judgment in preparing the individual reports. An editor consolidated the team reports into one master report, based on style rules developed by a subset of the groups. However, the voice, structure, and content of each section are author-based.

Each group generated a summary and recommendations section for the body of this report. The appendices contain a detailed description of each working group's work.

Printing

- To view the report summary, print only the first 17 pages of the document.

Title Page Image Credits

- Rocky Mountain National Park (USGS/K Walters)
- Modeling of Watershed Systems graphical output (USGS/Fort Collins Science Center)
- Resource for Advanced Modeling Working Group (USGS/Fort Collins Science Center)

Background

The IDMN project worked with 20+ organizations over 2 years to develop data management concepts, support technologies, and collaborative environments available to all LCCs.

Data management is an obvious need for a successful project. Researchers need to ensure their products are documented and presented in a way that allows other researchers and decision makers to use these products. However, most efforts at data management are long on policy and short on mechanics. How researchers ensure use of their products requires technology and people who support the use of that technology. These three connected concepts (policies that consider proven technical options, technology supported by people, people who can influence policy based on their experiences) provide the backbone to effective data management.

The Landscape Conservation Cooperative (LCC) Integrated Data Management Network (IDMN) worked with over 20 organizations over two years to bring coherence to that fractionated information management landscape. Specifically, the IDMN Network tried to address ways LCC partners implemented the basic building blocks of data management. Issues addressed included building and sharing science products with partners, securely storing those data for the long term, and evaluating ways to get those outputs to cooperators and eventually the public.

Over the course of the IDMN project, the scope was expanded to address ways to track projects that produced scientific outputs, as well as build communications products and tools to help community members work together more efficiently. These efforts considered and learned from other landscape level cooperative efforts, especially the National Climate Change and Wildlife Science Center network of Climate Science Centers (CSCs).

The IDMN project assembled six teams of information management, natural resources, and communications specialists to evaluate and test different approaches. From this effort, the teams defined options for different LCCs and other landscape-level project teams to not just coexist, but also constructively work together.

One critical element of the IDMN project was an exploration of partnerships organized around data sharing and integration. Concepts included facilitating the use of new integration options, collaborating on guidance documentation, and fostering hardware and software solution partnerships among LCCs and CSCs. The integration pathways focused on:

- assessing user priorities;
- utilizing metadata and data standards to help groups work together in specific domains;
- discovering and building tools that let projects take advantage of partner capabilities;
- assessing how groups could visualize outputs;
- and, promoting concepts to help users with project tracking and software tools.

The IDMN project ended in early FY15, but the team efforts continue. Thanks to the collaborations built as a part of this project, the LCCs are seeing more and more linkages among data management and integration professionals. It is clear that the efforts of the IDMN project have made an impact on every agency and organization working with the Landscape Conservation Cooperative network.

Survey and Synthesis

The IDMN project developed a survey to collect comments from 22 LCCs and then used those survey results to identify the priority data management challenges faced by LCC coordinators.

Support for data management must be based on the issues, the data, and the science that LCC data and science coordinators consider high priorities. The IDMN project implemented the Survey and Synthesis Team to determine these LCC priorities. This team surveyed all LCC data and science coordinators, receiving responses from representatives from 22 LCCs. The IDMN project then used these findings to inform the development of other IDMN focus groups.

The respondents identified the following needs to ensure they could successfully manage data acquired as a part of funded projects:

- Region-based or national LCC data portals, with user support
- Caching data on ScienceBase, Data Basin, and Geo Data Portal
- Tool that "harvests" metadata and data from ScienceBase into Data Basin
- LCC-branded data.gov presence for all LCCs populated by ScienceBase services (so we don't have to worry about this as individual LCCs)
- Mechanisms to push data from ScienceBase/LC MAP to data.gov
- Metadata conversion and import tools
- Tool to simplify project metadata authoring, especially ISO 19115 metadata
- Project tracking for active projects
- Network site for posting finished LCC projects
- Network project and data metadata standards
- Data Management Plan Editor
- Options for different LCCs to effectively share documents with LCC stakeholders
- A platform to conduct and record webinars

Geospatial data management received the bulk of attention from LCCs. However, the coordinators also identified the need to manage other forms of data, including tabular data, reports, and other communications. Additional data management issues identified included: the lack of cataloging, archival, and web service tools that facilitate awareness and access to this

content; training in the use of these tools; and consistent ways to discover and use published online content.

The Survey and Synthesis Team also documented actions that LCC coordinators wanted to avoid. They were especially sensitive toward imposing any "one-size-fits-all" or "one database to rule them all" data management solutions. Coordinators also did not want to work with data portals that failed to address common metadata standards, or work with data portals that limited acceptable data formats.

Thanks to the efforts of the Survey and Synthesis Team the IDMN project not only developed focus groups addressing the topics of most interest to LCC coordinators, but also organized solutions to many of the identified high priorities.

Project Tracking

The IDMN project worked with multiple LCC project tracking systems to ensure both a national project reporting capability and options for LCCs to maintain control over the system they used.

Data management starts at the project level. For LCCs, all products, data and publications alike, stem from a project designed to address cooperative science needs. Accomplishment reporting must always link back to the project that delivered the outputs.

The LCC data and science coordinators identified consistent and coherent project tracking as a priority need. However, these same coordinators also insisted on local control over the systems they could use.

The IDMN Project Tracking Team worked with all LCCs to develop a project metadata model – a list of elements that all LCCs collect, reported at a national level. Because each LCC has a unique character, the systems designed by the Project Tracking Team allowed local customization of the project metadata collected. The Team developed ways for each LCC to work with their existing project tracking system while still providing an LCC-wide picture of funded projects. The Team also introduced each LCC to already-developed options that would make project tracking relatively easy.

Over the course of the IDMN project, a number of LCCs migrated away from custom solutions and started working with some of the options presented by the Team. The most common systems adopted by LCCs were:

- The Simple National Project Catalog (SNPC)
- ScienceBase Project Management Tools
- Arctic LCC Project Tracking System (PTS)
- Great Northern Project Tracking System

Since none of the existing reviewed databases fully fit all specified needs of every LCC, the Team focused on building an exchange system. This involved defining and prototyping an aggregation mechanism for basic “project catalog” metadata collected from all LCCs. The Team then identified a range of software tools and approaches that supported the core project catalog

metadata and added support for more extensive “project tracking” functionality. The overall goal was to maintain a way to share common elements while allowing for different software choices by LCCs.

The effort successfully tested this metadata exchange model, aggregating information from the four primary systems listed earlier. In addition, the Team convinced the ScienceBase product owner to add LCC extensions to its project tracking toolset. The resulting workflow now allows LCCs to initiate project records in any of these systems and then build products tied to those projects in ScienceBase.

Even with these demonstrated successes, challenges still exist. The Team identified a number of potential issues with the project tracking workflow, including the potential for duplicate records among the various systems and uncertainty over how to treat projects with multiple funding sources. The California LCC, the leader of the project tracking effort, is researching solutions to these issues with the cooperation of the US Geological Survey, Arctic LCC, and Great Northern LCC. As technology evolves over the coming years, the extended team will continue collectively to evaluate innovative ways to address these and other project tracking challenges.

Visualization

The IDMN project identified data visualization strategies for LCC-sponsored projects, as well as provided examples of geospatial data expression by several LCCs.

Most data visualization in the conservation world ends up as maps. Fairly simple, on-line maps can express multiple conservation concepts. And while simplicity may be the ultimate sophistication, some projects need more complex map visualizations. For example, some data integration efforts have sophisticated audiences and can accommodate more open-ended approaches to visualizing data.

This IDMN Visualization team reviewed and documented simple, moderate, and complex map visualizations used by conservation professionals and the public alike. The Team's full report delivers multiple products and assessments, including a matrix that allows coordinators to evaluate each visualization options against a multitude of different viewer requirements. The team also documented exceptions and nuances in the three archetype viewers (simple, moderate, and complex) and reviewed some of the rapidly evolving technical areas associated with visualization: more sophisticated handling of larger datasets; spatial data animations; handling of time series more effectively; enhanced integration of other graphic elements into the mapping display; and more sophisticated collaboration tools.

The team reviewed how Data Basin and ScienceBase could work together to provide unique visualization options. These efforts paid dividends with the implementation of multiple integrated LCC data catalogs. Data and science coordinators no longer have to make a selection among multiple repositories for their project data. Instead, coordinators can feel comfortable that data products can be visualized the way they want, regardless of where those data reside.

The Data Basin-ScienceBase effort provides a model for closer connectivity between the two data systems beyond merely reading external map services. The integration of these systems provides the ability to run a number of real-time analytical tools via the Data Basin front end while accessing datasets stored and managed in ScienceBase. The full integration workflow for delivering visualization on one platform from data residing on a second, third, or fourth system could be explored in future projects using this model.

Tools

The IDMN project not only documented existing tools available to LCC staff but also actively developed new data visualization and analysis capabilities with ScienceBase and Data Basin.

Conservation specialists use on-line tools to access, process, view, and distribute data. The specialists generally employ these tools to inform resource management decisions. However, there are few resources that currently let LCC coordinators and researchers quickly find and learn about tools that are pertinent to their work.

The IDMN Toolshed Team explored a number of tools open to the conservation community and established a web-based Toolshed to help researchers find on-line applications. For the purposes of this report the team defined a tool as: a singular, self-contained piece of software used for data analysis that accepts as input spatial and non-spatial data, applies a series of operations within a defined process, and produces spatial and non-spatial outputs.

One of the clear trends the Team observed when considering requests from LCC coordinators was the need for an efficient and effective mechanism for sharing information about tool needs across the LCC network. An information sharing framework could let conservation researchers and practitioners leverage similar efforts across LCCs, reduce redundant work, and identify national cross-cutting measures.

To complete the evaluation, the Team conducted an in-depth analysis of the details associated with scoping, developing, deploying, and supporting software tools for conservation. The focus on code sharing, project management, frameworks, platforms, integration, and a host of other software lifecycle concerns allows LCC coordinators to make better decisions on whether or not to invest in new custom tools.

The Team also assembled a list of recommendations for the LCCs, including: a web-based wiki system that includes an editable and easily maintained listing of tools, along with some descriptive information, potential use cases, and pros and cons. This wiki can serve as a nexus for highlighting tools of particular relevance to the LCC community.

Another recommendation involved creating a forum that LCCs can use for discussing specific tool needs and ideas, limited to the LCC community to avoid potential conflicts with the contracting process.

Finally, the team strongly recommended the LCC network continue to support integration efforts between platforms and tools that are currently being used within the network. This effort should include a venue for direct communication between platform developers and representatives of the LCC community, to ensure that integration efforts address the needs of the LCCs. The IDMN project explored and developed connections among Data Basin, ScienceBase, the Geo Data Portal, the National Climate Change and Wildlife Science Center, Data.Gov, and several other smaller systems. Other integration targets could include: DOI GeoPlatform, USFWS ServCat, BLM Rapid Ecosystem Assessment (REA) system, and the NPS Integrated Resource Management Application (IRMA).

Outreach

The IDMN project developed a series of webinars, presentations, and publications on available tools and best practices as a part of an overarching outreach strategy.

The Integrated Data Management Network (IDMN) project investigated and documented numerous improvements to data analysis, distribution, description, and tracking. However, these improvements obviously cannot help conservation specialists if they are unaware of the efforts. The IDMN project developed an outreach strategy to bridge this information gap between LCCs and users about available products.

The IDMN Outreach Team addressed these needs in four phases:

- promote general awareness of IDMN and products,
- solicit input on IDMN products and services from targeted users,
- promote partnerships and use of IDMN products and services,
- and, train users on IDMN products and services.

During each phase, the Team involved potential users from multiple LCCs to ensure product relevance.

The work of the Outreach Team resulted in over 80 presentations to LCCs and potential user groups, a dozen videos, and three separate webinar series that cataloged the efforts of the IDMN project. The Team closely coordinated with the User Training Team on these activities and helped develop information materials for future outreach and communications efforts with new partners.

User Training

The IDMN project identified areas where user training and support were essential, and then developed materials and support networks to address questions and concerns about these IDMN products.

The conservation community works with a number of project tracking systems, integration tools, and visualization portals. And while experts in each of these systems have little trouble navigating their specialized system, the occasional user can experience difficulty getting started. For that reason, the IDMN established a User Training Team, designed to help users work through the various IDMN partner tools and other data services.

The original IDMN proposal identified the following goals for the User Training Team:

- develop materials (documents, webinars, and videos) to support end users – initially ScienceBase-Data Basin integration;
- and, establish documentation best practices for future LCC products.

The team went well beyond the original goals. They surveyed all LCCs on regularly used systems and tools. The survey suggested that LCC coordinators had deep concerns with the way most systems addressed user training. The team decided that a review of existing systems was necessary to determine the scope of the problem. The extended goals added the following:

- evaluate if training material is available and accessible for the IDMN-affiliated tools;
- develop standard training best practices;
- and, identify future training opportunities.

The User Training Team also built a matrix of available systems, evaluating each system on:

- target audience;
- technical expertise of audience;
- types of user support and training available;
- primary use of the tool;
- training and support challenges;
- and, training material currency.

The team developed a set of training best practices and recommendations for software systems based on these communications:

- design tool in a simple and intuitive manner, so the need for training materials is limited;
- provide step-by-step instructions for use and a description of an application using the tool;
- re-release training materials following the release of a new version of the tool – ensure changes in these versions are noted on the tool's download site and in the included documentation;
- ensure that all training materials are under version control;
- focus on webinars and workshops; use recorded webinars and screencasts to demonstrate the tool and minimize the use of written materials;
- and, test training materials with non-developers – ensure that non-developers can understand and follow instructions, both written documentation and instructional videos.

These product owners also shared how they would address training needs if users have varied levels of expertise, training solutions that have worked best for them, and ideas to improve their current training and support.

The Team noted that these inputs are dynamic. As new technologies become available, training methods will have to change. The Team recommended that the IDMN implement an LCC-wide website, providing access to the tools and training materials identified through the IDMN partnership. This portal would also allow archiving of training webinars and products.

IDMN Team Members

Core Team

- Sean Finn, U.S. Fish and Wildlife Service/Great Northern LCC (lead)
- Emily Fort, USGS
- Tim Kern, USGS
- Natalie Latysh, USGS
- BJ Richardson, U.S. Fish and Wildlife Service/North Atlantic LCC
- Jim Strittholt, Conservation Biology Institute

Survey and Synthesis Team

- Ed Laurent, Connecting Conservation (lead)
- Julie Prior-Magee, USGS
- BJ Richardson, U.S. Fish and Wildlife Service/North Atlantic LCC
- Zhahai Stewart, Point Blue/California LCC
- Brendan Ward, Conservation Biology Institute
- Lei Ann Wilson, USGS
- Sean Finn, U.S. Fish and Wildlife Service/Great Northern LCC

Project Tracking Team

- Deanne DiPietro, Point Blue/California LCC (lead)
- Josh Bradley, Arctic LCC
- Emily Fort, USGS
- Sally Holl, USGS/Desert LCC
- Sean Finn, U.S. Fish and Wildlife Service/Great Northern LCC
- Tim Kern, USGS
- Zhahai Stewart, Point Blue/California LCC
- Lei Ann Wilson, USGS

Visualization Team

- Jim Strittholt, Conservation Biology Institute (lead)
- Jennifer Jenkins, U.S. Fish and Wildlife Service
- Ed Laurent, Connecting Conservation
- Dell Long, USGS

- Thomas Miewald, U.S. Fish and Wildlife Service/North Pacific LCC

Tool Team

- Brendan Ward, Conservation Biology Institute (lead)
- Jocelyn Aycrigg, University of Idaho
- David Blodgett, USGS
- Don Brown, USGS
- Sean Finn, U.S. Fish and Wildlife Service/Great Northern LCC
- Matt Heller, U.S. Fish and Wildlife Service/Great Northern LCC
- Amy Keister, U.S. Fish and Wildlife Service
- Natalie Latysh, USGS
- John (Dell) Long, USGS
- Jeff Lonneker, University of Idaho
- Julie Prior-Magee, USGS
- Marc Rempel, Oregon State University Libraries and Press
- BJ Richardson, U.S. Fish and Wildlife Service/North Atlantic LCC
- Leonardo Salas, PointBlue
- Lei Ann Wilson, USGS

Outreach Team

- Janine Salwasser, Institute for Natural Resources at Oregon State University (lead)
- Tosha Comendant, Conservation Biology Institute
- Sean Finn, U.S. Fish and Wildlife Service/Great Northern LCC
- Tim Kern, USGS
- Lei Ann Wilson, USGS

User Training Team

- Tosha Comendant, Conservation Biology Institute (lead)
- Tim Kern, USGS
- Janine Salwasser, Institute for Natural Resources at Oregon State University
- Lei Ann Wilson, USGS

Appendix I – Survey and Synthesis Team Evaluation

No integration project among multiple scientific, academic and land management agencies can succeed without understanding the issues, data, and science that managers consider high priorities. The IDMN project developed a Survey and Synthesis Team to address this information need by implementing findings from this team across project working groups.

Introduction

The LCC Network funded the IDMN in 2012 to address data integration barriers to LCCs and large conservation networks. The IDMN Survey and Synthesis Team grew out of the Knowledge Management working group. The IDMN group recognized the need to identify LCC staff knowledgeable about data management needs and to synthesize their feedback. This team met bi-weekly over several months. The group organized, posted, and archived call notes and other team products at https://griffingroups.com/groups/profile/16351/lcc-idmn-survey-and-synthesis-team.

We targeted an initial survey of high-level questions to LCC Science Coordinators about their LCC's data management activities. In January 2014 the team distributed a survey to LCC Science Coordinators. The survey organized questions into the following sections:

- LCC data management point of contact
- General Questions
- Data Management Plans
- Data Management Policies and Standards
- Project Products:
    - Geospatial Data
    - Tabular Data
    - Documents and Reports
    - Websites
    - Miscellaneous

Representatives from all 22 LCCs responded to the survey. The questions and individual responses, as well as tabular and graphical summaries of responses, are archived here: https://griffingroups.com/file/view/50271/survey-1-responses

After the survey, the IDMN Survey and Synthesis Team continued to meet bi-weekly, reviewing the responses and synthesizing them in terms of integrated data management as an LCC network. This document was then drafted to summarize the findings with reference to specific questions and answers. We also used this opportunity to synthesize the survey results in terms of the state of integrated data management within and among LCCs, as well as offer observations, interpretations and suggestions for steps toward improving integrated data management within and among LCCs. We divided the recommendations into those that are relevant to the IDMN, LCC internal and funded projects, the LCC Network, and LCC partners. When possible, we also tagged recommendations with IDMN Team names in case these teams want to pursue them. This survey summary and synthesis report provided the IDMN Core Team distribution and use suggestions.

Summary Findings

General information

- Most LCC projects collect or produce data (Q5).
- LCCs want to be enabled to work together and not forced down any particular path. However, there is strong interest in guidance from the LCC Network regarding recommendations for what tools to use (Q52, Q53).
- LCCs have limited capacity for data management (Q7); some but not all participate in data management coordination across LCCs such as the LCC Data Management Working Group (Q9).
- Many, but not all LCCs recommend or require data management plans (Q10), but enforcement is difficult due to resource limitations. Several appear to be looking for more guidance regarding standardization of plans and whether or not to require them (Q14). Recommending or requiring data management plans appears to have occurred a few years into funding projects for several LCCs (Q12) rather than from the beginning.
- Each LCC is focusing on its own data management priorities at its own pace and often using its own tools. Most focus is on cataloging, archival, and web services for spatial data layers (Q22). Most LCCs are distributing spatial data via ScienceBase, Data Basin, Geo Data Portal or their websites (Q22); some are not.

    - The mentioned systems (ScienceBase, Data Basin, and Geo Data Portal) are all described in detail in the Glossary.
- The cataloging and digital archival of non-spatial data such as tabular data (Q26) and reports (Q33) is not yet being addressed by most LCCs.
- There are many issues of data integration that go beyond just awareness and access to data. For example, some LCCs have difficulty integrating data due to differences in classification systems (e.g., vegetation (Q45)), conversion of formats (Q46), need for training (Q47), project reporting (Q48), and lack of capacity to explore options and follow provide guidance (Q49, Q51).
- LCCs have not addressed the standardization of tabular data formats (Q28), classification systems (Q45), and grid reference systems (Q24), possibly because most LCCs are working semi-independently.
- Most LCCs are generating non-spatial data (Q25) but there does not seem to be a consistent approach (Q26, Q28, Q33).
- While many metadata creation tools are available standardized metadata formats, or even metadata itself, is often not seen a requirement (Q21).
- Nearly all reports are independently archived (Q33); there is no central repository or system for accessing reports.
- 95% of those polled said some of their projects resulted in websites (Q36), often for the same purposes (Q37). Practices are rarely in place to catalog and archive information posted on these websites (Q38). Follow-up is needed to determine if duplication of effort is occurring.
- Many apps and web tools are in the planning or production stage (Q40). Follow-up is needed to determine if there is coordination or collaboration occurring among LCCs during the design and development phase of these apps and web tools to facilitate their integration among LCCs and partners.
- People are requesting guidance on how to check for compliance of projects (Q17).
- Data review processes are mostly ad hoc (Q41), without guidance for consistency of the data review process (accuracy, content, completeness, metadata).
- LCCs use many different methods for communications (Q43 & 44). Most LCCs use email for communications, which does not support integration. There are also many other ways LCCs communicate. It appears users interpreted this question as "outreach" instead of internal project communication archival and access.

- Most LCCs are not evaluating the use of their products and services (Q47). Some have Google Analytics installed on websites, but follow-up is needed to determine if these web analytics are being analyzed and used for evaluation, and to inform decision-making.
- Several LCC's have stated difficulty in distributing products due to capacity, expertise, and lack of guidance (Q48 & 49).

Requested tools

- Regional based LCC data portals and websites or one national data portal, with support
- Data Management Plan editor
- Metadata conversion and import tools
- Project tracking for active projects
- National site for posting finished LCC projects
- Tools that address the ability to download an original dataset from ScienceBase as opposed to the Service Definition file that generated the web service
- Tool that "harvests" metadata from ScienceBase into Data Basin (already in development)
- Ensure that different LCCs can effectively share documents with LCC stakeholders
- Establish an LCC-branded data.gov presence for all LCCs populated by ScienceBase services (so we don't have to worry about this as individual LCCs)
- A platform to conduct and record webinars (e.g., http://lcc.webex.com) or another platform
- National project and data metadata standards
- Tool to generate metadata, especially ISO 19115/19139 metadata
- Tools to simplify project metadata authoring and setup on LC MAP that would also feed into data.gov (e.g., the ScienceBase Data Entry for Project Tracking and Highlighting, or DEPTH, application)
- Project tracking
- Ability to cache data on Science Base

Requested training

- Existing protocols, methods and storage options that are currently being used
- Metadata, climatology (NetCDF) and other large dataset training

- All topics included in this survey; we need to understand the different data formats listed in this survey, and need a LCC Network standard for all formats
- Geospatial and tabular data formats and standards
- Website product life cycle recommendations
- Tips and tricks for promoting future product integration such as selection of appropriate format standards
- DEPTH, LC MAP, Confluence, Griffin Groups – the greatest barrier is staff time and capacity to explore, there is a need to realize what already exists and how to best employ it and promote its use among PIs
- Metadata training
- Overview of tabular data management
- Clarification of rationale for consistent data management plans

Requested future focus of LCC Network and Partners

- Development and adoption of a centralized data management framework
- Creation of set protocol and standards of workflow from DMP through to data product delivery
- Creation of frameworks and tools that are flexible enough to allow customization, but still offer basic interoperability for "roll-ups" and compilation without database gymnastics
- Migrate lccprojects.org to LC Map; this would help seed project metadata records for each LCC
- Allow the LCC project metadata to be edited using DEPTH Give monthly updates to Science Coordinators answering, "How do I apply IDMN products to my daily workflow?"
- National project and data metadata standards; tool to generate metadata, especially ISO 19115/19139 metadata
- Ways to share data outside the DOI firewall
- Resolve issues of data ownership, licensing, and legal guidance regarding access
- Outreach training and guidance for community on existing tools
- Full life-cycle data management support
- Standardization across the LCCs, including common platforms for delivery
- Helping LCCs recognize and adopt IDMN products and processes

Requested: What should the IDMN NOT focus on?

- Non-DOI issues

- Trying to create a "one-size-fits-all" or "one database to rule them all" data management infrastructure
- Do not create additional data portals that LCCs are required to use without first addressing metadata standards
- Micro-managing data formats

Synthesis

In general, most LCCs appear to either have a strong approach to geospatial data management or have the capacity to develop one via existing tools such as ScienceBase, Data Basin, and Geo Data Portal. Awareness of geospatial data, and often access to the data, are provided through web services. Therefore, due to the abundance of ongoing development and use of existing geospatial data catalogs and portals resources should not be directed to the creation of new LCC-specific geospatial data portals. LCCs should consider whether or not they can enhance existing geospatial data portals instead of developing new ones, unless they have compelling reasons why existing portals are not sufficient.

The presence of tools to provide awareness and access to geospatial data are only one component of integrated data management. For example, several LCCs have expressed a lack of capacity to explore options, learn how to use existing tools, and implement tools as components of data management plans and policies. LCCs expressed a need for assistance in the development and enforcement of data management plans and policies. These groups also need help with the selection and use of tools to comply with these data management plans.

While geospatial data management has received the bulk of attention from LCCs, there is a general need for attention to the management of other forms of data and information such as tabular data, reports, and other communications. LCC staff lack the cataloging, archival, and web service tools that facilitate awareness and access to this content, as well as training in the use of these tools. There are few consistent ways to discover and use published online content without prior knowledge of their existence. Until the LCCs address these concerns, it is likely that much of these data and information may be unnecessarily duplicated or lost over time. Further, the lack of consistency in their management renders the use and integration of content among LCCs as essentially impossible except through individual communications. Our observations and interpretations from these responses, as well as some questions and

recommendations for follow-up, were provided to the IDMN Core team for distribution to other IDMN teams and partners.

To learn more please view the report “IDMN Survey: Observations, Interpretations, Questions, and Recommendations” at https://griffingroups.com/file/view/96107/idmn-survey- observations-interpretations-questions-and-recommendations.

Appendix II – Project Tracking Team Evaluation

Introduction

Building data products is a critical function of any landscape conservation effort. However, without the project context, the full provenance of the product set cannot be understood. To ensure that LCCs had a consistent way to understand the context around data products, the IDMN worked with multiple LCCs and partner projects to generate best practices for, and test implementations of, project tracking systems.

The Project Tracking Working Group addressed these challenges by developing the following objectives:

- Develop and design data exchange architectures that embrace the diversity of methods existing across the LCCs through the use of simple exchange standards
- Address the needs of the LCC Network for certain types of information that the LCCs may not necessarily generate otherwise
- Implement prototypes to refine the protocols and provide reference implementations
- Investigate possible project cataloging or project tracking systems that could be integrated into the overall framework, existing or in development

The team accomplished these goals via the following tasks:

Task 1. Describe a Framework for Project Metadata Exchange

- Develop a needs assessment, starting with Network office's requirements
- Establish an overall architecture
- Establish a content standard, a core set with optional additional complexity
- Establish an exchange format and protocol

Task 2. Review and Research Options for Potential LCC Adoption

- Review existing systems in use; look for applications that LCCs can use
- Provide descriptive information, low-tech versus high-tech requirements, feature comparison chart
- Make recommendations

Task 3. Prototype the Project Metadata Exchange Framework

- Create export functions for one or more existing project cataloging systems
- Develop prototype harvesting function to test and refine the protocol
- Test interoperability of the export and harvesting functions.

Task 4. Report with Recommendations

- Document the data exchange format and protocol
- Report on results of prototyping in task 3
- Recommendations for LCCs implementing their own Project metadata systems
- Recommendations for permanent metadata harvesting implementation
- Description and comparison of existing tools suitable for LCC adoption
- Recommendations for follow-up projects to move into full implementation, with funding estimates
- Recommendations for potential integration with other aspects of the IDMN project

Working Group Process

The team met once per month via GoToMeeting over the 18 month project time period. At the onset of the project we revised the work plan written into the original proposal to better reflect the needs of the LCC, as these needs had become more elucidated in the interim period between writing the proposal and beginning the project.

In order to work efficiently on the tasks, a coordinator was assigned to each task and team members chose to focus on one or more of the tasks. Our method was to check in as a team on all tasks in our monthly meetings, identify next steps and make assignments, work between meetings in our sub-groups, and then reconvene. The sequence of our work was to review needs and existing solutions, develop a draft architecture, begin working on the prototype and data formats, and then revise the architecture from lessons learned.

Review of sources and resources

Existing Project Metadata Catalogs

The technologists who comprise this team have extensive experience in creating and managing metadata catalogs and distributed data exchange systems. Some of us have also worked with project catalogs (such as California's Natural Resource Project Inventory, http://ice.ucdavis.edu/project/nrpi, and existing project records in ScienceBase) and the standards that support them, including national and international metadata standards. None of the existing reviewed databases fully fit the specific needs of the LCC Network. We decided we needed to build a data management and exchange system that was tailored to serve the needs expressed in the user needs assessment.

While building this custom standard and architecture we explored the feasibility of supporting metadata output in the international ISO 19115/19139 standard, using the experience of the Alaska Data Integration Working Group (ADIWG) projects to support this work. The one existing project metadata catalog that was useful to this effort was the preceding LCC project metadata effort, called the Simple National Project Catalog (SNPC).

The California LCC (CA LCC) developed the SNPC in spring of 2013 in response to a request from the LCC Network office. The primary goal of SNPC was to make information describing all the projects funded by the LCCs available for basic review and query to serve the information needs at the Network level.

This effort also accomplished the following:

- An improved understanding of the project information management capacity and needs of the LCCs,
- Progress toward ownership and management by LCCs of their project information shared at the national level,
- A common standard for aggregating project information across systems.

The standard project metadata record used in the SNPC, including the fields and controlled vocabularies, was developed by a collaborative process. We started with the CA LCC's project catalog record and negotiated with the Network office about how to modify it to meet their needs. We then negotiated with all LCCs regarding the fields that they were using or considered

important in the process of developing an LCC-wide standard. The public can access the resulting database at http://lccprojects.org. The LCC Network website also links to this site.

2013 LCC Project Metadata

The main goal of the SNPC effort was to get a cleaned and aggregated All-LCC projects database and website. A secondary goal was to standardize each of the LCC-maintained "master" datasets, whether maintained as a spreadsheet or database, to support an efficient process going forward. At the end of that project each LCC was given their project metadata in a standardized form in an Excel spreadsheet along with instructions about how to update it, giving them a way to respond easily to the next call for project metadata.

Draft Project Metadata Standard

The IDMN adopted the project metadata content standard created by the SNPC effort. That standard consists of 25 data fields, some of which require the use of a controlled vocabulary. The LCCs reviewed and commented on the developed fields and vocabularies. Due to the need to use Excel as a medium for people to enter their information, the content standard created a flat record. This necessitated compound fields that would normally be stored in a separate table in a relational database (example: Funding by Year and Source). The effort handled these inputs with delimiters within a single cell. Upgrading this standard to a more robust solution was a primary concern of the Project Tracking team.

Knowledge about how the LCCs currently manage project metadata

This Project Tracking effort is working forward from the knowledge gained about the LCCs from the SNPC effort. An important piece of information is the manner in which they managed their project metadata records at the time of the SNPC effort, and what they thought they wanted to do to improve upon their methods. The LCC Project Tracking Methods demonstrated that:

- most of the LCCs were using a spreadsheet and of these some of them said they wanted to migrate into another system such as ScienceBase
- several LCCs had developed database systems
- most of the LCCs already had ScienceBase communities

The team concluded that we need to develop a solution that incorporates this diversity of methods, as well as levels of technical capacity, and reflected desire for control and flexibility.

LCC-wide Survey

In August and September 2012, the Working Group developed a survey to understand the current status of, and plans for, the 22 LCCs regarding their data resources, information management, storage and dissemination, and project tracking.

In response to the question “How are you tracking projects” over half of respondents indicated they either had no tracking system (10.5%) or were using spreadsheets (42.1%). The remaining LCCs developed separate custom systems to track project data, including MS Access, SharePoint, and custom web applications. The survey clearly demonstrated the wide range of available information technology capacity and technical expertise among the LCCs.

LCC Needs Statement

From the materials described above, our team developed a Needs Statement expressing the need for project information by both the Network office and the individual LCCs. Thus, we have identified two distinct end-uses for project metadata and the systems that we develop to manage it:

- Cross-LCC: The LCC Network staff, the LCCs, and the public need to discover, explore and summarize the projects across the LCCs,
- Intra-LCC: Each LCC needs to manage its own projects.

The needs can be addressed separately with appropriate data management solutions. This effort does not attempt to address all of these needs but makes progress toward a long-term, robust solution with a solid foundation of design and a working prototype.

Cross-LCC needs for project information includes:

- Ability to see in one place the current comprehensive listing of all the projects being funded by the LCCs
- Provide descriptions to partners, Congress, other agencies
- Support the exchange of information about work being done across LCCs for the purpose of planning and creating partnerships
- Ability to summarize: e.g., reports of amounts and types of funding, grouping of projects into categories
- Find project outputs and additional information from the originating LCC

Intra-LCC needs for project information includes:

- Fiscal: Amounts of funding awarded, awarded purpose, pertinent award dates
- Contractual: Project PI contact information, contracts, project completion status
- Administrative: Project deliverables: their types, expected due dates, deliverable dates
- Public Relations: Present projects on LCC website
- Contribute updated information to a comprehensive listing of projects across LCCs

Additionally, these requirements for the National Project Catalog were defined by the LCC Network office:

- Can be updated by LCCs
- Doesn't present any unnecessary additional burden (avoids duplicate data entry)
- Does not constrain or require a change in software used by LCCs
- Takes advantage of previously collected information

Approach

In recognition of the diversity of LCCs in regard to needs and resources, the team has taken a dual approach. One effort has been to define and prototype an aggregation mechanism for basic “project catalog” metadata collected from all LCCs. The mechanism was kept simple to allow for universal adoption and successful aggregation.

The other main effort is to identify a range of software tools or approaches that support this same core project catalog metadata and in some cases can additionally support more extensive “project tracking” functionality. We intend to support different software choices with the understanding that different LCCs may adopt different tools.

This two-pronged approach attempts to balance the need for consistency across LCCs for core project information with flexibility for each LCC to handle its more elaborate local management needs (if any).

Project Metadata Exchange and Aggregation

Overview

The team considered various overall architectures for creating and maintaining a national aggregation of LCC project metadata. One approach would have been requiring all LCCs to use

the same software, such as Arctic Tracker or ScienceBase, as their master repository. Forcing a choice would not account for the diversity of LCC needs, or the systems they have already built in some cases; it was clear that no one system was going to be suitable for all LCCs. We instead defined an LCC-wide project metadata standard that all LCC-adopted software options could be expected to support. We also suggest optional architectures for aggregating that metadata on the national scale.

An early decision was to focus on a “Project Catalog” set of data designed for both consistent national aggregation, and for utility for the individual LCCs. We did not attempt to specify what additional data or features an LCC could implement. Those additional features could include tracking deliverables, deadlines, reports, contracts, and payments – enhancements we described as a “Project Tracking” system.

We did specify two standardization requirements:

- systems must include all of the defined common Project Catalog fields and vocabularies,
- and, systems must be able to export this subset of their data in a standard format for ingestion in a national aggregation system.

Our proposed aggregation architecture consists of distributed heterogeneous Project Cataloging or Project Tracking systems united by a common core of fields and vocabularies and a data exchange format and protocol for aggregation.

Some LCCs are using a more extensive Project Tracking system already or are poised to do so. A number of LCCs have built project catalogs into the databases underlying their websites or are planning such an implementation. The remaining LCCs use spreadsheets of varying complexity. Our goal was to be able to support all of these implementations, both initially and as LCCs upgrade their technologies.

This approach was inspired in part by the Simple National Project Catalog, which had addressed a similar diversity of LCC technologies and staff resource constraints (very short timescale and budget). The SNPC adopted an Excel spreadsheet format (or equivalent CSV file) as its data exchange format, which included constraining the standardized fields (columns) as well as limiting some fields to controlled vocabularies.

Below is a simplified diagram of the architecture, depicting the master copy managed by the LCC and the derived copy housed in the aggregation database. This also shows options for contribution to the aggregation database:

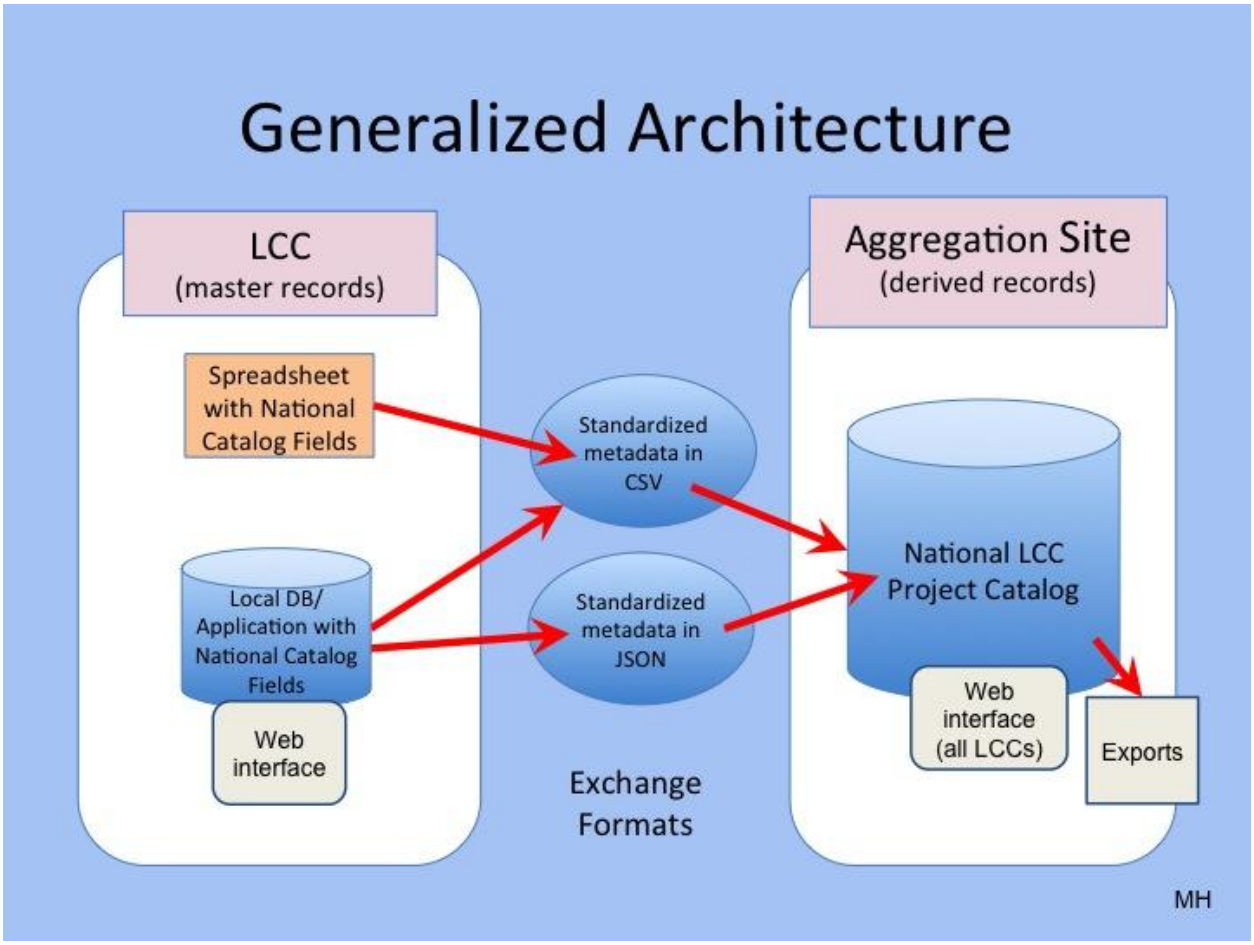


Figure 1. *Generalized architecture of LCC project data, master to derived records in the aggregate database*

A standard for LCC project catalog metadata

The fields and vocabularies adopted by the SNPC resulted from considering existing project cataloging efforts, negotiation with the sponsoring LCC Network office to meet their needs, and individual discussions and negotiations with all 22 LCCs via telephone interviews. As such these fields and vocabularies had a balance of usefulness and feasibility. The IDMN Project Tracking team decided to adopt the same SNPC fields and vocabularies as the basis for our aggregation architecture, although reserving the right to make any well-justified modifications.

One benefit of this decision is that the database of project metadata cleaned and aggregated in 2013 was available to the team for testing and pre-populating the prototype aggregation system. (Populating the prototype with real data had not been a requirement of the IDMN project, so this was a bonus).

Prototyping the Aggregation

The next major decision was choosing a central aggregation system for the prototype implementation. Three main candidates emerged: expanding the functionality of the existing SNPC website (lccprojects.org), adding this functionality to ScienceBase, or expanding Arctic Project Tracker to handle multi-LCC data and aggregation. While the Prototype aggregator implementation was not required to be the same as a final long-term production system, the team felt that there would be some advantage to implementing the Prototype within a candidate production site. ScienceBase emerged as the leading candidate for a long-term home for LCC project metadata aggregation. ScienceBase project staff was also willing to commit staff time to this effort. Based on this we chose ScienceBase as the prototype aggregation site.

The effort needed to define mappings between the LCC project metadata and ScienceBase structures. ScienceBase had existing project records that were not a direct structural or semantic match with the LCC projects, and any changes to accommodate LCC projects would need to avoid breaking existing functionality for other ScienceBase users. (By contrast, lccprojects.org was designed for LCC project metadata and had no other users to consider.)

Selecting ScienceBase as the prototype project metadata aggregator was found to be very useful. The cross-walk between the LCC project records and ScienceBase effort allowed us to discover a number of conceptual issues. Issues were either resolved for both systems or documented for future consideration, allowing a full picture of the production metadata aggregation system built upon ScienceBase. Figure 1 shows how the LCC Project Metadata (LPM) format works with ScienceBase and the Data Entry for Project Tracking and Highlighting (DEPTH) system.

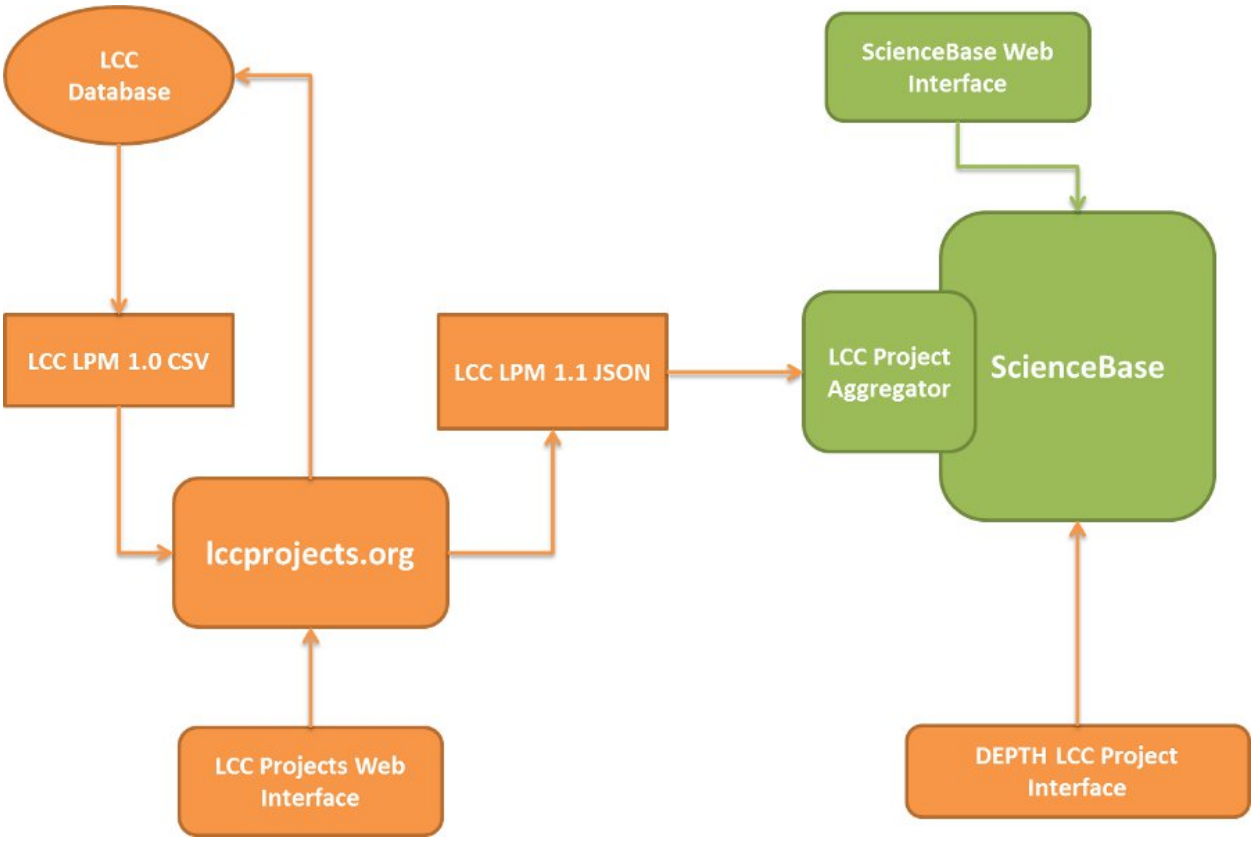


Figure 2. *Elements of the initial prototype*

Maintaining the initial prototype allowed development flexibility. When we defined an alternative JavaScript Object Notation (JSON) format for data exchange (to supplement and hopefully eventually replace the Excel/CSV formats) we were able to modify lccprojects.org to export in the new JSON format in order to smooth the import into ScienceBase for the prototype. The new workflow should result in lccprojects.org being fully supplanted by ScienceBase (Figure 3).

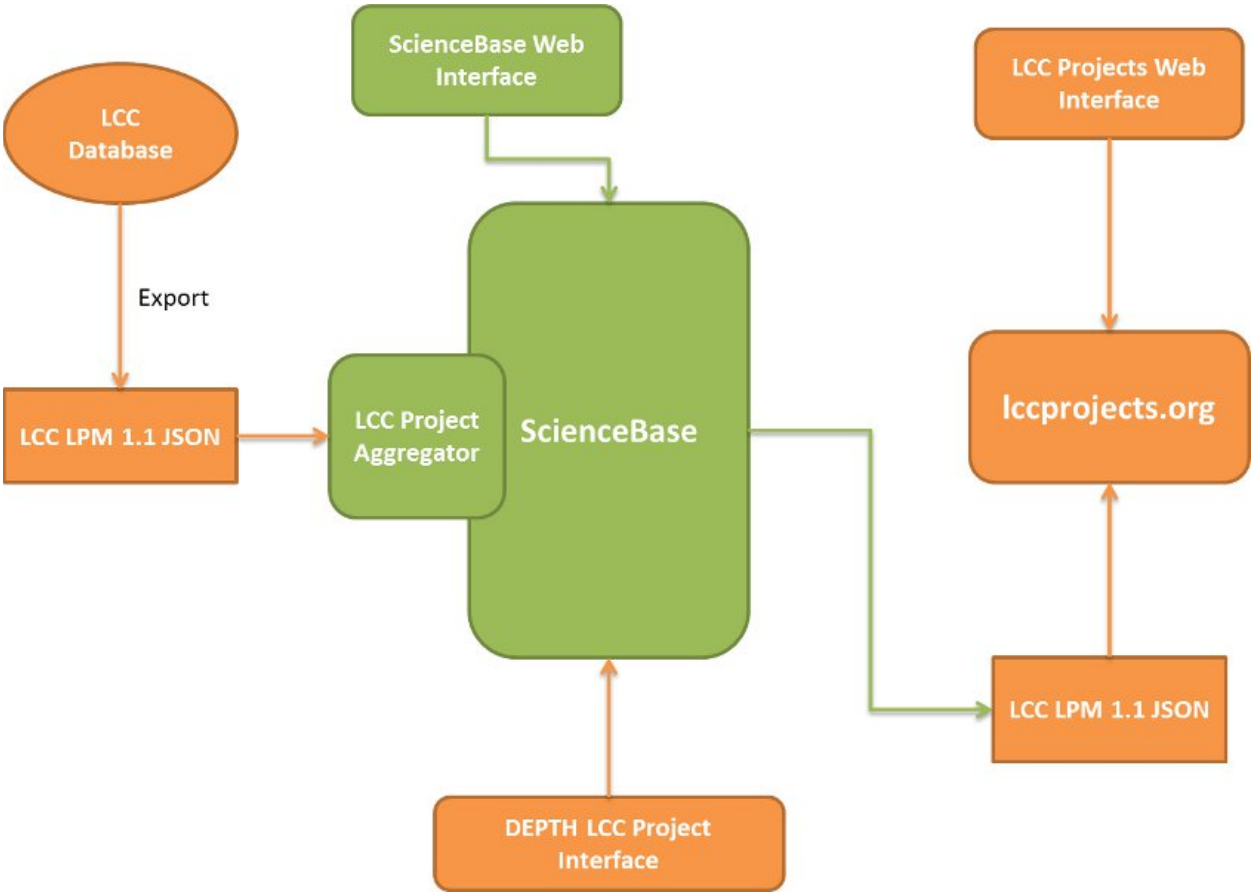


Figure 3. *System elements of interim arrangement*

A few LCCs continue to use spreadsheets rather than project tracking systems. To support these LCCs the production project catalog system will need to accept Excel spreadsheets (or CSV equivalents). LCCs that maintain their project metadata in databases would implement JSON exports for aggregation into ScienceBase (Figure 4).

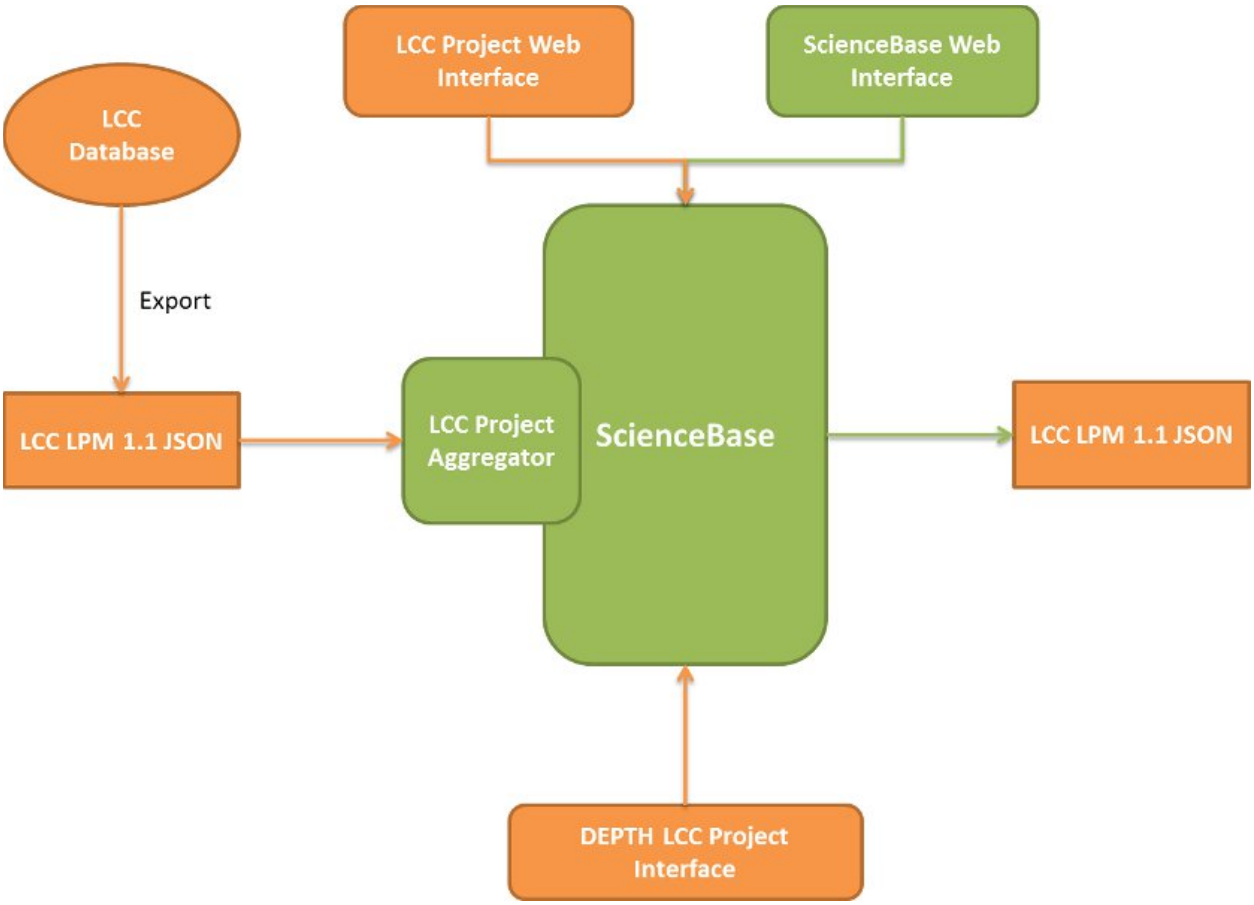


Figure 4. *Ultimate project metadata flow*

Avoiding duplicate or incompatible records

There are several ways duplication of project records can occur in the aggregation database:

- the same project could be reported twice by the same LCC (via two methods, for example),
- multiple LCCs could report the same project due to co-funding,
- both an LCC and CSC could report the same project due to co-funding.

There is also the possibility of the system describing related but distinct projects. Below we discuss the data management implications of each.

To distinguish original metadata from records duplicated from another source, we need to be clear about the concept of a “master” metadata record for each project and “derived” records.

The receiving systems would have to prohibit editing of incoming records to prevent future inputs from over-writing any changes.

The “master” and “derived” configuration, of project records of various systems aggregated into ScienceBase, calls for implementing the Source System of Record, or SSoR, concept. Identification of the authoritative source information system for each record provides a link back to the data source. For example, records created and edited in ScienceBase would have an SSoR of ScienceBase and would be identified as such through ScienceBase tagging. In addition, records imported from an individual LCC’s spreadsheet would have an SSoR of the corresponding spreadsheet. Then, records provided by an individual LCC’s system (i.e., Arctic Tracker) would have an SSoR of the respective LCC’s system.

Tagging ScienceBase project records with the SSoR allows users to aggregate master and derived records, as well as trace back to the authoritative record. The tagging option lets users identify derivations of original items and provides expanded query options on stored records.

Duplicate record: Same project reported by the same LCC

Another complication of using ScienceBase as the prototype national project metadata aggregator is that some LCCs have already entered some or all of their project metadata into the pre-existing ScienceBase project facilities. Some of the earlier entries were not entirely compatible with the LCC project metadata.

For example:

- some LCC project fields are missing in ScienceBase,
- some LCC project fields are similar to but not quite semantically the same as ScienceBase fields,
- the LCC project controlled vocabularies are not presented or enforced, and
- ScienceBase has additional fields (mostly not a problem but possibly confusing)

Ultimately we had to incorporate the concept of a “master” metadata record for each project – i.e. in the system where the record gets created and edited. Copies of this record (or at least of the common standardized national Project Catalog fields) may exist in “derived” records on other systems. Systems need to prohibit editing these records since any changes may be overwritten during future imports.

In the production ScienceBase system, an LCC-specific configuration of DEPTH provides a user interface (web based). This allows LCCs the option of creating and editing their “master” project records there, with the standard LCC fields, descriptions, and enforced controlled vocabularies. There is no need for any import since ScienceBase maintains a fully LCC-compatible project record.

Other LCCs may create and edit their “master” records on other systems – like Arctic Project Tracker, Great Northern Project Tracker, or in their website database. Those systems export a CSV file or JSON file for aggregation and import by ScienceBase. The system must prohibit editing of imported fields since these are “derived” records. It may also be possible for LCCs to use Excel spreadsheets for the master records, imported as derived records in ScienceBase.

This approach of importing project records from only the LCCs with external master systems should avoid this kind of duplicate in ScienceBase (same project as reported by the same LCC).

With the implementation of the production system, the team organized a one-time effort to merge LCC-compatible project and existing non-compatible project metadata records in ScienceBase.

Duplicate record: Same project reported by multiple (co-funding) LCCs

Some projects have been funded by multiple LCCs. In this case, each LCC submits a Project record for aggregation. Individual submissions are appropriate in some contexts. For example, each LCC has its own funding metadata. Records described in the system as “per LCC project reports,” have multiple records for the same project, since each LCC does independent reporting.

Project managers need to know whether several project records do indeed refer to “the same project” (as reported by different funders). In terms of what we mean by the same project, suppose it has the same proposal, timeframe, and deliverables. Two projects that use the same techniques to produce data for different geographic regions would have different deliverables and thus not be “the same project.” How best to deal with this overlap (through merging, or cross reference, or other methods) would require further study, beyond the scope of this team.

Duplicate record: Same project reported by (co-funding) LCC and CSC

The aggregator for LCC project metadata system, ScienceBase, also contains projects from the regional Climate Science Centers. LCCs sometimes co-fund CSC projects, resulting in projects with both LCC and CSC records in ScienceBase. In this case, however, the emphasized fields and vocabularies for CSCs and LCCs can differ, as well as the emphasis in the proposals, so the records would not necessarily be the same even in an ideal world. Nevertheless, there could be some utility in identifying that CSC and LCC records refer to “the same project” in some meaningful sense.

Duplicate record: Related but distinct projects

It is worth mentioning that sometimes an LCC funded “project” is a facet of a larger overall project. For example, an LCC might fund a $100,000 two-year offshoot or extension that is part of a larger $500,000 five-year project mostly funded by other sources. In this case the LCC funded “project” would be distinct from the overall project or other independent facets of that overall project. These separate items would present no duplication issue. Tracking relationships between these “different projects” is probably not needed by our metadata system.

Data Transfer Mechanisms

Our goal has been to define the aggregation methods for project metadata from LCCs (who maintain their own master records). We believe it would be useful if the production system were able to handle both uploads and harvesting. The upload scenario envisions an LCC data manager logging into the aggregation website, and uploading the latest snapshot of their LCC’s project metadata in one of the accepted standard formats: Spreadsheet, CSV or JSON. The harvest scenario involves the aggregation website automatically “fetching” data over the Internet from participating LCCs; this could involve retrieving a previously generated file from a web accessible folder, or it could involve accessing a web service which generates the data dynamically from a database. The result should be the same.

Supporting LCC Software Adoption

Software Option Review

No one software solution will fit the diverse needs of every LCC. Because of this diversity this project tracking document the project tracking software options were researched to meet the needs of a network system and the individual LCCs. To support individual LCC software

adoption, we considered all project tracking options through a SNPC survey. The final network aggregate recommendation is designed to work with a variety of individual LCC master database solutions so LCCs can use existing software and methodologies, entirely new options, or a blend of existing and new approaches.

Five systems used by individual LCCs were considered for the network aggregate architecture. We recognize that some LCCs are happy with the solutions they currently have and may not be interested in the five offered solutions. We feel confident these LCCs will be able to contribute with exports of their data to the SNPC as they have already done. The five software options listed below support the Network metadata compatibility standard.

Software Options Descriptive Information

SNPC spreadsheet: As part of the 2013 "Simple National Project Catalog" the most current information about all LCC projects was cleaned and aggregated into a combined national catalog. The updated data was sent back to each LCC (when appropriate) as a spreadsheet that could be maintained by the LCC and later reimported. This option assesses using that spreadsheet (or a version with additional local columns) as the master project catalog for a given LCC. Various LCCs that have their own project tracking systems can export data to this format as an intermediate step for national data submittals.

ScienceBase: ScienceBase has numerous project tracking and description capabilities. As part of the IDMN project, ScienceBase offers capabilities compatible with the LCC Network project catalog metadata standards. This means the system could be used by LCCs to comply with the network aggregation. ScienceBase has a front end data entry website, called DEPTH, to aid in the data entry of project information.

Arctic LCC Tracker PTS: Arctic Project Tracker is a full-featured project tracking system (PTS) implemented as a stand-alone tool for use by LCC staff. It was designed specifically to suit the needs and workflow of the Arctic LCC, but the codebase could be cloned and modified for use by other LCCs.

GNLCC Project Tracker PTS: Great Northern has begun implementing a desktop project tracker built on Microsoft Access with the database design modeled after the (web based) Arctic

Project Tracker and utilization of ScienceBase. This application was designed to meet the needs of the Great Northern LCC, taking into account the projected needs of other LCCs.

Customized Content Management Systems (CMS): Public facing web based catalog of datasets, documents, web resources and LCC projects can be developed by LCCs with their respective CMS (Content Management System). Project Catalog functionality using CMS technology has been incorporated into various LCC portals but was not necessarily designed to be portable to other LCCs.

	SNPC Spreadsheet	ScienceBase (DEPTH)	Arctic PTS	Great Northern PTS	Custom CMS
Public Web Display	No	Yes	No	No	Yes
Web Data Entry	No	Yes	No	No	Possible
Technical Expertise Needed for Administration	None	None	High	Low	High
Project Tracking	No	Yes	Yes	Yes	Possible
Customization Options (data, Reporting, etc.)	No	Yes	Yes	Yes	Possible
Automated National Data Integration	No	Yes	Planned	Planned	Possible
Ongoing Development	No	Yes	Yes	Yes	N/A
Contact Name	Zhahai Stewart, Deanne DiPietro	Lei Ann Wilson	Josh Bradley	Matt Heller	N/A

***Table 1**. LCC Project Tracking Software Options*

Individual LCC's Decision

This document does not endorse one software option over any other. When looking into options, consider all the elements in this summary matrix and your current and future LCC needs. Research the detailed matrix, detailed comparison report, demonstration material, and get in touch with the software contacts for discussion of your LCC and potential adoption of software.

We encourage your LCC to continue to enter project data as completely and thoroughly as possible for your LCC and the national aggregate database. Don't be afraid to make a decision on a software option. Most likely, the biggest time commitment is gathering comprehensive project data and entering data. No matter what software solution any LCC chooses, future software options are likely to accept data imports and exports.

Products/Outputs/Use Cases

The Project Tracking Team developed a number of products from their 18-month effort.

Products from Task 1: Project Metadata Exchange and Aggregation

The following outcomes resulted from Task 1:

- an overall architecture,
- a content standard, with a core set and optional additional complexity, and eventually a Dublin Core metadata standard crosswalk,
- a functional demonstration of the data exchange protocols, with documentation, including the following elements:
    - an exchange format and protocol (XML and JSON schema specification, still in progress),
    - and, a field crosswalk between the content standard and ScienceBase is currently underway.

Products from Task 2: Supporting LCC Software Adoption

The following documentation resulted from Task 2:

- Comparison of Software Options for Use as the Master Database for Managing Your Project Metadata
- Spreadsheet of Project Tracking Options for Possible National or Individual Use

Recommendations/Collaboration

It is the recommendation of this group that the LCCs coalesce around a common production system. This would allow all the LCCs to contribute project records to support project tracking. We recommend that the LCC Network office provide funding to build capability this in ScienceBase.

Below are details on needed technical development efforts:

- ScienceBase should be the eventual production aggregation server for all LCC project metadata.
- ScienceBase should distinguish between LCC project records with a master copy maintained by ScienceBase (native) and records externally maintained and periodically imported.
- The ScienceBase project editor, DEPTH, should continue to be enhanced with a mode that specifically supports LCC projects, guiding the user to fill out LCC project metadata fields and to use the LCC controlled vocabularies. This editor should be enabled only for ScienceBase native project records (master copies). LCCs maintaining their own master copies should edit using their own facilities, with the derived copy on ScienceBase overwritten with new imports.
- Enhancement of ScienceBase should continue and eventually take over all functionality currently provided by the lccprojects.org website:
    - checking and conversion of CSV/XLS LMP 1.0-format inputs
    - aggregation of multiple LCCs into a single database
    - export of aggregated data to LMP 1.0-layout spreadsheet format for external analysis
    - embedded structured search and review of LCC projects (see the lccprojects.org website)
- ScienceBase should add functionality for:
    - user upload of updated csv, xls, or json input files by an LCC
    - feedback to the uploader on errors and omissions
    - fetching of input files or equivalent web services from LCC systems
- Developers can refine the LMP 1.1 JSON format used for import to and export from ScienceBase (as well as export from lccprojects.org until system retirement). ScienceBase exports should not lose data from the originally imported information.

- Developers should stabilize and fully describe the LPM 2.0 JSON format (the extended Arctic Data Integration Working Group, or ADIwg, JSON). Developers should then alter ScienceBase to import LPM 2.0 JSON format.
- LCCs who maintain their own master project record systems (i.e., who do not use ScienceBase as their masters) should be encouraged to export in LPM 1.1 (json) if they can, LPM 1.0 (csv and xls) if not; when LPM 2.0 (json) is available that should be an option. Automatic harvesting would be desirable, but manual uploads to the ScienceBase aggregation facility would be acceptable given how seldom LCCs create or update projects.
- A future team should consider options and approaches to adding product or deliverables support to the project aggregation system. These could be modeled on the NCCWSC Data Management Plan template (https://nccwsc.usgs.gov/sites/default/files/files/DMPv3_0.docx) fields, or on Dublin Core, or on geospatial metadata standards, or some combination. The team needs to consider these tradeoffs, as well as defining data formats (e.g., LPM 2.1 based on ADIwg) for uploading, downloading, and harvesting this enhanced project metadata.
- The issue of various kinds of duplicate records should be further explored and tools or techniques developed to eliminate or mitigate any real problems caused by this.
- Reporting of funding could potentially use some further exploration.

Appendix III – Visualization Team Evaluation

Introduction

Conservation and natural resource management are largely about place; therefore, mapping is a natural cornerstone for the planning and implementation of numerous activities. An important consideration in managing spatial data to support the LCC mission is how the data are expected to be used by the LCC community. Understanding the target audience and what social processes they need is extremely important before making technical decisions regarding data visualization. This chapter is focused primarily on spatially explicit data visualization, although other forms of data (e.g., graphs, charts, and tables) is also recognized as important in the context of LCCs.

Visualizing data has always been extremely important for conservation planning and natural resource management. Traditionally, this function was the domain of paper maps and later desktop GIS. Now, with the advent of advanced web technology and a growing web savvy user base, new web mapping capabilities are being provided for both the novice and technical user. Visualizing spatial and non-spatial data via the Internet has become extremely popular. Web portals of every size, shape, and configuration are becoming commonplace and the use of infographics, which sometimes includes maps, is becoming increasingly popular (Cook, 2013; Meirelles, 2013). This chapter explores the topic of on-line visualization, particularly as it pertains to mapped data. Three levels of viewer complexity will be described highlighting design choices, associated functionality, and desired outcomes. For each broad visualization category, we provide a description, guidance in choosing a visualization option, examples, and a list of some recommendations and best practices. Finally, we briefly discuss the use case of integrating Data Basin with ScienceBase from the standpoint of data visualization.

Map Visualization on the Web

There is more than one way to build a viewer. A simple web search on the subject “map viewer” will yield hundreds of millions of results. In the context of conservation planning and resource management, we’ve simplified the vast number of potential viewers into four basic categories classified by level of viewer complexity. As we will discuss in this section, each has their own set

of advantages and disadvantages, but there are some best practices that apply to all web viewers. First and foremost, it is critical to have a clear understanding about the *target audience* for the viewer. Is the viewer intended for mapping specialists, resource managers, students, or the general public? The answer to this important question will help decide how complex a successful viewer needs to be. Other high-level considerations before building or choosing a viewer include:

- **Goal of the Viewer** – Is the primary goal to disseminate spatial information? Does it display raw spatial data, synthetic map products, or both? Is the viewer intended for polished presentations or for providing users with a toolbox for exploring data and creating new information?
- **Devices Supported** – Is the viewer for desktop computers only, or do you want it to work on tablets and other portable devices such as cell phones?
- **Web Browsers** – What web browsers do you want or can you afford the viewer to work with - Chrome, Safari, Firefox, Internet Explorer, etc.?
- **Data Management** – Do you want or need to control all of the content or are you interested in integrating with a variety of data sources?
- **Data Security**– How important is data security? Do you need to keep everything behind a firewall or is all or part of the objective is to make spatial data public?
- **Language Support** – Is the viewer to be in English only, or do you need to use multiple languages? If multiple languages, which ones?
- **Longevity** – what is the intended useful lifespan of the viewer? If it is lengthy (multiple years) is there a plan for keeping its functionality up to date with advancing platform versions?
- **Cost** – How much initial and ongoing funding do you have to dedicate to the viewer? Building a viewer is just the first step. Technology is advancing so rapidly that maintaining a viewer is an ongoing investment.

Simple Viewer

Simple viewers are the most common map visualization approach available on the Internet. Some websites elect to distribute spatial information in a static way by providing links (often displaying thumbnail images) to individual datasets or maps that are in jpeg or pdf file formats (e.g., Appalachian LCC). Simple viewers are effective if the goal is to redistribute spatial information in picture form that is easy for users to download and use in various types of printed documents.

However, a growing use of relatively simple viewers is to present information or analytical results via an easy-to-use interface that tells a story while allowing users to explore the content in a controlled manner. These applications are often a component of a much larger, more extensive website often aimed at the general public. Currently, they are usually housed and supported by a resident computer system connected to the Internet, but the trend is toward supporting these viewers using dedicated cloud services. Data interoperability between the creators of these types of simple viewers and other data services is simple, with unidirectional data and information transfer from the originator to the end user.

Simple viewers often provide a handful of stylized data layers supported by base maps (e.g., Esri, Google, or Bing). Simple viewers include common map functions (pan, zoom) but do not support more advanced functionality such as identification queries, layer styling, map or data layer transparency control, and content filtering. Because of the light weight nature of the simple viewer, response time for the end-user is very fast. Export of results is uncommon.

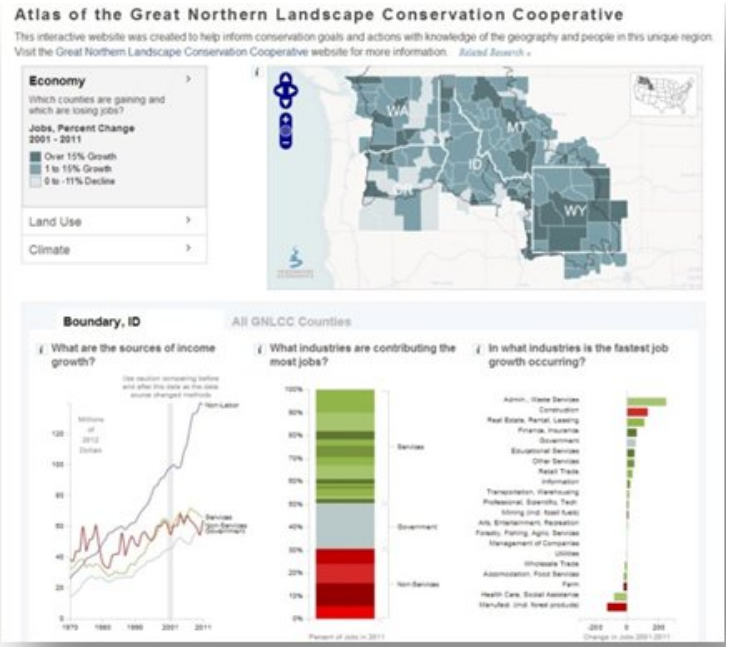


***Figure 5**. Example of a simple viewer (From Great Northern LCC website produced by [Headwaters Economics)](#)*

Some of the more effective simple viewers present analytical results allowing the user to explore the results by hovering over or clicking parts of the presentation with their mouse. Because most of the data processing is performed off-line, more effort is placed on packaging the results and

its presentation rather than on supporting real-time analytical tools. Integration of map data with other graphic elements (e.g., charts and graphs) is very common with simple viewers, but analytical results are not usually calculated real-time through user data queries. User-driven exploration is simply retrieving analytical results that had been generated previously from desktop applications and stored for easy access and viewing.

Simple viewer construction is not necessarily low effort and low cost. The new generation simple viewers can require considerable effort to construct – sometimes involving heavy data management, considerable preprocessing, thoughtful packaging of complex spatial data and information, clean storytelling design, and viewer construction. The goal of a simple viewer is to make the complex seem easy, which is often challenging in any context.

I want a simple viewer if I…

- Have a visual story to tell my audience
- Wish to provide results from often complex analyses in a compelling way
- Infographics using maps is desirable
- Wish to give my audience ways to explore results
- All analyses preprocessed
- Limited data management and storage needs
- Do not have to deal with data security issues
- Map customization by users not needed or desired
- Want to integrate the viewer with an existing website
- Want the presentation to respond rapidly under heavy use
- No (or very limited) export functionality necessary

Examples:

- News media outlets commonly embed simple viewers into data-rich news stories
- Interactive Atlas to Evaluate Economics, Land Use, and Climate – Great Northern LCC
- Headwater Economics
- Sockeye Assessment
- IUCN Polar bear specialist group
- International Centre for Integrated Mountain Development (ICMOD) - Pressure on forest resources of Nepal due to household firewood consumption
- Audubon Society - Pacific Flyway Marine Important Bird Areas

- Coalition for the Upper South Platte - Waldo Fire Recovery

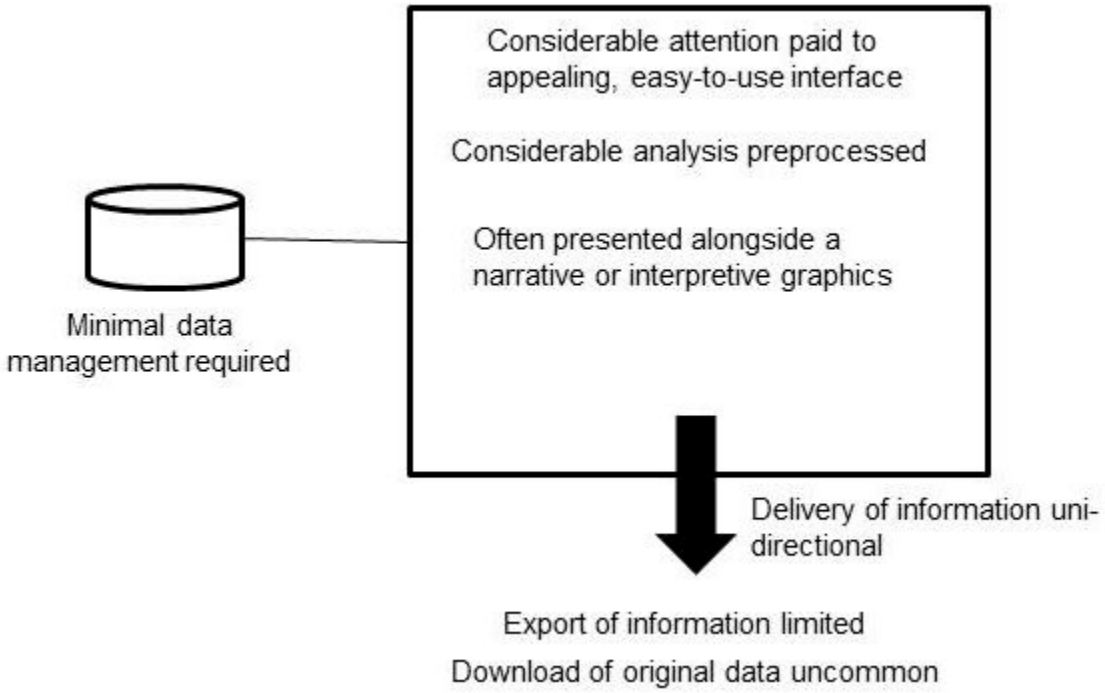


***Figure 6**. Diagram of an archetype simple viewer highlighting some of the key characteristics*

Best Practices:

- Choose/develop a system that is easy to maintain and update
- Build the viewer using programming languages that have substantial support networks
- Choose an application that will allow for easy updates of the source data.
- Design user interface that makes good use of graphic design principles (see Tufte 2001)
- Use a limited amount of explanation text on the viewer – link to supporting documentation when necessary
- Use reference maps to provide geographic context if necessary for clarity
- Use appealing and informative colors (e.g., soft colors preferred over bright colors) for maps and graphics
- Provide graphics that are easy to investigate and read
- Provide adequate supporting documentation and metadata in a format that your target audience can understand. Provide access to more detailed information for specialists via links
- Incorporate export of map snapshot where possible

Moderate Viewer

Moderate viewers are quite common and most popular when the originators want to show and often disseminate large collections of spatial data to a defined audience. Agency and academic research websites often use moderate viewers to provide the spatial data they manage to various user groups. In the past, this was handled through primarily tables with associated metadata and data download links of zipped GIS files. With the growing popularity and utility of web viewers, these services now often include a mapping interface to visualize the content library. For example, the USGS Gap Analysis Program has developed three online viewers to connect the public with the spatial data they are managing. Search for data is usually managed through drop-down menus as opposed to keyword search functionality.

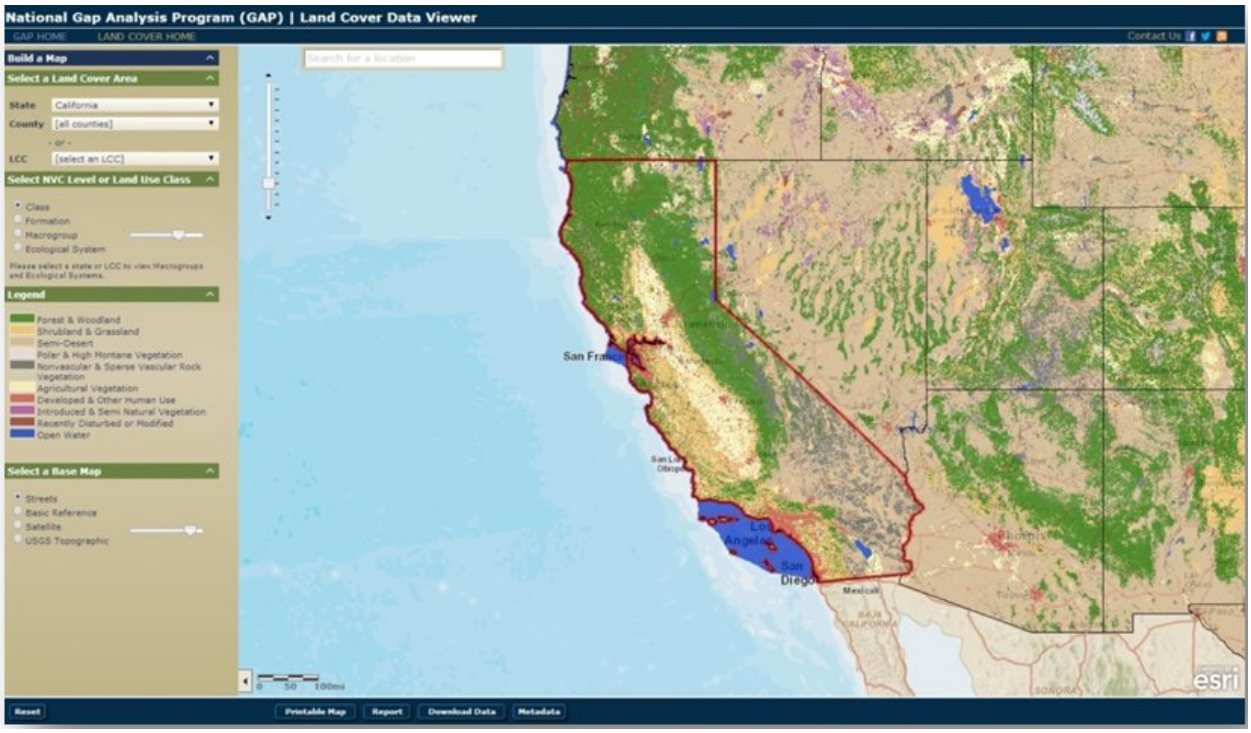


***Figure 7**. Example of a moderate map viewer (From USGS Gap Analysis Program)*

Although the general public is often the audience, most users are researchers and managers who wish to acquire data or information for their own work.

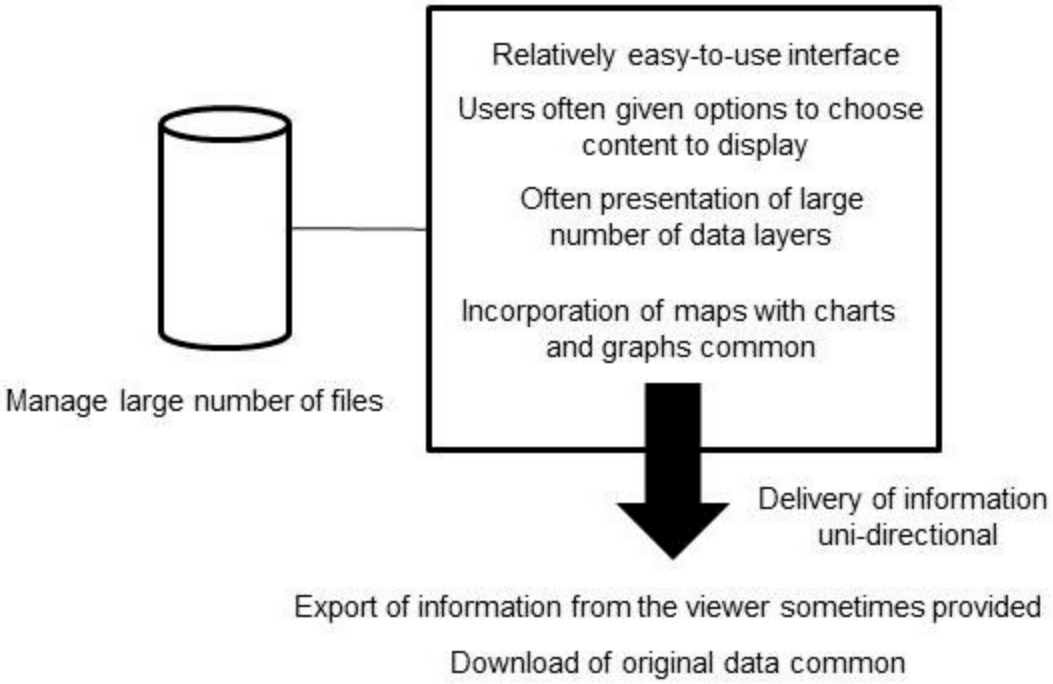


Figure 8. *Diagram of an archetype moderate viewer highlighting some of the key characteristics*

Content supported by moderate map viewers is managed either by one or more robust resident servers or via cloud hosting. Some systems are also moving toward providing web mapping services (WMS) as a delivery option, which opens up possibilities for broader scale interoperability. Movement of data and information is still uni-directional for most moderate viewers – from the site developer to the end user.

The maps provided with moderate map viewers are often of a library of numerous stylized data layers usually supported by base maps (e.g., Esri, Google, or Bing). Common map functions like adding and removing datasets plus typical pan and zoom functionality are standard, but more advanced functions such as identification queries, layer styling, and content filtering are usually not supported. Although moderate map viewers are more complex than simple ones, response time is still usually quite fast. Often, the site is built to facilitate data downloads (typically one file at a time). Integration of graphic elements (e.g., charts and graphs) sometimes accompanies map data.

Use a moderate viewer when you…

- Want to facilitate locating and distributing spatial and associated non-spatial data and information
- All associated analyses preprocessed
- Significant but constrained data management needs
- Do not have to deal with data security issues
- Map customization by users not needed or wanted
- Want to integrate the viewer with an existing website
- Download functionality desired
- Want to limit the amount of data considered.

Examples

- Multivariate adaptive constructed analogs (MACA) statistical downscaled climate data from CMIP5
- National Gap Analysis Program – Species Viewer
- Western Governor's Crucial Habitat Assessment Tool
- National Map - National Map Viewer

Best Practices

- Choose/develop a system that is easy to maintain and allows for implementation of database management standardization, including content taxonomies
- Build user interface that is not too complex and is highly intuitive
- Use a limited amount of explanation text on the viewer – link to supporting documentation when necessary
- Use appealing and informative colors (e.g., soft colors preferred over bright colors) for maps and graphics
- Do not overwhelm the user with too many choices to make
- Provide user help as needed to guide them through the use cases the site is built to provide
- Provide adequate supporting documentation and metadata in a format that your target audience can understand
- Incorporate export of content from the viewer where possible
- Streamline data downloading by providing bundling of data requests – avoid having to download data one file at a time

- When affordable, consider hosting some or all of the spatial data as web map services to encourage interoperability.

Complex Viewer

There are only a few complex viewers. The users of such views, including large software vendors, large government research facilities, and a handful of independent enterprises, want to promote spatial data sharing in some manner. Organizations behind complex viewers have the capacity to manage large volumes of data and information, but expand their data library by connecting to the increasing number of external web map services. Number of datasets runs into the many thousands making an efficient “search” function a necessity. While these organizations sometimes identify the general public as the user audience, these more complex viewers are more suitable to the more advanced user. And unlike simple and moderate viewers, which focus largely on the *presentation* of existing spatial data and information, complex viewers focus on providing the means for their users to create and publish new content from an ever-expanding spatial data library.

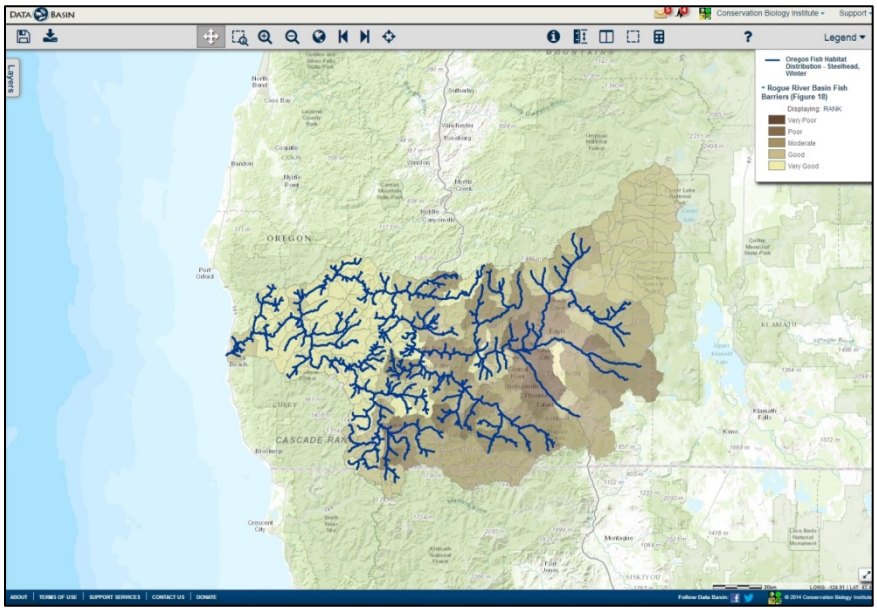


***Figure 9**. Example of a complex viewer (From Data Basin by Conservation Biology Institute)*

Content supported by complex viewers is managed either by an extremely robust resident computing system or increasingly via cloud services, which provides more flexible scaling. Complex viewers also rely more heavily on a data disaggregation model that populates the

system via external web map services. Working towards *data* interoperability is a common primary goal. Furthermore, encouragement of and support for user-generated data adds challenges to a complex system, but the potential rewards are huge. Movement of data and information is bi-directional in most cases although the overwhelming majority of movement is still from the data producer to the data consumer. Data security is an important consideration in providing complex viewers.

The mapping features of complex viewers are rich allowing for standard and advanced functionality, including identification queries, layer styling, map or data layer transparency control, and content filtering. Incorporation of numerous base maps is common, and some systems allow users to add drawings and text. Because the emphasis is on helping users build their own content (mostly maps), the incorporation of other graphics like charts are less common. Simple map viewers find integrating non-map graphics considerably easier since the presentation is highly controlled. With complex viewers, the programming needed to allow users to construct graphics from data is far more open-ended and, therefore, more challenging. Response times for complex viewers are also slower compared with simple or moderate ones because much more data is being transferred in real-time. Because of this, internet connections need to be higher quality to maintain good performance. Slow or crowded band width can significantly impede performance and in some cases make them unresponsive. Complex viewers usually provided map export functionality.

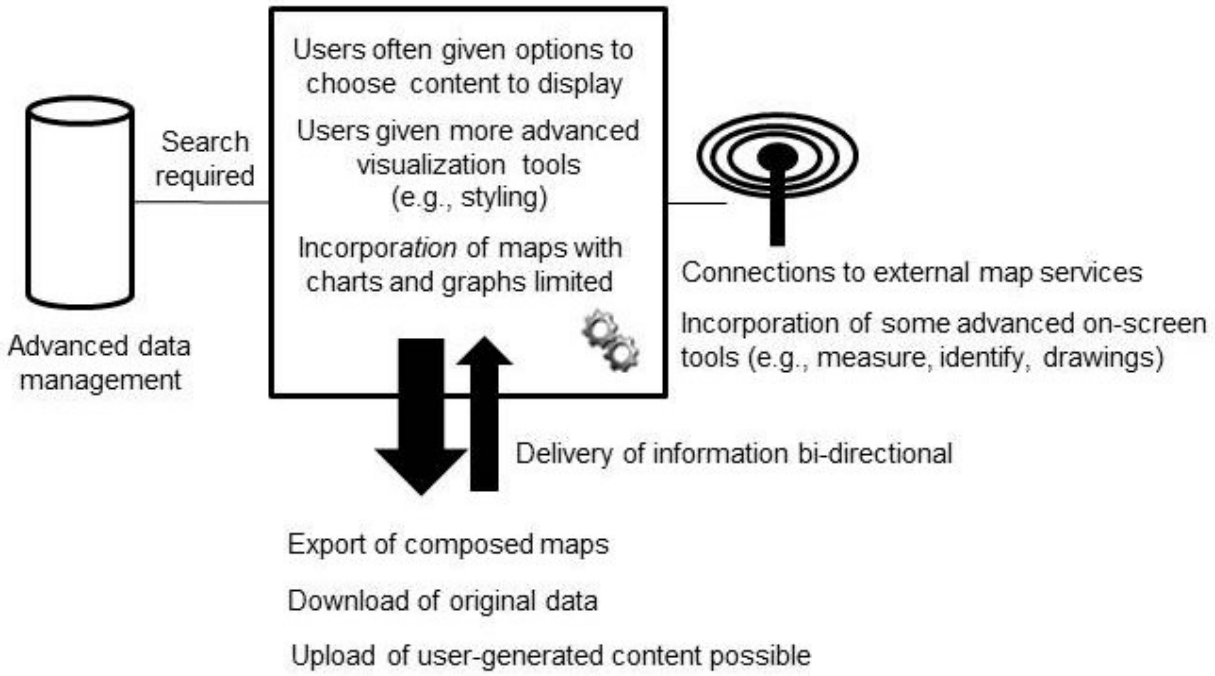


Figure 10. *Diagram of an archetype complex viewer highlighting some of the key characteristics*

Use a complex viewer when you…

- Want to facilitate the creation of new content from existing data libraries
- Want a system that allows users to contribute new data
- Users have more sophisticated on-line mapping skills
- Data security is an important issue
- More emphasis on a disaggregated data management model
- Map customization highly desirable
- Incorporation of non-spatial graphics as part of the viewer less important
- Want the system to accommodate multiple data types and large numbers of files
- Download of content including customized maps important

Examples

- Data Basin
- LCC Conservation Planning Atlases
- Arc Online by Esri
- Oregon Explorer
- Koordinates

Best Practices

- Choose/develop a system that is easy to maintain and is easily scalable
- Build user interface that is highly intuitive
- Use a limited amount of explanation text on the viewer – link to supporting documentation when necessary
- Do not overwhelm the user with too many choices to make without guidance
- Provide user help as needed to guide them through the use cases the site is built to provide
- Provide adequate supporting documentation and metadata in a format that your target audience can understand
- Incorporate export of content from the viewer where possible
- Streamline data downloading

Matrix of System/Viewer Requirements and Viewer Complexity

Summarizing the previous sections, Table 2 provides a means to evaluate the three visualization options against 35 different viewer requirements. We fully recognize that there are exceptions and nuances in the three archetype viewers (simple, moderate, and complex) discussed. We also realize that this technical area is rapidly evolving. First, there are the rapid advancements in the hardware underpinnings of electronic data management and display, including but not limited to faster computational speeds, larger storage capacity, and faster Internet speeds. These items greatly improve the handling of existing datasets and models, but also open the door for meaningful big data applications in the conservation arena. With regards to visualization, these advances are allowing for: more sophisticated handling of larger datasets, spatial data animations, handling of time series more effectively, enhanced integration of other graphic elements into the mapping display, and more sophisticated collaboration tools.

Even though the information presented here is certain to have a short shelf-life, we feel that there is valuable information to be gained by approaching the topic as we have at this point, and we believe that this simple construct can provide practical guidance for the LCC network and others.

Viewer Requirements	Simple	Moderate	Complex
System management	Resident/Cloud	Resident/Cloud	Networked cloud
Level of interoperability (other sites)	None	None/Limited	High
Data security handled	No	Limited	Limited
Need for GIS server support	None	None/Limited	High
Level of user customization	None	None	Mod-High
Number of file formats supported	Very Limited	Limited	Mod-High
Allows for spatial data download	No	No-Limited	Yes
Supports individual user accounts	No	No	Yes
Adding content to maps by user	Very simple	Simple	Advanced
Saving of user content	No	No	Sometimes
Export of maps	No	Simple	Advanced
Provides user collaboration tools	No	No	Limited
Support for 'search'	No	Limited	Yes
Allow data uploads by users	No	No	Yes
Sharing of datasets and maps	No	No	Limited
Direction of data flow	Out	Out	In/Out
System use analytics	Possible	Possible	Possible
Use of base maps	Usually one	One or two	More than two
Number of viewers	One	One	One
Web response time	Very fast	Fast	Moderately fast
Time series data	No	No	Yes
Links to priority content	Yes	Yes	Yes
Identify features on a dataset	Maybe	Maybe	Yes
Filter content	No	No	Sometimes
Style content	No	No	Limited
Setting transparency	No	Sometimes	Yes
User-generated drawings	No	No	Sometimes
Supports conversion of drawings to data layers	No	No	Sometimes
User-generated annotation	No	No	Sometimes
Incorporation of charts and graphs	Very Common	Common	Rare

Supports commenting and review	No	No	Yes
Supports presentation of results from processes conducted off-line	Yes	Yes	Yes
Supports simple analytics real-time (e.g., buffer, summarize, intersect)	No	No	Sometimes
Supports complex analytics real-time	No	No	Sometimes
Ability to tell a story	Yes	No	No

***Table 2**. Matrix of possible viewer requirements by general viewer type.*

Data Basin and ScienceBase Integration

Close integration of Data Basin and ScienceBase was an important component of the IDMN project. Effective integration of the two systems was fundamentally important to several LCCs.

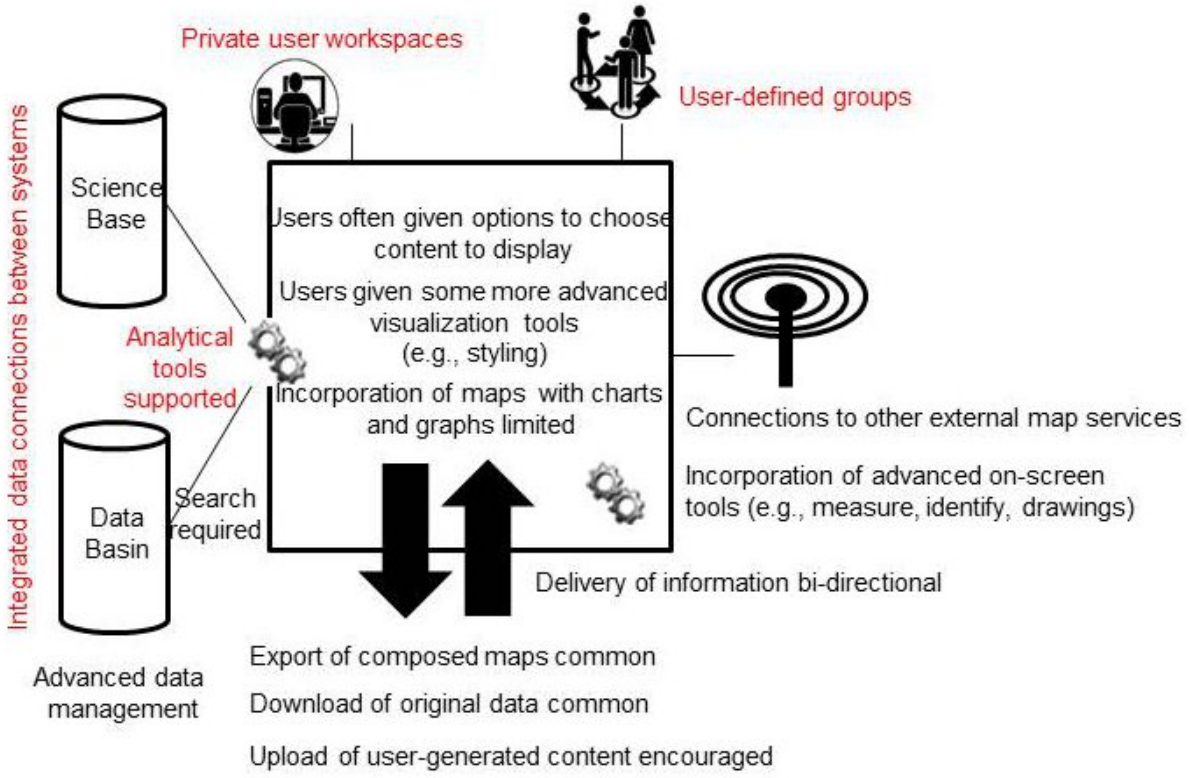


***Figure 11**. Diagram of the integration of ScienceBase and Data Basin highlighting some of its functions*

In these instances, ScienceBase is being used as the primary data storage and management system, and the customized Data Basin sites (called Conservation Planning Atlases in this context) are being used to provide the user interface to support regional conservation activities. Data Basin is also playing a small, but important role of hosting non-government spatial datasets. Multiple LCC Conservation Planning Atlases take advantage of this workflow. With Data Basin being a complex viewer, the schematic for the Data Basin-ScienceBase integration is very similar to the Complex Viewer diagram presented in with a number of special components added.

The additional components (highlighted in red) include two standard Data Basin features including private user workspaces and user-defined groups. Unique to the integration is the closer connectivity between the two data systems beyond simply reading external map services. The integration of these systems provides the ability to run a number of real-time analytical tools via the Data Basin front end while accessing datasets stored and managed in ScienceBase. The full integration required that developers address several technical areas, including:

- the direct import of ScienceBase records into Data Basin,
- use of Data Basin styling and tools on ScienceBase map services,
- use of Data Basin spatial analysis tools on ScienceBase map services,
- browsing of ScienceBase records from within Data Basin, and
- and, harvesting Data Basin records into ScienceBase.

Appendix IV – Tools Team Evaluation

Introduction

By expanding the functionality of existing resources in novel ways, as well as integrating disparate information resources, the Landscape Conservation Cooperative (LCC) Network can ensure immediate impact on the information management needs of various cooperatives. Due to the large number of information resources in development and use, the need exists for formalized guidance on how integration should occur.

Each LCC requires knowledge of and accessibility to a variety of analytical tools and platforms to support LCC projects and communication of results to stakeholders. A centralized and easily discoverable list of tools and platforms available for LCC use did not exist at the outset of the IDMN project. Furthermore, there was a lack of guidance on technical standards to support integration of data and analytical tools across projects, geographic boundaries, and platforms. Although each LCC is independent and self-directed, it is more efficient to share information about common tools, platforms, methods and standards across the LCC Network. Sharing efforts will ensure that each LCC has access to the best tools and standards currently available and avoid duplication of work for identified needs.

The IDMN Toolshed team was created specifically to provide guidance for data integration across platform and analysis tools, and demonstrate said guidance in practice by integrating two platforms already in use by the LCCs: ScienceBase and Data Basin. Composed of representatives from the USGS, USFWS, Oregon State University, Gap Analysis Program at the University of Idaho, Conservation Biology Institute, and PointBlue, the team met semi-regularly since 12/11/2012, and nearly monthly since 5/31/2013.

Key Concepts

To enhance compatibility and integration of information among LCCs, the Toolshed Team focused on the following infrastructural elements: Data Sharing, Metadata, Data Catalog, Tools, Map and Data Viewer and Platform.

Metadata

Metadata describe data, providing valuable insight about the data so they may be understood, used by others, and integrated with other data. Metadata includes information about what the

data describes, where the data came from, who is responsible for the data, methods to access the data, and more. While metadata often describes datasets, it may also describe non-file entities, such as websites and projects. Metadata standards provide structure to ensure key attributes are provided to support the data. Separate metadata schemes are used to structure metadata of a particular type or provide coherency for metadata intended for a specific purpose. For example, Ecological Metadata Language (EML) is a metadata standard including a set of XML schema that allows a rich description of ecological information. Every element name and its meaning in the schema are specified to ensure data providers generate consistent information.

Metadata described using different schemas may be integrated and made interoperable by mapping metadata elements to a common data model, a core set of metadata attributes, and transforming one metadata format into another. Metadata mapping into a common data model allows information indexing. This indexing enables faceted queries, such as the ability to search by the author's name or keyword. Mapping metadata into a common data model is laborious, and while interoperability, queries, and reuse of data will improve, unique elements may not be wholly captured in the data model. Many data catalogs promote and use one metadata standard, such as Federal Geographic Data Committee - Content Standard for Digital Geospatial Metadata (FGDC-CSDGM), which requires a specific data model to incorporate homogenous metadata elements without vast variability.

Data Sharing

Data sharing over the internet is typically accomplished by websites providing downloadable files or in the form of web services through browser, desktop, or mobile applications. Web services are used to deliver metadata components to other applications, such as websites, visualization platforms, online data-gathering forms, and others. Data obtained through download or web services may be amalgamated with other data to create derivative products, which may also be cataloged and shared with others.

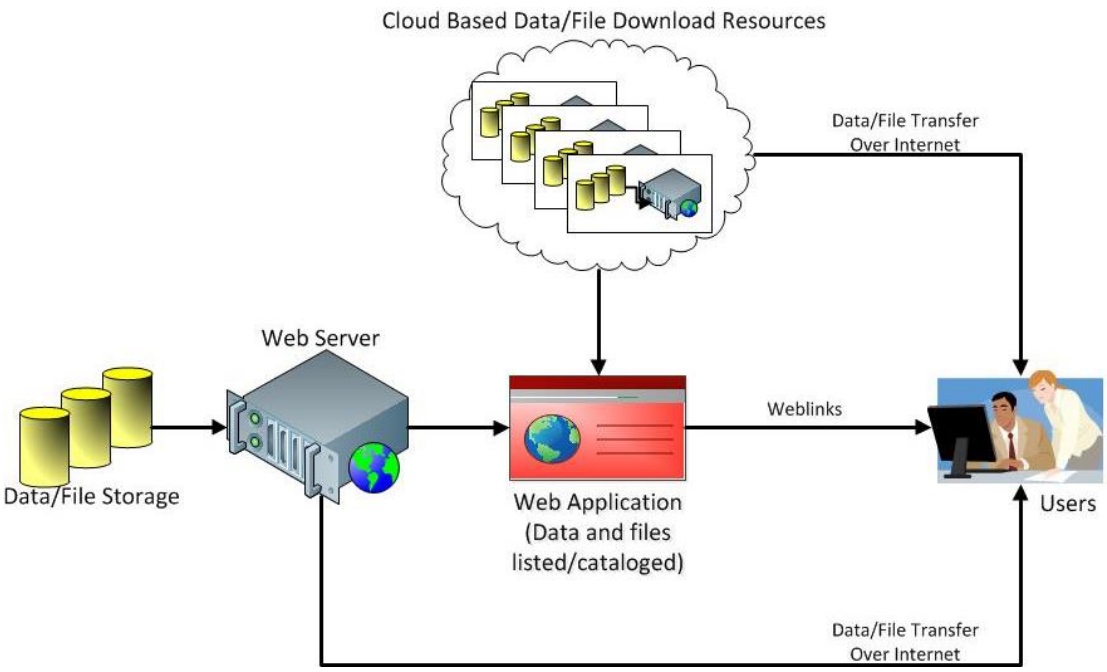


Figure 12. *Websites provide downloadable files. Web applications and web pages allow for data discoverability and provide web links to downloadable data files from a variety of web servers*

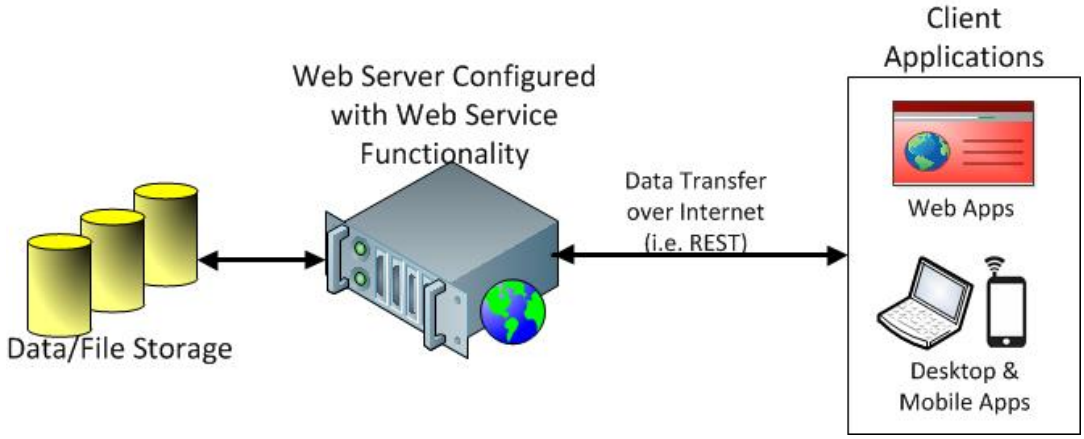


Figure 13. *Web services provide internet resource sharing. From a data provider's web server, data are transferred via the internet for various web, desktop, mobile, and other applications to consume*

Data Catalog

Primary Purpose: to provide access to metadata and associated data files

Key Properties:

- allows users to browse or search for metadata records
- may allow users to associate other data files with metadata records
- may serve as an authoritative source of information records, allowing cross-indexing in other catalogs

Data catalogs are vital for archiving data and enabling access to and sharing of information, which saves resources by reducing duplication and contributing to greater knowledge by enhancing scientific acuity. Cataloging metadata in one location – a data catalog – and disseminating information to other applications is helpful for maintaining consistent records and precluding existence of multiple record versions.

For example, ScienceBase is a data management platform that provides a data catalog, which includes metadata records harvested from data providers' web accessible folders or sitemaps. Metadata records and associated data files can be added directly to ScienceBase by users. Provenance information recorded for each record informs users of record origin, which is important for determining authoritative sources. ScienceBase is not an authoritative source for many of its cataloged records and serves instead as an aggregator of information deemed important by ScienceBase users who include the items in the catalog. Web services allow vast record exchange among catalogs. Records may be manipulated and modified producing derivative data products.

Tools

Primary Purpose: analysis of spatial or non-spatial data to create new data and information

Key Properties:

- takes spatial and non-spatial data as input
- performs analysis
- produces both spatial and non-spatial data as output

In practice, the term "tool" covers a broad range of concepts from a specific operation within a user interface, such as font size control in a text editor, to a full-featured, self-contained web platform. We use the term "tool" to refer to a singular, self-contained piece of software used for data analysis that accepts as input spatial and non-spatial data, applies a series of operations within a defined process, and produces as output spatial and non-spatial data, for example, a

user performs an action *using* a tool. A simple tool (e.g., Fig.14) could provide a summary table of land cover types and acreages within a designated geographic area of interest.

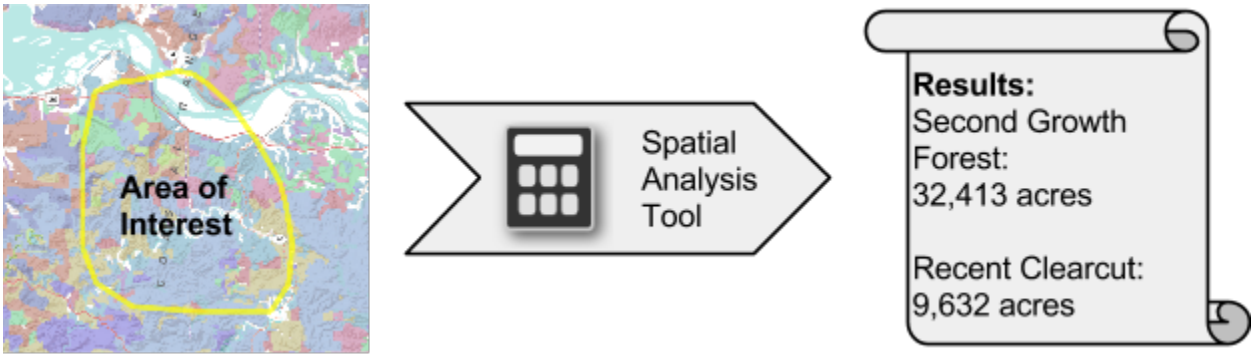


Figure 14. *Spatial analysis workflow*

Map and Data Viewers

Primary Purpose: interactive visualization of spatial or non-spatial data

Key Properties:

- Map viewer
  - allows viewing of one or more themes on a map
  - includes base maps or reference data
  - may provide tools for analyzing data available within the viewer
- Data viewer
  - allows viewing of charts and tables
  - charts may be interactive and allow the user to explore the data
- Maps and charts and tables may be combined to communicate and show information about a place
- Typically limited to the maps and data selected by the organization that authored the viewer

Map and data viewers are typically an integrated suite of tools that collectively allow the user to visualize and explore one or more spatial and non-spatial data sources; in essence a user *visualizes* information *using* a map or data viewer. These may be components of a larger platform, and may be a primary mechanism for launching analysis tools described above. For example, a map viewer could present a series of wildlife habitat layers overlaying a base map

that provides geopolitical boundaries or place names and biogeophysical (hill shade, forest, nonforest or water) context, such as the GAP Species Viewer.

Figure 15. *GAP species viewer*

Platform

Primary Purpose: access to spatial and non-spatial data resources in a system that may provide additional capabilities using those resources

Key Properties:

- supports data discovery through search and browse capabilities
- includes both a data and metadata catalog
- may allow user creation or publishing of maps, data, or other resources
- may provide dynamic, user-controlled visualization of maps, charts and tables
- may provide tools that operate using data available within the system

In the context of this document, the term “platform” is used to refer to a self-contained piece of software that provides a suite of capabilities that a user may use to perform a variety of tasks; in essence the user can perform several actions *within* a platform. These tasks could include visualizing spatial data using a map viewer, executing a specific spatial analysis tool, performing a search to find resources or downloading data.

A platform typically integrates a catalog, a set of tools, and map or data viewing functionality. For example, a platform could be a web-based catalog of spatial data resources which users may query for specific resources, display data resources through map viewers, access

metadata records, and organize these metadata and data records within a folder structure. The ScienceBase data management platform provides such functionality.

Toolshed Team Objectives

- Identify a list of known platforms and analytical tools available for use by LCCs and categorize them according to standardized characteristics. Make these tools easily discoverable by LCCs
- Identify hosting options for an open-source toolshed
- Share knowledge about and capture lessons learned about each of the major platforms represented by team members:
    - ScienceBase
    - Data Basin
    - Climate Science Center Geo Data Portal
    - Western Landscapes/Oregon Explorer
    - USGS GAP data viewers
- Integrate ScienceBase and Data Basin to support LCC needs and capture lessons learned from the integration process
- Define integration requirements for ScienceBase and Data Basin pilot program
- Implement specific technical tasks to integrate ScienceBase and Data Basin
- Identify best practices to support data integration and sharing of tools across platforms and analytical tools, and provide these best practices in an extensible manner
- Identify needs and opportunities for additional, general tools that could be contributed to open-source Toolshed repositories

Products, Use Cases, and Lessons Learned

The team assessed a number of technical efforts to develop lessons learned to share with the LCCs.

Catalog of Tools, Map & Data Viewers, and Platforms

One of the key deliverables from the Toolshed Team is a catalog of existing data analysis tools, map & data viewers, and data sharing & visualization platforms. These records were cataloged in ScienceBase within the Natural Resource Data Analysis Tools folder.

These records were characterized according to:

- category (tool, map viewer or platform)

- availability for download (and location)
- purpose
- generality or specificity of the tool
- point of contact
- open source or not
- server location/hosting
- data input and output properties
- standards compliance of results or metadata
- pros and cons

Several of the platforms and tools reviewed above were directly represented by members of the Toolshed team. The team also discussed a future survey of the LCCs to add additional entries to this catalog.

Use Case

The primary role of the Tools, Map & Data Viewers Platforms catalog is to help LCC staff discover available tools that may meet their needs. The additional characteristics we collected, which are listed above, may help guide further decisions about which tools to use for particular needs.

Lessons Learned

Lack of easy discoverability

Catalogs of tools available to LCCs are usually not centralized, if they exist at all. While they are sometimes listed within resources that foster discoverability (e.g., ScienceBase), no uniform organization scheme has been put into practice. There is a wide range of resources, and typically they are known only to collaborators and project staff in the cases where tools are purpose-built for a particular project. Our initial cataloging exercise is limited to those tools that Toolshed Team members directly interact with or have some knowledge of, and is far from comprehensive.

Lack of available open-source tools

One of the challenges in identifying resources for an open-source Toolshed is that many available tools are not open-source, nor are they likely to become so in the future. While the tools themselves may be available for download, access to the source code within a version control repository system can provide additional power to allow utilization and integration of the

tools with other tools or platforms, as well as contribution of improvements back to the original tool. The choice to open source any project should remain with the project collaborators and this team considered it outside the scope of its objectives to attempt to re-host these tools elsewhere, such as within a centralized code repository.

There are multiple reasons tools are not often available as open-source:

- tools not specifically required as part of project scope;
- extra effort involved in releasing open source tools, including significant documentation effort and governance of updates to source code;
- developers lack knowledge and guidelines for hosting open source tools;
- scientists and other research staff are not familiar with developer-oriented hosting platforms (e.g., GitHub);
- and, some tools are based on or executed within proprietary tools or platforms

Data and map viewing platforms are rarely designed for interoperability. Many platforms and viewers are built to achieve a specific set of objectives. For example, a map viewer may be designed primarily as a means of sharing spatial information as an interactive publication rather than as static images within a website or publication; the data services may not have been designed for integration with other applications. As such, interoperability is often an afterthought in design and architecture. Where interoperability is available, it may be limited to specific subsets of functionality within the broader range of activities offered by the platform. While viewers often display map services hosted by an authoritative source outside of the viewing platform, the user usually cannot perform higher level functions within that view. That includes creating custom symbology, altering the ordering of layer, and performing spatial analysis of those data from within the viewer.

Hosting Options for an Open Source Toolshed

At the outset of this project, we discussed some of the key issues around managing and hosting an open source toolshed. As originally described in the IDMN proposal, the objective was to create one or more open source repositories that would host analytical tools to be used by the LCC network. However, upon reviewing the existing tools above, we found that very few were appropriate to attempt to migrate into one or more centralized repositories. We, therefore, refined our objectives to provide guidance about developing new tools and sharing existing tools

as open source projects, as well as centralizing references to them for discoverability within the LCC Network.

LCC staff and projects funded by LCCs need to make several key decisions regarding where and how to host open source tools. One of the most important considerations is using version control for source code. It is necessary to use version control to accumulate and manage the history of changes to the codebase over time and effectively share tools using open source methods.

We reviewed several publicly-accessible hosting platforms intended for sharing source code or pre-built tools, and found the following:

General repositories

- Cloud-based platforms:
    - GitHub
        - used by Climate Science Center Geo Data Portal
        - provide authoritative control of codebase, and restricted commit (write) access
        - provides easy ability to fork repository, so that non-authoritative users can modify their own copy
        - forks provide a clear link back to authoritative repository
        - provide easy ability to compare different versions of source code, and review contributions of individuals
        - Git is the only supported repository type
    - BitBucket
        - used by Data Basin
        - similar functionality to above
        - supports Git and Mercurial
        - very nice code review tools
- Internally-hosted systems:
    - Stash
        - self-hosted version of BitBucket
        - used by ScienceBase team for some projects
    - Subversion

        - only example of a centralized repository system listed here
        - relatively easy to break the application build
    - Language and platform specific repositories:
        - Comprehensive R Archive Network (CRAN): R statistics platform
        - Python Package Index (PyPI): Python packages
        - hosts built packages; source code hosted on one of the repository platforms above
        - Macports / Yum: Unix/Linux package managers

Within the open source software development arena, we have noticed a very strong trend toward GitHub as the favored repository hosting platform.

Lessons Learned

Requires specific domain knowledge

Use of open source repository hosting platforms is challenging for non-software developers as they rely heavily on specialized domain knowledge for sharing and obtaining source code and related files. This includes:

- use of repository software (specialized applications, syntax, and workflow)
- management of branches and tags
- synchronizing changes made by different developers

However, the trend for several of these platforms is to support easy download of source code as well as built versions of the tools, making them more accessible to non-developers. These platforms continue to adapt toward ease of use, and we expect to see increased usage of them by developers and non-developers alike.

Authoritative control of open source projects

A wide range of organizations and individuals develop tools that should be managed by their authors. Creating repositories of tools for many individuals and organizations presents several challenges, including:

- authoritative control of additions and updates to source code (open source project maintainers)
- process for contributing to open source projects (e.g., “Pull Request”)

- agency requirements about location of hosting, and creation of organizational or personal accounts
- update frequency and ability to respond to issues as they arise (generally does not happen unless project has sustaining funding or through dedication of maintainers)
- for large projects, maintainers may lack specific domain knowledge addressed by a particular analytical tool

Representative Tools, Map & Data Viewers, and Platforms

One of our objectives was to share detailed knowledge, and capture lessons learned from the platforms represented by team members. Because each of these platforms represents a major, ongoing effort on behalf of their respective organizations and clients, we feel that this information is particularly relevant to the discussion of integration of spatial and non-spatial data as well as tools across independent efforts.

Data Basin

Data Basin is a publicly-accessible, web-based spatial data sharing, visualization, and analysis platform developed by the Conservation Biology Institute (CBI). CBI developed Data Basin for decision makers, technical staff, researchers, students, and the general public. It provides a searchable catalog of spatial datasets and externally-hosted map services, all visualized within a feature-rich map viewer. Data Basin includes group and personal workspaces for access control and organization of resources and provides content types intended for sharing large collections of information (galleries) and communicating results and approaches (guides & case studies). Data Basin has been used to build Conservation Planning Atlases for several LCCs (e.g., South Atlantic LCC), with more planned for the future. While the core of Data Basin is not planned to become open source, it is based on many 3rd party open source components, and original components are being redeveloped to release as open source projects in the mid-term.

Key functionality relevant to the Toolshed

- spatial, time-series, and non-spatial data visualization within a feature-rich map viewer
- able to consume and display externally-hosted map services
- able to consume catalog entries from other systems
- area of interest analysis tool
- provides catalog application programming interface (API) for datasets

Lessons Learned

- It is difficult to achieve the proper balance between accessibility of the user interface to support non-technical users as well as making powerful tools available to highly technical users.
- Continuous improvement and integration of new features are essential to the sustainability and maintainability of large data management platforms. This new feature integration requires on-going development support.
- Use of distributed version control and associated repository tools (e.g., pull requests) are essential in achieving high productivity of the development team as well as ensuring higher quality source code and associated tangible benefits (e.g., ease of maintenance, fewer bugs reported by users, etc.)
- Popular proprietary map server software is not designed for user-based publishing of many thousands of datasets, presenting unique scaling and data management challenges at the architecture level.

Opportunities for Integration

- Much of the API is for internal use; additional APIs could be developed to support integration with external platforms.
- Integration with additional external APIs is highly desirable. For instance, being able to access data from eBird or the Avian Knowledge Network, to show bird observations for a particular species across a defined timeframe on the map, would provide value to the users of this system.
- Support for additional types of map and data services is highly desirable.
- Support for 3rd party hosting and control of secure data is highly desirable for some users (e.g., ability to host data securely on ScienceBase, and access those through Data Basin).

ScienceBase

ScienceBase is a collaborative scientific data and information management platform developed by the U.S. Geological Survey (USGS) for use by scientists, data practitioners, decision-makers, and the public. ScienceBase provides access to aggregated information derived from many data and information domains; including feeds from existing data systems, metadata catalogs, and scientists contributing new and original content. The architecture of ScienceBase is designed to help users centralize their data and information resources, by coupling metadata and datasets, to create a foundation needed for their work. ScienceBase web services disseminate information from the ScienceBase catalog to other custom applications for a broad set of uses.

ScienceBase is an open source project which promotes involvement from the larger scientific programming community both inside and outside the USGS.

Key functionality relevant to the Toolshed

- allows users to create new records with metadata, attach additional data files, and link to external resources
- harvests information from external data sources, and can serve as an authoritative source for data and metadata
- standards-compliant metadata (FGDC - CSDGM, ISO, EML) are available for most records
- hosted on a federal system, allowing control of data location, security, and federal standards compliance
- allows users to create and host map services from spatial data sources uploaded into the system, making them available for consumption by a wide range of users and partners
- and, provides a RESTful API that allows comprehensive interactions with the system for internal and external applications.

Lessons Learned

- the many features and capabilities available may overwhelm a novice user
- features and capabilities are difficult to articulate simply
- popular proprietary map server software is not designed for user-based publishing of many thousands of datasets, presenting unique scaling and data management challenges at the architecture level
- and, low-quality data and metadata generated by data providers may reflect poorly upon the catalog.

GAP Viewers

The U.S. Geological Survey Gap Analysis Program (GAP) has developed three publicly-accessible online viewers that allow users to explore, download, and create maps: the GAP Landcover Viewer, the Protected Areas of the US (PAD-US) Viewer, and the GAP Species Viewer. These viewers are intended to simplify access and understanding of data for all interested data users (e.g., decision makers, technical staff, researchers, managers, and students) so that end users can easily apply the data to their conservation projects. GAP data can be queried and downloaded by administrative boundaries (e.g., Landscape Conservation

Cooperative, state, etc.) as well as taxa and land owner. Metadata and web map services are also available.

GAP data supports numerous conservation planning projects including:

- State of the Bird's reports,
- Crucial Habitat Assessment Tool for the Western Governors' Association,
- NPScape,
- NOAA's State of the Coast report,
- America's Great Outdoors Initiative,
- and assessing the representation of ecological systems within Landscape Conservation Cooperatives.

Key functionality relevant to the Toolshed

- Spatial and non-spatial data visualization within a feature-rich map viewer, and ability to generate presentation quality maps.
- Users can download spatial and non-spatial data, including ancillary data layers designed specifically for species modeling across the entire United States
- Toolshed uses analysis tools that summarize data by administrative unit (Landscape Conservation Cooperative or state).
- Ability to generate species modeling reports, including descriptions of all variables used within the species distribution model and ecological system affinities of each species.
- Ability to generate land cover reports of area and percent area of each land cover class within an LCC or state.

Lessons Learned

- Simplicity of user interface is important to allow data users to focus on understanding the data; however, users will continue to demand additional functionality.
- Using multiple viewers simplified interface so that each theme could be explored easily, given that the data structure varies between these data products.
- Aspects of data that GAP deems important to view are different from what data users deem important; data users view and query the data in different ways than GAP personnel envisioned when developing these online viewers.
- Maintaining current data and technology is challenging.

- Need to provide contact information to data users to address questions and input about viewers and data.
- Web-service data organization is important, and it starts at data development. For example, rather than creating a web-service for each species' range, the species' range web-services are dynamically generated on-the-fly.
- Access to web mapping services, raw data, and downloadable data in addition to viewers is important to meeting user needs.

Opportunities for Integration

- API is currently for internal use only. Currently working towards creating a publicly accessible API.
- Adding additional support for common map service types (e.g., Web Feature Service) would provide better integration support with platforms and data providers using these services.
- Support for additional external APIs is highly desirable.
- Providing metadata that are well organized, structured, searchable, and web-accessible is highly desirable.

Western Landscapes Explorer

The Western Landscapes Explorer provides access to information used to evaluate large landscapes in the Western United States over time. This information includes data and information developed by the Integrated Landscape Assessment Project (ILAP), as well as other landscape-level information designed to facilitate decision-making, research, and exploration.

Through the Western Landscapes Explorer you can

- explore western landscapes and their diversity
- use the Western Landscapes Map Viewer
- access data, maps and models from ILAP
- learn about landscape-level collaborations and initiatives
- discover additional landscape resources
- and, submit information to the online archive

Key functionality relevant to the Toolshed

- publically accessible map services and data
- ability to integrate external map and feature services dynamically into a map viewer

- integration with ESRI Geoportal Server for cataloging and searching of metadata with ability to harvest data from and federate searches with external repositories
- ability to integrate with ArcGIS Online
- extensible mapping framework built with third party Latitude Geographics Geocortex Essentials framework

Lessons Learned

- Data sharing standards via map services is a simple form of inter-application leveraging of data and resources. Exposing easily-discoverable map services and connected within applications like the Western Landscapes Map Viewer is desirable.
- Lack of a common standard or protocol for finding and searching other repositories is a problem. Developing applications against custom APIs is cumbersome and not sustainable. Common cataloging standards such as Catalogue Services for the Web (CSW) would be an easier way to share data and services.
- The feasibility of sharing geoprocessing services can be limited due to the nature and specificity of tasks. Many of these geoprocessing tasks require direct access to the data that in a distributed toolshed scenario does not readily happen.
- The need exists for tools and services that pull in the layer metadata within map services. The current ArcGIS Map service model does not expose layer metadata; only map service level metadata are available.

Geo Data Portal

The Geo Data Portal (GDP) is a USGS-led initiative to provide scientists and environmental resource managers access to downscaled climate data and other data resources that are useful in landscape modeling and analysis of climate impacts to natural resources. The GDP provides a web user interface that allows the user to extract summary data from a growing collection of historical climate baselines, downscaled climate model projections, and other datasets with similar structure. The GDP is also available via simple script tools that execute the underlying web service processes that underpin the system.

Key functionality relevant to the Toolshed

- users can calculate summary statistics for areas of interest using weighted or unweighted area grid statistics, returning a simple text file of summary time series
- users can analyze against catalog of existing datasets or compatible new datasets

- available for implementation in a wide variety of applications and workflows.
- free web processing utilities for all users
- free and open source code available on GitHub for the core Geo Data Portal project and add on script client interfaces
- R and Python script client interfaces available

Lessons Learned

- While many datasets are available via the basic web service interfaces that the Geo Data Portal implements, there are many service implementation details that cause the Geo Data Portal to fail or work in a sub-optimal manner.
- Lack of standardization of variable names and service structure complicates processing with the Geo Data Portal.
- Users must do their own research to determine if datasets on the Geo Data Portal is appropriate for their application.

Opportunities for Integration

- Web processing services are available for any application to implement and work with remotely held data hosted via applicable standard interfaces.
- Data served on Geo Data Portal servers are available as standard data-providing web services.

All Platforms Assessed

Commonalities of the Platforms

- All platforms are trying to reach a broad audience, with significant variability in users' technical capabilities.
- All are trying to make complex data and information resources more easily available to their audience.
- All are adding value by integrating multiple data resources into a single platform.
- All platforms use map services to consume or publish a map or spatial data resources.
- All expose at least some of their resources via a public API.

Common Lessons Learned

- Communicating the complexities of different data representations and capabilities to a broad audience is challenging; many visual conventions do not exist for representing data

type, structure, or capability (e.g., an icon that clearly represents “Raster” to all users regardless of the platform being used).

- Many of these systems are implemented to serve multiple audiences with varied understandings of the terminology and methods offered by the system. Bridging the divide between these disparate disciplines is a particularly difficult challenge, especially when project repurposes a system for a new, more diverse community.
- Providing a user interface that allows a user to perform many complex tasks, while remaining simple for non-specialists to use is very challenging. Focusing on a specific audience segment, such as specialists or non-specialists, risks alienating others.
- Architecting a generalized system to provide uniform capabilities across a broad range in data types and formats is challenging, especially if the system’s chosen software frameworks only support a subset of those data formats.
- It takes substantially more effort to building analysis tools that must be generic with respect to inputs. In other words, it is easier to build an analysis workflow that manipulates a designated, static set of data, whereas it is more difficult to build a tool that performs a generic operation to one or many user-defined datasets of different formats or data structures based on standards.
- Common data and metadata structures and standards have not been widely or rigidly adopted across the above platforms. The lack of continuity complicates some elements of information transfer. For example, granularity of information – a single resource versus a collection of resources – presents challenges sharing records between ScienceBase and Data Basin and for process automation within the Geo Data Portal.
- Web map services are highly variable in functionality and data representation across map service types, versions, and even within types (e.g., WMS), which complicates building systems to provide similar functionality to users.
- A syntactically and functionally equivalent implementation of standards-based data services may not be compatible with different specialized use-cases.
- ArcGIS is an often used core technology but does not provide access to metadata for data layers within map services. This lack of metadata connection to data is a major constraint, as most platforms require metadata to catalog or consume those map services.
- Integration of ScienceBase and Data Basin

To capture specific technical lessons learned and provide a concrete example of integrating together data and tools between two complex platforms, we created several integration points between ScienceBase and Data Basin.

Import of ScienceBase Records into Data Basin

One of the most important points of integration between the ScienceBase and Data Basin platforms is providing the end-user with the capability to import a record from within ScienceBase to Data Basin. ScienceBase provides strong cataloging capabilities coupled with the ability to publish new map services on federally managed systems (often a requirement for federal agencies), but has not, to date, developed sophisticated map presentation and analysis capabilities. Data Basin provides sophisticated mapping and analysis capabilities for resources hosted both internally and externally on top of a user-friendly spatial data publishing and sharing platform but does not deliver map services to external clients. . To date, Data Basin only meets federal security requirements for hosting of non-sensitive, non-internal spatial data. While users have been able to import map services hosted elsewhere into Data Basin for some time, the current workflow required them to enter metadata manually to import the dataset into Data Basin (due to the lack of metadata available from the map service itself). The new workflow provides a fast and easy mechanism for importing ScienceBase records into Data Basin.

This integration point specifically provides the following:

- The user enters the URL to a compatible ScienceBase record into the import page in Data Basin.
- Supported resources are limited to map services published directly through ScienceBase (both ArcGIS and WMS), as well as map services hosted outside ScienceBase that meet Data Basin requirements for data types, projections, and other criteria. The record must include at least the title, description, and tags.
- Data Basin automatically retrieves metadata from ScienceBase API and locks the authoritative source of this information to ScienceBase. Users are not allowed to edit any of the metadata within Data Basin.
- Provided all these requirements met, Data Basin presents a fully-functioning dataset to the user.
- Data Basin periodically checks for updates to these records in ScienceBase and automatically updates the dataset record in Data Basin. It is, therefore, easy for data

publishers to update the information in a single location (ScienceBase) without having to re-synchronize this information across downstream platforms such as Data Basin.

To best use ScienceBase resources in Data Basin:

- read the ScienceBase Help Page on the topic, especially the best practices section
- when uploading or editing data in ScienceBase, make sure to fill out required fields
    - title
    - body
    - tags
- fill out optional fields to improve record in Data Basin
    - Provenance - Annotation field (on “From” tab): used for “Data Provided By” section in Data Basin
    - Contacts (on the “Who” tab), used for “Data Provided By” where contact type is Author, Co-Investigator, Data Owner, Lead Organization, Originator, Principal Investigator when Provenance - Annotation field is empty
    - Contacts, used for “Contact Organizations” when contact refers to an organization (drop down menu to the right of contact type), this should include an email address
    - Contacts, used for “Contact Person(s)” when contact refers to a person
    - Dates (on “When” tab), used for “Content Date” field on all layers within the dataset
    - Purpose (on “What” tab), used for “Purpose” field or layer within the dataset if there is only one layer present; field is otherwise ignored

Import Use case

The primary use case of this integration point is to allow federal agency staff and LCC staff to publish spatial resources in ScienceBase. Publishing in ScienceBase will achieve broad discoverability within the catalog, hosting on a federally controlled system, and then importing those resources into Data Basin to provide user-friendly mapping and presentation components. One of the common cases for this in practice is through the Conservation Planning Atlases developed by several LCCs to date (such as South Atlantic LCC’s Conservation Planning Atlas).

Lessons Learned

- A clean, well-documented API makes it easy to implement requests into a platform to obtain information.
- This approach relies on public access to a record in ScienceBase and is not currently functional for secure (e.g., password protected) resources. Discussions are underway to determine the technical requirements and user workflow for importing secure resources from one system to the other. There are large technical and workflow challenges to handling security of resources between multiple independent account systems.
- Providing authoritative control and single-source-of-information is critical to the success of this component; the user should only have to enter the information into the originating platform to ensure that it is propagated to downstream platforms.
- Different data models may hinder the ability of users to document their datasets. For example, ScienceBase allows users to create a single metadata record for a series of data layers within a multi-layer map service. Data Basin allows users to edit metadata records for each layer in this situation, including a description, methods, and citations of each layer. However, Data Basin is unable to allow user editing of these individual layers if the authoritative location of the metadata is ScienceBase.

Use of Data Basin Styling and Filtering Tools on ScienceBase Map Services

Data Basin previously allowed users to change the symbology of datasets only for those hosted directly within the system, because this required custom tools to be available on the map servers. These tools are useful because they allow users to change the presentation of datasets to better overlay with other datasets in the system in ways not expected when the dataset uploaders originally defined the default symbology. Likewise, Data Basin allows users to display a subset of features within datasets using filter criteria for one or more attributes and their values. It is highly desirable to make these same capabilities available for ScienceBase-hosted map services to achieve a more meaningful integration between the two systems. Due to recent improvements in the core technology used by both systems (ArcGIS Server 10.1+), we were able to develop these capabilities without requiring the deployment of custom tools to ScienceBase servers.

This integration point will specifically provide the following for ArcGIS map services published on ScienceBase:

- User can filter features visible on the map

- Custom styling (not yet fully implemented)

Styling Use Case

The primary use case of this integration point is to allow users to customize the presentation of spatial data within the map viewer. The integration point can help focus presentation of information and interpretation by users on select features within a dataset. It will also make it easier to use map services provided by many different sources within a cohesive map visualization experience.

Lessons Learned

- Functionality provided by the base technology is greatly preferred to custom tools
- Providing custom filtering and styling makes both the data and the map viewer more useful to map creators and advanced users, as well as making the information easier to interpret by end users.

Use of Data Basin Spatial Analysis Tools on ScienceBase Map Services

It is important to provide similar spatial analysis tools (to those provided for native datasets in Data Basin) in order to effectively utilize spatial datasets hosted on ScienceBase. The current spatial analysis tool (*Site Assessment Tool)* provides the capability for a user to define an area of interest, and perform an intersection against features from one or more layers visible within their map. They can then see a summary report of the characteristics of those features including aggregation by particular attribute values (e.g., administrative unit name), as well as the number of features, area, and length within the area of interest. Previously, these analyses required custom geoprocessing tools on Data Basin ArcGIS servers that were system specific and not easily portable. Spatial analysis tools design centered on internal Data Basin use.

The Data Basin team completely rebuilt the Site Assessment Tool to be portable to other ArcGIS servers and versions, and open-sourced it so that it is easily available to other users (available here). The new version of the tool also includes support for performing analysis against raster data, including area-weighted statistics within the area of interest.

Use Case

The primary use case of this integration point is to allow users to perform detailed area-of-interest summaries against compatible map services hosted by ScienceBase.

Lessons Learned

- The analysis software must execute as close to the data source as possible; streaming large amounts of data from a data source machine to an analysis machine results in poor performance.
- Performance issues in the ArcGIS Server and geoprocessing components used by this analysis tool can result in slow response time for performing relatively simple analyses.
- The core technology does not readily support accessing the data sources behind a map service; custom methods were required to obtain access to the data sources to use as inputs to the analysis.
- Several assumptions had to be hardwired into the analysis tool to keep complexity within reasonable limits (e.g., support for a limited number of spatial reference systems). These assumptions constrain the data sources that are compatible with this tool.
- A detailed test suite is required to ensure that the tool performs calculations correctly and to verify that changes to the source code do not generate invalid data.

Browsing ScienceBase Records from within Data Basin

To better support users and organizations wanting to access ScienceBase resources from within Data Basin, we provide the ability for users to specify a collection of external resources in a Data Basin gallery. A gallery in Data Basin is a user-created item that makes it easy to collate, organize, and share datasets, maps, and other galleries available within Data Basin as a single collection. This new functionality allows gallery creators to show resources from one or more pre-configured searches against ScienceBase, allowing users to browse the results of those searches. Users can then leverage a much broader range of spatial and non-spatial resources available in ScienceBase, such as a GIS data inventory of cataloged resources. Previously, users could access these records only by downloading from their authoritative source. These resources could not be directly represented in Data Basin without providing them as either map services or uploading directly into the system.

For example, within the North Pacific LCC's Conservation Planning Atlas - Coastal and Marine Resources Gallery, the external resources tab provides access to search results for Coastal,

Estuarine, and B.C. Marine resources from the NPLCC GIS Data Inventory in ScienceBase. These searches are configured to select results that have been specifically tagged by NPLCC staff.

Use Case

The primary use case of this integration point is to expose an organized, controlled search experience against a broader collection of resources from ScienceBase than can be represented directly in Data Basin due to data type restrictions.

Lessons Learned

- It is better not to provide a single search interface into very different catalogs (e.g., Data Basin and ScienceBase) as this can easily confuse users and result in a highly complex search tool. Instead, providing search functionality against external catalogs from a very specific context simplifies the interface and improves usability.
- Allowing access to additional resources from another catalog adds value to the system, when a secondary objective is to expose end users to a broader range of resources.

Harvesting Data Basin Records into ScienceBase

Data Basin has several thousand datasets that have been directly uploaded into ScienceBase. Adding these items to ScienceBase adds value to both systems because it enhances the discoverability of resources. While most of the functionality for these resources is only available within Data Basin (e.g., viewing on a web map), it is desirable to help users discover these items. Interoperability and metadata exchange depends on metadata element crosswalks between the two systems. For example, ScienceBase's body field is akin to Data Basin's description field. Mapping metadata elements among the systems allows the cataloging, displaying, and querying of common records in both ScienceBase and Data Basin. Not all fields map cleanly between the two systems; thus, it is necessary to make a link back to the authoritative catalog clear in downstream data catalogs.

This integration point provides:

- simple public catalog API of spatial datasets
- periodic harvesting of Data Basin datasets into ScienceBase, available within this ScienceBase item

Harvest Use Case

The primary use case of this integration point is to make Data Basin datasets available within the ScienceBase catalog. A secondary use case could involve the addition of these datasets to other catalog systems not specifically identified here.

Lessons Learned

- Web standards work best where they streamline and clarify information exchange. If web standards complicate or add or require significant effort to adopt (e.g., Catalog Services for the Web), they are often ignored in favor of much simpler non-standard APIs. In this case, it was highly inefficient to adopt CSW because neither platform had previously adopted it.

Related Integration Activities

Integration of ScienceBase and Geo Data Portal: Use of ScienceBase as a content management system for the Geo Data Portal's processing artifacts

The Geo Data Portal's processing tools generates summaries of large gridded data sources for a given set of input polygons. In this integration, ScienceBase houses the input polygons as an item and provides them to the GDP via web services. The user can execute a GDP process and store the results of that process as a child-item of the input polygons. The GDP does not provide cataloging for either input polygons or processing results. This integration adds a web-catalog of these processing artifacts to the overall capability of the GDP.

The workflow implemented by the two systems is as follows:

- A user creates a public ScienceBase item with their polygon area(s) of interest uploaded as a shapefile and published as a Web Feature/Map Service.
- A link from ScienceBase re-directs the user to the Geo Data Portal with a reference to the Web Feature/Map Service and associated ScienceBase item ID.
- The user chooses a dataset and configures their process in the Geo Data Portal.
- Upon process completion, a link is provided which allows the user to direct ScienceBase to associate the results with the original item.
- ScienceBase retrieves process metadata and results from the Geo Data Portal server and saves them in a child-item of the original input-polygon item.

Geo Data Base – ScienceBase

ScienceBase users and teams coordinating projects with ScienceBase can use this integration to manage Geo Data Process results associated with various areas of interest. The alternative is an ad-hoc, local management scheme invented by an individual scientist or project team. This formality ensures more clear documentation and less time spent managing various project data resources.

Lessons Learned

- Integrations such as this rely on specific agreed-upon implementation details. If one or both systems make changes to aspects of their system, there is potential for integrations such as these to break. The relationship should be well defined, tested, and understood by both systems.
- Documenting and the publicizing availability of such integrations take significant effort. Planning should include realistic outreach and training goals for users of both systems.

Use of ScienceBase-hosted services as a data-source for Geo Data Portal processing

Because ScienceBase can host Web Coverage Services (one of the gridded-data-source web service protocols the GDP can process), developers pursued this integration. While, technically possible, it has not been used in a significant way for practical reasons. The Geo Data Portal's catalog of metadata, pointing at various web-service resources, was migrated to ScienceBase to take advantage of this integration.

The GDP processing engine does not host data. Rather, it retrieves *any suitable* data based on a web service URL given to it by a client executing a processing request. A metadata catalog of web service sources compatible with the Geo Data Portal is used to help client software and users identify datasets to use as inputs to GDP processing. ScienceBase is metadata catalog in that it holds descriptions of data and, for some items, URLs for web services that allow access to that data. In a previous version of the Geo Data Portal web client software, the capability to search the ScienceBase catalog was implemented with some success. Few resources of great interest for processing are available, so the capability was de-commissioned. However, the

demonstration of capability was of great value and plans to use it in a more meaningful way are forming.

Raster Use Case

ScienceBase holds references to web service datasets that are of interest for processing with the GDP. Rather than users having to make the connection between the two systems, we need a search interface that allows a user to identify suitable datasets from ScienceBase.

Lessons Learned

- Just because a metadata search returns a potentially useful URL does not mean that the data are useful or usable for the user of processing system respectively.
- The Geo Data Portal's search works against a particular 'profile' of ISO metadata. ScienceBase can provide that profile, but alternative, more flexible solutions may be available in the future.

Integration of GAP Web Services into ScienceBase

Currently, the Gap Analysis Program (GAP) provides access to a variety of web map services for their data sets on the GAP Web site. Whereas six web mapping services are available for the GAP Land Cover product and two web services are available for the Protected Areas Database of the U.S., the species' data set provides thousands of web services for species ranges and distribution models. During the course of Toolshed meetings, the group discussed the need to integrate these web map services into the ScienceBase data platform to provide enhanced access to these services for the LCC community.

GAP and ScienceBase are both part of the USGS Core Science Analytics, Synthesis, and Libraries (CSAS&L) program. CSAS&L will be addressing integration of GAP's web services into ScienceBase. The first step of this process will address enhancing and improving GAP web services to result in better service-level metadata. CSAS&L is using a "sprint development process across many of its products. Planning for the GAP sprint will occur during spring 2014 with the sprint planned for summer 2014. After the program addresses GAP web service enhancement during the first sprint, follow-up activities will begin focusing on integration of web services into ScienceBase. Both GAP and ScienceBase are devoted to ensuring the most efficient access to GAP data for the LCC community.

Integration of ScienceBase and Data.gov

On May 9, 2013, the White House released an Executive Order, "Making Open and Machine Readable the New Default for Government." This Order, built upon an earlier interagency memorandum released by the Office of Science and Technology Policy (OSTP), was accompanied by an Office of Management and Budget (OMB) Policy and site hosted on GitHub called Project Open Data, to guide implementation. To meet the requirements outlined in these initiatives, USGS has developed a data catalog, in concert with the Department of the Interior (DOI). ScienceBase serves as an aggregator for USGS metadata records. These records are delivered to the USGS Science Data Catalog, which serves as a gateway for publicly-available USGS resources and delivery to data.gov. Active discussions are identifying a workflow for delivering LCC data resources to data.gov, and investigating the feasibility of ScienceBase being the data conduit. LCC participants must determine their agency's preferences for delivering data to data.gov.

Recommendations & Best Practices

Data & Metadata

Consideration of the data lifecycle and data management should occur at project inception. Data and metadata have finite usage and relevance; data standards become antiquated, and data formats become obsolete. Data and metadata require stewardship and maintenance to ensure data remain relevant, useful, and comprehensive. The USGS Data Management program provides excellent guidance for managing information during the data lifecycle.

Missing data has become a high-profile issue for the scientific community. Documented research indicates that 80% of data used in peer-reviewed science publications have been irretrievably lost within 20 years of publication (Gibney and Van Noorden 2013; Vines et al. 2013). Management of data and metadata must be addressed directly as part of data-producing projects to mitigate this data loss.

Data

Data may contain information about simple to complex, and may be represented using a variety of formats ranging from simple, easily portable data formats to complex, proprietary formats. Data derived from many inputs (e.g., model results) broaden the scope of data management to include those inputs, as well as outputs. Data may remain useful many years after completion of

the project that created them, thus proper data management and documentation are essential for ensuring that data continue to be accessible for later uses.

Good Data Practices

The following metadata and data best practices highlight key elements of proper management of these valuable resources:

- Management
  - Make data management a priority in any project that uses or creates data.
  - Follow an established data management plan wherever possible, or create a standardized data management plan for use across multiple projects. The quality and longevity of data are highly influenced by adherence to a proper data management plan.
  - Assign a properly-trained data steward to implement data management activities, including metadata development, data maintenance, and data archival.
  - Back-up data following standard protocol. Do not rely on data catalogs alone for backup purposes; data producers should always possess copies of their data.
  - The user should become familiar with data disposition policy outlined in the data lifecycle documentation. From a data manager's perspective, data considered obsolete by some users may still have use, whether data reside in a static file or records within a live database. When managing files preserve old versions (e.g., make new file versions, archive old versions). For records in a live database being edited, configure the database to preserve all data (e.g., versioning, active record fields, timestamp, etc.)
  - Use established data standards wherever possible (some guidelines are available at USFWS GIS Contracting Guidelines).
- Documentation
  - follow metadata recommendations (below)
  - track and document input data (e.g., clipped LandFire vegetation data) used to create derived data products (i.e., model output datasets)
  - whenever possible, reference publicly accessible data, or store and catalog input data when cataloging output
  - document all non-trivial processing steps, such as transformation of data
- Sharing
  - to maximize the use of data, provide the following everywhere possible:

        - data for download
        - data and map services for consuming the data in other applications
        - metadata fully documenting the data
        - records in one or more data catalogs,, widely disseminated
    - ensure data have undergone appropriate review and approval for sharing and distribution
    - ensure correct distribution rules for cataloging and dissemination data (public versus private authorization)
    - determine the rules of the catalog, such as longevity of the catalog, stewardship for cataloged data, and data maintenance requirements
    - assume that publicly-cataloged data may be disseminated to other applications, including other catalogs
    - other applications may reuse and manipulate cataloged data for derived products
    - sharing of large datasets (e.g., NetCDF climate data) may require special considerations for posting data to data catalog platforms and access by end users
    - data catalogs may have limits on file size or format; review data catalog platform documentation or contact platform administrators
    - consider cataloging multiple data formats for a dataset to ensure longevity and easier transformation from antiquated formats; for example, a csv file containing x, y coordinates may be transformed to other future formats while transforming a shapefile with the same coordinates may be more difficult
    - for final data that should not be manipulated (for example, a presentation), provide data in a format such as PDF
    - be aware of proprietary or open-source formats and APIs for data and metadata transfer; proprietary formats or APIs may require special licenses or become inaccessible as platforms evolve

Metadata

Metadata standards provide rules for describing data, including structure, content, and file format. Standards enable interoperability between systems by standardizing the sharing and semantic interpretation of documentation. Within the United States, the main metadata standard is transitioning from the Federal Geographic Data Committee - Content Standard for Digital Geospatial Metadata (FGDC CSDGM) to the International Organization for Standardization (ISO) suite of metadata standards. Eventually, the ISO 19115 and its accompanying standards

will replace the FGDC CSDGM as the official metadata standard for Federal agencies. For more information about the FGDC endorsed ISO metadata standards, please see: http://www.fgdc.gov/metadata/geospatial-metadata-standards#fgdcendorsedisostandards

Good Metadata Practices

- make the documentation of data a priority in any project that creates or uses data
- strive to be as complete and detailed as possible; information that is easily at hand while implementing any project is often unobtainable once data have been cataloged and archived elsewhere
- use accepted metadata standards; use the ISO standards everywhere possible for new metadata
- consider using available tools to generate standards-compliant metadata
- use available resources to produce current and relevant metadata (examples at: USFWS GIS Contracting Guidelines)
- assign a properly-trained metadata steward to maintain metadata records
- assume that publicly-cataloged data may be disseminated to other applications, including other catalogs
- metadata ownership, provenance, metadata update date (version control) information is vital for tracking the records
- when submitting data to any data catalog (Data Basin, ScienceBase, data.gov, etc.), verify original metadata (i.e., FGDC XML metadata) have been properly transferred and translated into data catalog elements or fields within that catalog

Tools

General Considerations

There are a few key questions that groups should consider when planning to find and use existing or developing new tools:

- What are the goals of the tool?
    - What questions are you (or your users) trying to answer with this tool?
        - Are these questions thematically or geographically specific (e.g., species-specific decision support tool)?
        - Are these questions more general (e.g., summary of feature and raster datasets within area of interest)?
        - What level of technical complexity does it require?

- Who is your audience?
    - What level of technical complexity is appropriate for the audience (e.g. what level of technical capacity do the users have)?
    - How many users will use this tool?
    - How often will they use this tool?
    - How will you train them to use the tool?
- What is the mid and long-term plan for this tool?
    - What is the use timeframe for this tool?
    - Is there a maintenance plan?
    - How is the tool going to be maintained given staff turnover?
- Within what contextual constraints must this tool operate?
    - Are there requirements that control the location where the tool can be run (e.g., desktop operating systems, browsers for web-based tools, etc.)?
    - Are there data format limitations?
    - Are there other security or infrastructure requirements?
    - Are programming language or code frameworks required?

Finding Tools

Before beginning any tool development, we highly recommend surveying available tools to determine if what you need has already been built. We have occasionally invested significant effort in building a tool only to find that someone has already built one just like it, but we did not easily discover it during our initial searches.

Good resources to use when looking for tools include:

- Google Search
- IDMN Tools Catalog
- ScienceBase
- ESRI Code Galleries, organized by toolkit / programming language
- GitHub is a rich resource particularly for these accounts:
    - ESRI
    - MapBox
    - Toblerity
    - PyData
    - OpenClimateGIS project

    - Data Visualization Showcase
- Programming language / platform specific portals, including:
    - Comprehensive R Archive Network (CRAN): R statistics platform
    - Python Package Index (PyPI): Python packages
    - Macports / Yum: Unix/Linux package managers

If you are building tools, please list them on centralized, easily discoverable system such as ScienceBase. Quite often, projects and their results are showcased by LCCs on their web pages and other distribution media, but the tools developed or refined for those projects are not in easily searchable systems.

Discussing Tools & Tool Needs

A Tool Design and Evaluation group has been established by Griffin Groups to support discussion of tool ideas, needs, use cases, and more during the design phase of creating tools.

Stack Overflow is a particularly active question and answer site for well-established tools and platforms. Its content and focus is typically quite technical and is often more accessible to software development staff and may not be the most appropriate for general LCC staff. However, it is important for LCC staff to be aware of this site to direct more technical staff to it.

GIS Stack Exchange is a question and answer site focused on Geographic Information Systems and the broad range of open source and proprietary technologies used therein, with content intended for the GIS professional or software developer. Again, it is important for LCC staff to be aware of this site in order to direct GIS staff to it.

Industry or platform specific forums are also a potential source of information for discussing or troubleshooting tools. For example, ESRI provides forums for each of their major software products: ArcGIS Forums. In practice, these forums are often more focused on troubleshooting existing tools than discussing new tool design and use cases.

Establish networks with other experts, organizing discussions about technical topics of common interest.

Using Existing Tools or Building Your Own

Once you have an existing tool, a key decision is to determine if the tool sufficiently meets your needs. Often, this process leads to building a new tool from scratch, rather than leveraging existing resources (see this discussion). If you build a new tool, this can result in redundant effort, significant expense, and lack of a clear maintenance plan beyond the initial buildout. We recommend considering the following:

- Is the tool based on proprietary software or only functional within a proprietary platform? If so, what are the licensing costs and use constraints?
    - For relatively isolated applications (e.g., a single user executing it on their desktop), using seemingly expensive proprietary software may still be more cost effective than the investment required to build up and maintain a comparable toolkit. However, this is beginning to change with the proliferation of open source tools and components.
    - There is a tradeoff between risks of known products versus building something new, which depends in part on the amount of resources and timeframe for delivering new tools.
- Can the tool be adapted to meet your highest priority needs now and into the future?
    - Often, a tool may initially fall short of your wish list. The reality is that you may not be able to get some of the things on your wish list, even if you build it yourself, unless you have access to nearly unlimited resources. Instead, focus on the highest priorities that must be functional in the short term and the flexibility to grow to meet your next highest priorities in the future. Do not unnecessarily penalize a tool because it does not meet your complete vision more than a year out. Keep in mind your priorities will most likely change in the meantime.
- How much flexibility do you require within the tool?
    - Often, tools appear less flexible than desired by users during the design phase, leading many potential tool-builders to build their own tool. However, when a tool is designed to be too flexible it may increase user interface complexity. This complexity can reduce usability and therefore usage of the tool. We recommend changing the emphasis from what you want the tool to do (which usually leads to poorly defined, broad-brush statements) to what the tool must do in order to satisfy your highest priority requirements.
- Does the tool have an established community?
- Does the tool have active development and maintenance?

- What expertise does the tool currently require to maintain or extend it?
- What expertise does the tool currently require to operate?

Building Your Own Tool

If you are going to build your own tool, we highly recommend you review the Best Practices for Scientific Computing (Wilson et al. 2014).

Key questions to consider when developing your own tool:

- Who will build the tool?
  - Will it be existing staff, external or internal contractors, volunteers?
- Who will maintain the tool?
  - If maintenance cannot be guaranteed at the outset by the tool builder, then version control, documentation, and thorough use of style guides from the beginning are critical to ensuring the maintainability.
- Who will iteratively review the tool?
  - Frequent review of the tool in progress is important to keep development on track and achieve technical objectives while maintaining usability. Reviewers have a large impact on the quality of the review process.

Software development practices

Agile development

We recommend following an agile development process involving iterative and incremental development (Schaller et al. 2010). Rather than defining all the details required to engineer a tool in advance and rigidly sticking to them throughout the development process, it is better to focus on defining an initial set of priorities and measures of success. Divide the development process into smaller steps that produce interim progress toward the main priorities of the tool. Development teams can then focus on a smaller problem space within each cycle and include frequent review and reevaluation of the desired outcomes. This approach can refine the project direction to meet the main priorities of the tool (Wilson et al. 2014).

Programming languages

We recommend using modern programming languages that are appropriate for the problem being addressed. Using newer protocols will aid developer productivity without sacrificing code performance.

- Python is a clear favorite for scripting geoprocessing and spatial analysis operations.
- R is a clear favorite (among non-proprietary toolkits) for statistics.
- JavaScript is becoming a favorite for web-browser based toolkits.

Essential principles and style guides

We recommend following a consistent set of principles and style guides throughout the tool development process. Style guides help simplify development, ease documentation, and are essential for maintainability. Three critical principles are:

- Don't repeat yourself: in essence, there should only be one authoritative representation of the data or a given operation (Wilson et al. 2014, Wikipedia).
- Don't repeat others: many excellent 3rd party packages and tools exist; many of which undergo more rigorous testing than is feasible to replicate within a new tool. Look for opportunities to leverage existing components into new tools.
- Don't work alone: A developer working in isolation can have a tremendous negative impact on the maintainability of a given codebase. Use peer review of source code wherever possible, pair programming where appropriate, and actively seek out a network of similar experts to share ideas.

Style guides should describe conventions for:

- folder structure and file names
- code structure
    - naming conventions within code: variables, classes, etc.
    - modular design
- documentation
- language specific style guides
    - Wherever possible, use an established style guide. Example: Python PEP 8-style guide
    - Documentation or examples of existing packages can be a good starting point for elements of a style guide.

Testing and continuous integration

Testing is critical to high-quality tools. Often, testing is limited to informal tests performed manually during tool development. We recommend developing formal test suites as part of the development process. Test suites provide more value than the effort used to create them.

Properly constructed tests assure that changes to the code will continue to produce valid results and can help detect software and algorithmic errors earlier in the development.

We also highly recommend the use of continuous integration (CI) coupled with a version control system and test suite because it will help to ensure that smaller changes to the code are included in the main codebase more frequently. CI also reduces the number and complexity of integration challenges when components are developed in isolation for a long time. Automatically run test suites can catch bugs early. Services such as Travis-CI can provide this for repositories hosted in GitHub.

Designing Tools for Integration

In order to integrate tools, there must be a well-defined interface, called an API (Application Programming Interface) and clear understanding of tool inputs and outputs. The nature of an API depends on the tool context. For example, it could be a:

- command-line tool: called via command-line arguments
- web service (e.g., REST, SOAP, etc.): executed on the web
- code library or toolkit: called from within other source code

To best support integration between tools

- use existing API standards (e.g., Web Map Service), where possible
- document the API in a publicly accessible location to ensure that usage is clear
- simplify the format of data exchanges between tools to reduce the number and complexity of data transformations required within a chain of tools
- use principles of service-oriented architecture where appropriate

Sharing Tools

Sharing tools promotes tool use by fostering reusability and helps avoid redundant efforts across projects. To inform decisions early on about hosting, repository system, use of standards and style guides, etc., we recommend planning early in the development process of the tool.

Hosting

We recommend sharing open source tools on a publicly accessible code repository. In particular, we recommend the following platforms:

- GitHub: has the best package discoverability tools at the time of this writing
    - public repositories are free - private ones are not

- BitBucket: has the best code review tools at the time of this writing
    - public and private repositories are free; these systems base charges on the number of collaborators

Centralized Versus Decentralized Version Control

There are two types of code repository systems: centralized and distributed. A simple video that explains centralized versus distributed version control is available here. A centralized version control system, such as Subversion, is characterized by a single instance of the full codebase history, hosted at a single location. Users synchronize against this instance, and must pull any changes to be able to add their own. The single shared code base can be clumsy when multiple people attempt to fix bugs or add functionality in parallel. It is also easy for one user to check in bad code that stalls a project until it is later fixed and pushed to all developers. Only trusted users can contribute 'commits.' Thus, while a new user may be able to download the source code, non-team users have limited options for contributing.

In contrast, each user of a distributed version control system can have a personal copy of the entire codebase history. Social networking technologies are then used to allow multiple copies of the same repository to stay synchronized. Teams setup one authoritative repository, which only receives changes through a peer-reviewed request to pull new changes into the canonical repository. There are excellent tools provided for merging changes and addressing merge conflicts that arise. This peer reviewed 'pull request' model allows any user, trusted or not, to submit changes to the canonical repository. Users not previously known to the project's main developers can make their own copy of the project and develop changes to it. They can then submit those changes to the canonical repository with a 'pull request' to merge those changes. In this model, the core development team maintains authoritative control of the canonical repository and vets the changes from outside parties before merging them.

Documentation

Documentation is critical not only to the maintainability of a tool, but also to its ease of use and adoption within other tools or platforms. In addition to code-level documentation, we recommend following a documentation standard that includes a description of the following information about the tool:

Required elements:

- Tool name
- Tool version and last modification date
- Tool authors
- Tool point of contact
- Tool license or reference to that license
- Tool overview / purpose
- Tool requirements (required software, operating system, etc.)
- Tool installation process (if applicable)
- Tool inputs and outputs (format, interpretation, etc.)
- Tool execution example

Optional elements:

- Intended audience
- Documentation of internal methods
- Code samples
- Instructions for running tests
- Known issues and limitations

We recommend that the documentation be kept directly alongside the source code. The repository hosting platforms listed above provide good support for this through web pages generated from documentation files within the repository as well as wiki and hosting of static web pages that accompany the repository. Other tools are available to provide hosting of documentation for open source tools, such as Read the Docs.

Supporting Discoverability

In addition to the discoverability features of the above hosting platforms, we recommend the following to increase the likelihood that users will find the tool:

- add the tool to the IDMN Tools Catalog
- add the tool to ScienceBase
- add the tool to the packages catalog for the appropriate programming language (e.g., PyPI for Python)

Map & Data Viewers

The Visualization Team created an overview and decision matrix related to map viewers. The focus of this section is to provide additional technical details and options to consider when developing simple map and data viewers for use within a web browser. However, our general recommendation is to use existing platforms where possible.

General Considerations

- See tool general considerations above
- Is the viewer:
    - Policy oriented?
    - Research oriented?
    - Decision making oriented (e.g.; conservation oriented, funding prioritization, land acquisition)?
    - Public/stakeholder/partner oriented?
    - Does the viewer present the data such that it is easy for the user to understand and apply?
    - Does it allow users to ask questions?
    - Does it allow drilling down to source data with metadata?
    - Does it allow the user to ask questions outside the original intent of the viewer?
- Can the data be downloaded to address more complex questions?
    - How does it allow the user to consume the data in an efficient and informative manner?
    - Are web mapping services available?
- How familiar are your users with basic web mapping functionality and concepts?
- Are there data sensitivity issues? For example, is access to the data restricted to authorized users?

Hosting Spatial Data for Browser-Based Mapping

In general, there are three approaches to hosting spatial data for use in a map viewer:

- Image-based map services:
    - The map service returns a rendered image of the spatial data. Images may be pre-rendered (cached) to significantly improve performance, especially for complex data or styling. Examples: ArcGIS Map Service, Web Map Service and derivatives.

    - Dynamically rendered images require dedicated map server technology (e.g., GeoServer, ArcGIS Server)
    - Pre-rendered images require web-accessible folders in the simplest case, or map server technology in more complex cases.
    - We recommend this option if:
        - you already have access to map server technology
        - data or styling approaches are not simple
        - consuming in a map viewer that expects a map service
    - We do not recommend this option if:
        - data are very simple and can be displayed entirely within the browser
- Data services:
    - The web service returns structured data that can be displayed on a map within the browser. Examples: ArcGIS Feature Service, Feature Service.
    - We recommend this option if:
        - the data can be easily rendered within the browser (e.g., points)
        - a small amount of data can be manipulated in the browser at one time
        - the user needs to control data filtering and styling, as well as make additional queries against the data server
    - We do not recommend this option if:
        - visualizing large volumes of data at a single time
        - data are complex
        - displaying on a device (e.g., phone) that is not sufficiently powerful to render data
- Data files:
    - Data files hosted in web-accessible folders and can be displayed on a map within the browser.
    - Examples: GeoJSON, TopoJSON, or CSV files.
    - Simplified features can be created using existing tools, such as MapShaper.
    - We recommend this option if:
        - data files are small and easy to render within the browser
        - you do not require server side manipulation or data querying
        - using a map viewer that can easily render data from files
    - We do not recommend this option if:
        - data are complex or large volume

    - you require a map or data service
    - displaying on a device that is not sufficiently powerful to render data

Several map and data service hosting platforms exist, including:

- MapBox
- ArcGIS Online
- CartoDB
- ScienceBase

Otherwise, the following open source and proprietary platforms can be used to host map or data services:

- GeoServer
- ArcGIS Server
- MapServer
- TileStache

An evaluation of the pros and cons of each of the above platforms and server technologies is outside the scope of this document. We recommend that you explore the capabilities and licensing of each and choose a hosting solution that provides the appropriate balance between ease of use, cost, maintainability, performance, and scalability. We also recommend that you seek out input from experts in one of the above platforms, because they are likely able to highlight usability or scaling concerns not immediately apparent in the marketing materials or help documentation.

Overall recommendations:

- Use established map service standards to increase interoperability.
- Use the latest version of the standard, if possible. Several key updates have been made to the standards in recent years.
- Where standards permit customization or do not define a single specification for data transfer, we highly recommend that you focus on transferring data rather than transferring distilled representations of data. Follow existing patterns for usage, where possible.
- Where possible, register services with appropriate metadata into a data catalog (e.g., ScienceBase) to support the discoverability.

Browser-Based Map Viewer Toolkits

Several open source and proprietary map viewers exist, including:

- Leaflet
- OpenLayers
- Google Maps API
- ArcGIS API for JavaScript (note: other browser-based APIs provided by ESRI are not recommended, per ESRI web developers roadmap)
- ModestMaps

Each map viewer listed above works with both map services and data files, either natively or through 3rd party plugins. Each has a unique community, design direction, and pace of releasing new features. We recommend exploring examples of the above and comparing them against the requirements and visual design style of your viewer and website. The choice of viewer toolkit depends on the type and variety of displayed map and data services.

Browser-Based Data Visualization Toolkits

Several browser-based data visualization toolkits exist, but none currently match the flexibility and breadth of applications demonstrated by D3. We highly recommend exploring the example's directory. Many additional toolkits have been developed using D3, including the Dimensional Charting JavaScript Library.

For interactive charting, several toolkits are available, including:

- HighCharts
- Dygraphs
- NVD3
- Rickshaw

The data visualization showcase on GitHub includes many additional resources.

Similar to the above section on hosting spatial data resources, there are a couple of alternatives for hosting non-spatial data for use in visualization toolkits in the browser:

- data files
    - hosted on web-accessible folders
    - recommended if data volume and complexity are low

- data services
    - hosted on specialized servers that expose data via a web services API recommended if data volume or complexity are high

We recommend choosing a data visualization toolkit that best:

- adds clarity to the data using intuitive display of information, including color schemes, labels, and legends
- adds value to the data through interaction; presenting data dynamically presents several opportunities not easily realized in static charts, such as tooltips, mouse or touch interactions, and user-driven filtering of data
- guides the user through exploring the data; charts often do not stand entirely alone but can complement interpretive information

Where possible, we recommend seeking out input from data visualization experts to help refine the vision for data presentation, even if they are not being asked to build it. Data visualization is very challenging to get right for a diverse audience, and experts can help guide you through picking the right questions to use for exploring the data to thinking through the best way to create an integrated data experience for your users.

Platforms

Given the complexity involved in architecting and developing a platform, which is outside the scope of this document, this section will instead focus on recommendations related to integration between platforms. By using existing platforms, you can take advantage of existing and ongoing integration work: integration builds upon integration. Using existing resources also mitigates having to building platforms, viewers, and tools on your own.

Key factors that contribute to integration:

- institutional culture that recognizes that the value added by integration is greater than the values of the parts being integrated
- providing publicly accessible APIs
- adherence to established web standards for consuming and publishing data, map, and catalog services
- adoption of similar data models and representations for similar data types

- clear documentation of APIs available within the platform; supported web standards, custom extensions to web standards, or original data structures
- documentation specifically developed with software developers in mind, for purposes of integration and application development, goes a long way
- strategic initiatives which directly call for and support integration across platforms (such as was done within the IDMN)
- using similar underlying technology for key components (e.g., ArcGIS Server, GeoServer) provides an extra level of standardization
- clear communication via a blog or release notes that documents recent changes to the platform

To better support integration across platforms, we recommend organizations, such as LCCs:

- work with existing platforms and core technologies to develop a vision that will meet their needs across multiple platforms rather than developing something original
- support building additional functionality to existing platforms to better standardize consumption and publication of information resources according to established data and web service standards
- enlist technical specialists to provide expertise when evaluating existing platforms; often constraints and tradeoffs may not be easily apparent to non-specialists

With respect to the platforms discussed in detail above, we also recommend:

- registering resources (spatial data, tools, etc.) in ScienceBase for centralized discoverability; this includes:
    - registering all GAP map services with complete metadata in ScienceBase
    - register all original map services provided by the Western Landscapes Explorer in ScienceBase
- where appropriate, visualizing those spatial resources in Data Basin (Example: importing spatial resources to use within a given LCC's Conservation Planning Atlas in Data Basin)
- using the GAP online viewers for easy exploration and access to GAP species, land cover, and protected areas data across the contiguous U.S., Alaska, Hawaii, Puerto Rico, Virgin Islands, and U.S. Territories
- using the Western Landscapes Explorer for easy exploration and access to spatial resources within the Western US

- using the Climate Science Center Geo Data Portal and associated tools for summarizing climate data resources by areas of interest

In general, we recommend exploring the above options with careful consideration to the breadth and technical sophistication of your audience. Some may require greater expertise with technical concepts whereas others may be more easily accessible to a broader range of users.

Recommendations for IDMN Project and LCC Network

One of the clear trends we observed in the results of the survey conducted by the Survey and Synthesis team is the need for an efficient and effective mechanism for sharing information about tools, needs, and ideas across the LCC network. If implemented properly and integrated into existing communication mechanisms (e.g., Data Management Working Group), this could better leverage similar efforts across LCCs, reduce redundant work, and identify national cross-cutting measures.

We recommend that the larger IDMN project setup a web-based wiki system that includes an editable and easily maintained listing of tools, along with some descriptive information, potential use cases, and pros and cons. This wiki can serve as a nexus for highlighting tools of particular relevance to the LCC community.

We also recommend creating a forum system that LCCs can use for discussing tool needs and ideas, which may need to be limited to the LCC community to avoid potential conflicts with the contracting process.

We further recommend that the LCC network continue to support integration efforts between platforms and tools that are currently being used within the network. This effort should include a venue for direct communication between platform developers and key representatives of the LCC community, to ensure that integration efforts address the needs of the LCCs.

Appendix V – Outreach Team Evaluation

Introduction

The IDMN has involved users throughout the project to ensure products would have practical relevancy. Outreach included two primary tasks assigned to two teams. The IDMN User Training group oversaw IDMN training and extension activities. The IDMN Use Case team will be conducting use cases with real users in a pilot area to assess the value and relevance of IDMN products and outcomes.

The IDMN User Training group reported to the full IDMN team on progress made toward developing the outreach strategy. All IDMN groups were invited to shape the outreach strategy and assist with the implementation.

Outreach Goals

- improve awareness, trust and accurate understanding of project goals
- increase collaboration and communication efforts with potential partners
- promote the use of project products and feedback from end-users

General Outreach Objectives

- inform the development of IDMN principles
- encourage the use of these IDMN principles
- form a broader network of interest and support
- share resources and exchange ideas
- solicit new ways of collaborating that provide mutual benefits
- promote partnerships that extend beyond the life of the IDMN project (as funded by the National LCC)
- identify the level of technology transfer required by the targeted user groups and end-users
- identify the technical capacity of partners, user groups, and end users

Specific Outreach Objectives

- Focus on the use of IDMN to support specific projects and use cases that are appropriate to IDMN:
    - Interagency assessments and plans at regional and national level
    - Resource Management Plans and Ecoregion Assessments
    - 2015 State Forest and Resource Assessments
    - Landscape Conservation Design projects
    - Collaborative Landscape Forest Restoration projects
    - State Wildlife Conservation Strategies and Action Plans
    - National Forest Plans
    - Habitat Conservation Plans
    - Ecological Integrity Assessments
- Provide trainings so that users know the scope and appropriate use of data

Guiding Principles for IDMN Outreach

- Expand visibility and awareness
- Actively seek user participation and input
- Make an extended effort to include representation from all primary and secondary targeted user groups
- Employ a multi-faceted approach that utilizes a variety of strategies for outreach, communication, marketing, and training
- Provide multiple ways in which individuals can participate and contribute ideas
- Support continual updating and modifications of IDMN that reflects current technology and user needs

Generic User Groups

The Outreach team identified a number of user groups as IDMN targets.

Primary

The first-level user group included the following:

- decision-makers from public agencies (e.g., forest supervisors, refuge managers, etc.)
- conservation specialists from NGOs, public agencies, private industries, tribes
- land managers from public agencies, NGOs, private industries, tribes
- collaborative groups working at landscape-levels

Secondary

A secondary user group includes the following:

- community leaders
- scientists and researchers
- general public
- elected officials

Other user groups may emerge as the IDMN tools and data become more widely used.

Specific Groups to Target

The team evaluated specific organizations within these identified target groups.

Primary

Through multiple user interactions:

- Landscape Conservation Cooperatives
- USGS Climate Science Centers
- Agencies who are working in collaborative environments

Secondary

Through annual (or opportunistic) IDMN briefings:

- Statewide forest assessment groups (state forestry divisions, TNC)
- Federal landscape-oriented project and program groups, such as
    - BLM Rapid Ecoregion Assessments
    - LandFire
    - USFS Collaborative Forest Landscape Restoration Program
    - USFS Priority Watersheds Integrated Resource Restoration Program
    - USGS Climate Science Centers
- Congressional staff visits

These user groups contain staff in the following roles:

- Resource Managers (E.g., Resource chiefs, program leaders)
- Program leaders (E.g., Forest supervisor, superintendent, complex leader, regional staffs, agency heads)
- Support Staff (e.g., representing budget support, planning, LCC coordinators, Centers of excellence, NGOs)

- Scientists and other subject matter specialists
- Private landowners and concerned citizens

Outreach Phases

The team implemented outreach for the IDMN in phases:

- promote general awareness of IDMN and products,
- solicit input on IDMN products and services from targeted users,
- promote partnerships and use of IDMN products and services,
- train in use of IDMN products and services.

Phase 1 – IDMN Awareness

The general awareness phase included the following steps:

- Produce publicly accessible website with relevant materials
- Develop and promote the use of fact sheets for IDMN and project modules
- Develop and promote the use of 2-3 minute IDMN video on YouTube
- Conduct and make accessible IDMN “how-to” webinars
- Develop a constituent database for use by IDMN team members
- Produce and distribute external e-newsletter
- Produce posters
- Attend and present at appropriate conventions (e.g., NLCC workshops)
- Maintain a shared calendar
- Develop, archive and make accessible project presentations
- Conduct speaking engagements during and after project ends
- Author articles and guest columns
- Assemble media packets
- Develop key points
- Publish in peer reviewed literature
- Prepare a General Technical Report on the IDMN
- Create a matrix of project products (data, tools, analyses, etc.) and user types (technical, non-technical user)
- Create a web portal that provides access to IDMN data, models, tools, products, expertise

Phase 2 – Solicit Input

Once the general awareness phase completed, the team solicited input on the approach:

- Use project advisory groups to provide input on use and usability of products at multiple scales
- Use pilot project partners to provide input on use and usability of IDMN products and services at landscape scales
- Develop a list of questions (i.e. user needs) that we want to ask of potential project partners and supporters

Phase 3 – Promote Partnerships and Use

The next phase incorporates inputs from phase 2 and begins promotion activities:

- Provide project briefings to target organizations working on landscape assessments and conservation plans across "all lands."
    - USFS
    - BLM
    - NRCS
    - USGS
    - FWS
    - TNC
- Collaborative Landscape Forest Restoration Groups
- State forest and wildlife agencies
- Establish a project description and links on the national LCC website
- Develop and maintain partnership database

Phase 4 – Provide Trainings

The next phase, training, involved development of multiple activities:

- Produce training videos on products with scripts (ten minute)
    - What is LC MAP?
    - How to upload a map service to LC MAP/ScienceBase
    - Where do you navigate in LC MAP/ScienceBase to get the URLs
    - How to import and register your map service in Data Basin
    - Where do you put in the links to Data Basin
- Post in YouTube (databasin.org/videos)
- Conduct webinar training sessions and post on National LCC site
    - How to use LC MAP/ScienceBase
    - How to use Data Basin

    - How to use tools in the IDMN Toolshed
    - How to use Confluence
    - How to use Griffin Groups
    - How to become part of the IDMN
- Conduct educational meetings
- Conduct workshops

Phase 5 – Recommend enhancements based on user input

The last phase integrates the previous phases:

- Take the input from previous phases (1-4) and summarize in a semi-annual report
- Develop a formal process and mechanisms, including prioritization for making IDMN enhancements and adjustments
- Recommend changes to the IDMN leadership on an annual basis that can enhance IDMN and the user experience; this includes identification of new user groups.

Appendix VI – User Training Team Evaluation

Introduction

The original IDMN proposal identified the following goals for the User Training team:

- Develop materials (documents, webinars, videos) to support end users – initially ScienceBase-Data Basin integration
- Set precedent and best practices for documentation of LCC products for the future

In addition to the above goals, the User Training team identified the following deliverables:

- Evaluate if training material is available and accessible for the IDMN-affiliated tools
- Develop standard training best practices
- Develop training materials to support end-users with the ScienceBase-Data Basin Integration
- Develop an IDMN outreach strategy
- Identify future training opportunities

Review of Sources and Resources

Of relevance to the work of the IDMN User Training Team, the IDMN Survey and Synthesis Team (SST) conducted a survey in January 2014 with 22 LCCs participating. This survey highlighted the need for training on tools that encouraged data integration across the LCC landscape. Specifically, SST survey question 51 asked respondents to describe any training topics that they would find helpful for LCC's Data Management. Several of the respondents indicated they had difficulty distributing products due to capacity, expertise and lack of guidance. Question 52 asked respondents what data management issues they would like the IDMN to work on. Responses indicated the need for outreach and training on existing IDMN partner tools as well as the need for standardization across the LCCs using common platforms for delivery.

The IDMN User Training team developed an Outreach Strategy to address several needs in the SST survey. This outreach strategy creates awareness of the IDMN project, promotes collaboration with partners, promotes the use of IDMN products and provides training recommendations on IDMN partner affiliated tools.

We also leveraged the work of the IDMN Toolshed team by surveying the points of contact of IDMN partner tools identified by the Toolshed team. The purpose of this survey was to evaluate if training materials were available and accessible for tools affiliated within the IDMN. We have used the results of this survey to identify opportunities for training best practices.

Lastly, the IDMN Tool Development team worked to develop integration between ScienceBase and Data Basin. This integration imports ScienceBase services directly into Data Basin to enable LCC users to use Data Basin's analysis tools for enhanced visualization. In support of this implementation, the IDMN User Training Team provided training opportunities in the form of webinars, short-how-to videos, and a quick guide.

Approach

As a team, we identified five deliverables:

- Evaluate if training material is available and accessible for the IDMN-affiliated tools
- Develop standard training best practices
- Develop training materials to support end-users with the LC MAP-Data Basin Integration
- Develop an IDMN outreach strategy
- Identify future training opportunities

Deliverable 1: Evaluate training material

In February 2014, we surveyed the points of contact of the tools identified by the Toolshed team. This Training survey, administered via Survey Monkey, had a 68% response rate (17 respondents). The survey objectives included:

- Evaluate whether training material is available and accessible for tools affiliated with the IDMN
- Identify methods of training
- Identify opportunities for training best practices
- Identify future training opportunities to improve training

As part of this deliverable, we created a Training Matrix to catalog the IDMN affiliated tools and training available (Appendix VII).

Deliverable 2: Develop standard best practices

Another objective of the Training survey was to identify opportunities for training best practices. Respondents shared the following summarized list of best practices for maintaining training materials:

- Re-release training materials following release of a new version of the tool
- Provide step-by-step instructions for use and a description of an application using the tool
- Focus on webinars – use recorded webinars and screencasts to demonstrate the tool in workshops
- Avoid written materials because they go out of date so quickly
- Track updates and official version releases to the tool; changes in these versions are noted on the tool's download site and in the included documentation
- Design tool in a simple and intuitive manner so the need for training materials is limited
- Utilize standard templates that come with a tool platform (often the templates come with their own user support materials)
- Test training materials with non-developers – it is useful for non-developers to read and attempt to follow documentation written by developers to ensure clarity and comprehensive information

Respondents also shared how they would address training needs if users have varied levels of expertise:

- Workshops and user groups provide opportunities to tailor training to a variety of users
- User guides provide step-by-step instructions, and interested parties can contact the author for help if needed
- Offer introductory tours as well as more advanced sessions
- Offer topically focused workshops where learning the tool is a component
- Encourage mentorship and networking among users to support lower expertise users - users that are unable to use the tool with the supplied materials are encouraged to seek assistance from peers
- For users with various levels of expertise, provide support via phone and email
- Engage multiple staff that represent different audiences
- Utilize existing help templates and videos from other similar tools that offer a variety of help levels from how-to videos to in-depth searchable help
- Vary interfaces to the same services that are more or less 'opened-up' and keep documentation for each

- Simplify material and provide thorough information on the website

Respondents indicated the following to improve their current training and support:

- Create or update on-line videos describing how to get started, various components of the tool and typical workflows
- Improve written material and add examples to the user guide to help with interpretation of output and contextualization
- Include more contextual or embedded help with the tool
- Develop a list of suggested external (additional, but related) resources
- Provide more external links to tools to increase exposure so that it can be shared via 'word of mouth' and from one peer to another
- Publicize the release of the tool and providing detailed installation instructions and references
- Provide more frequent and smaller trainings
- Obtain expertise in-house to address small requested changes to the online tool
- Design the tool to make its use more self-explanatory

Deliverable 3: Develop training materials

The User Training team hosted an informal demonstration of the integration between ScienceBase and Data Basin webinar to the IDMN extended team on February 13, 2014. In addition to this informal webinar, the Integrated Data Management Network (IDMN) is hosting a series of webinars from June 12 to August 28, 2014 at 2 pm EDT/1 pm CDT/12 pm MDT/11:00 am PDT. Working team staff recorded and transcribed this webinar series and published the series through the LCC YouTube Channel.

- *ScienceBase* (Dell Long, Lei Ann Wilson & Don Brown, USGS): ScienceBase provides a collaborative data management platform (metadata catalog, data repository, product and project metadata editing and creation tools and web services) for LCCs.
- *Data Basin* (Tosha Comendant and Brendan Ward, CBI): Data Basin is science-based mapping and analysis platform currently in use by a number of LCCs, agencies, and organizations.
- *Integration in Action* (Brendan Ward, CBI; Tom Miewald & BJ Richardson, USFWS): For landscape-level analyses to succeed, researchers need data portals and catalogs to interact at many levels. This way users can discover and use distributed science products.

- *IDMN Toolshed* (Brendan Ward, CBI; Dell Long, USGS ; Ed Laurent, Connecting Conservation): The IDMN project developed several integration points between ScienceBase and Data Basin, a catalog and framework for evaluating tools available for the LCC Network, and best practices for developing tools to support integration
- *Science Project Tracking* (Deanne DiPietro, PointBlue Conservation): Successful project tracking requires agreements on data standards and integration points, allowing LCCs to identify and assess funded research projects at regional and national scales
- *Next Generation Data Integration Challenges* (Tim Kern, USGS; Sean Finn, GNLCC; Jim Strittholt, CBI): Future landscape-scale analyses will rely on more and larger datasets, more computationally intense models, and an ever increasing rate of data inputs and outputs. Integration activities must keep pace with these needs.

To augment the ScienceBase and Data Basin webinars, we have developed a "Quick Guide" and a ScienceBase ArcGIS Service video that is available on the USGS YouTube Channel. This video demonstrates how to upload a Service Definition file into ScienceBase and consume a REST service in outside applications such as Data Basin.

Deliverable 4: IDMN Outreach Strategy

The IDMN User Training developed an outreach strategy that is described in a previous section. The general outreach goals are:

- To improve awareness, trust and accurate understanding of project goals
- To increase collaboration and communication efforts with potential partners
- To promote the use of project products and feedback from end-users
- To disseminate information about the project

Deliverable 5: Identify future training opportunities

The Survey and Synthesis identified several training opportunities requested by the LCCs. Respondents of this survey requested training for:

- Data management best practices
- Guidance on how to check for compliance
- Metadata creation
- How to distribute data
- How to use the USGS data management lifecycle tools
- How to work with large data

Products, Outputs & Use Cases

We delivered the following products:

- Training Matrix to catalog the IDMN partner tools and training
- Summary of training survey results
- Training best practices
- Outreach Strategy
- End-user Training Materials for the LC MAP – Data Basin Integration
    - Quick guide
    - ScienceBase ArcGIS Service video
- Webinar series in (June – August 2014)

Recommendation

We recommend that the IDMN implement the proposed Outreach Strategy that includes developing an IDMN website providing access to the tools and training materials identified through the IDMN partnership. Training webinars on the various IDMN tools as well as identified training products should be archived and made accessible through the IDMN website.

Appendix VII – IDMN Training Survey Summary

Training Survey Objectives:

- Evaluate whether training material is available and accessible for tools affiliated with the IDMN
- Identify methods of training
- Identify opportunities for training best practices
- Identify future training opportunities and how to improve training

The survey was conducted in late February 2014, with a 68% response rate. The survey responses are available at the User Training space in myUSGS.

Status of Training Materials

We asked the POCs about the status of their tool training materials: completed, planned, or not available. Responses indicated that 70% of the tools have some form of training materials completed and available. These include:

- BISON
- Data Basin including Site Assessment Tool
- Dose response calculator for ArcGIS
- Geo Data Portal
- GRAIP
- Linkage Mapper Tool Box
- North Pacific Forest Landscape Connectivity Project
- ScienceBase
- Software for Assisted Habitat Modeling- SAHM
- Western Landscape Explorer Map Viewer

Four tools indicated that they did not have training materials available. Of these, the GAP Protected Areas Viewer, GAP Protected Species Viewer, and GAP Land Cover Viewer indicated that their tool was designed in a simple and intuitive manner so that users would not need training materials. However, they do provide assistance via phone and email for those

users who want additional support. The Metadata Wizard tool is currently in the planning phase and will have training available.

Target Audience

Participants were asked to identify the target audience for their tool. The majority of the tools indicated that they served multiple target audiences. The survey indicated that these tools are being used by subject matter experts, such as scientists, academics, consultants, GIS/data managers, etc.

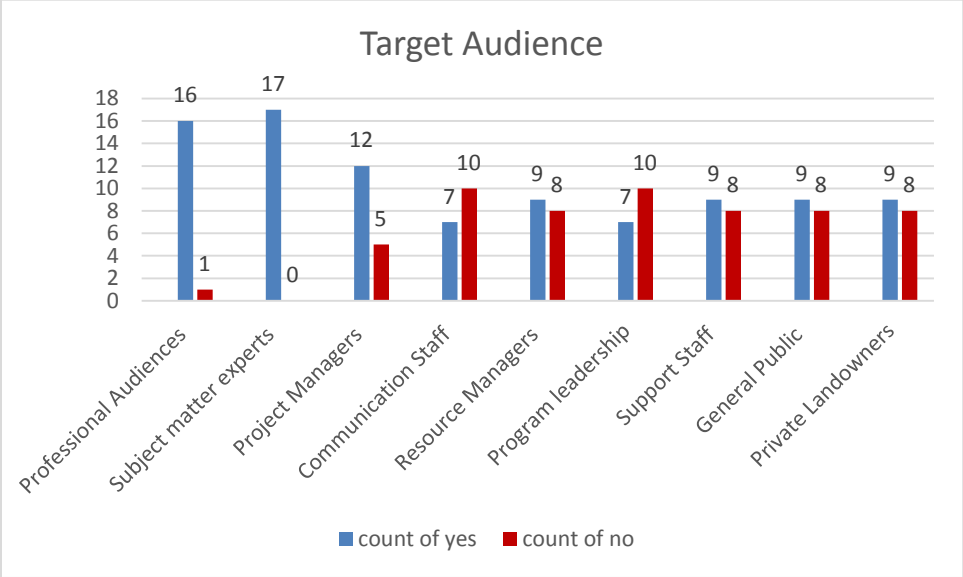


Figure 16. *Target audience survey results*

When asked to identify one or two primary target audiences, 83% (14) of the tools identified subject matter experts. Examples of subject matter experts included scientists, academics, and consultants, GIS/data managers who develop monitoring protocols, analyze and publish data. Secondary target audiences included professional audiences such as technicians and resource staff and support staff such as budget, planning, LCC coordinators as important secondary target audiences.

Technical Expertise of Audience

The majority of the tools classified that their users' level of technical expertise as average or advanced while only two tools classified their users as novice. An advanced user is one that is comfortable with information technology to use for analysis. An average user is one that uses information technology for browsing. We define a novice user as one new to information technology.

Tool Name	Level of Technical Expertise
Biodiversity Information Serving Our Nation	Average
Data Basin	Novice, Average, Advanced
Data Basin Site Assessment Tool	Advanced
Dose response calculator for ArcGIS	Average, Advanced
Excel® Spreadsheet Tools for Analyzing Groundwater Level Records and Displaying Information in ArcMap™	Advanced
GAP Land Cover Viewer	Average
GAP Protected Areas Viewer	Average
GAP Species Viewer	Average
Geo Data Portal	Advanced
GRAIP - Geomorphic Roads Analysis and Inventory Package;	Advanced
Linkage Mapper Toolbox	Average, Advanced
Metadata Wizard - An Easy to Use Tool for Creating FGDC-CSDGM Metadata in ESRI Arc Desktop	Novice, Average, Advanced
MODIS Phenology Image Service ArcMap Toolbox	Advanced
North Pacific Forest Landscape Connectivity Project	Advanced
ScienceBase	Average
Software for Assisted Habitat Modeling	Advanced
Western Landscapes Explorer Map Viewer	Average

Table 4. *Expertise required for tool use*

Recommendation: For the tools surveyed, we recommend training methods targeted toward the average or advanced user because they tend to seek out the information to solve a particular issue. Providing training documentation in the form of Frequently Asked Questions (FAQs), online help, tool tips, and short how-to videos should suffice for an average or advanced user. More novice users tend to require a more formal training delivery method such as a classroom, workshop, or one-on-one training.

Types of User Support and Training available

The majority of IDMN-affiliated tools supply some level of support and training. The most commonly used methods include email support, tool tips, pop-ups and user guides. The next most frequently used methods are phone support, FAQs, webinars, and short how-to videos.

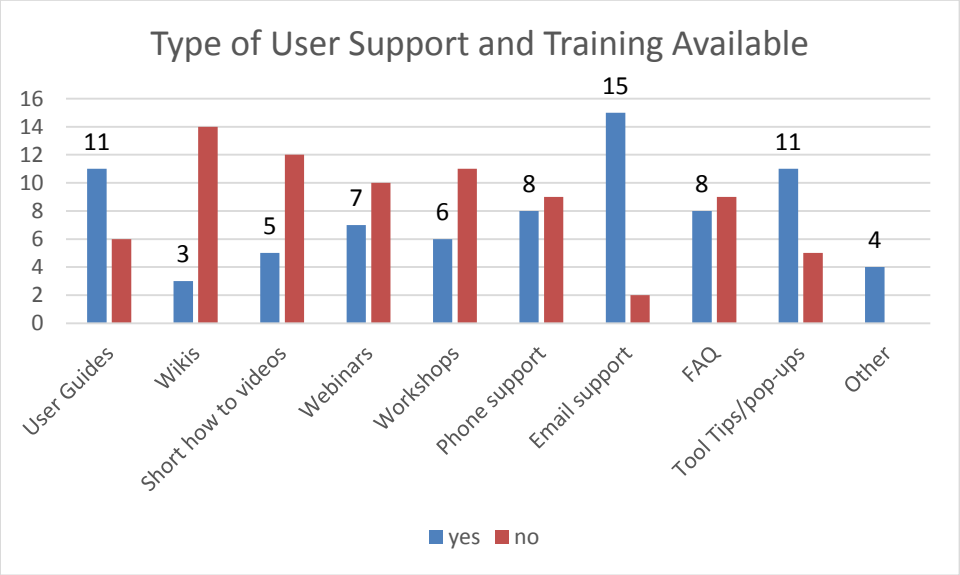


Figure 17. *User support available for tools*

Tool Name	Method of User Training
Biodiversity Information Serving Our Nation	Webinars
Data Basin	Webinars
Data Basin Site Assessment Tool	Webinars
Dose response calculator for ArcGIS	User guides
Excel® Spreadsheet Tools for Analyzing Groundwater Level Records and Displaying Information in ArcMap™	Published techniques and methods report
GAP Land Cover Viewer	Email support
GAP Protected Areas Viewer	Email support
GAP Species Viewer	Email support
Geo Data Portal	Don't know
GRAIP - Geomorphic Roads Analysis and Inventory Package;	User guides
Linkage Mapper Toolbox	User guides
Metadata Wizard - An Easy to Use Tool for Creating FGDC-CSDGM Metadata in ESRI Arc Desktop	Don't know
MODIS Phenology Image Service ArcMap Toolbox	Tool tips and pop-ups
North Pacific Forest Landscape Connectivity Project	User guides
ScienceBase	Email support
Software for Assisted Habitat Modeling	Workshops
Western Landscapes Explorer Map Viewer	Don't know

Table 5. *Tool support available*

Primary Tool Uses

Survey respondents also described the primary uses of their tool.

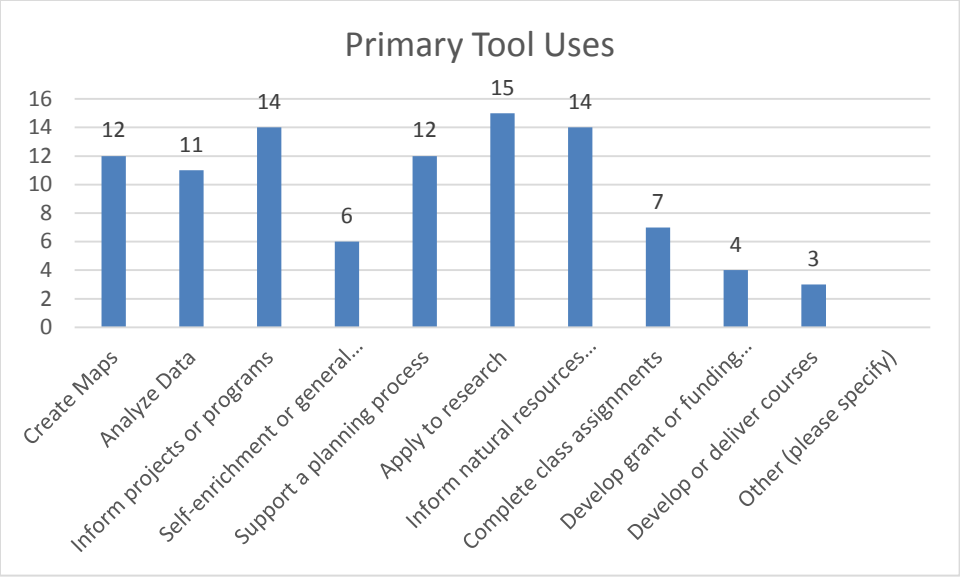


Figure 18. *Primary uses for tools*

We found that the IDMN affiliated tools were developed to serve multiple uses. These include, but are not limited to the following:

Tools that Create Maps (12): SAHM, Excel Spreadsheet Tools for Analyzing Groundwater Level Records, Data Basin, GRAIP, Linkage Mapper Toolbox, GAP PA Viewer, GAP LC Viewer, GAP Species Viewer, BISON, Western Landscapes Explorer, ScienceBase, MODIS

Tools that Analyze Data (11): SAHM, Dose Response Calculator for ArcGIS, Excel Spreadsheet Tools for Analyzing Groundwater Level Records, Data Basin, Data Basin Site Assessment Tool, GRAIP, Linkage Mapper Toolbox, BISON, Western Landscapes Explorer, GDP, MODIS

Tools that Inform Projects or Programs (14): SAHM, Dose Response Calculator for ArcGIS, Excel Spreadsheet Tools for Analyzing Groundwater Level Records, Data Basin, Data Basin Site Assessment Tool, Metadata Wizard, GAP PA Viewer, GAP LC Viewer, GAP Species Viewer, GRAIP, BISON, Western Landscapes Explorer, GDP, ScienceBase

Tools that provide Self-enrichment or general interest (6): GRAIP, Data Basin, BISON, Western Landscapes Explorer, GDP, ScienceBase

Tools that Support a Planning Process (12): SAHM, Data Basin, Data Basin Site Assessment Tool, GRAIP, Linkage Mapper Toolbox, GAP PA Viewer, GAP LC Viewer, GAP Species Viewer, BISON, Western Landscapes Explorer, GDP, ScienceBase

Tools that Apply to Research (15): SAHM, Dose response calculator for ArcGIS, Excel Spreadsheet Tools for Analyzing Groundwater Level Records, Data Basin, Metadata Wizard, GRAIP, Linkage Mapper Toolbox, GAP PA Viewer, GAP LC Viewer, GAP Species Viewer, BISON, Western Landscapes Explorer, GDP, North Pacific forest Landscape Connectivity Project, ScienceBase

Tools that Inform natural resources decision-making (14): SAHM, Dose response calculator for ArcGIS, Data Basin, Data Basin Assessment Tool, GRAIP, North Pacific forest Landscape Connectivity Project, Linkage Mapper Toolbox, GAP PA Viewer, GAP LC Viewer, GAP Species Viewer, BISON, Western Landscapes Explorer, GDP, ScienceBase

Tools that assist with class assignments (7): Data Basin, Linkage Mapper Toolbox, GAP PA Viewer, GAP LC Viewer, GAP Species Viewer, BISON, Western Landscapes Explorer

Tools that assist with developing grants or funding proposals (4): Data Basin, GRAIP, BISON, Western Landscapes Explorer

Tools that assist with developing or delivering courses (3): Data Basin, BISON, Western Landscapes Explorer

Other: GRAIP helps staff meet regulations. Users incorporate the Metadata Wizard in their spatial data development workflow, to build metadata.

Training and Support Challenges

The majority of respondents indicated that staffing and funding are the major challenges when it comes to training and support. Keeping training current, addressing multiple levels of expertise and communicating distinctions between similar tools was less of an issue.

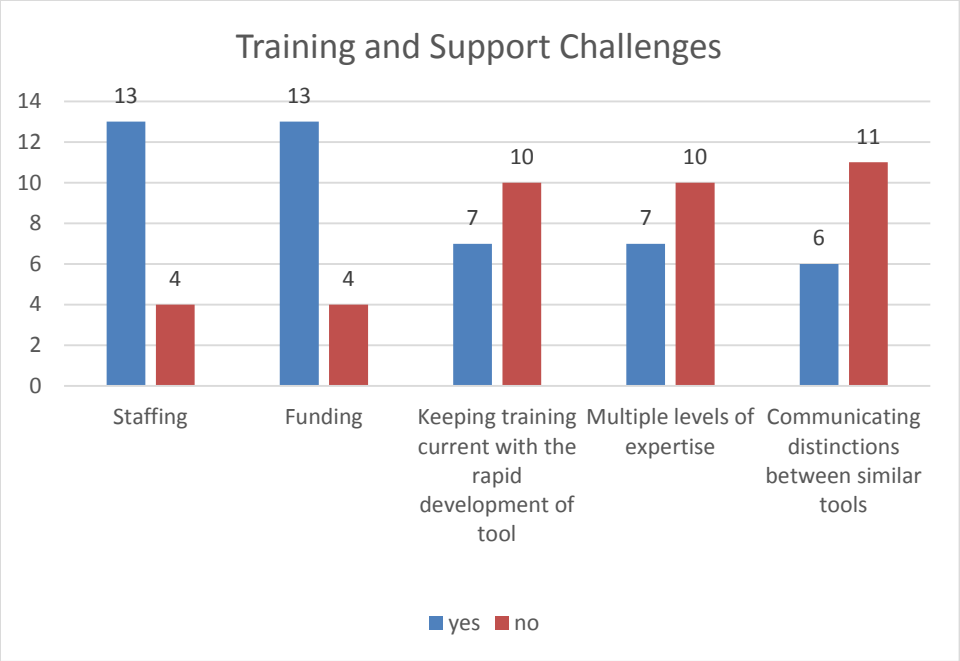


Figure 19. *Training challenges*

Currency of Training Materials

According to respondents, approximately 35% of the tools have tightly aligned training materials with software versions, while 41% of the tools have partial alignment with training materials. Additionally, 24% indicated no alignment of training materials with the current version of the tool.

Recommendation: To keep training materials current, we suggest selecting training formats that are dynamic and more easily updated. For example, host, record, and make accessible webinars that demonstrate the tool in its current version.

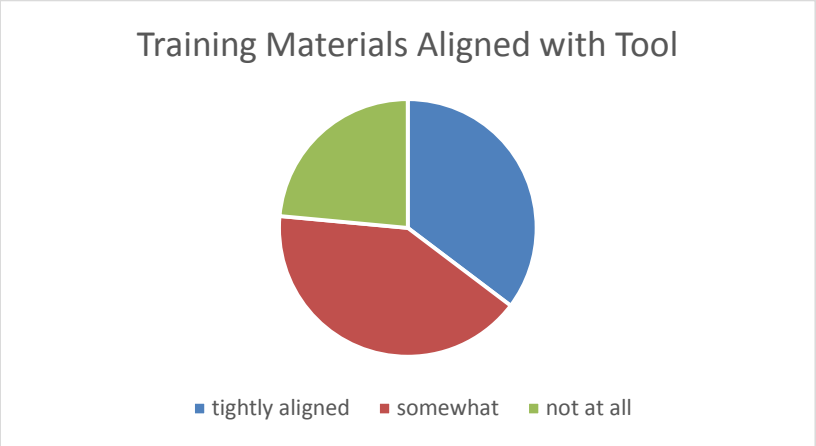


Figure 20. *Alignment of training materials with software version*

Best Practices

Respondents shared the following summarized list of best practices for maintaining training materials:

- Re-release training materials following release of a new version of the tool
- Provide step-by-step instructions for use and a description of an application using the tool
- Focus on webinars; use recorded webinars and screencasts to demonstrate the tool
- Demonstrate the tool in workshops and avoid written materials because they go out of date so quickly
- Track updates to the tool and official version releases; changes in these versions are noted on the tool's download site and in the included documentation
- Ensure that all materials are under subversion control on Google Code
- Design tool in a simple and intuitive manner so the need for training materials is limited
- Utilize standard templates that come with a tool platform (often the templates come with their own user support materials)
- Test training materials with non-developers; (it is useful for non-developers to read and attempt to follow documentation written by developers to ensure clarity and comprehensive information for a novice)

Respondents shared how they would address training needs if users have varied levels of expertise:

- Workshops and user groups provide opportunities to tailor training to a variety of users
- User guides provide step-by-step instructions. Interested parties can contact the author for help if needed
- Offer introductory tours as well as more advanced sessions
- Offer topically focused workshops where learning the tool is a component
- Encourage mentorship and networking among users to support lower expertise users.
- Users that are unable to use the tool with the supplied materials are encouraged to seek assistance from peers
- For users with varied levels of expertise, provide support via phone and email
- Engage multiple staff that represent different audiences
- Utilize existing help templates and videos from other similar tools that offer a variety of help levels from how-to videos to in-depth searchable help
- Vary interfaces to the same services that are more or less 'opened up' and keep documentation for each

- Simplify material and provide thorough information on the website to direct users

Respondents shared the following training solutions that have worked best for them:

- Conduct small group (10-15 people) workshops in which users are using the software hands-on. This discussion can get people off and running very effectively
- Develop on-line user guides
- Record webinars and make available online
- Include comprehensive and clearly written instructions with software products
- Promote one-on-one screen share sessions to set up software, configure validation schemas, and discuss process workflows and ways to address common challenges.
- A vested interested on the part of the trainee (i.e., they are being required to do something and need to understand how to do it properly) has also been a very critical piece to successful training efforts as well
- Prepare user guides and tutorials, plus email support via Google Groups
- Provide support via phone and email for each user
- Prepare website documentation - saves time not having to re-explain material

Respondents indicated the following to improve their current training and support:

- Create or update on-line videos describing how to get started, various components of the tool and typical workflows.
- Improve written material and add examples to the user guide to help with interpretation of output and contextualization.
- Include more contextual or embedded help with the tool.
- Develop a list of suggested external (additional, but related) resources.
- Provide more external links to tools to increase exposure and encourage peer sharing.
- Finalize the official publication announcing the release of the tool and providing detailed installation instructions and references.
- Provide more frequent and smaller trainings.
- Obtain expertise in-house to address small and quick requested changes to the online tool.
- Design the tool to make its use more intuitive.

Proposed tools to include in the IDMN Toolshed

Name of Resource (tool, platforms, viewers)	In IDMN toolbox?	Does your organization have the capability and vested interest in maintaining these tools and associated documentation?
Biodiversity Information Serving Our Nation	Yes	Yes
Data Basin	Yes	Yes, if funding/resources were available
Data Basin Site Assessment Tool	Yes	Yes
Dose response calculator for ArcGIS	Yes	No, the product does not change
Excel® Spreadsheet Tools for Analyzing Groundwater Level Records and Displaying Information in ArcMap™	Yes	Yes, if funding/resources were available
GAP Land Cover Viewer	Yes	Yes
GAP Protected Areas Viewer	Yes	Yes
GAP Species Viewer	Yes	Yes
Geo Data Portal	Yes	Yes
GRAIP - Geomorphic Roads Analysis and Inventory Package;	Yes	Yes, if funding/resources were available
Linkage Mapper Toolbox	Yes	Yes, if funding/resources were available
Metadata Wizard - An Easy to Use Tool for Creating FGDC-CSDGM Metadata in ESRI Arc Desktop	Yes	Yes, if funding/resources were available
North Pacific Forest Landscape Connectivity Project	Yes	Yes, if funding/resources were available
ScienceBase	Yes	Yes

Software for Assisted Habitat Modeling	Yes	Yes
Western Landscapes Explorer Map Viewer	Yes	Yes

Table 6. *Potential toolshed participants*

References Cited

Black, A. and S. Perin, Delivering the Science Synthesis: FuelsTools. 2007. http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1058&context=usdafsfacpub (accessed 5-14-10)

Gibney, E. and R. Van Noorden. 2013. Scientists losing data at a rapid rate. Nature News. [http://www.nature.com/news/scientists-losing-data-at-a-rapid-rate-1.14416?WT.ec_id=NEWS-20131224]

https://github.com/Snugug/north

Lankow, J., J. Ritchie, R. Crooks. 2012. Infographics: The Power of Visual Storytelling. John Wiley and Sons, Inc., Hoboken, NJ. 263 pp.

LCC IDMN Knowledge Transfer Team, Define Knowledge Transfer Audiences and Needs. 2013. https://griffingroups.com/pages/revision/7671 (accessed 1-23-14)

LCC IDMN Survey & Synthesis Team, IDMN Survey: Observations, Interpretations, Questions and Recommendations. 2014. https://griffingroups.com/groups/profile/16351/lcc-idmn-survey-and-synthesis-team (accessed 6/4/2014)

LCC IDMN Toolshed Team, Tool Catalog. 2014. https://docs.google.com/spreadsheet/ccc?key=0AqpTqTEO5LmDdG1iNHFFNlq3V2xsZzFnRTVhZ3A2Umc&usp=sharing#gid=0 (accessed 6/2/2014)

Meirelles, I. 2013. Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Rockport Publishers, Beverly, MA. 224 pp.

NOAA fisheries Northeast Region. Outreach Strategic Plan. http://www.nero.noaa.gov/nero/outreach/outreachplan-final.pdf (accessed 4-20-10)

Open Source Initiative. 2014. The Open Source Definition. [http://opensource.org/osd]

Scheller, R.M., B.R. Sturtevant, E.J. Gustafson, B.C. Ward, and D.J. Mladenoff. 2010. Increasing the reliability of ecological models using modern software engineering techniques. Frontiers in Ecology and the Environment 8(5): 253-260. [pdf]

Toman, E and B. Shindler, Wildlife Fire and Fuel Management: Principles for Effective Communication. 2006. http://www.treesearch.fs.fed.us/pubs/18689 (accessed 4-20-10)

Tufte, E.R. 1990. Envisioning Information. Graphics Press LLC, Cheshire, CT. 126 pp.

Tufte, E.R. 2001. The Visual Display of Quantitative Information (Second Edition). Graphics Press LLC, Cheshire, CT. 197 pp.

Vines, T.H., A.Y.K. Albert, R.L. Andrew, F. Debarre, D.G. Bock, M.T. Franklin, K.J. Gilbert, J.S. Moore, S. Renaut, and D.J. Rennison. The availability of research data declines rapidly with article age. 2013. Current Biology 24(1): 94-97. [DOI:http://dx.doi.org/10.1016/j.cub.2013.11.014]

Wilson, G., D.A. Aruliah, C. Titus Brown, N.P. Chue Hong, M. Davis, R.T. Guy, S.H.D. Haddock, K.D. Huff, I.M. Mitchell, M.D. Plumbley, B. Waugh, E.P. White, and P. Wilson. Best practices for scientific computing. 2014. PLOS Biology 12(1): e1001745. [http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.1001745]