



Long-Term Data Management Task Force Report

Natural Resource Report NPS/NRSS/IMD/NRR— 2017/1548



ON THE COVER

Aspen leaves on water. ©2005 James Disney

Long-Term Data Management Task Force Report

Natural Resource Report NPS/NRSS/IMD/NRR— 2017/1548

¹Marianne Tucker, ²Paula Capece, ¹Simon Kingston, ³Mark Hart, ⁴Bonnie Campbell, ⁵Nicole Athearn,
⁶Rob Daley, ⁶Kristin Legg

¹National Park Service
Inventory and Monitoring Division
1201 Oakridge Drive, Fort Collins, CO 80525

²National Park Service
Southeast Coast Inventory and Monitoring Network
135 Phoenix Road, Athens, GA 30605

³National Park Service
Great Lakes Inventory and Monitoring Network
2800 Lake Shore Drive East, Ashland, WI 54806

⁴National Park Service
Northern Colorado Plateau Inventory and Monitoring Network
Arches National Park, Building 11, Moab, UT 84532

⁵National Park Service
Great Rivers Cooperative Ecosystem Studies Unit
University of Missouri
105I Anheuser-Busch Natural Resources Building
Columbia, MO 65211

⁶National Park Service
Greater Yellowstone Inventory and Monitoring Network
2327 University Way, Suite 2, Bozeman, MT 59715

November 2017

U.S. Department of the Interior
National Park Service
Natural Resource Stewardship and Science
Fort Collins, Colorado

The National Park Service, Natural Resource Stewardship and Science office in Fort Collins, Colorado, publishes a range of reports that address natural resource topics. These reports are of interest and applicability to a broad audience in the National Park Service and others in natural resource management, including scientists, conservation and environmental constituencies, and the public.

The Natural Resource Report Series is used to disseminate comprehensive information and analysis about natural resources and related topics concerning lands managed by the National Park Service. The series supports the advancement of science, informed decision-making, and the achievement of the National Park Service mission. The series also provides a forum for presenting more lengthy results that may not be accepted by publications with page limitations.

All manuscripts in the series receive the appropriate level of peer review to ensure that the information is scientifically credible, technically accurate, appropriately written for the intended audience, and designed and published in a professional manner.

This report received informal peer review by subject-matter experts who were not directly involved in the collection, analysis, or reporting of the data.

Views, statements, findings, conclusions, recommendations, and data in this report do not necessarily reflect views and policies of the National Park Service, U.S. Department of the Interior. Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the U.S. Government.

This report is available in digital format from the [Natural Resource Publications Management website](#). To receive this report in a format optimized for screen readers, please email irma@nps.gov.

Please cite this publication as:

Tucker, M., P. Capece, S. Kingston, M. Hart, B. Campbell, N. Athearn, R. Daley, and K. Legg. 2017. Long-Term Data Management Task Force report. Natural Resource Report NPS/NRSS/IMD/NRR—2017/1548. National Park Service, Fort Collins, Colorado.

Contents

	Page
Figures.....	v
Executive Summary	vi
Acknowledgments.....	vii
List of Terms.....	viii
Introduction.....	1
Overview of Process and Methods	1
Fundamental Objectives.....	4
Program accountability - maximize.....	4
Utility to parks and networks - maximize	4
Broad science utility - maximize	4
Development-related effort and cost - minimize	4
Post-development effort and cost - minimize.....	4
Requirements	4
Alternatives	6
Status Quo	6
Minimally Compliant	6
Park Focused	6
Remodeled Data Store.....	6
LTER/PASTA Node.....	6
Centralized Interface and Administration	7
Earthsoft (EQuIS).....	7
Primary Findings.....	8
Remodeled Data Store Alternative	8
Recommendations.....	10
Limitations and Risks	11
Limitations.....	11

Contents (continued)

	Page
Limited details identifying project constraints	11
Limited time to evaluate all objectives, requirements, and possible alternatives.....	11
Limited details for LTER/PASTA and EQuIS	11
Differing understandings of alternatives and their capabilities	11
Lack of cost and time limits	12
Relative effort estimates	12
Risks	12
Risk of externally developed software	12
Risk of missing important scientific features	12
Risk of implementation delays	13
Additional Considerations and Findings.....	14
Several actions will be necessary regardless of which alternative the IMDLT chooses	14
Policies need to be clearly defined and accessible as well as enforceable	14
Metadata must be clearly defined for tabular and non-tabular data; standardization aids in this effort.....	14
Establishing a single, authoritative source for protocol ecological data is critical to ensure the most accurate data are offered.....	14
Requirements and recurring themes should be leveraged heavily for the design and development of any solution	14
Engaging IMD network personnel before arriving at policy and process decisions is one of the best signs of decision success.....	15
Open communication and collaboration is key to successful implementation.....	15
Use of small, focused groups facilitated reaching Task Force milestones	15
Implementation	16
Uncertainties.....	17
Cultural Change.....	17

Figures

	Page
Figure 1. Long-Term Data Management Task Force timeline.....	2

Appendices

Appendix A. Structured Decision Making – Summary Report	A-1
Appendix B. List of requirements developed by the Long-Term Data Management Task Force	B-1
Appendix C. Details for Alternatives.....	C-1
Appendix D. Workflow Diagrams for Alternatives.....	D-1
Appendix E. Long-Term Data Management Task Force Membership.....	E-1

Executive Summary

From November 2016 through July 2017, a group of National Park Service (NPS) Inventory and Monitoring Division (IMD) staff convened as a Long-Term Data Management Task Force to recommend a future path for managing the division's scientific data and information.

Data are core to the IMD program. The need to assess long-term data management planning arose from concerns about data preservation and documentation, as well as the challenge of increasing the discoverability, accessibility and usability of IMD data sets for broader conservation initiatives. The desire to incorporate existing and emerging best practices for science information management, and the need to meet new governmental policies and mandates, also served as impetus for the formation of the Task Force.

The Structured Decision Making (SDM) process allowed the Task Force to use a proven tool to arrive at decisions, maintain transparency, and document each step of the process from initial problem definition through to final recommendations.

Of the seven alternatives that the Task Force developed and considered during the final evaluation, the Remodeled Data Store alternative consistently achieved the highest scores among individuals and across differing exploratory analyses. Remodeled Data Store leverages the existing Data Store application (<https://irma.nps.gov/DataStore>) and allows networks to continue many of their current practices while meeting long-term goals. Through a phased implementation of policies and enhancements, the Data Store becomes IMD's centralized and authoritative data source. Phased implementation would include policy establishment, development of tools to facilitate policy compliance, and modification of the Data Store application to accommodate those policies and tools.

This report and the Remodeled Data Store recommendation were presented to the IMD Leadership Team at their meeting in September, 2017. After review and discussion, the team unanimously approved the recommendation. The team also authorized the immediate formation of two workgroups: one to define metadata standards; the other to define authoritative source requirements for IMD.

The division will now direct its focus on implementation. Key considerations include: an initial focus on establishing policies; a staged implementation approach; reliance on requirements already defined by this Task Force; the creation of a similar implementation task force; and a solid communication strategy both internally and IMD-wide throughout all phases. The methods used to implement the recommendations must be carefully evaluated, realistic, and achievable in light of limited staff and budgets. Assembling and maintaining collegial, coordinated teams with representation across IMD disciplines and locations will be imperative for the success of this effort.

Acknowledgments

The Long-Term Data Management Task Force was a successful collaboration among its members, who each brought their valuable time, attention, and experience to the project.

We acknowledge and thank Isabel Ashton, Margaret Beer, John Boetsch, Peter Budde, Bonnie Campbell, Rob Daley, Michael DiSibio, Brent Frakes, Mark Hart, Angela Jarding, Bill Johnson, Dan Kocol, Adam Kozlowski, Kristin Legg, Scott Miller, Lisa Nelson, Judd Patterson, Tom Philippi, Geoffrey Sanders, and Marianne Tucker for their long hours and hard work.

The expertise and dedication of Paula Capece and Simon Kingston, in particular, were fundamental to the success of the project. Nicole Athearn gave her time and knowledge of Structured Decision Making generously and with enthusiasm. Thank you.

Our appreciation also extends to the supervisors of Task Force members, who gave them the time and latitude to participate in this project, and to the Inventory and Monitoring Division Leadership Team for their support.

List of Terms

Some of the terms in this report are used in the specific context of the Structured Decision Making process and are defined below for clarity.

Alternative - A possible management action or approach available to the decision maker to solve a problem. If selected, an alternative becomes the decision.

Attribute - An element of the decision model used to assess achievement of an objective.

Consequences - The results of different management actions in terms that are relevant to the fundamental objectives of the problem at hand.

Decision model - A tabular or diagrammatic representation of the full suite of factors affecting a decision for the purpose of displaying, exploring, or manipulating important relationships among the decision factors. A decision model is intended to incorporate only the information most relevant to help a decision maker gain a clear picture of the problem, complete with fact-based and values-driven components.

Elicitation - The process of obtaining, deriving, and synthesizing opinions of experts and authorities on a subject, used when there is uncertainty due to data being insufficient, unattainable, or otherwise lacking. Expert elicitation allows for parameterization of an "educated guess" for the respective topic or issue. It is a commonly used consensus methodology in formal decision making.

Fundamental objective - An explicit statement of desired outcome—the ultimate goal of a decision. The fundamental objective is the end that a decision will be directed to achieve. It is often stated as the thing that matters with a direction of action (i.e., minimize, maximize).

Performance measure - Direct means by which an alternative can be evaluated against attributes. Performance measures provide the units or scales for evaluation.

Requirement - A need, a stipulation, or a necessary condition. Solution requirements refer to expected features or functional specifications.

Structured Decision Making (SDM) - A type of normative decision process and a formal method for analyzing decisions by breaking a problem into a series of fact-based and values-driven components, which are selected and weighted in a transparent, systematic, and replicable manner. A key feature of SDM is its iterative approach, allowing for incorporation of new information or parameters as the process unfolds.

Tradeoff - A balance achieved between two or more desirable but incompatible features; a compromise.

Other terms used in this report:

Ecological Metadata Language (EML) - A metadata standard developed by and for the ecology discipline.

EQuIS - A commercial software product developed by EarthSoft, Inc., and designed for managing environmental sampling data.

IMD - The National Park Service Inventory and Monitoring Division, part of the Natural Resource Stewardship and Science Directorate.

IMDLT - The Inventory and Monitoring Division Leadership Team

LTDM - Long-Term Data Management

LTER - Long Term Ecological Research Network. Created by the National Science Foundation in 1980 to conduct research on ecological issues that can last decades and span large geographical areas.

PASTA - Provenance Aware Synthesis Tracking Architecture. Provides a repository for all LTER data, including both metadata and data.

Introduction

This document presents the findings of Inventory and Monitoring Division (IMD) data managers, scientists, and program managers on their investigation of the best alternative for long-term management of IMD inventory and monitoring data.

In 2016, IMD conducted an evaluation of the EQuIS commercial software product. The process and end results made clear that a broader review of alternatives (e.g., open-source or other commercial systems, revisions to existing resources and infrastructure) was required to address the key data management needs for the division. The EQuIS evaluation report ¹ recommended the creation of a Long-Term Data Management (LTDM) Task Force.

In September 2016, the IMD Leadership Team (IMDLT) authorized a task force to develop recommendations for a long-term solution to manage IMD's heterogeneous ecological data.

The recommended solution had to fulfill several key needs:

- support the informatics needs of scientists
- meet the legal and policy mandates for information technology
- be sustainable
- ensure both the preservation of and access to IMD science data.

The solution was expected to be a combination of policy, process, technology, and training that would be re-evaluated annually. The Long-Term Data Management Task Force (“Task Force”) was created in October, 2016.

Overview of Process and Methods

The Task Force comprised over 20 members, including IMD scientists, data managers, and program managers. Both field and central office staff were represented. Early on, the Task Force decided to use the Structured Decision Making (SDM) process as a way to arrive at a recommended solution. Using SDM, Task Force actions were transparent, consistent, and reproducible. This methodology provided a way to simplify the evaluation of multiple alternatives among a relatively large group whose members worked remotely.

Major milestones established in the SDM process were:

- define the problem
- determine the objectives (including attributes and performance measures)
- identify alternatives

¹ Frakes, B., M. Tucker, and D. Kocol. 2016. EarthSoft EQuIS evaluation. National Park Service, Inventory and Monitoring Division, Fort Collins, Colo. [[Data Store record](#)]

- predict the consequences
- evaluate tradeoffs.

The Task Force also spent considerable time developing solution requirements as a necessary item for an effective recommendation (see Figure 1). The end products delivered to the IMDLT on August 2, 2017, consisted of the final Task Force recommendation included in this report, the final SDM decision model, and supporting materials.

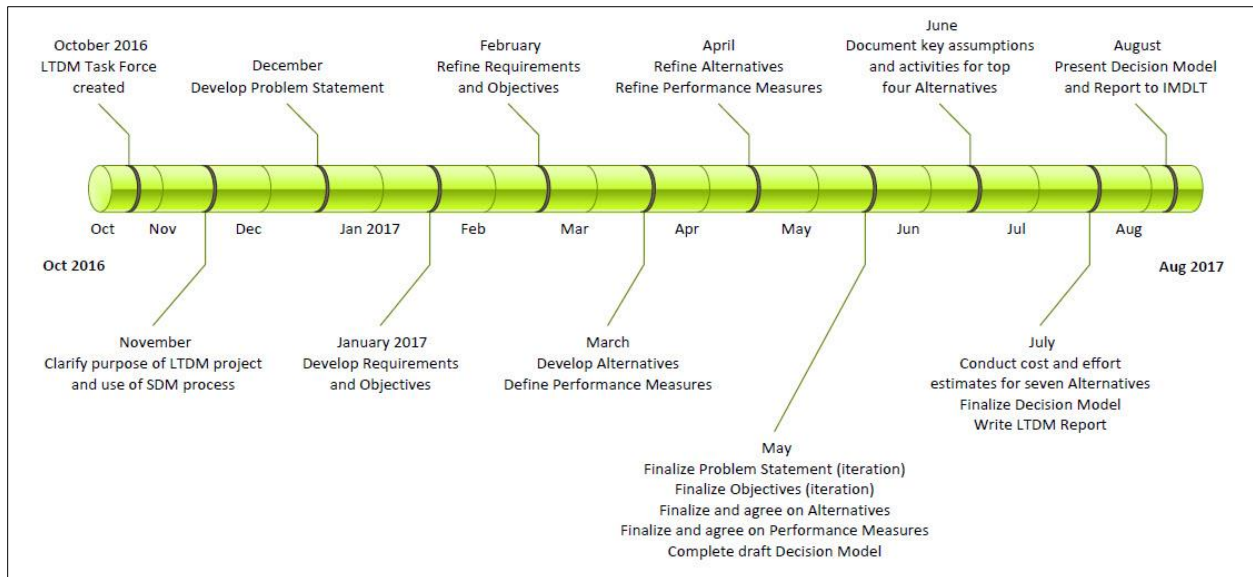


Figure 1. Long-Term Data Management Task Force timeline.

The IMDLT presented six long-term data management considerations and asked the Task Force to develop concrete and realistic requirements to address each. These considerations were:

1. data are discoverable
2. data are accessible
3. data are usable (documented for proper use)
4. data are of known quality
5. data are preserved
6. data are secure
7. *data support IMD science needs.*

The seventh item was repeatedly advocated by science representatives in the Task Force. While this overlapped to some degree with the six initial considerations, it was important as a stand-alone consideration due to the unique needs inherent in using data for long-term ecological analyses and was included as part of this review.

The Task Force developed the following problem statement to concisely guide the decision-making process toward a recommendation:

The LTDM Task Force will recommend an effective solution for managing long-term ecological data to support park management and broader conservation science.

During the first six months of the project, the Task Force defined requirements, as well as three of the five common SDM decision elements: problem statement, objectives (including performance attributes and measures), and alternatives. The process of defining each element was iterative, in that the deeper the problem was explored and examined, the clearer some elements became. This clarification required re-evaluating each previous element for each iteration. The iterative revising of the problem to incorporate details learned throughout the process is one of the primary advantages of using SDM.

The final two common SDM decision elements, consequences and tradeoffs, were performed using formal elicitation techniques.

An in-depth explanation of the SDM process used to develop this recommendation is presented in Appendix A.

Fundamental Objectives

The Task Force determined that there were five fundamental objectives for determining a recommended data management solution to the IMDLT. They are (in no particular order):

Program accountability - maximize

The solution supports efficiently answering questions about the who/what/when/where of IMD operations, achieves compliance with departmental and bureau laws and policies, and maximizes adherence to industry best practices.

Utility to parks and networks - maximize

The solution supports the ability of networks to provide parks management-relevant scientific information on vital signs and inventories.

Broad science utility - maximize

The solution supports uses of IMD data and information at broader scales, and beyond those uses captured by the core objectives of vital signs and inventories.

Development-related effort and cost - minimize

The solution minimizes effort and cost required to develop and deploy the new data management solution, from concept to fully operational status.

Post-development effort and cost - minimize

The solution minimizes the effort and cost required for ongoing maintenance and operation of data management and the solution.

Since the fundamental objectives were somewhat intangible, the Task Force formulated several performance attributes and associated performance measures to better define the objectives and provide a concrete mechanism to evaluate potential alternatives. In essence, the performance attributes and performance measures were the judging criteria for the decision (see Appendix A).

Requirements

At the inception of the Task Force, members were asked to discuss and collaborate on generating requirements (including definitions) for a long-term data management solution. Initially, these requirements were intended to be used to assess solutions.

The process of generating requirements was introduced early in the LTDM process. At the November, 2016 Data Management Training, ideas for long-term IMD data management were solicited using breakout group exercises. This generated a significant number of requirements for the Task Force to consider. Network participants on the Task Force also reached out to staff in the regions they represented to gather additional requirements. Soliciting requirements in this way served as an effective outreach tool, established critical buy-in from network staff, and provided an opportunity for stakeholders to be heard.

Once the SDM process was decided upon, various small groups reviewed, refined, dropped, de-duplicated, and added new requirements, aiming to arrive at an explicit requirement statement and to document underlying assumptions for each remaining requirement.

While this review occurred, the groups were intent on the outcomes of (1) documenting all of the work for transparency; and (2) preserving the original intent of any modified requirements. A science requirements group reviewed the draft “science vision” document and further distilled new science requirements. Another small group classified all of the remaining requirements as mandatory or desired.

Several efforts were made to group requirements under objectives. However, we recognized that as both requirements and objectives evolved, the idea of grouping requirements into objectives and using requirements to score objectives (and subsequently, our alternatives) was not feasible. Each alternative could fulfill a requirement in various ways. How requirements are met depends, to some degree, on the overall approach a particular alternative will take in addressing the full complement of fundamental decision objectives.

As a result, requirements were used as “touch points” to assess alternatives, and they helped reveal where an alternative description inadequately addressed a requirement and could be improved. Through this process, an overall picture developed of whether an alternative “passed” (met the needs of a requirement), “failed,” or fell somewhere in between.

Additionally, requirements served as a seed for generating actions that needed to occur to implement the alternatives.

Over the course of the Task Force’s work, the number of requirements was eventually reduced from over 180, to 79 (56 designated as mandatory, 23 designated as desired). This was accomplished through a process of consolidating duplicates and removing requirement candidates that were not or could not be used effectively in assessing the viability of a solution (see Appendix B).

Alternatives

Once requirements and objectives were refined and actions to address requirements were created, the Task Force turned its attention to generating alternatives. Another small group convened to draft alternative “packages.” Packages were composed using the aforementioned actions, and also consisted of policies, processes, technologies, and training that described paths we might take to achieve our objectives. The group aimed to describe the alternative packages sufficiently to distinguish them from one another.

The Task Force refined alternatives in its face-to-face meeting in Fort Collins in May 2017. It arrived at a final suite of seven alternatives for further consideration and scoring using performance measures (see Appendix C). The Status Quo alternative was retained as a point of comparison with the other alternatives. The seven alternatives are:

Status Quo

No changes to current IMD data management operations. Every network does its best; central office issues limited guidance and provides assistance to networks when needed.

Minimally Compliant

Satisfy compliance (meet legal mandates, regulations, and directives pertaining to the federal government, DOI, NPS, NRSS, and IMD), but with minimal impact or change to network operations.

Park Focused

Meet minimally compliant requirements while maximizing delivery to parks in the individualized manner parks may request. De-emphasize central office overhead. Implementation may include lightweight enhancements to the existing Data Store application.

Remodeled Data Store

Enhance and revise the Data Store to include the best elements of LTER/PASTA and other systems. Networks can maintain their own databases and add these to Data Store as files, with the option of having them converted to a queryable data source. This queryable source conversion requires additional metadata (e.g., schema definition) and results in structured web services that promote broader consumption. Flags in the Data Store (e.g., visibility, protected status) identify content to expose as services. All systems for metadata, tabular data, and non-tabular data are linked and presented as a single source.

LTER/PASTA Node

Leverage existing PASTA open-source code and related tools (e.g., Morpho, Excel metadata templates) to stand up an NPS PASTA site on the Windows platform. Networks would maintain their own databases and would upload periodic data packages (with Ecological Metadata Language metadata) into PASTA in the form of flattened .csv files. Opt in as a DataOne node to exchange information with LTER, the National Ecological Observatory Network, universities, and other end-

users, with flexibility on the look and feel of public and internal pages.² Handle metadata and packages of tabular and non-tabular data through the PASTA system. Develop other NPS tools to hold and manage relationships among associated non-tabular files, protocol documents, and products (e.g., published reports).

Centralized Interface and Administration

Store network data associated with monitoring protocols in data center databases that are designed and maintained collaboratively with central office staff. Field databases and applications may be used, but networks must connect to centralized databases to load data (network-located data entry databases would be for temporary storage only). Networks use the Data Store to catalog the centralized databases, and Data Store holds and manages relationships among associated non-tabular files, protocol documents, and products (e.g., published reports).

Earthsoft (EQuIS)

Environmental Quality Information System (EQuIS) is commercial software from Earthsoft, Inc., with separate data tables enforced via electronic data deliverable (EDD) definitions for every protocol (i.e., Survey). Data file submissions are checked for formatting and valid values before being accepted into EQuIS. Automatic notifications and tools help users identify and correct submitted data not conforming to EDD parameters. EQuIS can process data files submitted in multiple file formats (e.g. txt, csv, xlsx). The central office helps networks configure Surveys in EQuIS, data tables and associated EDDs, and helps load data for existing protocols. Networks have the flexibility to use existing or adjusted tools and workflows to collect field data, but networks must promptly transform and submit or load field data to EQuIS. Networks use the Data Store to catalog the authoritative data sets stored in EQuIS, and Data Store holds and manages relationships among associated non-tabular files, protocol documents, and products (e.g., published reports).

At the May 2017 meeting, four alternatives consistently rose to the top of the scoring: Remodeled Data Store, LTER/PASTA Node, Centralized Interface and Administration, and EQuIS. Small groups provided additional details for the top four alternatives, and crafted diagrams that illustrated workflow and provided more information on general descriptions, assumptions, and necessary activities related to effort expended during development and long-term use. Another small group concurrently developed effort estimates for the final seven alternatives.

² The option to become a DataOne Node is available to IMD under any of the alternatives.

Primary Findings

The Remodeled Data Store alternative consistently achieved the highest weighted scores among all members of the Task Force. The pattern also held up across different exploratory analyses. The Task Force examined patterns in the highest-scoring alternatives under the following scenarios:

- individual scores with individual weights
- mean scores with equal weights including all attributes
- mean scores with equal weights excluding effort and costs
- individual scores with equal weights, and equal scores with individual weights.

In all cases a very similar pattern emerged, which helped bolster confidence in the decision model and process.

In decision analysis, models and their outputs are tools; however, the Task Force makes the recommendation, not the model. Model results always need to be checked against intuition—the unquantifiable understanding and experience each team member brings to the process. The Task Force members agreed that this particular result of the decision model did, indeed, conform with their intuition.

More variation was seen in the second- and third-highest-scoring alternatives. In most cases, the Centralized Interface and Administration alternative came in second place, with the LTER/PASTA Node alternative in third place. In the minority of cases in our analyses, this pattern shifted with LTER/PASTA Node in second place and Centralized Interface and Administration in third place. A majority of members of the Task Force were comfortable with this result. A few individuals expected LTER/PASTA Node to score higher than Centralized Interface and Administration more consistently.

Remodeled Data Store Alternative

The Remodeled Data Store alternative leverages the existing Data Store application to allow networks to continue many of their current practices, while meeting the long-term data management goals through a phased implementation of policies and enhancements.

The Data Store is enhanced to pull in the best elements of other ecological data management systems and provides full integration with the metadata standard IMD adopts. Networks already maintain their own databases; they can add these to the Data Store as files. Enhancements add the option of having datasets converted to a queryable data source (i.e., extracted into an enterprise relational database), which can power web services and external visualization and analysis tools. Existing functionality is extended such that all metadata, data files (tabular and non-tabular), reports, and downloads are linked together, discovered, and presented in a unified fashion. The Data Store becomes the authoritative source for all IMD vital signs and inventory data.

The Remodeled Data Store assumes a phased implementation that consists of policy establishment, development of tools to facilitate policy compliance, and modification of the Data Store application

to accommodate those policies and tools. As policies develop, networks would be required to meet metadata standards, meet compliance policies, and use consistent versioning and change-tracking procedures.

The Task Force has identified a set of standards, policies, and tools, best handled by focused workgroups, which would be essential for beginning implementation work for the Remodeled Data Store alternative. These tasks include:

- develop or adopt one or more metadata standards for describing the tabular and non-tabular data collected by IMD
- define ‘Authoritative Source’ and what it includes
- develop policy defining how a snapshot of data can or should be connected to protocols and SOPs at any given point in time; deliver enhancements and tools that allow Data Store to support such policy
- define compliance (protected records, data quality, record management) requirements
- develop tools and define the Data Store structures to present metadata, tabular and non-tabular data in a consistent and unified fashion
- develop policy and tools for correct use of the versioning and change-tracking features in Data Store for data, reports, and metadata.

Recommendations

The Long-Term Data Management Task Force presents the following recommendations:

- **Unanimously recommends** the Remodeled Data Store as the preferred solution for long-term data management for IMD.
- **Agrees** that implementation details, which will be determined in the course of project planning and implementation, will have a significant impact on long-term success and functionality of the solution. Many details either not relevant or not available during the decision process will need to be concretely defined leading up to and during project development. As a cautionary note, the Task Force emphasizes that *how* the alternative is implemented is at least as important as *which* alternative is implemented.
- **Stresses** reliance on the objectives, and especially the requirements, as important foundational items when proceeding toward implementation.
- **Recommends** phasing-in different components of the recommended alternative in a logical manner across the time period of development.
- **Recommends** building in a brief annual decision review to the project management cycle. As time proceeds, this will help ascertain whether the decision maintains relevancy and currency in light of new legal mandates, availability of new software products or other changes, and will incorporate anything learned about the alternative in the process of implementing its initial phases.
- **Recommends** a structured decision process (such as SDM) be used in any major future decision in IMD as a way to effectively leverage expertise from across the program and to build consensus.
- **Recommends** continuing the highly effective way of collaborating that the Task Force found so successful, and that this spread to other significant IMD projects. Leveraging open and highly transparent collaboration and communication, among all disciplines and stakeholders, helps ensure sound decision making.

Limitations and Risks

Any recommended solution includes both risks and benefits. As with any effort of this nature, recommendations also entail limitations.

Limitations

We recognize the following limitations that have bearing on our recommendations:

Limited details identifying project constraints

As noted previously, the Task Force was convened to review and recommend sustainable solutions for managing IMD's long-term ecological data. While key solution requirements were provided to the Task Force, specific goals or objectives for the project were not. As a result, the Task Force took on the task of defining these project objectives and other missing elements. While consulted, it is possible that not all members of the IMDLT agree with the project objectives identified by the Task Force.

Limited time to evaluate all objectives, requirements, and possible alternatives

While the Task Force dedicated significant time and effort in establishing requirements, objectives, and alternatives, time was limited. New requirements, objectives, and alternatives could possibly be identified with an extended period of reassessment. However, in the opinion of the Task Force, key requirements and objectives have been identified and the best alternative recommended.

Limited details for LTER/PASTA and EQuIS

In determining which alternative would most effectively meet IMD LTDM needs, the Task Force assessed a variety of factors. Although these factors were broad in scope—assessing technical components as well as development feasibility—our evaluation was, by necessity, limited.

The LTER/PASTA and EQuIS alternatives were initially evaluated based on in-house knowledge of these platforms; however, this knowledge was primarily limited to information gleaned from other users (Task Force members had limited daily working experience with these alternatives). Although the previous EQuIS evaluation provided some Task Force members with greater knowledge of this software's capabilities, not all Task Force members were part of that evaluation and there was uneven understanding of capabilities across the Task Force.

To establish a clearer understanding of their functionality, further details were gathered by eliciting feedback from active users—particularly those using active LTER/PASTA nodes. This information was critical in evaluating these alternatives, but clearly was not exhaustive and not at the level of familiarity with Data Store alternatives.

Differing understandings of alternatives and their capabilities

While significant effort was made to clearly define and delineate the features of each alternative, the sheer number of requirements (approximately 80) by which they were initially assessed meant that not all alternatives were fully detailed, even after multiple definition refinement iterations by several small groups. This variation in detail led to differing understandings on whether and how each alternative performed on various performance measures.

This unevenness in understanding was partially addressed through extensive discussion both during and after the Fort Collins LTDM in-person meeting in May 2017. In addition, both cost-and-effort and performance-measure small groups further refined alternative definitions. Perhaps the greatest clarifier was the development of workflow diagrams by the diagramming group (See Appendix D). Unfortunately these diagrams were developed late in the overall project, during field season, when few Task Force members were able to devote as much time as desired to fully digest their implications.

Lack of cost and time limits

The Task Force lacked clear cost and time limits. As a result, alternatives were created without a definitive timeline or budget. This led to some of the flexibility (“mushiness”) in alternative definitions noted above because features could not be ruled in or out based on solid time or cost constraints. These boundaries would have provided further clarity in the alternatives, although to be fully used would require more definitive estimates of the time and cost required for alternative features and functionality.

Relative effort estimates

A small group took on the significant task of making relative effort estimates for a variety of actions necessary for each alternative. The group’s work was limited by the lack of definite functional detail within the alternatives. With full functional specifications, it would be possible to determine relatively accurate estimates for cost, time, and effort. While fairly detailed, none of the alternatives are sufficiently specific to gather more than rough estimates of relative effort.

Risks

In addition to evaluating each alternative’s performance against key objectives and requirements, Task Force evaluations attempted to identify and mitigate risk.

Risk of externally developed software

Outsourcing the development of core software carries both benefits and risks. Outsourcing can be beneficial, in that the software can contain core functionality without requiring time to design and build it internally; however, there are also risks. Third-party software such as EQuIS and LTER/PASTA generally carry higher collaborative costs for integration into existing systems and for feature development. They also carry a higher risk of project abandonment by the vendor or developer due to vendor insolvency or inattention. By nature, third-party software is focused on a broad user base and IMD’s needs or requests may not be the highest priority. As a result, critical features for IMD may be delayed or deferred indefinitely.

Risk of missing important scientific features

The Task Force attempted to incorporate scientific views from a variety of sources: field ecologists, data managers with science backgrounds, park resource managers, and program and network quantitative ecologists. Due, in part, to the time required for Task Force activities and field season timing, ecologists had limited time to weigh in on Task Force discussions. Despite our efforts, some scientific features may have been missed.

Risk of implementation delays

If anticipated budget reductions occur in FY18 and FY19, project implementation will move forward at a pace modified to accommodate financial constraints. Aside from financial shortfalls, the key limitation is likely to be availability of network resources. In this scenario, establishing clear priorities of changes is critical to continue implementation; in particular, changes required to preserve the science that IMD delivers to parks; to maintain science quality; and to show good faith in working towards compliance of national policy mandates. As stated earlier, many high-priority changes, such as the development of a metadata policy, can continue at minimal cost and result in more focused and efficient use of existing resources.

Additional Considerations and Findings

In addition to the recommendations noted above, the Task Force offers the following additional recommendations based on its analyses.

Several actions will be necessary regardless of which alternative the IMDLT chooses

While the IMDLT may defer its decision to a later date—effectively choosing the Status Quo alternative—several actions are necessary to maintain compliance with new IT policies and procedures. Policies, workflow processes, and standards should move forward independently of the Leadership Team’s decision.

Policies need to be clearly defined and accessible as well as enforceable

Clearly delineating required policies and processes for QA/QC, protocol publication, and other network tasks will clarify duties as our collective responsibilities increase. IT mandates will need clearly defined policies and processes. Moving documents from draft stages to finalized policies will aid this process.

Establishing viable enforcement methods—whether manual processes or automated tools—will increase data value. Establishing clearly communicated consequences (both “sticks” and “carrots”) can also encourage adherence to policy.

Metadata must be clearly defined for tabular and non-tabular data; standardization aids in this effort

Metadata is the glue that holds together the various components of any alternative. Without clearly defined metadata, identifying data’s appropriate use and collection methodology is difficult. We rely on metadata to provide everything from descriptions to user access. Determining metadata that is required or optional, as well as metadata formats, should be one of several first steps.

Establishing a single, authoritative source for protocol ecological data is critical to ensure the most accurate data are offered

In addition to providing clarity that data are the most current, defining a single authoritative source, along with its associated metadata, informs data provenance. This can be used by internal and external scientists to clarify the depth and breadth of data collected as well as their appropriate use.

Requirements and recurring themes should be leveraged heavily for the design and development of any solution

Requirements identify critical features for the desired alternative. As a result, they should be heavily integrated into the design and development process. Recurring themes should be used in a similar manner. Task Force documentation should be referenced to leverage its insights.

Engaging IMD network personnel before arriving at policy and process decisions is one of the best signs of decision success

While this is common sense, too often we make decisions based on expediency. The Task Force exemplified what can be accomplished through broad, cross-disciplinary engagement in a collegial environment.

Open communication and collaboration is key to successful implementation

The Task Force relied heavily on electronic collaboration tools, and audio and video conferencing. Task Force members were remote, and many meetings took place via teleconference and webinars. Google's suite of tools—Hangouts, Sheets, Docs and Gmail, as well as SharePoint, maximized collaboration among Task Force members.

Regular communication with IMD network personnel was accomplished both informally, via emails from Task Force members at periodic intervals and direct discussions with network personnel, and more formally through periodic email updates sent from the IMD Central Office.

Once the Leadership Team's decision is made, communication will be an essential ingredient to ensure program-wide buy-in.

Use of small, focused groups facilitated reaching Task Force milestones

Breaking down the overall task into smaller components addressed by small, highly focused task groups was very effective. Groups were reasonably agile and able to adjust to new discoveries and insights while still moving the overall process forward.

Implementation

The key to implementing the selected LTDM alternative will be designating a project leader and creating a workgroup with diverse membership analogous to the LTDM Task Force. This implementation group would manage the project, develop detailed specifications, and elaborate on elements of the selected alternative that were left open-ended.

The implementation group should rely heavily on the requirements and other materials created by the original Task Force and consult the Task Force members where questions arise. Materials should be readily available, transparent, and thoroughly documented. All project artifacts should be repositied and curated for long-term access.

Much of the initial implementation work will center on creating policies and the associated workflows and processes for enacting them. Technology and training will follow.

Implementing technology will require careful and deliberate phasing. Features should be introduced iteratively and build upon prior stages. Formal user interface and user experience testing, potentially by a third party, should be conducted to maximize system usability and ensure smooth workflows that meet stakeholder needs and satisfy compliance requirements.

Training should be developed that communicates via multiple channels and formats. Webinars, user manuals, interactive help, recorded video, and support staff would all provide needed assistance. An array of options allows users to select the training method that best suits their learning style.

Essential throughout the implementation of the selected alternative is frequent, clear communication on developments, scheduling, opportunities for input and feedback, and impacts to workflow.

Guiding IMD into the future requires clarity of vision, a common purpose, and unified leadership. Change is a challenge in the best of times, and an even greater challenge in these times of budget uncertainty.

In creating the Task Force, the IMDLT established a dedicated, knowledgeable, and highly effective group. These factors laid the foundation for the success of the group:

- including representatives from all facets of IMD (science, data management, and program management from both central and network staff) to broaden the perspectives within the group;
- establishing and maintaining trust by giving the Task Force latitude to accomplish its charter without interference or micromanagement.

Successful implementation of the chosen alternative will continue to require effective leadership. An overall strategy leveraging Task Force recommendations, findings, and considerations will complement that effort.

Uncertainties

Any change carries with it a level of uncertainty. The IMD Leadership Team's LTDM decision is no exception. Each alternative carries some level of organizational, technical, and personal uncertainty as design details are developed. Data management policies and their implementation will need to be thoughtfully and collaboratively crafted as will an overall implementation strategy. Personnel will question how their roles and responsibilities will change with new or modified processes, policies, and technology. What new skills will they need to acquire? What training will be made available? Will they have sufficient support to do their job well?

Each question is an opportunity—an opportunity for IMD to reaffirm its commitment to high-quality, scientifically sound ecological data, and to the team of dedicated professionals who collect, analyze, and manage the data, both now and in the future.

Cultural Change

The cultural changes necessary for implementing the Leadership Team's decision will be significant. The Task Force, through their commitment to the SDM process and its outcomes, can serve as essential advocates for change and help mitigate a natural resistance to it.

The LTDM project has provided inroads toward engaging all IMD personnel and extra effort was made to ensure the scientific viewpoint was effectively represented. While each personnel group within IMD makes unique contributions, these are contributions to the whole of the organization—the IMD *team*.

The Task Force recognized, throughout its process, the importance of cross-disciplinary teamwork. Working through the changes necessary for long-term data management will require heightened teamwork within IMD, with all disciplines and expertise working together toward common organizational goals.

Appendix A. Structured Decision Making – Summary Report

Selecting an appropriate solution for long term data management for the NPS Inventory & Monitoring Division

Structured Decision Making Summary Report

Authors: Paula I. Capece³ and Nicole Athearn⁴

Decision Problem

The LTDM Task Force will recommend an effective solution for managing long-term ecological data to support park management and broader conservation science.

Background

The NPS Natural Resource Inventory & Monitoring (I&M) Program was created by congressional mandate in 1998, with the purpose of providing park managers with a broad-based understanding about the status and trends of the National Park System’s natural resources to be used in management decision making, in working with other agencies, and in communicating with the public (Frakes et al. 2016).

Historical Context

The Inventory & Monitoring Program is a decentralized organization consisting of 32 networks, coordinated through the Inventory and Monitoring Division (IMD). To date, the planning, development, and implementation of data management systems used for natural resource monitoring data across the program has been delegated to networks, with minimal central office IT governance or coordination. Individual data management approaches across the program tend to utilize MS Access-based solutions, with a few networks using MS SQL Server and a few networks using other technologies based on the preference or experience of individual data managers (Frakes et al. 2016).

The cumulative result of decentralized governance is an ever-increasing number of discrete technical approaches and solutions for managing IMD inventory and monitoring data. Over the years, the number, size, and complexity of datasets and applications across the program has grown in excess of 400 customized databases supported by just as many custom applications, tools, and processes. Some solutions are highly complex, and therefore fragile and difficult to maintain. The degree to which each successfully serves data management needs varies widely. All network data are part of a long-term ecological monitoring (LTEM) effort, where the expectation is that the data will be collected, maintained, and managed for decades. Accordingly, all data management systems require the resiliency to exist and to be utilized without the institutional memory of individuals. In addition, there is an increasing demand from NPS and DOI to use IMD data to support science-based

³ SDM Facilitator and Ecologist/Data Manager, NPS, Southeast Coast Network, Athens, GA.

⁴ SDM Facilitator and Research Coordinator, NPS, Great Rivers Cooperative Ecosystems Studies Unit, Columbia, MO.

management at landscape scales; that is, beyond the boundaries of parks and individual networks. When considering both long-term sustainability and the need to integrate IMD data with partners, program leadership has identified severe limitations in the current strategy (Frakes et al. 2016).

As an initial effort to address these concerns, IMD conducted a lengthy evaluation of a software product called EQuIS by Earthsoft to determine its capabilities as a central repository to preserve and allow for consistent management of long-term scientific data. During the course of the evaluation in 2016, it became clear that many members of IMD staff did not understand the need for a central office solution, and that even the evaluation participants had varying understanding of how such a solution would satisfy data management needs program-wide (Frakes et al. 2016). Based on this outcome, IMD Leadership chartered a new group, the Long Term Data Management (LTDM) Task Force, to delve deeper into the problem. Made up of over 20 IMD staff members across the IMD Central office and I&M Networks, the group assured program-wide representation including data managers from each region, and several scientists.

Regulatory & Policy Context

Recent statutes, directives, and policies on information technology have prompted major changes to how data and IT systems will need to be managed over the long-term across federal agencies and at the department, bureau, and divisional levels. The legal and policy mandates affecting the LTDM decision have been listed extensively elsewhere (see Tucker 2016b). Most will require significant changes to IMD's current data management practices; as such, the LTDM Task Force was asked to take these into account in developing their recommendation.

Primary Considerations

The Long Term Data Management (LTDM) Task Force was appointed to define long-term data management requirements, identify a range of possible solutions, and recommend the most appropriate solution for implementation (Tucker 2016a). IMD Leadership presented six initial data management considerations for long-term data management and asked the Task Force to develop concrete and realistic solution requirements to address each. These considerations are: (1) Data are discoverable, (2) Data are accessible, (3) Data are usable (as in documented for proper use), (4) Data are of known quality, (5) Data are preserved, and (6) Data are secure (Tucker 2016b). An especially important seventh consideration, (7) Data support IMD science needs, was repeatedly advocated by science representatives in the LTDM Task Force. While this overlapped to some degree with the six initial considerations, it was important as a stand-alone consideration due to the unique needs inherent in using data for long-term ecological analyses.

Solution Requirements

As a first step, the LTDM Task Force worked to develop a comprehensive list of requirements that any solution under consideration would need to meet. These were repeatedly refined and clarified in order to derive the best and most useful set of requirements. They were then characterized into two groups: mandatory requirements and desired characteristics. Initially composed of requirements derived from the initial six primary considerations, the list was expanded in January 2017 and subsequently re-categorized to incorporate additional requirements focusing on the seventh

consideration, science needs. In total, the final list of requirements includes 56 mandatory requirements and 23 desired characteristics (NPS 2017). These were meant to serve as a basis for constructing and implementing the selected data management solution.

Decision Structure

Following the initial assessment of requirements, the Task Force focused attention on developing its recommendation. Structured Decision Making was selected as the approach the group would use to frame interactions and guide the determination of a recommended data management solution.

Normative Decision Processes are considered best management practices for decision-making (Smith et al. 2016). One such process, Structured Decision Making (SDM), is a formal method for analyzing decisions by breaking a problem into components, which are selected and weighted in a transparent, systematic, and replicable manner (Runge et al. 2016). SDM is an iterative approach, allowing repeated “prototyping” of a decision model. As more is learned about the problem and its components, new information or parameters can be incorporated into subsequent iterations and different approaches of tackling a problem may be attempted. The goal is a decision model that incorporates only the information most relevant to a decision to help the decision maker gain a clear and objective picture of the problem, with its value-driven as well as fact-based components. The specific framework that the LTDM Task Force used to facilitate structuring its decision was the ProACT framework (see Figure A-1). ProACT is an acronym which stands for the five primary elements of a decision: (1) Define the **P**roblem, (2) Determine the **O**bjectives, (3) Identify **A**lternatives, (4) Predict the **C**onsequences, (5) Evaluate **T**radeoffs (Hammond et al. 1999).



Figure A-1. The ProACT framework used in Structured Decision Making.

The Task Force treated the problem as a *Multiple Objective* class of decision problem. Sometimes referred to in the literature as Multiple Criteria Decision Analysis (MCDA), the challenge with multiple objective problems is to determine how much performance in one objective should be

sacrificed in order to gain performance in another objective (Belton & Stewart 2002). Though the task force recognized that there were other issues involved in the LTDM decision, such as portfolio selection (a large number of component parts that could foreseeably be mixed and matched across a large number of possible alternative solutions) as well as risk (making a selection in the face of uncertainty), MCDA was nevertheless chosen as the most tractable way of dealing with the problem. This was considered the most appropriate classification because performance trade-offs was the dominant issue in selecting a long-term solution.

Decision Problem

The recommended solution needed to address the following needs: 1) Support the informatics needs of scientists, 2) meet legal and policy mandates for information technology, and 3) ensure preservation and access to IMD science data. An overarching goal of implementing a new data management solution is to manage all data collected in association with Vital Signs protocols and Inventories as a strategic asset and preserve IMD's legacy.

Problem definition was complicated by the vast complexities inherent in addressing long-term data management needs. To keep the decision process tractable, it was necessary to limit the scope of the problem with appropriate sideboards within the LTDM task force's control. To this end, the Task Force focused on only those details relevant for formulating a definitive recommendation on a data management platform and approach. This was done with the understanding that additional implementation details (both associated with and independent of the recommended solution) would be important for ultimate success. After several iterations of defining the problem, the task force developed the following problem statement to concisely guide the decision-making process toward a recommendation:

The LTDM Task Force will recommend an effective solution for managing long-term ecological data to support park management and broader conservation science.

Fundamental Objectives

Fundamental objectives are the basic or essential reasons for caring about the decision; that is, they reflect the values of the decision maker. Several characteristics of a fundamental objective distinguish it from other types of objectives (e.g. means objectives). While other types of objectives may be important overall, only the fundamental objectives are immediately relevant. A fundamental objective is controllable in the decision context –it represents an outcome that can be directly influenced by the decision at hand. A fundamental objective cannot be substituted, and is relevant to every alternative solution that can be reasonably considered. Perhaps the most important feature of a fundamental objective is that it is instrumental in distinguishing among alternatives (Runge 2016); the relative performance of an alternative on the full suite of fundamental objectives forms the basis for comparison with other alternatives.

The Task Force defined five fundamental objectives for determining a recommended data management solution to the IMD Leadership Team. In no particular order these solution characteristics are:

Program Accountability:

The solution supports efficiently answering questions about the who / what / when / where of IMD operations. The solution achieves compliance with departmental and bureau laws and policies, and maximizes adherence to industry best practices.

[Maximize]

Utility to Parks & Networks:

The solution supports networks providing their parks with management-relevant, scientific information on official Vital Signs and Inventories.

[Maximize]

Broad Science Utility:

The solution supports uses of IMD data and information beyond those captured by the core objectives of Vital Signs and Inventories.

[Maximize]

Development-related Effort & Cost:

Effort and cost required to develop and deploy the new data management solution, from concept to fully operational status.

[Minimize]

Post-development Effort & Cost:

Effort and cost required for ongoing maintenance and operation of data management and the solution.

[Minimize]

Attributes

In this problem, the attributes are subcomponents of the fundamental objectives. Since the fundamental objectives were somewhat intangible, the Task Force formulated several performance attributes to better define and evaluate potential alternatives (see Table A-1). The Task Force used influence diagramming to ascertain the relationships between the objectives, the means of achieving them, and the core performance attributes that would determine whether an objective is fulfilled. In essence, these are the criteria for judging the decision. Note that some attributes are repeated because they support multiple objectives.

Table A-1. LTDM Fundamental Objectives and Performance Attributes.

Objective	Description of Objective	Performance Attributes
Program Accountability	The solution supports efficiently answering questions about who/what/when/where of IMD operations. The solution achieves compliance with departmental and bureau laws and policies, and maximizes adherence to industry best practices.	• Method for providing overall IMD operational information about what specific natural resource parameters are monitored and inventoried by which networks, where, and how often • Achieving compliance with all legal, policy, and other requirements, including permanent archiving of electronic and physical records • Discovering and accessing data and related items from a previous point in time • Metadata & documentation checks
Utility to Parks & Networks	The solution supports networks providing their parks with management-relevant, scientific information on official Vital Signs and Inventories.	• Access control granularity (pertaining to non-NPS users) • Solution flexibility: What limits software solution development? • Bandwidth requirements • Customizable querying/sub-setting of data • Data flagging
Broad Science Utility	The solution supports uses of IMD data and information beyond those captured by the core objectives of the Vital Signs and Inventories.	• Metadata format consistency • Metadata & documentation checks • Data flagging • Customizable querying/sub-setting of data • Discovering and accessing data and related items from a previous point in time
Development-related Effort & Cost	Effort and cost required to develop and deploy the new data management solution, from concept to fully operational status.	• Internal Central Office development effort and cost • Outsourced Central Office development effort and cost • Network development effort and cost • Initial development license cost
Post-development Effort & Cost	Effort and cost required for ongoing maintenance and operation of data management and the solution.	• Internal Central Office post-development effort and cost • Outsourced Central Office post-development effort and cost • Network post-development effort and cost • Post-development license costs

Performance Measures

Performance measures are the direct means by which alternative solutions can be evaluated against the attributes. They provide the units (in the case of natural measures, e.g. dollars) or scales (in the

case of constructed or relative measures, e.g. 0 – 5) for evaluation. The Task Force developed both constructed scales and relative scales, where appropriate, for use as performance measures in this decision.

For attributes associated with non- effort and cost related objectives, the task force developed constructed scales as performance measures (see Table A-2).

Table A-2. Constructed-Scale Performance Measures. Associated performance attributes presented in alphabetical order.

Performance Attribute	Desired Direction	Performance Measure	
		Scale Value	Value Definition
Access control granularity (pertaining to non-NPS users)	Maximize	3	Selected non-NPS users can add/edit records and #2.
	Maximize	2	Selected non-NPS users can read non-public records and #1.
	Maximize	1	All non-NPS users can read public records.
	Maximize	0	No access for non-NPS users.
Achieving compliance with legal, policy, and other requirements, including permanent archiving of electronic and physical records	Maximize	4	Central Office extracts and reports compliance using data and metadata from materials that networks post to the solution (no additional reporting required by networks).
	Maximize	3	Central Office extracts and reports compliance using metadata from materials that networks post to the solution, and networks independently fill any remaining gaps in reporting.
	Maximize	2	Central Office provides directions and standard reporting for compliance that networks independently follow.
	Maximize	1	Central Office provides directions for compliance that networks independently follow using no particular reporting standard.
	Maximize	0	Networks independently learn and comply with requirements
Bandwidth requirements	Maximize	2	Solution workflow favors local processing of data and reduced use and reliability on bandwidth. Performance/user experience would not be hindered given current bandwidth capacity available to networks.
	Maximize	1	Solution workflow includes mixture of remote and local processing. Minor upgrades to networks' (5-15%) bandwidth infrastructure would be required to implement solution.

Table A-2 (continued). Constructed-Scale Performance Measures. Associated performance attributes presented in alphabetical order.

Performance Attribute	Desired Direction	Performance Measure	
		Scale Value	Value Definition
Bandwidth requirements (continued)	Maximize	0	Solution workflow favors remote processing requiring increased use and reliability on bandwidth. Significant upgrades to networks' (15-30%) bandwidth infrastructure would be required to implement solution.
Customizable querying/sub-setting of data	Maximize	4	Solution provides an advanced user interface with interactive controls to subset and query data (e.g. access to all permissible fields with robust functionality).
	Maximize	3	Solution provides a basic user interface with interactive controls to subset and query data (e.g. access to some but not all fields with less functionality).
	Maximize	2	Solution supports access to rows and columns of data using web services.
	Maximize	1	Solution supports access to rows of data using web services.
	Maximize	0	Users must download data and perform their own sub-setting and querying.
Data flagging	Maximize	4	Solution supports #3 and cell/value level flags.
	Maximize	3	Solution supports #2 and row level flags.
	Maximize	2	Solution supports #1 and table/object level flags using object metadata.
	Maximize	1	Solution communicates/leverages database-level flags using database metadata elements.
	Maximize	0	Solution does not require and/or does not communicate/leverage flags of any kind.
Discovering and accessing data and related items from a previous point in time	Maximize	3	Data sets and related items are available from a previous point in time by providing date and/or time criteria.
	Maximize	2	Data sets are available from a previous point in time by providing date and/or time criteria; items related to the data set are not included.
	Maximize	1	Manual effort is required to recreate data from a previous point in time using information contained in data sets, metadata, and related documentation.
	Maximize	0	No capability exists to discover and access data or related items from any previous point in time; only offers most current data.

Table A-2 (continued). Constructed-Scale Performance Measures. Associated performance attributes presented in alphabetical order.

Performance Attribute	Desired Direction	Performance Measure	
		Scale Value	Value Definition
Metadata & documentation checks	Maximize	4	Required metadata elements completed and checked; additional metadata elements are checked and reported on, but not required; metadata structure/format is standardized and checked.
	Maximize	3	Required metadata elements completed and checked; metadata structure/format is standardized and checked.
	Maximize	2	Metadata and links to documentation required and checked; metadata structure/format is standardized and checked.
	Maximize	1	Metadata and links to documentation are required and checked; metadata structure/format NOT standardized or checked.
	Maximize	0	No checks on presence or format of metadata and links to documentation (protocols, SOPs, QAPs, etc.).
Metadata format consistency	Maximize	4	Solution provides a single metadata format for each file type, including tabular, geospatial, image, video, and audio files.
	Maximize	3	Solution provides a single metadata format for tabular and geospatial files; Solution allows/supports multiple metadata formats for other file types.
	Maximize	2	Solution provides a single metadata format for tabular and geospatial files; metadata format for other file types is not a consideration.
	Maximize	1	Solution allows/supports multiple metadata formats for tabular and geospatial files; metadata format for other file types is not a consideration.
	Maximize	0	No requirement for consistent metadata format for any data/file type.
Method for providing overall IMD operational information about what specific natural resource parameters are monitored and inventoried by which networks, where, and how often	Maximize	3	Information is automatically extracted from data and metadata.
	Maximize	2	Information is automatically extracted from metadata.
	Maximize	1	Information is manually extracted by networks.
	Maximize	0	No information is provided.

Table A-2 (continued). Constructed-Scale Performance Measures. Associated performance attributes presented in alphabetical order.

Performance Attribute	Desired Direction	Performance Measure	
		Scale Value	Value Definition
Solution flexibility: What limits software solution development?	Maximize	2	IMD limits solution development. (Solution limited only by IMD whose processes/technology limits its ability to make desired updates/improvements.)
	Maximize	1	Science community limits solution development. (Solution involves a greater science community whose processes/technology limits IMDs ability to make desired updates/improvements.)
	Maximize	0	Commercial vendor limits solution development. (Solution involves an outside vendor whose processes/technology limits IMD's ability to make desired updates/improvements.)

Based on the uncertainties in effort and cost required for a data management solution at such an early stage of consideration, constructing scales with defined threshold values for effort and cost related attributes may have amounted to inaccurate guesswork. Instead, the Task Force decided to utilize relative scales to gage the performance of alternatives on effort and cost related performance attributes. This means that alternatives would be assigned *points* values for each attribute; when taken together, approximate magnitudes of difference expected across alternatives with respect to an attribute could be derived. In essence, this approach allowed the Task Force to indicate that one alternative was several times more costly in terms of effort or cost than another.

Constraints

As the Task Force worked through the decision problem, several constraining factors affecting the decision were identified. The Task Force was instructed from the start of the effort that their recommendation must be implementable within a 10 year time period from the time a decision is made by the IMD Leadership Team. In early iterations of developing an accurate decision model, the concept of project *Completion Time* was treated as an objective to be minimized. However, it became clear that project completion time had a direct but inverse relationship with the *Development-related Effort & Costs* objective. The group determined that the most efficient way of handling these objectives during the decision analysis would be to treat project completion time as a hard constraint, removing it from the list of objectives, thereby simplifying the decision model and allowing the task force to concentrate on the remaining objectives.

The Task Force assumed there would be inherent constraints on the amount of effort and cost that could be expended to develop the recommended data management solution. Since the Task Force was not provided enough detail to set a top-limit constraint on this objective, the task force decided

to focus on developing relative values of effort and costs required to implement each possible alternative.

Mandatory solution requirements developed by the Task Force at the start of the process were also intended as constraints. They could not be treated as hard constraints in this decision model for a number of reasons, including the sheer complexity and number of requirements. Initially, the Task Force attempted to group these requirements by the fundamental objective each supported, and use the fulfillment of a requirement as a way to determine the performance of each alternative. While logical, this approach proved to be problematic. The question of *how* a requirement is fulfilled is equally as relevant as *whether* a requirement is fulfilled. How requirements are accomplished depends to some degree on the overall approach a particular alternative solution follows to address the full suite of fundamental objectives. This can vary dramatically from alternative to alternative while still technically fulfilling requirements. Due to this realization, the requirements were instead used as factors in the development and description of Alternatives, with the expectation that any of the alternatives can be implemented in such a way as to meet all mandatory requirements. Thus, they are also intended to serve as a foundation for project planning and implementation.

Alternatives

Prior to identifying alternatives, the Task Force held several assumptions regarding what a decision alternative for IMD long term data management should include. The main assumption was that each alternative should be composed of a combination of policy, process, and technology that would together make up a comprehensive data management solution.

Using the initial primary considerations, the solution requirements, and the fundamental objectives as references, the task force first developed a thorough list of all possible individual choices, actions, and products (called the *Action List*) to use as source material in formulating comprehensive *Alternative Solution Packages*. Several noteworthy points were achieved through this effort. The Task Force succeeded in generating a large volume of potential ideas that were useful in composing the complete Alternative Solution Packages. An unexpected benefit was that a new category – *training* – was added as a necessary component of each alternative. Policy-related content generated in this effort was of particularly high value and is expected to be immediately useful for determining priorities for which policy and guidance gaps need to be filled by IMD. Furthermore, a number of “common themes” presented themselves when summarizing the list of possible individual choices, actions and products. These common themes proved especially instrumental for describing the ultimate Alternative Solution Packages.

Next, the task force formulated comprehensive Alternative Solution Packages– discrete combinations of policy, process, technology, and training actions which together made up individual decision alternatives. Given that alternatives could be composed of varying combinations of different components, the primary challenge in identifying viable alternatives was to narrow the near infinite number of possible portfolios to a reasonable number of distinct choices. The Task Force accomplished this by considering a realistic range of general approaches to data management from status quo all the way to a comprehensive, highly integrated approach, defining the “character” of each approach, and then describing each in detail.

The task force fully described eleven distinct Alternative Solution Packages. Alternatives which were similar in nature, failed to meet requirements, or were otherwise unattainable were dropped from the final list. This reduced the number of alternatives to seven for final consideration used within the decision model (Table A-3).

Table A-3. Seven finalized Alternative Solution Packages used in the LTDM Decision Model.

Alternative	General Description
Status Quo	Don't change anything about current IMD data management operations- every network does their best; central office issues limited guidance and provides assistance to networks when needed.
Minimally Compliant	Satisfy compliance (meet legal mandates, regulations, and directives pertaining to the federal government, DOI, NPS, NRSS, IMD), but with minimal impact/change to network operations.
Park Focused	Meets minimally compliant requirements, while maximizing delivery to parks in the individualized manner parks may request, and de-emphasizing central office overhead. Implementation may include lightweight enhancements to the existing Data Store application.
Remodeled Data Store	Data Store is enhanced and revised to pull-in the best elements of LTER/PASTA and/or other systems. Networks can maintain their own databases and add these to Data Store as files with the option of having them converted to a query-able data source. This query-able source conversion requires additional metadata (e.g. schema definition), and results in structured web services that promote broader consumption. Flags in Data Store (e.g. visibility, protected status) are used to determine what content to expose as services. All systems for metadata, tabular data, and non-tabular data are linked together and presented as a single source.
LTER/PASTA Node	Leverage existing PASTA open-source code and related tools (e.g. Morpho, Excel metadata templates) to stand up an NPS PASTA site on the Windows platform. Networks would maintain their own databases and would upload periodic data packages (with EML metadata) into PASTA in the form of flattened .csv files. Opt in as a DataOne node to exchange information with LTER, NEON, academics, etc., with flexibility on the look and feel of public/internal pages. The PASTA system handles metadata and packages of tabular and non-tabular data. Other NPS tools would be developed to hold and manage relationships between associated non-tabular files, protocol documents, and products (e.g. published reports).
Centralized Interface & Administration	Network data associated with monitoring protocols are stored in data center databases that are collaboratively designed/maintained with Central Office staff. Field databases and applications may be used, but networks must connect to centralized databases to load data (i.e. network located data entry databases would be for temporary storage only). Networks would use NPS Data Store to catalog the centralized databases, and Data Store would hold and manage relationships between associated non-tabular files, protocol documents, and products (e.g. published reports).

Table A-3 (continued). Seven finalized Alternative Solution Packages used in the LTDM Decision Model.

Alternative	General Description
Earthsoft (EQuIS)	Environmental Quality Information System (EQuIS) is commercial software from Earthsoft with separate data tables enforced via Electronic Data Deliverable (EDD) definitions for every protocol (i.e. Survey). Data file submissions are checked for formatting and valid values before being accepted into EQuIS. Automatic notifications and tools help users identify and correct submitted data not conforming to EDD parameters. EQuIS can process data files submitted in multiple file formats (e.g. txt, csv, xlsx). The Central Office would help networks configure Surveys in EQuIS, data tables and associated EDD's, and help load data for existing protocols. Networks have the flexibility to use existing and/or adjusted tools and workflows to collect field data, but networks must promptly transform and submit/load field data to EQuIS. Networks would use NPS Data Store to catalog the authoritative data sets stored in EQuIS, and Data Store would hold and manage relationships between associated non-tabular files, protocol documents, and products (e.g. published reports).

Considerable thought, effort, and research was necessary to compose and articulate the Alternative Solution Packages. This effort was required in order to ensure each member of the Task Force had the same understanding of what each alternative would offer IMD as a final solution. To facilitate this common understanding, the task force worked to come to agreement on a general characterization of each alternative (see Table A-3), clearly articulating how each alternative addressed “common themes” that arose in the Action List, and determine how each alternative fulfilled the 56 mandatory solution requirements. Full details pertaining to each Alternative Solution Package are available in the Long Term Data Management Task Force Report (NPS 2017).

Although the *Status Quo* alternative does not meet many of the mandatory requirements for long-term data management, the task force included this alternative in the decision model for several reasons. First, “no change to current operations” is almost always a reasonable outcome for any decision process, albeit typically not the ideal solution. Even in cases where there is no deliberate intention to select the *Status Quo* option, if a decision should be delayed, the decision that is being executed in the meantime is the *Status Quo* alternative. The Task Force decided that it was particularly important to point out that failing to take active action is, in itself, a management decision, complete with its own set of consequences. Including the *Status Quo* alternative in the decision model allowed the Task Force to forecast those consequences transparently. The Task Force also found this alternative important to include as a point of comparison with the other identified alternatives.

Decision Analysis

The main output of the Task Force’s formal decision structuring process was a *Decision Model*. This is a tabular or diagrammatic representation of the full suite of factors affecting the decision. It is a tool set up such that important relationships among the decision factors are apparent and able to be manipulated.

Consequences

A key component of the decision model is the prediction of consequences. Consequences may be predicted through empirical estimation or expert judgement. Predicting consequences of the alternatives is intended to be an objective step that allows for quantitative or semi-quantitative evaluation of alternatives relative to decision objectives (Larson et al. 2016).

Consequence scores for non-effort and cost attributes were predicted using expert judgement for each alternative. To elicit consequence predictions, the LTDM Facilitators initially used four-point interval elicitation (Speirs-Bridge et al. 2010) on pairs of Task Force members. For each alternative, Task Force experts were asked to provide their lowest realistic score for each performance attribute, the highest realistic score, their best guess for the score, and their confidence that the true value falls within the lowest to highest range they offered. Even though the score values pertaining to constructed scales were defined at the integer level, selection of partial values was allowed in the elicitation, to better capture experts' notions of scores that might be intermediate of the defined scale values. After all pairs of task force members completed their scoring, they then participated in a nominal group technique (Van de Ven & Delbecq 1974), modified for this exercise. This was done to achieve consensus on a single set of scores via group interaction. Consensus scores resulting from this interaction were very similar to the average values of individual scores.

The Task Force found this exercise to be difficult for a variety of reasons. Likely the most problematic issue was a lack of confidence in the relevance of the attributes and soundness of the performance measure scale values in the earliest iterations of this elicitation. Additional time and work was incorporated into the process to refine both the attributes and performance measure scales to satisfy this issue. In the final elicitation of consequence predictions, the LTDM facilitators abandoned the four-point interval elicitation in favor of eliciting probabilities of whole value scores for each attribute under each alternative. This was accomplished by providing each individual Task Force expert with each allowable score for each attribute and asking him/her to distribute 100 points across the range of allowable scores (Figure A-2). The 100-point allocation spread was then converted into a single weighted score. Individual weighted scores for each expert were then averaged to obtain a composite group score. Because scores differed among experts, individual scores were retained to examine separately in a sensitivity analysis.

Objective		Attribute																					
Hypothetical Objective	Hypothetical Attribute	Constructed Scale 0-4 (See scale reference for value definitions)		Maximize	Possible Scores (Place points for each score within the boxes below)					Total Check (needs to equal 100)	We can easily derive:												
		Alternatives ↓	Scores →	0	1	2	3	4		Most Likely Score	Lowest Reasonable Score	Highest Reasonable Score	Weighted Score										
		Status Quo		95	5				100	0	0	1	0.05										
		Minimally Compliant		10	85	5			100	1	0	2	0.95										
		Park Focused			90	10			100	1	1	2	1.1										
		Remodeled Data Store				15	70	15	100	3	2	4	3										
		LTER/PASTA node			10	75	15		100	2	1	3	2.05										
		Centralized Interface & Administration (CIA)					60	40	100	3	3	4	3.4										
		EQulS				10	70	20	100	3	2	4	3.1										

Figure A-2. Example elicitation exercise that accompanied the actual LTDM elicitation sheets used to elicit consequence predictions from task force experts.

Consequence scores for effort and cost attributes were predicted using expert judgement of a subgroup of four Task Force members. Approximately 164 types of activities necessary for implementation and maintenance of a solution were listed out by the Task Force during the development of detailed alternatives. The effort/cost subgroup catalogued these 164 activities and an additional 36 activities into 16 activity categories (12 Development; 4 Post-development). For each alternative, the activity categories received assignments of relative scores based upon the anticipated level of effort at the *Central Office – Internal (C)*, *Central Office – Outsource (O)*, and *Network (N)* organizational levels (Budde, pers. comm.). License costs necessary for development and post-development implementation were estimated separately (Figures A-3 and A-4). Final group consequence prediction scores for all attributes can be found in Table A-4.

Activity Category requiring effort		Resource								Multiplier
			Status Quo	Minimum Comp	Park Focused	Remodel DS	LTER/PASTA	CIA	EQUIS	
Post Development-										
Implement changes due to operational	IC	Central (Internal)	84	52	52	55	0	0	0	3
	IC	Network	84	84	84	8	0	0	0	2
	IC	Outsource (Central)	52	52	52	29	0	0	0	1
Infrastructure and System support (IF)	IF	Central (Internal)	20	32	32	26	29	37	42	
	IF	Network	52	32	52	11	16	16	16	
	IF	Outsource (Central)	32	32	32	44	73	73	52	
New / Additional Data Processing requirements	NR	Central (Internal)	56	96	96	0	0	8	10	
	NR	Network	156	188	336	355	462	339	416	
	NR	Outsource (Central)	108	160	116	35	58	84	34	
Operations and Governance (OP)	OP	Central (Internal)	84	52	84	42	113	50	58	
	OP	Network	136	84	136	34	47	42	34	
	OP	Outsource (Central)	32	32	32	39	52	39	58	
Ongoing License Costs (k\$)	OL	Central	0	0	0	0	0	160	350	
Central Office Total			244	232	264	123	142	95	110	
Network Office Total			428	388	608	408	525	397	466	
Outsourced Total			224	276	232	147	183	196	144	
Total Post-Dev Effort			896	896	1104	678	850	688	720	
Post Dev License and Maintenance Costs (k\$)			0	0	0	0	160	0	350	

Figure A-3. LTDM calculation sheet for Post-development Effort and Cost activity categories and relative scores.

Long-Term Data Management Task Force
Appendix A – Structured Decision Making – Summary Report

Activity Category requiring effort		Resource	Status Quo	Minimum Comp	Park Focused	Remodel DS	LTER/PASTA	CIA	EQUIS	
Development	Abbrev									Multiplier
Develop Policy (DP)	DP	Central (Internal)	72	84	84	378	341	451	470	3
	DP	Network	40	48	60	228	190	240	224	2
	DP	Outsource (Central)	0	0	0	58	21	47	0	1
Develop IMD Wide Standards (DS)	DS	Central (Internal)	32	84	48	112	120	104	120	
	DS	Network	20	20	32	32	32	32	32	
	DS	Outsource (Central)	0	0	0	0	0	0	0	
Define Requirement Details (RD)	RD	Central (Internal)	0	0	0	55	89	44	123	
	RD	Network	0	0	0	21	44	0	42	
	RD	Outsource (Central)	0	0	0	21	39	39	39	
Develop Enterprise App or Update Current (DA)	DA	Central (Internal)	84	304	304	159	445	618	565	
	DA	Network	32	176	304	170	110	595	322	
	DA	Outsource (Central)	148	304	188	813	573	1598	1377	
System Setup and configuration (SC)	SC	Central (Internal)	0	0	0	83	39	88	112	
	SC	Network	0	20	32	5	5	5	5	
	SC	Outsource (Central)	0	0	0	0	0	0	0	
Plan Network Impact (NI)	NI	Central (Internal)	0	0	0	39	39	39	39	
	NI	Network	0	32	52	37	37	50	42	
	NI	Outsource (Central)	0	0	0	0	0	0	0	
Update Network tools/dbs (NT)	NT	Central (Internal)	32	84	84	52	49	0	0	
	NT	Network	52	52	84	39	57	47	47	
	NT	Outsource (Central)	84	64	32	332	207	280	147	
Develop Park/ Network Visualizations (DV)	DV	Central (Internal)	0	0	0	42	50	45	0	
	DV	Network	0	0	0	34	34	39	0	
	DV	Outsource (Central)	0	0	0	76	97	97	0	
Develop Training material (TM)	TM	Central (Internal)	0	32	0	42	63	50	94	
	TM	Network	0	0	32	31	34	34	47	
	TM	Outsource (Central)	0	0	0	0	0	0	76	
Rollout planning (RP)	RP	Central (Internal)	0	20	0	34	31	44	55	
	RP	Network	0	0	20	0	0	0	0	
	RP	Outsource (Central)	0	0	0	0	0	0	0	
Oversight / Governance (OG)	OG	Central (Internal)	0	32	32	0	68	0	152	
	OG	Network	0	32	52	0	0	0	0	
	OG	Outsource (Central)	0	0	0	0	0	0	63	
Hardware, Software, ATO, Java (HS) Not In ALT	HS	Central (Internal)				0	60	0	60	
	HS	Network				0	8	0	8	
	HS	Outsource (Central)				0	45	0	58	
Software License (in k\$)	SL	Central (Internal)	0	0	0	0	32	0	1,800	
Central Office Total			220	640	552	996	1394	1483	1790	
Network Office Total			144	380	668	597	551	1042	769	
Outsourced Total			232	368	220	1300	982	2061	1760	
Total Dev effort			596	1388	1440	2893	2927	4586	4319	
Total Dev Lic Costs (k\$)			0	0	0	0	32	0	1,800	
Total Dev & Post-Dev Software License Costs Total			1492	2284	2544	3571	3777	5274	5039	
			0	0	0	0	192	0	2150	

Figure A-4. LTDM calculation sheet for Development-related Effort and Cost activity categories and relative scores.

Table A-4. Consequences table for LTDM decision model, using composite group scores.

Objective	Attribute	Direction	Alternative								
			Status Quo	Minimally Compliant	Park Focused	Remodeled Data Store	LTER/PASTA node	Centralized Interface & Administration (CIA)	EQUIS	Maximum value	Minimum value
Program Accountability	Method for providing overall IMD operational information on specific natural resource parameters monitored and inventoried by which networks, where and how often.	Max	0.25	0.5	0.25	2.75	2.5	2.75	2	2.75	0.25
	Achieving compliance with legal, policy, and other requirements, including permanent archiving of electronic and physical records.	Max	0	1.75	1.75	3.25	3.25	3.5	3	3.5	0
	Discovering and accessing data related items from a previous point in time	Max	0.5	0.5	1.99	2.75	2.75	2.75	2	2.75	0.5
Broad Science Utility	Metadata format consistency	Max	0	1.3	2.75	3.75	4	2.25	2.1	4	0
	Metadata / documentation checks	Max	0	0.25	2.5	3.5	3.5	3.01	2.75	3.5	0
	Data flagging	Max	0.1	1	1.75	4	4	4	4	4	0.1
	Customizable querying/sub-setting of data	Max	0	0	0.5	1.5	0	1.75	1.25	1.75	0
Utility to Parks & Networks	Access control granularity for non-NPS users	Max	1	1	1	2	3	1.5	1	3	1
	Solution Flexibility: Who drives software solution direction?	Max	2	2	2	2	2	2	0	2	0
	Bandwidth requirements	Max	2	2	2	1.75	1.8	0	0	2	0
Development-related Effort & Cost	Internal Central Office Effort & Cost	Min	220	640	552	996	1394	1483	1790	1790	220
	Network Effort & Cost	Min	144	380	668	597	551	1042	769	1042	144
	Outsourced Central Office Effort & Cost	Min	232	368	220	1300	982	2061	1760	2061	220
	Licensing Cost	Min	0	0	0	0	32	0	1800	1800	0

Table A-4 (continued). Consequences table for LTDM decision model, using composite group scores.

Objective	Attribute	Direction	Alternative								
			Status Quo	Minimally Compliant	Park Focused	Remodeled Data Store	Status Quo	Centralized Interface & Administration (CIA)	EQUIS	Maximum value	Status Quo
Post-development Effort & Cost	Internal Central Office Effort & Cost	Min	244	232	264	123	142	95	110	264	95
	Network Effort & Cost	Min	428	388	608	408	525	397	466	608	388
	Outsourced Central Office Effort & Cost	Min	224	276	232	147	183	196	144	276	144
	Licensing Cost	Min	0	0	0	0	160	0	350	350	0

Tradeoffs

Since decisions with multiple objectives cannot be resolved by focusing on only one objective (Hammond et al. 1999), achieving balance or “trading-off” among competing objectives is necessary. Tradeoffs are driven by values of the decision-makers (or in the case of LTDM, the recommending body [Task Force]), and define how much performance in one attribute should be sacrificed to achieve some level of performance in another attribute.

The Task Force employed the Simple Multi-Attribute Ranking Technique (SMART) to perform direct tradeoffs across attributes (Goodwin and Wright 2004). To execute the SMART technique, the group Consequences table was first normalized (Table A-5) by converting all scores for each attribute to a normalized scale (i.e., 0-1). Because some attributes influenced more than one fundamental objective, preference weights for each of the 18 attributes were rated from 0 to 100 by each task force expert using swing weighting (Figure A-5). Swing weighting is a method of eliciting preference weights that takes into account both the general importance of the attribute and the range over which performance of the attribute varies across alternatives. Ratings were converted to weights by dividing each attribute’s rating by the sum of all 18 ratings (all weights for each task force member totaled to 1.0).

Swing Weighting			Hypothetical Alternatives																				
Objective	Attribute	Goal	Worst	Best	Bench mark	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Program Accountability	Method for providing overall IMD operational information about what specific natural resource parameters are monitored and inventoried by which networks, where and how often.	Max	0.46	2.40	0.46	2.40	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
2	Program Accountability	Achieving compliance with legal, policy, and other requirements, including permanent archiving of electronic and physical records.	Max	0.53	3.16	0.53	0.53	3.16	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53
3	Program Accountability	Discovering and accessing data related items from a previous point in time*	Max	0.64	2.43	0.64	0.64	0.64	2.43	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64
4	Broader Science Utility	Metadata format consistency	Max	0.36	2.89	0.36	0.36	0.36	0.36	2.89	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36
5	Broader Science Utility	Metadata / documentation checks*	Max	0.09	2.96	0.09	0.09	0.09	0.09	0.09	2.96	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
6	Broader Science Utility	Data flagging*	Max	0.14	2.98	0.14	0.14	0.14	0.14	0.14	2.98	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14
7	Broader Science Utility	Customizable querying/subsetting of data*	Max	0.11	2.68	0.11	0.11	0.11	0.11	0.11	0.11	2.68	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
8	Utility to Parks & Networks	Access control granularity for non-NPS users	Max	1.13	2.25	1.13	1.13	1.13	1.13	1.13	1.13	1.13	2.25	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13
9	Utility to Parks & Networks	Solution Flexibility: Who drives software solution direction?	Max	0.34	1.76	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	1.76	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34
10	Utility to Parks & Networks	Bandwidth requirements	Max	0.53	1.65	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	1.65	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53
11	Development-related Effort & Cost	Internal Central Office Effort & Cost	Min	1,790	220	1,790	1,790	1,790	1,790	1,790	1,790	1,790	1,790	1,790	1,790	220	1,790	1,790	1,790	1,790	1,790	1,790	1,790
12	Development-related Effort & Cost	Network Effort & Cost	Min	1,042	144	1,042	1,042	1,042	1,042	1,042	1,042	1,042	1,042	1,042	1,042	144	1,042	1,042	1,042	1,042	1,042	1,042	1,042
13	Development-related Effort & Cost	Outsourced Central Office Effort & Cost	Min	2,061	220	2,061	2,061	2,061	2,061	2,061	2,061	2,061	2,061	2,061	2,061	2,061	220	2,061	2,061	2,061	2,061	2,061	2,061
14	Development-related Effort & Cost	Licensing Cost	Min	1,800	0	1,800	1,800	1,800	1,800	1,800	1,800	1,800	1,800	1,800	1,800	1,800	0	1,800	1,800	1,800	1,800	1,800	1,800
15	Post-development Effort & Cost	Internal Central Office Effort & Cost	Min	264	95	264	264	264	264	264	264	264	264	264	264	264	264	264	95	264	264	264	264
16	Post-development Effort & Cost	Network Effort & Cost	Min	608	388	608	608	608	608	608	608	608	608	608	608	608	608	608	388	608	608	608	608
17	Post-development Effort & Cost	Outsourced Central Office Effort & Cost	Min	276	144	276	276	276	276	276	276	276	276	276	276	276	276	276	144	276	276	276	276
18	Post-development Effort & Cost	Licensing Cost	Min	350	0	350	350	350	350	350	350	350	350	350	350	350	350	350	0	350	350	350	350
* Attribute supports more than one Objective.																							
			Rank (1 to 19)	19																			
			Score (0 to 100)	0																			
			Weight (0-1)	0																			

Figure A-5. Example swing weighting elicitation tool.

Table A-5. Normalized consequences table for LTDM decision model, using composite group scores re-scaled between 0 and 1.

Objective	Attribute	Direction	Alternative								
			Status Quo	Minimally Compliant	Park Focused	Remodeled Data Store	LTER/PASTA node	Centralized Interface & Administration (CIA)	EQUIS	Maximum value	Minimum value
Program Accountability	Method for providing overall IMD operational information on specific natural resource parameters monitored and inventoried by which networks, where and how often.	Max	0	0.1	0	1	0.9	1	0.7	1	0
	Achieving compliance with legal, policy, and other requirements, including permanent archiving of electronic and physical records.	Max	0	0.5	0.5	0.93	0.93	1	0.86	1	0
	Discovering and accessing data related items from a previous point in time	Max	0	0	0.66	1	1	1	0.67	1	0
Broad Science Utility	Metadata format consistency	Max	0	0.33	0.69	0.94	1	0.56	0.53	1	0
	Metadata / documentation checks	Max	0	0.07	0.71	1	1	0.86	0.79	1	0
	Data flagging	Max	0	0.23	0.42	1	1	1	1	1	0
	Customizable querying/sub-setting of data	Max	0	0	0.29	0.86	0	1	0.71	1	0
Utility to Parks & Networks	Access control granularity for non-NPS users	Max	0	0	0	0.5	1	0.25	0	1	0
	Solution Flexibility: Who drives software solution direction?	Max	1	1	1	1	1	1	0	1	0
	Bandwidth requirements	Max	1	1	1	0.88	0.9	0	0	1	0
Development-related Effort & Cost	Internal Central Office Effort & Cost	Min	1	0.73	0.79	0.51	0.25	0.2	0	1	0
	Network Effort & Cost	Min	1	0.74	0.42	0.5	0.55	0	0.3	1	0
	Outsourced Central Office Effort & Cost	Min	0.99	0.92	1	0.41	0.59	0	0.16	1	0
	Licensing Cost	Min	1	1	1	1	0.98	1	0	1	0

Table A-5 (continued). Normalized consequences table for LTDM decision model, using composite group scores re-scaled between 0 and 1.

Objective	Attribute	Direction	Alternative								
			Status Quo	Minimally Compliant	Park Focused	Remodeled Data Store	Status Quo	Centralized Interface & Administration (CIA)	EQUIS	Maximum value	Status Quo
Post-development Effort & Cost	Internal Central Office Effort & Cost	Min	0.12	0.19	0	0.83	0.72	1	0.91	1	0
	Network Effort & Cost	Min	0.82	1	0	0.91	0.38	0.96	0.65	1	0
	Outsourced Central Office Effort & Cost	Min	0.39	0	0.33	0.98	0.7	0.61	1	1	0
	Licensing Cost	Min	1	1	1	1	0.54	1	0	1	0

Each Task Force member’s attribute weights were used to generate a weighted average for each alternative. This generated a personalized outcome for each task force member based on mean group consequence scores.

Decision Model Results

In addition to observing personalized outcomes for each Task Force member who responded, results were examined in several different ways in an effort to detect patterns and variation from patterns. When setting all tradeoff weights equal to one another and analyzing against the mean group consequence scores, the *Remodeled Data Store* alternative scored the highest, with *Centralized Interface & Administration* in second position, followed by *LTER/PASTA Node* (Figure A-6).

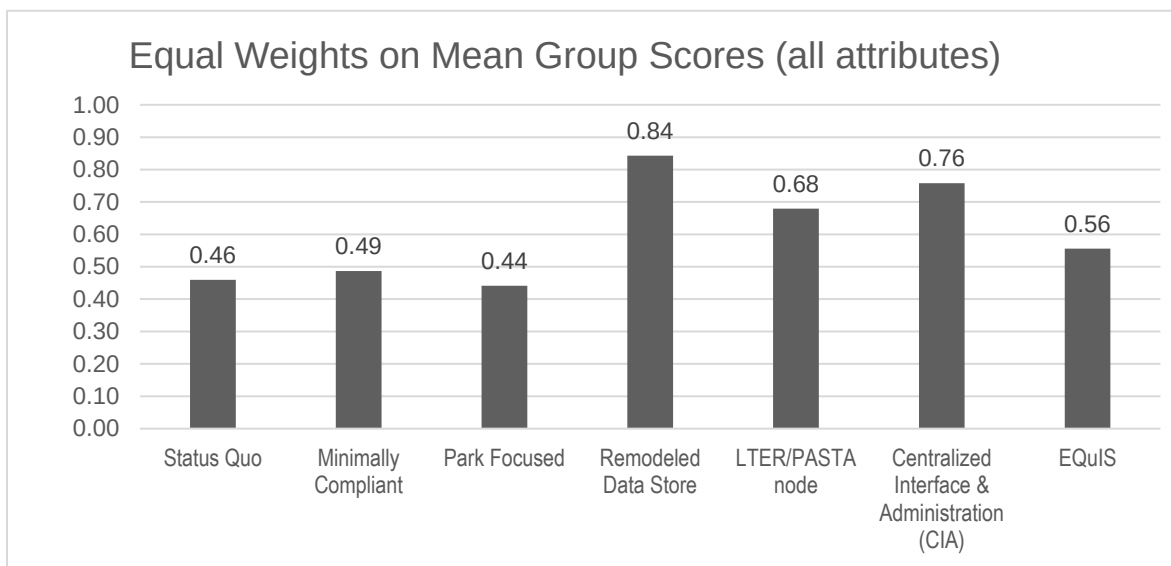


Figure A-6. LTDM decision model outcome of equal weights on mean group scores, considering all performance attributes.

A similar pattern holds when effort and cost-related attributes are excluded from the results (Figure A-7). The pattern differs in that the *Park Focused* alternative changes position with *Minimally Compliant*. *Remodeled Data Store* and *Centralized Interface & Administration* are much more closely matched, although *Remodeled Data Store* remains in first position.

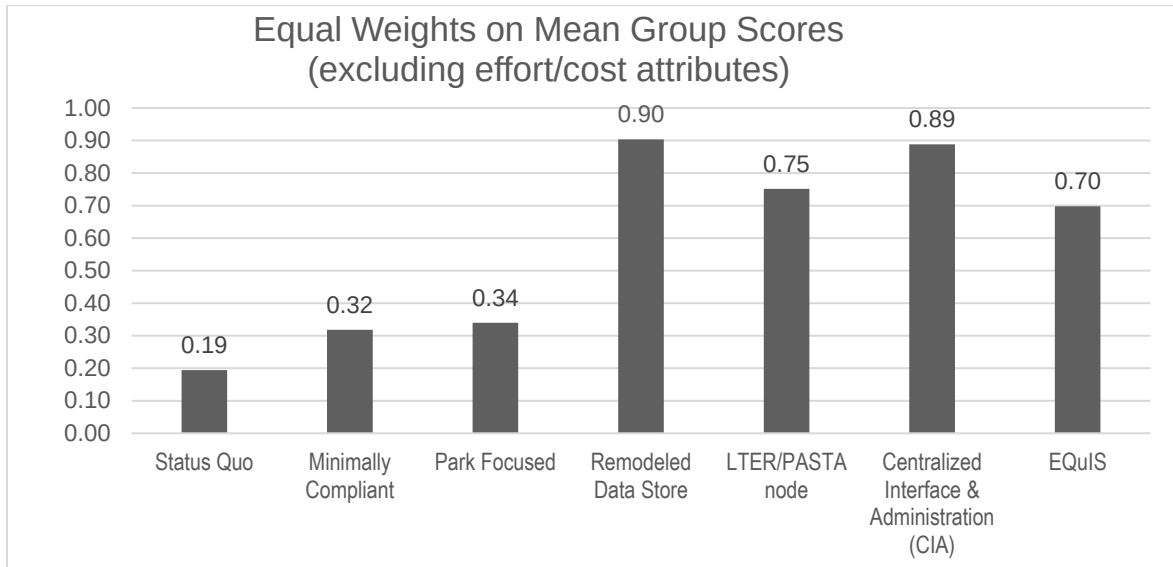


Figure A-7. LTDM decision model outcome of equal weights on mean group scores, excluding effort and cost related performance attributes.

When examining individual Task Force member tradeoff weights against mean group consequence scores, there were few differences in the highest-scored alternative, but the second-position alternative was sometimes affected (Figure A-8). Of the 12 sets of tradeoff weights submitted for the final decision model, 11 responses resulted in *Remodeled Data Store* in first place; while the last response resulted in a tie between *Remodeled Data Store* and *Centralized Interface & Administration*. Ten responses resulted in the *EQUIS* alternative in fourth place, while two responses had *Minimally Compliant* in fourth place.

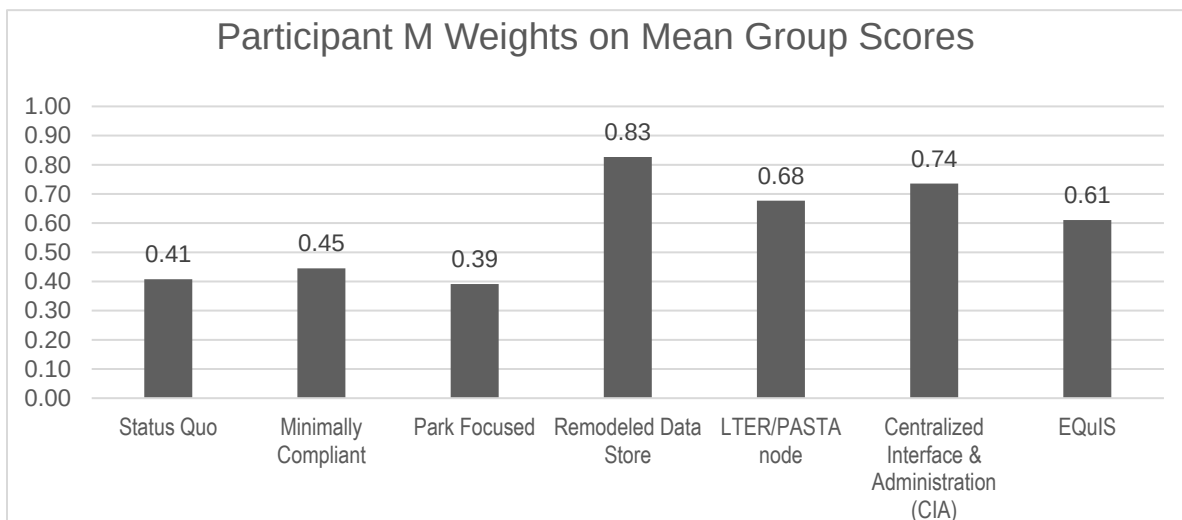


Figure A-8. LTDM decision model outcome: A typical individual task force member weighting result against mean group scores.

Of the 12 sets of tradeoff weights submitted for the final decision model, 10 responses resulted in *Centralized Interface & Administration* in second position; one response saw *Centralized Interface & Administration* tied for first place with *Remodeled Data Store*, and one response saw *LTER/PASTA Node* in second place slightly ahead of *Centralized Interface & Administration* (Figure A-9).

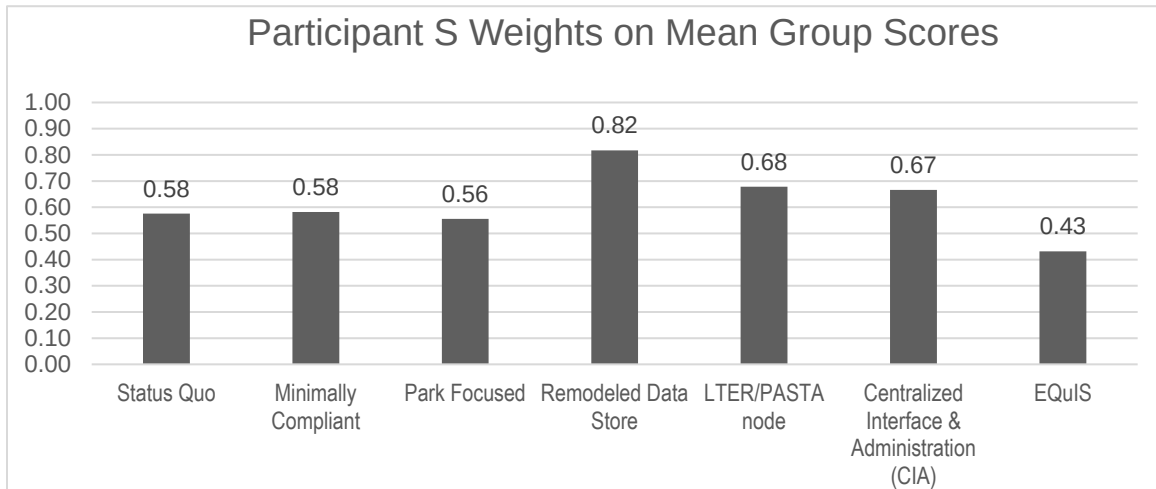


Figure A-9. LTDM decision model outcome: The single task force member response that placed *LTER/PASTA Node* in second place (evaluated using mean group scores).

Of the 12 sets of tradeoff weights submitted for the final decision model, 11 responses resulted in *LTER/PASTA Node* in third position. A single response placed *LTER/PASTA Node* in second position, ahead of *Centralized Interface & Administration*. Three task force member responses saw *EQUIS* placed very closely behind the third position, yet still in fourth position (Figure A-10).

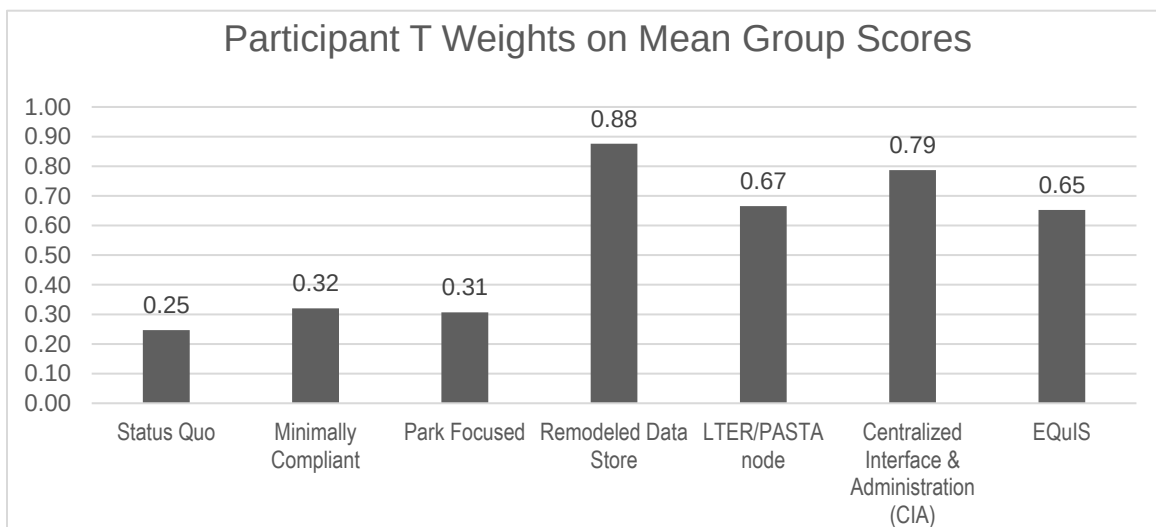


Figure A-10. LTDM decision model outcome: *EQUIS* close behind *LTER/PASTA Node*.

Looking at the model outcome in another way, results of equally weighted tradeoff weights on individual task force member consequence scores also demonstrated similar outcomes in the majority of cases (Figure A-11 and Figure A-12). Individual consequence scores do not seem to have a large effect on the first, second, and third place alternatives overall. The notable exception to this is that *LTER/PASTA Node* is elevated from third place to second place in 3 of 12 responses (25%).

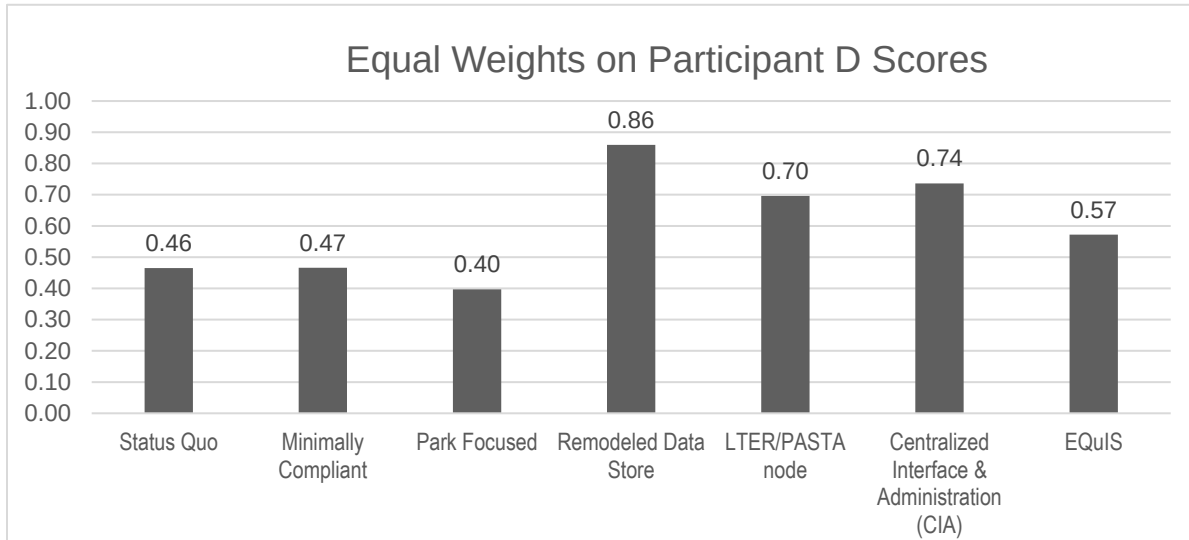


Figure A-11. LTDM decision model outcome: A typical individual task force member scoring result against equal tradeoff weights.

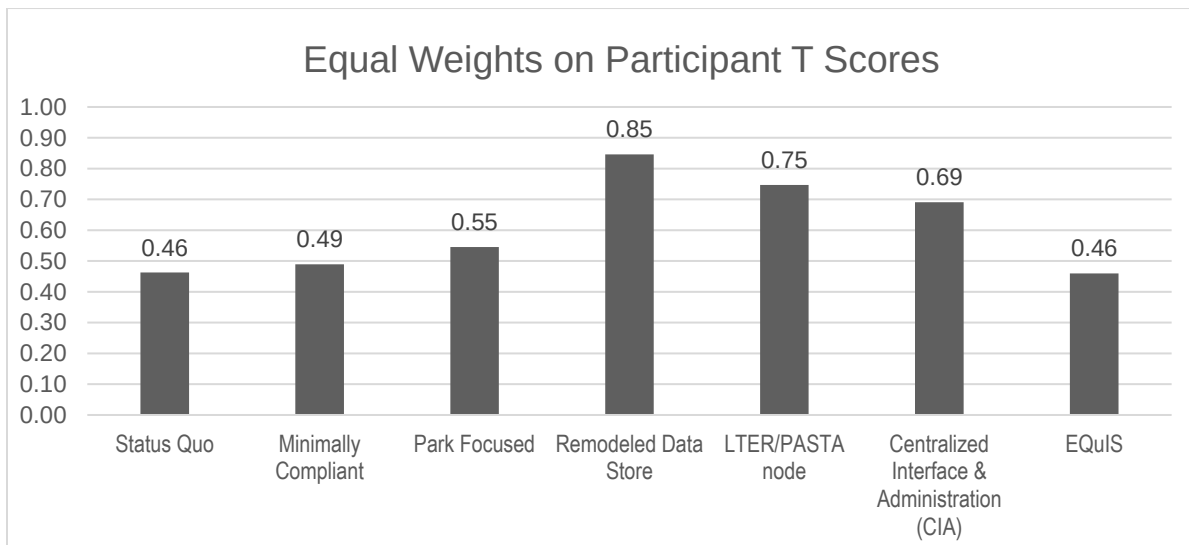


Figure A-12. LTDM decision model outcome: *LTER/PASTA Node* elevated to second place (individual scoring against equal tradeoff weights).

Sensitivity Analysis

Several different analyses were performed across the multiple iterations while developing the LTDM decision model. Performing sensitivity analyses helps to identify potential bias and determines the

impact of uncertainty on the model outcome. These analyses allow for better understanding of model robustness and help determine which issues are critical to resolve in a subsequent iteration of the decision model. Where resolution is not possible, sensitivity analyses help identify which issues must be accounted for in some other manner.

To explore uncertainty in the group's understanding of the alternatives, the Task Force ran several different versions of an early iteration of the decision model. The standard version was a model that used consequence predictions achieved under consensus (i.e., the group's consensus scores indicating how well each alternative performed against each attribute). Another version was created to shift scores to utilize all lowest reasonable scores reported in the elicitation exercise. Yet another version was created to shift scores to all highest reasonable scores reported in the elicitation. Mixed versions were also possible. This exercise helped the group to visualize how the model changes if the lowest or highest performance values contributed by the task force were treated as the "true" values. The exercise also led the group to suspect that the attributes and performance measures in this early model iteration seemed biased to favor certain alternatives over others. As a result, the task force executed a plan to revise the list of attributes and performance measure scales to create attributes and measures that would more appropriately reflect the fundamental objectives and help distinguish the alternatives from one another to a greater degree. Following several revisions, this yielded a new set of attributes and measures used in the final decision model.

Since the Task Force was less confident in the effort and cost consequence predictions, sensitivity of the model to effort and cost related attributes was also explored by including or excluding those particular attributes. Robust to this change, *Remodeled Data Store* typically identified as the top scoring alternative.

The effect of different preference weighting strategies and scoring values in the final decision model was also explored. The Task Force examined patterns in the highest scoring alternatives under the following scenarios: individual scores with individual weights, mean scores with equal weights including all attributes, mean scores with equal weights excluding effort and costs, individual scores with equal weights, and equal scores with individual weights. In all cases, a very similar pattern emerged, providing additional confidence in the robustness of the decision model.

Discussion

The results of the decision analysis demonstrate a unanimous pattern. *Remodeled Data Store* is the alternative that performed the best in the Task Force's analysis; it is expected to provide the most optimal solution for IMD long-term data management.

Though the *Remodeled Data Store* is anticipated to require greater effort and cost as compared with *Status Quo*, *Minimally Compliant* and *Park Focused*, it is expected to perform better on attributes of high importance to long-term data management. It fulfills a broader suite of solution requirements and best addressed the fundamental objectives of this decision. Overall, it is expected to improve on current preservation practices, encourage sustainable workflows, provide for data provenance, and greatly improve efficiency overall, especially as the complexity and size of IMD datasets grow over time.

One of the most important messages the Task Force offered for the IMD Leadership Team in making the final decision and proceeding into the next steps of this process is:

How this solution is implemented is at least as important as which alternative is chosen.

Sources and Types of Uncertainty

A number of sources of uncertainty affected the LTDM decision problem. Several important considerations needed to be taken into account in the decision, some complementary and some competing. While a list of general considerations was provided by the decision-makers, a clear set of goals was lacking, so the Task Force first needed to develop and agree on objectives. Indeed, much time was spent addressing this “framing uncertainty”. This also contributed to uncertainty in how to express and measure the objectives in ways that would be meaningful to a decision outcome (i.e., “values uncertainty”). The Task Force also knew that the data management solution could include a large number of component parts that could foreseeably be mixed and matched across possible alternative solutions, creating what could be a large number of possible solution portfolios. Determining how to best select an appropriate range of alternatives came in to play in this decision.

Perhaps the most influential source of uncertainty in the LTDM decision model was lack of clarity in the alternatives themselves. The technical capacity of different software options was far from explicitly understood. The Task Force needed to not only identify suitable options, but evaluate the merits of each. Given the timeframe for submitting a recommendation, the Task Force was forced to operate with uncertainty in the actual performance capacity of different options. While a great deal of time and energy was spent reducing this source of uncertainty to the extent possible, it remained as the greatest source of unresolved uncertainty in the LTDM project. This was the primary reason for different assignment of performance scores across Task Force members, and was the premise for the degree of sensitivity testing that was used to evaluate different possible outcomes. The task force believed that differences in individual understanding about what different alternatives could or would do likely exceeded the inherent variation in preferences for which attributes were more important. But despite the differences in individual scores, the model outcomes were relatively robust (i.e., the same pattern emerged even with the Task Force members’ variable understanding of alternative solutions and their resultant scores). The group agreed that regardless of what solution is chosen as the final decision, the specifics of how it will be implemented can vary greatly and can significantly impact its performance and long-term success.

Value of Decision Structuring

Decision structuring was valuable to the LTDM Task Force effort. The process offered a collaborative space for close interaction among central office and network staff and the leveraging of existing expertise across IMD. The iterative nature of the SDM process allowed the Task Force to fully explore and incorporating what was learned about the problem as the work proceeded.

The LTDM decision problem is an unusually complex one in both scope and scale. The recommended solution will need to be implemented on a national scale across 32 individual I&M networks and the central IMD office, requiring hundreds of NPS staff members to modify key aspects of how they do business. Each I&M network has unique needs and capacities with existing

local systems and processes that must be modified over time to conform to a universal data management solution. Potential data management solutions include a large number of components addressing 56 mandatory requirements that can be addressed in nearly infinite ways - only a fraction of which are within the immediate purview of the initial recommendation. Structuring this decision allowed for the necessary simplification, enabling this problem to be managed in a tractable way.

Recommendations

The LTDM Task Force unanimously recommends the Remodeled Data Store as the preferred solution for long-term data management for IMD.

The LTDM Task Force agrees that implementation details, which will be determined in the course of project planning and implementation, will have a strong impact on long-term success and functionality of the solution. Many details either not relevant or not available during the decision process will need to be concretely defined leading up to and during project development. As a cautionary note, the task force reminds the leadership team that *how* the alternative is implemented is at least as important as *which* alternative is implemented.

The task force stresses reliance on the Solution Requirements List as important foundational item when proceeding toward implementation.

The task force recommends phasing-in different components of the recommended alternative in a logical manner across the time period of development.

The task force recommends building-in to project management cycles an annual decision review, to ascertain whether the decision is maintains relevancy and currency as time progresses given any new legal mandates, availability of new software products, etc. and to incorporate anything learned about the alternative in the process of implementing its initial phases.

The LTDM Task Force recommends structured decision processes (such as SDM) be used in any major future decision in IMD as a way to effectively leverage expertise from across the program and to build consensus.

Acknowledgements

The LTDM structured decision process would not have been possible without the tireless efforts and dedication of the other LTDM Task Force members. In alphabetical order by last name, they are: Isabel Ashton, Margaret Beer, John Boetsch, Peter Budde, Bonnie Campbell, Rob Daley, Michael DiSibio, Brent Frakes, Mark Hart, Angela Jarding, Bill Johnson, Simon Kingston, Dan Kocol, Adam Kozlowski, Kristin Legg, Scott Miller, Lisa Nelson, Judd Patterson, Tom Philippi, Geoffrey Sanders, and Marianne Tucker.

Thanks are also due to the many individuals throughout IMD who attended LTDM webinars, reviewed materials on the LTDM website, and otherwise remained engaged with this process and provided feedback to the task force members. The task force members thank their supervisors for allowing the work time to participate on this project. Thanks are in order to the IMD Leadership Team, for giving the task force the creative space necessary to develop its recommendation.

The authors thank Bonnie Campbell, Simon Kingston, and Peter Budde for providing comments on an earlier version of this report.

Literature Cited

- Belton, V., and T. J. Stewart. 2002. Multiple criteria decision analysis. Kluwer Academic Publishers.
- Frakes, B., M. Tucker, and D. Kocol. 2016. EarthSoft EQuIS evaluation: Evaluation summary. NPS. Fort Collins, CO.
- Goodwin, P., and G. Wright. 2004. Decision analysis for management judgement. 3rd Ed. Wiley and Sons, New York, NY.
- Hammond, J. S., R. L. Keeney, and H. Raiffa. 1999. Smart choices: A practical guide to making better life decisions. Broadway Books, New York, NY.
- Larson, M. A., D. R. Smith, S. J. Converse, J. A. Szmanski, and J. E. Lyons. 2016. Consequences. Module 7 in: M. C. Runge, A. M. Romito, G. Breese, J. F. Cochrane, S. J. Converse, M. J. Eaton, M. A. Larson, J. E. Lyons, D. R. Smith, and A. F. Isham. Introduction to structured decision making, 2016 edition. US Fish and Wildlife Service. National Conservation Training Center, Shepherdstown, WV.
- Runge, M. C., A. M. Romito, G. Breese, J. F. Cochrane, S. J. Converse, M. J. Eaton, M. A. Larson, J. E. Lyons, D. R. Smith, and A. F. Isham. Introduction to structured decision making, 2016 edition. US Fish and Wildlife Service. National Conservation Training Center, Shepherdstown, WV.
- Smith, D. R., J. A. Szymanski, M. C. Runge, J. F. Cochrane, and S. J. Converse. 2016. Normative decision making. Module 2 in: M. C. Runge, A. M. Romito, G. Breese, J. F. Cochrane, S. J. Converse, M. J. Eaton, M. A. Larson, J. E. Lyons, D. R. Smith, and A. F. Isham. Introduction to structured decision making, 2016 edition. US Fish and Wildlife Service. National Conservation Training Center, Shepherdstown, WV.
- Speirs-Bridge, A., F. Fidler, M. McBride, L. Flander, G. Cumming, and M. Burgman. 2010. Reducing overconfidence in the interval judgement of experts. *Risk Analysis* 30(3):512-523.
- Tucker, M. 2016a. Long Term Data Management Task Force Charter. NPS. Fort Collins, CO.
- Tucker, M. 2016b. Long-Term Data Management Task Force: Considerations for defining requirements. NPS. Fort Collins, CO.
- Van deVen, A. H., and A. L. Delbecq. 1974. The effectiveness of Nominal, Delphi, and Interacting group decision making processes. *The Academy of Management Journal* 17(4): 605-621.
- National Park Service. 2017. Long-Term Data Management Task Force Report: Recommendations for a long-term data management solution. Fort Collins, CO.

Appendix B. List of requirements developed by the Long-Term Data Management Task Force

Table B-1. Mandatory requirement statements and assumptions.

Statement	Assumptions
The solution shall make data accessible in efficient, non-proprietary, common, machine-readable formats	Efficiency is measured by the ability to convey the same information with the least amount of bytes.
The solution shall use delivery systems that meet the requirements of the Rehabilitation Act of 1973 (specifically 508 compliance)	–
The system shall ensure that data are registered with DATA.GOV and data.doi.gov	–
The solution shall support common citation formatting	IMD will define a common/standard citation format
The solution shall provide clear guidelines for data product publications	–
The solution shall include ways to facilitate discovery by key stakeholders	key stakeholders are defined by IMD or in monitoring protocols
Solution will ensure discoverability of all authoritative data sets.	• "All authoritative data sets" includes 'all natural resource data collected by I&M activities'. • A 'single authoritative source' is identified for the dataset/project. • Data sets' may include annual and cumulative data sets. • 'Discoverability': ○ includes but is not limited to: APIs, web services, web crawlers (online methods), ○ implies indexing (incl. filtering on spatial, temporal, other properties) and cataloging of available data (data sets, metadata, standard products) ○ may be at the data set or record level as appropriate. • Assumes definition of structured workflow (process) for changes/updates of various kinds to the authoritative source.
Authoritative datasets are made available in a discoverable repository whose availability is shared with external providers.	–
Repository must include DOIs for datasets and IMD must mint DOIs that point to holdings on the NPS archive file server	–

Table B-1 (continued). Mandatory requirement statements and assumptions.

Statement	Assumptions
The solution shall accommodate/store hierarchical data products resulting from data standardization or aggregation efforts	–
The solution shall ensure that all finalized records are sent to NARA within the appropriate time frame and with the correct retention schedules	–
The solution shall follow industry standard data backup, disaster recovery, and business continuity practices.	–
The solution shall provide access to all versions of data and related documentation (e.g. protocols and SOPs)	Specifically, what was the authoritative version of the data on any prior date can be recovered.
The system shall routinely check for and fix errors introduced in transmission and/or storage to prevent degradation	–
The solution shall ensure that archived data remains accessible	TP Note: This is about previous versions of the data being made available.
The solution shall comply with federal, NPS, and IMD data preservation policy and timeframes	Clear guidance, workflows, and tools will be provided to meet this need; Specific preservation policy and timeframes will need to be explicitly stated; Solution needs to be built with functionality in place to perform audits or trigger notifications.
Solution supports restricting data availability based on defined and standardized data processing levels (including quality, objectivity, and utility thresholds). (AK)	Assumes clear business logic/rules for identifying if quality, objectivity, and utility measures are met. (Clear threshold values.) (BC)
Solution includes administrative mechanisms for data users to obtain corrected information. (BC)	Assumes agency issues implementing guidelines that allow affected people to seek & obtain corrected information (BC)
Solution's data certification processes can accommodate QA activities completed at point of collection. (AK)	AK (4/13/17): Assume automagical steps for this.
Solution clearly defines and communicates data quality standards (AK)	AK: Clarification of vocabulary around QA/QC/Data Processing Level
Dataset quality assignment is independent of protocol version or changes in technology. (BC, AK)	AK: Want to ensure not a moving target for quality assignment due to new tech.
Solution provides guidance for versioning & change management. (BC)	–
Solution's data quality must align with NPS quality policies. (BC, AK)	Assumes I&M quality policies are aligned with NPS quality policies.
Support corrections to authoritative data after analyses and reporting and prior to public release of data	n.b. 2 pieces: fixes to old data, and embargo before public release. TP2 suggests scoring this only on part 1
Solution identifies and mitigates privacy and security risks as federally mandated. (BC)	Assumes identification & mitigation strategies are clearly defined. (BC)

Table B-1 (continued). Mandatory requirement statements and assumptions.

Statement	Assumptions
Data publishers need to be able to define appropriate users	–
Solution provides information security protections commensurate with the risk and magnitude of harm resulting from unauthorized access, use, disclosure, modification or destruction of information collected or maintained on behalf of the agency. (BC)	Assumes risks appropriately assessed & mitigations identified. (BC)
Solution requires clear documentation and communication of system security. (BC)	Clear is defined w/r/t documentation & communication of system security. (BC)
Solution provides clear standards/guidelines for identifying and securing sensitive data. (AK)	• "Clear" is defined w/r/t standards/guidelines. (BC) • Includes but is not limited to sensitive locations, species. (BC) • Appropriate expertise is consulted in establishing/changing standards/guidelines. (BC)
Solution is in compliance with broader NPS IT policies/limitation that impact data security and access. (BC, AK)	SK: We're thinking about this in terms of Authority to Operate (ATO) to have a framework for thought.
System limits read/write access to appropriate users	–
Solution provides access to appropriate users in the appropriate format or scale	4/14/17 TP: assuming personalized settings for what products are presented
Solution will include a defined timetable by which data are published and made available to various stakeholders. (AK)	• Timeliness is defined. (BC) • Number of years after collection is defined. (BC) • Timetable takes into consideration that data needs to be delivered in a timely manner to be valuable for decision making. (AK) • Timetable takes into consideration work load (e.g. QAQC) for the data producer/manager. (AK) • Timetable takes into account data quality level (different release times for raw vs. certified, etc.)? (AK) • Timetable defines specific timelines for various data quality "releases". (BC)
Solution addresses/mitigates barriers to timeliness (e.g. transitions to new system; suitable staffing resources, cultural barriers to sharing data). (BC, AK)	• Appropriate QA/QC & sufficient staff levels are defined. (BC) • Barriers to timeliness are known (AK) • Barriers includes but is not limited to bottlenecks, cultural barriers. (BC) • Solution must be assessed for barriers. (BC)
System must support discovery of data by scientific topic, location, etc. by NPS (park or network) staff	–

Table B-1 (continued). Mandatory requirement statements and assumptions.

Statement	Assumptions
Repository supports archiving and authoritative versions of cumulative monitoring data, not only of annual increments of data.	• Networks will episodically submit data (& metadata) to central system as chunks (data packages, EDDs, etc.). • Process is in place to manage updates to previously submitted data and rebuild cumulative dataset. (sk)
Exported data are in schemas/themes/structures standard for that subject: such as WQX, VegX, ...	–
System must support upload of data.	–
Click-through disclaimer about suitability for any specific use is on the end user, or else pointing all liability at Chief Information Officer.	We may need some combination of appropriate use disclaimer as click-through for public download, and associated dataset documentation warning against most obvious mis-uses of the data (e.g., treating non-probability sample plots as representative or proportional to the entire park) (tp2)
Solution should accommodate non-tabular data (e.g., images, sounds, spatial).	–
The system provides solutions to ensure local user access to authoritative data even with central server outages, network or vpn outages, IT network security blocks, or other foreseeable events.	^\$#! happens. Networks and parks need uninterrupted ability to work with their data. If they can't work with the central repository data because of server or network issues, they will work with local (non-authoritative) data despite policies.
Software version and data format addressed? Should software version and perhaps data format be addressed with data records. In simply the metadata or the product?	Data formats and versions evolve/change
System will accommodate intermediate products (replaced by 510)	some derived products such as population estimates, IBIs and other indices, will be re-used by other projects, may require hundreds of CPU hours to re-compute, and should be treatable as if they were data
System will support network-specific ancillary tables such as management or regulatory thresholds, including indexing, versioning, and archiving such tables if networks submit them.	networks have various ancillary tables such as WQ thresholds that they use in analyses & reporting. These are part of the analyses, reports need to point to them in an archive.
System enforces a structured process for 'authoritative data source' changes/updates/archiving/logging of changes/corrections.	• Authoritative data source' is defined. • Structured process is defined for changes, updates, archiving, logging of changes/corrections.

Table B-1 (continued). Mandatory requirement statements and assumptions.

Statement	Assumptions
METADATA CONTENT: Solution provides metadata from appropriate recognized scientific metadata standards that require a minimum set of responses and include: contact information for data source, quality information, usage/suitability (i.e., to determine applicability for addressing novel scientific or management questions), data dictionaries, location, time, species names (and other scientific domain-specific attributes), and keywords.	Sufficient usage metadata is needed to determine suitability and applicability for addressing scientific or management questions; such usage metadata may include target population, sample design, stratum boundaries, inclusion probabilities, etc.
METADATA SEARCH FUNCTIONALITY: Solution supports discoverability by allowing search by metadata attributes of: location, time, species names (and other scientific domain attributes), keywords.	–
OTHER METADATA FUNCTIONALITY: Solution vends metadata with downloadable products.	4/17/2017: tp: Could be met by downloading protocol & SOPs
The solution shall identify and provide access to a dataset's originating protocol package (narrative, SOP's, QAP, version). In other words, for any data point, it is possible to trace back to the suite of methods used.	Methods and standards (including QAQC) as documented in the protocol package are sufficiently clear to facilitate 'substantial reproducibility' when original/supporting data are independently analyzed using identical methods.
The solution must allow users to see the authoritative data as it was at a point in time.	–
The solution must audit progress of data lifecycle against a plan.	An accepted, and enforceable 'data management lifecycle' exists at the network level. (RD) system can determine what data are not yet in system? So is the audit report by protocol, by data table, or by something else? (TP); • Timeliness is defined. (BC) • Accountability includes both sticks and carrots (rewards) for timeliness. (BC) • Rewards & reward levels are identified. (BC)
The solution must allow for entry and reporting of data lifecycle stage as data are processed, archived, and published.	–
Authoritative datasets in their native schema can be directly queried and subset	–
Solution supports metadata using domain-standard tags (FGDC, Dublin Core, Darwin Core, EML, etc.).	–

Appendix C. Details for Alternatives

Alternatives vs. Mandatory Requirements⁵

A set of 10 initial Alternatives were assessed against the list of mandatory Requirements that the Task Force created using ratings of: Pass (P), Fail (F), Add (A) [could be added], or Maybe (M) (see Table C-1).

Original Brief Descriptions

These are the brief descriptions that were in use at the time that Alternatives were assessed against mandatory Requirements.

- Status Quo: Don't change anything - every network does what they want and central office doesn't issue much guidance, but just provides assistance when needed.
- Minimally Compliant: Satisfy compliance, but with minimal impact (change) to network operations
- Park Focused: Park-focused to low WASO overhead and maximize delivery to parks.
- Adapt/Adopt Topic-specific Systems: Take ideas for applications and add to ours - USFS FIA & Plots, NEON, AKN, WRD EQuIS/Aquarius. Adapt tools and workflows into our own applications. No assumption of pushing data back to original organization's version.
- Data Store+more SQL Server: Data Store opens up posted data files/linked SQL databases and stands up structured web services on top of the data for broader consumption, using the flags in Data Store (visibility, protected status) to determine what to expose.
- LTER/PASTA node: Team up with LTER/ecoinformatics community and become active partners. Our data node fully exchanges information with LTER, NEON, academics, etc.
- Build Deluxe: Policies, actions, infrastructure set up in a way that not only fulfills our needs, but models best practice to other branches of government. May involve adopting current best practice in some cases, but crafting best practice in others.
- Everything Central: All network databases are absorbed into centrally stored, Data Center databases and networks must connect to those to load data. No more office data entry databases - enter directly into centralized database.
- Phased Approach: trajectory possibly leading to something like "at your service", but with viable stop points along the way.
- EQuIS: EQuIS (n.b. This is for separate tables & EDD definitions for different protocols; Munged into common schemas isn't a good fit for Network Utility) - assume it includes WRD EQuIS and Aquarius

⁵ Copied from Google Sheet "[LTDM Alternative Solution Packagers \(ASP\) - Draft Alternatives for Review](#)" tab "Alternatives vs. Requirements" on 7/27/2017.

Table C-1. Ratings for alternatives to mandatory requirements. The ratings are Pass (P), Fail (F), Could be Added (A), or Maybe (M).

ID	Statement	Assumptions	Status Quo	Minimally Compliant	Park Focused	Adapt/Adopt Topic-specific systems	Data Store + more SQL Server	LTER/PASTA Node	Build deluxe	Everything central	Phased Approach	EQulS
24	The solution shall make data accessible in efficient, non-proprietary, common, machine-readable formats	Efficiency is measured by the ability to convey the same information with the least amount of bytes.	F	F	F	P	P	P	P	P	P	P
33	The solution shall use delivery systems that meet the requirements of the Rehabilitation Act of 1973 (specifically 508 compliance)	–	A	A	F	A	A	M	P	A	M	A
39	The system shall ensure that data are registered with DATA.GOV and data.doi.gov	–	M	P	M	M	M	A	P	P	P	A
45	The solution shall support common citation formatting	IMD will define a common/standard citation format	F	F	F	M	F	P	P	M	A	A
152	The solution shall provide clear guidelines for data product publications	–	M	M	M	M	M	M	P	M	M	M
157	The solution shall include ways to facilitate discovery by key stakeholders	key stakeholders are defined by IMD or in monitoring protocols	P	P	P	P	P	P	P	P	P	A
306	Solution will ensure discoverability of all authoritative data sets.	• "All authoritative data sets" includes 'all natural resource data collected by I&M activities' • A 'single authoritative source' is identified for the dataset/project. • 'Data sets' may include annual and cumulative data sets. • 'Discoverability': ◦ includes but is not limited to: APIs, web services, web crawlers (online methods), ◦ implies indexing (incl. filtering on spatial, temporal, other properties) and cataloging of available data (data sets, metadata, standard products), ◦ may be at the data set or record level as appropriate. • Assumes definition of structured workflow (process) for changes/updates of various kinds to the authoritative source.	F	P	F	M	P	P	P	P	P	A
501	Authoritative datasets are made available in a discoverable repository whose availability is shared with external providers.	–	F	P	F	M	P	P	P	0	–	A
511	Repository must include DOIs for datasets and IMD must mint DOIs that point to holdings on the NPS archive file server	–	A	P	A	A	A	P	P	A	P	A

Table C-1 (continued). Ratings for alternatives to mandatory requirements. The ratings are Pass (P), Fail (F), Could be Added (A), or Maybe (M).

ID	Statement	Assumptions	Status Quo	Minimally Compliant	Park Focused	Adapt/Adopt Topic-specific systems	Data Store + more SQL Server	LTER/PASTA Node	Build deluxe	Everything central	Phased Approach	EQulS
37	The solution shall accommodate/store hierarchical data products resulting from data standardization or aggregation efforts	–	P	P	P	P	P	M	P	P	P	P
61	The solution shall ensure that all finalized records are sent to NARA within the appropriate time frame and with the correct retention schedules	–	F	P	M	P	P	A	P	M	P	A
62	The solution shall properly manage IMD data throughout its life cycle, including all stages through which the information passes, including: creation, collection, use, processing, storage, dissemination, and disposition.	DO 11A includes explicit and measurable requirements	–	–	–	–	–	–	–	–	–	–
63	The solution shall follow industry standard data backup, disaster recovery, and business continuity practices.	–	M	M	A	P	A	P	P	A	A	A
64	The solution shall provide access to all versions of data and related documentation (e.g. protocols and SOPs)	Specifically, what was the authoritative version of the data on any prior date can be recovered.	F	P	M	A	M	P	P	A	P	A
66	The system shall routinely check for and fix errors introduced in transmission and/or storage to prevent degradation	–	F	F	F	A	F	M	P	A	A	A
159	The solution shall ensure that archived data remains accessible	TP Note: This is about previous versions of the data being made available.	F	P	M	A	M	P	P	A	P	A
300	The solution shall comply with federal, NPS, and IMD data preservation policy and timeframes	Clear guidance, workflows, and tools will be provided to meet this need; Specific preservation policy and timeframes will need to be explicitly stated; Solution needs to be built with functionality in place to perform audits or trigger notifications.	F	P	A	A	A	M	P	A	P	A
88	Solution supports restricting data availability based on defined and standardized data processing levels (including quality, objectivity, and utility thresholds). (AK)	Assumes clear business logic/rules for identifying if quality, objectivity, and utility measures are met. (Clear threshold values.) (BC)	F	A	A	A	A	A	P	A	A	P
92	Solution includes administrative mechanisms for data users to obtain corrected information. (BC)	Assumes agency issues implementing guidelines that allow affected people to seek & obtain corrected information (BC)	F	F	P	A	A	A	P	A	A	A
108	Solution's data certification processes can accommodate QA activities completed at point of collection. (AK)	AK (4/13/17): Assume automagical steps for this.	M	M	P	A	M	M	P	A	A	A

Table C-1 (continued). Ratings for alternatives to mandatory requirements. The ratings are Pass (P), Fail (F), Could be Added (A), or Maybe (M).

ID	Statement	Assumptions	Status Quo	Minimally Compliant	Park Focused	Adapt/Adopt Topic-specific systems	Data Store + more SQL Server	LTER/PASTA Node	Build deluxe	Everything central	Phased Approach	EQulS
168	Solution clearly defines and communicates data quality standards (AK)	AK: Clarification of vocabulary around QA/QC/Data Processing Level	F	M	M	M	M	A	P	M	A	M
170	Dataset quality assignment is independent of protocol version or changes in technology. (BC, AK)	AK: Want to ensure not a moving target for quality assignment due to new tech.	P	P	P	P	P	P	P	P	P	P
171	Solution provides guidance for versioning & change management. (BC)	–	F	A	A	A	A	A	P	A	A	A
172	Solution's data quality must align with NPS quality policies. (BC, AK)	Assumes I&M quality policies are aligned with NPS quality policies.	F	P	M	P	P	P	P	P	P	P
504	Support corrections to authoritative data after analyses and reporting and prior to public release of data	n.b. 2 pieces: fixes to old data, and embargo before public release. TP2 suggests scoring this only on part 1	M	P	P	A	P	A	P	A	P	P
700	The solution shall accept and vend all IMD data, including metadata, and standard products, with the appropriate qualifications of quality (e.g., raw, provision, accepted, etc.) to appropriate audiences	–	–	–	–	–	–	–	–	–	–	–
29	Solution identifies and mitigates privacy and security risks as federally mandated. (BC)	Assumes identification & mitigation strategies are clearly defined. (BC)	F	P	P	A	P	A	P	A	P	P
60	Data publishers need to be able to define appropriate users	–	P	P	P	P	P	M	P	P	P	A
67	Solution provides information security protections commensurate with the risk and magnitude of harm resulting from unauthorized access, use, disclosure, modification or destruction of information collected or maintained on behalf of the agency. (BC)	Assumes risks appropriately assessed & mitigations identified. (BC)	P	P	P	P	P	P	P	P	P	P
140	Solution requires clear documentation and communication of system security. (BC)	Clear is defined w/r/t documentation & communication of system security. (BC)	F	F	F	F	F	A	P	F	F	F
144	Solution provides clear standards/guidelines for identifying and securing sensitive data. (AK)	"Clear" is defined w/r/t standards/guidelines. (BC) - Includes but is not limited to sensitive locations, species. (BC) - Appropriate expertise is consulted in establishing/changing standards/guidelines. (BC)	F	A	A	A	A	A	P	A	A	A
145	Solution is in compliance with broader NPS IT policies/limitation that impact data security and access. (BC, AK)	SK: We're thinking about this in terms of Authority to Operate (ATO) to have a framework for thought.	A	P	A	P	P	A	P	P	P	P

Table C-1 (continued). Ratings for alternatives to mandatory requirements. The ratings are Pass (P), Fail (F), Could be Added (A), or Maybe (M).

ID	Statement	Assumptions	Status Quo	Minimally Compliant	Park Focused	Adapt/Adopt Topic-specific systems	Data Store + more SQL Server	LTER/PASTA Node	Build deluxe	Everything central	Phased Approach	EQulS
303	System limits read/write access to appropriate users	–	P	P	P	P	P	M	P	P	P	A
304	Solution provides access to appropriate users in the appropriate format or scale	4/14/17 TP: assuming personalized settings for what products are presented	F	F	P	A	A	F	P	F	A	A
94	Solution will include a defined timetable by which data are published and made available to various stakeholders. (AK)	• Timeliness is defined. (BC) • Number of years after collection is defined. (BC) • Timetable takes into consideration that data needs to be delivered in a timely manner to be valuable for decision making. (AK) • Timetable takes into consideration work load (e.g. QAQC) for the data producer/manager. (AK) • Timetable takes into account data quality level (different release times for raw vs. certified, etc.)? (AK) • Timetable defines specific timelines for various data quality "releases". (BC)	F	A	A	A	A	A	A	P	P	A
113	Solution addresses/mitigates barriers to timeliness (e.g. transitions to new system; suitable staffing resources, cultural barriers to sharing data). (BC, AK)	• Appropriate QA/QC & sufficient staff levels are defined. (BC) • Barriers to timeliness are known (AK) • Barriers includes but is not limited to bottlenecks, cultural barriers. (BC) • Solution must be assessed for barriers. (BC)	F	A	A	A	A	A	P	A	A	A
10	System must support discovery of data by scientific topic, location, etc. by NPS (park or network) staff	–	M	A	A	A	A	P	A	A	P	A
12	Solution needs to support the Summarization and publication of data and findings to resource managers.	• ASP (4/14/17): We assume this is not simply annual reports. • From Isabel Ashton in "Considerations" doc. The original Reference or Issue was "I don't know many resource managers that "analyze" data. I think of their need as 'summarized and published data' to support park planning, EAs, etc."	F	F	–	–	–	–	–	–	–	–

Table C-1 (continued). Ratings for alternatives to mandatory requirements. The ratings are Pass (P), Fail (F), Could be Added (A), or Maybe (M).

ID	Statement	Assumptions	Status Quo	Minimally Compliant	Park Focused	Adapt/Adopt Topic-specific systems	Data Store + more SQL Server	LTER/PASTA Node	Build deluxe	Everything central	Phased Approach	EQulS
65	Repository supports archiving and authoritative versions of cumulative monitoring data, not only of annual increments of data.	• Networks will episodically submit data (& metadata) to central system as chunks (data packages, EDDs, etc.). • Process is in place to manage updates to previously submitted data and rebuild cumulative dataset. (sk)	F	P	P	P	P	A	P	P	P	P
101	Exported data are in schemas/themes/structures standard for that subject: such as WQX, VegX, ...	–	F	F	F	P	A	A	P	A	P	A
106	System must support upload of data.	–	P	P	P	P	P	P	P	P	P	P
137	Click-through disclaimer about suitability for any specific use is on the end user, or else pointing all liability at Chief Information Officer.	We may need some combination of appropriate use disclaimer as click-through for public download, and associated dataset documentation warning against most obvious mis-uses of the data (e.g., treating non-probability sample plots as representative or proportional to the entire park) (tp2)	F	A	A	A	A	A	A	A	A	A
166	Solution should accommodate non-tabular data (e.g., images, sounds, spatial).	–	A	A	P	A	A	P	P	A	P	F
191	The system provides solutions to ensure local user access to authoritative data even with central server outages, network or vpn outages, IT network security blocks, or other foreseeable events.	^\$#! happens. Networks and parks need uninterrupted ability to work with their data. If they can't work with the central repository data because of server or network issues, they will work with local (non-authoritative) data despite policies.	F	F	A	A	A	P	P	A	P	M
211	Software version and data format addressed? Should software version and perhaps data format be addressed with data records. In simply the metadata or the product?	Data formats and versions evolve/change	F	F	F	A	A	P	P	P	A	P
222	System will accommodate intermediate products (replaced by 510)	Some derived products such as population estimates, IBIs and other indices, will be re-used by other projects, may require hundreds of cpu hrs. to re-compute, and should be treatable as if they were data	P	P	P	P	P	P	P	A	P	A
223	System will support network-specific ancillary tables, including versioning archiving (replaced by 510 per TORG)	Networks have various ancillary tables such as WQ thresholds that they use in analyses & reporting. These are part of the analyses, reports need to point to them in an archive.	P	P	P	P	P	P	P	A	P	A
510	System will support network-specific ancillary tables such as management or regulatory thresholds, including indexing, versioning, and archiving such tables if networks submit them.	Networks have various ancillary tables such as WQ thresholds that they use in analyses & reporting. These are part of the analyses, reports need to point to them in an archive.	P	P	P	P	P	P	P	A	P	A

Table C-1 (continued). Ratings for alternatives to mandatory requirements. The ratings are Pass (P), Fail (F), Could be Added (A), or Maybe (M).

ID	Statement	Assumptions	Status Quo	Minimally Compliant	Park Focused	Adapt/Adopt Topic-specific systems	Data Store + more SQL Server	LTER/PASTA Node	Build deluxe	Everything central	Phased Approach	EQulS
307	System enforces a structured process for 'authoritative data source' changes/updates/archiving/logging of changes/corrections.	- Authoritative data source' is defined. - Structured process is defined for changes, updates, archiving, logging of changes/corrections.	F	P	A	P	A	A	P	A	P	P
309	METADATA CONTENT: Solution provides metadata from appropriate recognized scientific metadata standards that require a minimum set of responses and include: contact information for data source, quality information, usage/suitability (i.e., to determine applicability for addressing novel scientific or management questions), data dictionaries, location, time, species names (and other scientific domain-specific attributes), and keywords.	Sufficient usage metadata is needed to determine suitability and applicability for addressing scientific or management questions; such usage metadata may include target population, sample design, stratum boundaries, inclusion probabilities, etc.	F	A	F	P	A	P	P	A	P	A
310	METADATA SEARCH FUNCTIONALITY: Solution supports discoverability by allowing search by metadata attributes of: location, time, species names (and other scientific domain attributes), keywords.	–	M	M	A	P	P	P	P	P	P	P
311	OTHER METADATA FUNCTIONALITY: Solution vends metadata with downloadable products.	4/17/2017: tp: Could be met by downloading protocol & SOPs	F	F	F	A	A	P	P	A	P	A
313	The solution shall identify and provide access to a dataset's originating protocol package (narrative, SOP's, QAP, version). In other words, for any data point, it is possible to trace back to the suite of methods used.	Methods and standards (including QAQC) as documented in the protocol package are sufficiently clear to facilitate 'substantial reproducibility' when original/supporting data are independently analyzed using identical methods.	M	M	M	A	A	P	P	A	P	A
400	The solution must allow users to see the authoritative data as it was at a point in time.	–	M	M	M	A	A	P	P	A	P	A
401	The solution must audit progress of data lifecycle against a plan.	An accepted, and enforceable 'data management lifecycle' exists at the network level. (RD) system can determine what data are not yet in system? So is the audit report by protocol, by data table, or by something else? (TP); - Timeliness is defined. (BC) - Accountability includes both sticks and carrots (rewards) for timeliness. (BC) Rewards & reward levels are identified. (BC)	F	A	F	A	A	A	P	A	A	A

Table C-1 (continued). Ratings for alternatives to mandatory requirements. The ratings are Pass (P), Fail (F), Could be Added (A), or Maybe (M).

ID	Statement	Assumptions	Status Quo	Minimally Compliant	Park Focused	Adapt/Adopt Topic-specific systems	Data Store + more SQL Server	LTER/PASTA Node	Build deluxe	Everything central	Phased Approach	EQUIS
402	The solution must allow for entry and reporting of data lifecycle stage as data are processed, archived, and published.	–	F	A	F	A	A	A	P	A	A	A
503	Authoritative datasets in their native schema can be directly queried and subset	–	M	M	P	F	M	P	P	P	P	P
505	Solution supports metadata using domain-standard tags (FGDC, Dublin Core, Darwin Core, EML, etc.).	–	M	P	M	P	P	P	P	P	P	P

Description, Policy, Process, Technology, Training, Assumptions, and Activities

This section covers the seven final Alternatives. Note that Assumptions and Activities were not detailed for Status Quo, Minimally Compliant, and Park Focused Alternatives.

Remodeled Data Store

Table C-2. Description, policy, process, technology, and training for the final alternatives for the remodeled Data Store.

Aspect	Description
Brief Description	Data Store is enhanced and revised to pull in best elements of LTER/PASTA and/or other systems. Networks can maintain their own databases and add these to Data Store as files with the option of having them converted to a queryable data source. This queryable source conversion requires additional metadata (e.g. schema definition), and results in structured web services that promote broader consumption. Flags in Data Store (visibility, protected status) are used to determine what to expose as services. All systems for metadata, tabular data, and non-tabular data are linked together and presented as a single source.
Policy	- choose metadata standard(s) - finalize QA/QC policy - define authoritative data and make Data Store official repository for authoritative data - require posting data in certain timeframe - protected records defined clearly
Process	Phased Approach to Versioning: - just post it - Narrative of changes in text file goes with data - Details of changes in text file goes with data - Change tracking in database table Visualization on basic protocol metadata rather than analysis.

Table C-2 (continued). Description, policy, process, technology, and training for the final alternatives for the remodeled Data Store.

Aspect	Description
Process (continued)	Viable stop points along the way: Phase 0: Leadership codifies metadata & authoritative data policies; networks submit enhanced metadata and network-designated authoritative data file to Data Store before reporting (2 versions: full & redacted); enhanced tech for authentication for sensitive data & embargoed data; data discovery returns data AND all relevant usage metadata (Protocols, SOPs, documentation) Use Data Store flags for data exposure/hiding. Networks submit redacted + full data sets as holdings to Data Store (with metadata as DS records) in Access or whatever before reporting, but don't turn on public access to the holdings until after reporting. Networks use tools to fortify their metadata. Separate citations for redacted vs. full data (policy & possibly modification to Data Store). Clear guidance on what "turns on" services on datasets and exposes them Phase 1: Tabular data migrated from Access/whatever files to centralized Data Store SQL Server container, embargo and protecting sensitive data driven by AD authentication credentials; networks "subscribe" to subsets of data in central repository so they get all changes mirrored to their local SQL Server Express Same network behavior as phase 0, but tabular data into SQL Server container in network-native schema Phase 2: Networks with data on topic X "register" their authoritative dataset as relevant to that topic; additional "meta-metadata" pointing to tables & variables with: species names, locations, etc.; tool uses that meta-metadata to automatically build lists of species & locations in metadata, etc. Subset of most common variables harvested across relevant protocol datasets for easier rollup & landscape-scale applications. Phase 3: Subsets of most common attributes rolled up into common schema tables Phase 4: Dashboards, visualizers, etc. built upon rollup or meta-metadata

Table C-2 (continued). Description, policy, process, technology, and training for the final alternatives for the remodeled Data Store.

Aspect	Description
Technology	- open up posted holdings for datasets, convert to SQL Server, and build structured web services on top of SQL databases (assuming database and service per dataset) - put links to web service on Data Store citation Phase 0: tools for generating better metadata; Data Store cataloging includes links to related protocols & SOPs, improved permissions module for protected/sensitive data & embargoed data Phase 1: 1 (or more) big SQL Server Instance as container for sets of relational tables for each network protocol Phase 2:
Training	- protected records flagging - how to post to Data Store to protect record appropriately - provide training on consuming web services from Data Store (incl. integration with nps.gov websites)

Table C-3. Assumptions and activities for the final alternatives for the remodeled Data Store.

Aspect	Assumptions	Activities
Metadata Format Consistency	- Metadata standard will be required. - IMD will develop policy on metadata formats with science input (FGDC or EML) - IMD will develop minimum set of required metadata fields - Solution will include non-tabular data formats - Metadata are managed in Data Store for posted items; management of metadata for items not posted to Data Store is ad hoc - Metadata (including linked protocols) are available at time of download. - Metadata manually updated in conjunction with modifications to database. - Data Store is essentially the "one-at-a-time" (i.e. not batch) metadata creation tool by filling in the fields for a citation	- Choose/develop IMD-wide metadata standard(s) to be applied to all tabular and nontabular (ex. GIS, photos) data. - Central office modifies or develops tools to ease metadata generation & maintenance - Data Store updated to accommodate properties of metadata standard(s) for tabular and non-tabular data for both upload and download. - Develop a group to define details of how to use metadata format (one-time) - Maintaining and enforcing (ongoing) - Update metadata records on existing Data Store records to match chosen standard

Table C-3 (continued). Assumptions and activities for the final alternatives for the remodeled Data Store.

Aspect	Assumptions	Activities
Metadata Format Consistency (continued)	9) Morpho or other existing tools could be used to generate EML metadata locally; entering records into Data Store is another way to generate metadata 10) Technology is not the limiting factor in metadata handling for Data Store.	1)
Metadata / documentation auditing	1) Metadata auditing will be implemented over time. 2) Enforcement is achieved by more required fields in Data Store; automated auditing becomes a possibility because metadata are entered into Data Store 3) Metadata required to upload data. 4) Metadata schema as defined in metadata could optionally be used to check against data submitted 5) Line-of-Sight supported/easily discovered (Protocol, SOPs, QAPs, etc.) 6) Auditing would only be applied to records marked as IMD or of a specific reference type	2) Enhancement of Data Store to support metadata schema check against data provided. 3) Development of data auditing tool. 4) Network to provide additional Metadata to describe Schema 5) Enhancement to improve visibility and discoverability of Line-of-Sight documents associated with datasets. 6) Networks will be required to manage and audit all metadata documentation (ongoing). 7) Define metadata auditing levels (presence/absence, structure/format, etc.)
Cumulative (all years) data & consistent structure	1) IMD will develop a policy on providing tabular and non-tabular data in a consistent structure. 2) Tabular data sets will be provided in one of two methods a) Cumulative data sets to replace currently posted datasets b) Non-overlapping incremental datasets 3) Data store will manage relationships to associated non-tabular items 4) Non-overlapping, incremental datasets sharing a schema could be accumulated through the conversion to a queryable data source	1) Central Office will develop a policy to insure authoritative tabular dataset are stored in a structure consistent across time. 2) Data Store enhanced to handle nontabular relationships. 3) Data Store enhanced to handle upload of non-overlapping, incremental datasets 4) Data accumulation will be an activity primarily completed at the network level

Table C-3 (continued). Assumptions and activities for the final alternatives for the remodeled Data Store.

Aspect	Assumptions	Activities
Data quality flagging	1) Metadata records in Data Store contain quality flagging for the entire dataset only. 2) Data flagging of individual rows will be done prior to uploading the dataset by the networks. 3) Specific quality assurance guidelines are developed at the protocol level. 4) Networks will use the most current IMD Certification definitions	1) Central Office finalizes, with network input, guidelines for row/field level flagging. 2) Networks will be required to implement and audit flagging of row-level data (ongoing). 3) Networks will develop quality control guidelines by protocol 4) Central office finalizes Certification/Data Processing Level Guidelines (in particular ensure alignment with the draft QAP guidance)
Customizable querying/subsetting of data	1) Coarse queries based on values stored in metadata. 2) Dynamic subsetting and/or visualization is only available for datasets where the option to convert to stored queryable source was chosen.	1) Data Store enhanced to convert datasets to queryable source (i.e. data extraction from Access or other upload containers to RDBMS) 2) Data owners would need to provide additional metadata attributes and standards to assist in converting data to queryable source. 3) Data owners would need to provide usage metadata to assist in using data from queryable source. 4) Central office and networks will develop visualization tools against records that have been converted to queryable data sources 5) Determine how to reveal relevant usage metadata when dataset is subset or visualized
Connectedness to SOPs/protocols	1) Central office establishes policy to define minimum requirements for Line-of-Sight. 2) Networks will maintain necessary documents to complete Line-of-Sight in Data Store. 3) IMD will establish a consistent model for managing protocol versioning and SOP's.	1) Workgroup formed to establish Line-of-Sight policy and requirements. 2) Data Store would be augmented to improve visibility between data products and documentation. 3) Networks maintain documents in Data Store (ongoing). 4) Data Store will be updated to manage versioning of Protocols and SOPs.

Table C-3 (continued). Assumptions and activities for the final alternatives for the remodeled Data Store.

Aspect	Assumptions	Activities
Capability to interact with unique (external) solutions	1) Capability to interact with unique/custom solutions linked to ability to support customizable querying and subsetting of data.	1) Additional network level training to facilitate making connections to data. 2) Establish an ongoing work-group to determine which unique external solutions to pursue (these will evolve over time in # & functionality) and to what level integration/interaction should occur.
Continuous connection to network data	1) WAN at network is suitable for use of Data Store 2) Ability to have local mirroring is currently restricted by DOI policy	1) Examine network bandwidth program wide and compare to solution needs 2) IMD will invest time and effort into improving Physical Network Connectivity 3) Increase failover capacity at Data Center level
Degree/ease with which data is made available to external systems (e.g. ability to push data to external systems)	1) Degree/ease with which data is made available to external systems is linked to ability to support customizable querying and subsetting of data. 2) Metadata can be shared with anyone (internal and external). 3) Tabular data can only be shared externally if it has flags identifying nonsharable data. 4) Tabular data can only be pushed if data are converted to queryable data source. 5) Enhancements have been made to Data Store (see Metadata performance metrics above) to harvest complete metadata.	1) Identify and develop support for appropriate exchange formats to facilitate tabular data pushes 2) Networks ensure that nonshareable data are flagged
Degree of efficiency for providing IMD operational information (who, what, when, & where of the program)	1) IMD will develop policy on elements of dataset publication that need to be tracked 2) Enhancements have been made to Data Store (see Metadata performance metrics above) to harvest complete metadata. (metadata format needs to be augmented to compensate for the fact that all data are not stored as queryable source).	1) Enhance Data Store to include elements defined in IMD protocol tracking policy. 2) Enhance interaction between Data Store and Protocol Tracker

Table C-3 (continued). Assumptions and activities for the final alternatives for the remodeled Data Store.

Aspect	Assumptions	Activities
Compliance facilitation (e.g. record management)	1) IMD develops and issues policies and guidance on compliance standards. 2) Data Store is enhanced to meet all compliance standards for records management. 3) Networks are not responsible for meeting compliance (they are responsible for meeting IMD reporting requirements which should automatically be harvested and used to meet compliance).	1) IMD will develop compliance standards 2) Enhance Data Store to support compliance requirements
QA/QC performance reporting	1) Whatever metadata format is chosen must support QA flags for auditing 2) Data Store will be enhanced to comply with IMD QAQC requirements. 3) QA for row-level data occurs before uploading 4) 4) Automatic comparison of data to "standards from documentation" is possible but unrealistic	1) Create a clear policy on which QA metrics are auditable and to what extent they are audited
Versioning (dataset-level) capabilities	1) This versioning deals with the entire dataset 2) IMD will develop a consistent policy for all networks 3) Versioning applies to both cumulative and non-overlapping incremental datasets. 4) Whatever metadata format is chosen must support versioning. 5) Versioning will involve some level of system enforcement	1) Create a subteam to establish how versioning will be defined and work (including enforcement) in IMD 2) Networks manage consistent Protocol, Database, Datasets versions (Ongoing)

Table C-3 (continued). Assumptions and activities for the final alternatives for the remodeled Data Store.

Aspect	Assumptions	Activities
Change Tracking (record-level modifications to data) capabilities	1) Change-tracking deals with the record-level mods to data. 2) Change-tracking only required for changes made to data with a Data Processing Level (DPL) of "Provisional" or above. 3) IMD will develop a consistent policy for all networks 4) Change-tracking applies to both cumulative and non-overlapping incremental datasets. 5) Change-tracking occurs before uploading to Data Store (i.e. Network level). 6) If dataset is converted to queryable source, Data Store could handle change-tracking internally	1) Create a subteam to establish how change tracking will be defined and work in IMD 2) Data Store will need to be modified to support change tracking requirements 3) Queryable data sources would be built to support change tracking management

LTER/PASTA Node

Table C-4. Description, policy, process, technology, and training for the LTER/PASTA Node.

Aspect	Description
Brief Description	Leverage existing PASTA open-source code and related tools (e.g. Morpho, Excel metadata templates) to stand up NPS PASTA site on the Windows platform. Networks would maintain their own databases and would upload periodic data packages (with EML metadata) into PASTA in the form of flattened .csv files. If we opt in as a DataOne node we could exchange information with LTER, NEON, academics, etc. We would have flexibility on the look and feel of public/internal pages. The PASTA system handles metadata and packages of tabular and non-tabular data. Other NPS tools would hold and manage relationships between associated non-tabular files, protocol documents, and products (e.g. published reports).
Policy	• cooperative agreement (incl. decision on keeping up w/their dev. or snapshotting application) • how much we tweak/extend software to meet our additional needs • EML as metadata standard • decision on whether cataloging front-end is Data Store w/EML or MetaCat w/EML • backend storage local server/SQL server or full peer DataOne node • posting data in certain timeframe • protected records • finalize QA/QC policy

Table C-4 (continued). Description, policy, process, technology, and training for the LTER/PASTA Node.

Aspect	Description
Process	- instructions on posting to MetaCat if that's used instead of Data Store - confirm/improve passing feedback from data consumers to our data providers - guidance on posting datasets - guidance on doing EML metadata (fill in a lot more metadata than we are currently)
Technology	- potentially improve services/granularity of data vending (dependent on storage [DataOne vs. ?]) - already has built in mirrors and subsets - tools to generate metadata (typically need improvement) - Non-tabular data can be loaded - generic media files - metadata-driven (file type listed in metadata) - searchable by keyword (ontology)
Training	Existing training modules can be used for: - EML metadata - posting data - PASTA, DataOne

Table C-5. Assumptions and activities for the LTER/PASTA Node.

Aspect	Assumptions	Activities
Metadata Format Consistency	1) Metadata standard will be EML 2) IMD will develop minimum set of required metadata fields 3) Metadata are stored in PASTA for posted items; management of metadata is done locally in external tools (e.g. Morpho) and edits would need to be resubmitted 4) Additional policy will be required if IMD chooses to participate with DataOne 5) PASTA supports non-tabular datasets as simple file artifacts (packages)	1) Networks required to generate metadata with an external tool (e.g. Morpho) prior to submitting data 2) Networks required to maintain metadata with external tools and resubmit any changes 3) IMD workgroup created to do a deep dive into the EML standard and establish IMD required fields (IMD EML standard) 4) Standing up and implementing front end 5) IMD workgroup created to do a deep dive into DataOne to determine merits and costs/risks

Table C-5 (continued). Assumptions and activities for the LTER/PASTA Node.

Aspect	Assumptions	Activities
Metadata Format Consistency (continued)	6) Metadata manually updated in conjunction with modifications to database 7) No customization to existing LTER/PASTA code would be required for metadata management 8) Modifications/customizations would be limited due to possible connection to DataOne	6) Migrate all historical records from Data Store into PASTA
Metadata / documentation auditing	1) Enforcement is achieved by required fields in PASTA; automated auditing becomes a possibility because metadata are uploaded into PASTA 2) Metadata required to upload data. 3) Database schema as defined in metadata is required to be used to check against data submitted (only applicable to certain dataset/data types) Questions: 1) How are relationships between protocols, SOPs, and related documents handled/linked?	1) Network required to provide additional Metadata to describe Schema 2) Networks will be required to manage and audit all metadata documentation (ongoing).
Cumulative (all years) data & consistent structure	1) IMD will follow LTER/PASTA policy on providing tabular data in a consistent structure. 2) Tabular data sets will be provided in one of two methods a) Cumulative data sets to replace currently posted datasets b) Non-overlapping incremental datasets 3) For full functionality (e.g. schema comparison) datasets would be provided in flat file formats (e.g.csv) 4) PASTA does not currently support automatic accumulation of packages; will be handled at network level Questions: 1) What format(s) for non-tabular data?	1) Central Office will develop a policy stating that authoritative tabular dataset are stored in a structure consistent across time. 2) Convert relational databases to flat files for upload (ongoing) (a.k.a. "Operation Pancake")

Table C-5 (continued). Assumptions and activities for the LTER/PASTA Node.

Aspect	Assumptions	Activities
Data quality flagging	1) Metadata records in PASTA contain quality flagging for the entire dataset only. 2) Data flagging of individual rows will be done prior to uploading the dataset. 3) Specific quality assurance guidelines are developed at the protocol level. 4) Networks will use the most current IMD Certification definitions	1) Central Office finalizes, with network input, guidelines for row/field level flagging. 2) Networks will be required to implement and audit flagging of row-level data (ongoing). 3) Central office finalizes Certification/Data Processing Level Guidelines (in particular ensure alignment with the draft QAP guidance)
Customizable querying/subsetting of data	1) Coarse queries based on values stored in metadata. 2) Subsetting can be done against flat files, but we are unsure of level of effort/customization required Question: 1) What are the specifics on the interplay between metadata and uploaded data to support customizable queries?	1) Data owners would need to provide additional metadata attributes and standards to assist in converting data to queryable source. 2) Data owners would need to provide usage metadata to assist in using data from queryable source. 3) Central office and networks will develop visualization tools to the PASTA portal
Connectedness to SOPs/protocols	1) Central office establishes policy to define minimum requirements for Line-of-Sight. 2) Networks will maintain necessary documents to complete Line-of-Sight in PASTA. 3) IMD will establish a consistent model for managing protocol versioning and SOP's. Questions: 1) How are relationships between protocols, SOPs, and related documents handled/linked?	1) Workgroup formed to establish Line-of-Sight policy and requirements. 2) Networks maintain documents in PASTA (ongoing).
Capability to interact with unique (external) solutions	1) Capability to interact with unique/custom solutions linked to ability to support customizable querying and subsetting of data. 2) Opportunity to interact as a DataOne node is a key feature	1) Additional network level training to facilitate making connections to data. 2) Establish an ongoing work-group to determine which unique external solutions to pursue (these will evolve over time in # & functionality) and to what level integration/interaction should occur. 3) If joining DataOne community, form an IMD workgroup

Table C-5 (continued). Assumptions and activities for the LTER/PASTA Node.

Aspect	Assumptions	Activities
Continuous connection to network data	1) WAN at network is suitable for use of PASTA solution 2) Leveraging DataOne mirroring capabilities would not violate DOI policy	1) Examine network bandwidth program wide and compare to solution needs 2) IMD will invest time and effort into improving Physical Network Connectivity
Degree/ease with which data is made available to external systems (e.g. ability to push data to external systems)	1) Degree/ease with which data is made available to external systems is linked to ability to support customizable querying and subsetting of data. 2) Metadata can be shared with anyone (internal and external). 3) Tabular data can only be shared externally if it has flags identifying nonsharable data. 4) DataONE and Data.gov are not connected 5) Data sharing with DataONE is inherent to the PASTA system	1) Identify and develop support for appropriate exchange formats to facilitate tabular data pushes 2) Networks ensure that nonshareable data are flagged
Degree of efficiency for providing IMD operational information (who, what, when, & where of the program)	1) IMD will develop policy on elements of dataset publication that need to be tracked 2) 2) EML is capable of accommodating operational reporting policy	1) Use or adjust EML to accommodate operational reporting policy 2) Enhance interaction between Data Store and Protocol Tracker
Compliance facilitation (e.g. record management)	1) IMD develops and issues policies and guidance on compliance standards. 2) PASTA meets all compliance standards for records management. 3) Networks are not responsible for meeting compliance (they are responsible for meeting IMD reporting requirements which should automatically be harvested and used to meet compliance).	1) IMD will develop compliance standards 2) Obtain NARA-certification for PASTA content or develop data exchange system to a NARA-approved records management system
QA/QC performance reporting	1) EML will support IMD QA flags for auditing 2) PASTA complies with IMD QA requirements 3) QA for row-level data occurs before uploading 4) Automatic comparison of data to "standards from documentation" is possible but unrealistic	1) Create a clear policy on which QA metrics are auditable

Table C-5 (continued). Assumptions and activities for the LTER/PASTA Node.

Aspect	Assumptions	Activities
Versioning (dataset-level) capabilities	1) This versioning deals with the entire dataset 2) IMD will develop a consistent policy for all networks 3) Versioning appears to be user-enforced	1) Create a subteam to establish how versioning will be defined and work (including enforcement) in IMD
Change Tracking (record-level modifications to data) capabilities	1) Change-tracking deals with the record-level modifications to data. 2) Change-tracking only required for changes made to data with a Data Processing Level (DPL) of "Provisional" or above. 3) IMD will develop a consistent policy for all networks 4) Change-tracking applies to both cumulative and non-overlapping incremental datasets. 5) Change-tracking occurs before uploading to PASTA (i.e. Network level). 6) PASTA adequately supports change-tracking requirements	1) Create a subteam to establish how change tracking will be defined and work in IMD

EQulS

Table C-6. Description, policy, process, technology, and training for EQulS.

Aspect	Description
Brief Description	Environmental Quality Information System (EQulS) is commercial software from Earthsoft with separate data tables enforced via Electronic Data Deliverable (EDD) definitions for every protocol (i.e. Survey). Data file submissions are checked for formatting and valid values before being accepted into EQulS. Automatic notifications and tools help users identify and correct submitted data not conforming to EDD parameters. EQulS can process data files submitted in multiple file formats (e.g. txt, csv, xlsx). The Central Office would help networks configure Surveys in EQulS, data tables and associated EDD's, and help load data for existing protocols. Networks have the flexibility to use existing and/or adjusted tools and workflows to collect field data, but networks must promptly transform and submit/load field data to EQulS. Networks would use NPS Data Store to catalog the authoritative data sets stored in EQulS, and Data Store would hold and manage relationships between associated non-tabular files, protocol documents, and products (e.g. published reports).
Policy	• EQulS is authoritative data source • must decide on a metadata system • define metadata standard

Table C-6 (continued). Description, policy, process, technology, and training for EQuIS.

Aspect	Description
Policy (continued)	- determine which modules to include - decision on how to pass data to WRD EQuIS - finalize QA/QC policy - require posting data in certain timeframe
Process	- central office works with networks to get schemas loaded to EQuIS as surveys - central office works with networks to develop EDDs for loading data to EQuIS - need way to "unmunge" data from EQuIS for analysis - need to define roles - protected records need to be properly managed - assumes it includes WRD EQuIS and Aquarius.
Technology	- get IMD EQuIS installed and running on data center servers - develop user interface tools for working with data - Data Store as metadata catalog - provide access to input/output data by non-NPS entities (cooperator, partners, public) as appropriate
Training	- EQuIS training

Table C-7. Assumptions and activities for EQuIS.

Aspect	Assumptions	Activities
Metadata Format Consistency	1) Metadata standard will be required. 2) IMD will develop policy on metadata formats with science input (FGDC or EML) 3) IMD will develop minimum set of required metadata fields 4) Solution will include non-tabular data formats 5) IMD's metadata standard for tabular data can probably be accommodated in EQuIS Survey Designs.	1) Choose/develop IMD-wide metadata standard(s) to be included with all surveys/data-tables defined in EQuIS and applied to all non-tabular files stored in Data Store. 2) Central office modifies or develops tools to ease metadata generation & maintenance 3) Data Store updated to accommodate properties of metadata standard(s) for non-tabular data for both upload and download.

Table C-7 (continued). Assumptions and activities for EQUIS.

Aspect	Assumptions	Activities
Metadata Format Consistency (continued)	6) Data download or uploaded from EQUIS would include links to Data Store items for associated information, e.g. protocol and SOP docs. 7) Metadata elements required for data discovery (object-level metadata) could be harvested or exported from EQUIS and loaded to Data Store.	4) Develop a group to define metadata format (one-time) 5) Maintaining and enforcing (ongoing) 6) Develop EQUIS-Data Store link for metadata interchange 7) Update metadata records on existing Data Store records to match chosen standard
Metadata / documentation auditing	1) Metadata for tabular data are predefined in EQUIS, so auditing happens during data loading 2) Protocol-level metadata for nontabular data is achieved by including more required fields in Data Store; automated auditing becomes a possibility because metadata are entered into Data Store 3) Line-of-Sight supported/easily discovered (Protocol, SOPs, QAPs, etc.) with a link between Data Store and EQUIS	1) IMD creates subteam to govern and establish how Surveys will be established and managed. 2) Central Office, with network interaction creates 'Surveys' in EQUIS to define data table parameters, including required metadata, for every protocol. 3) Enhance Data Store to improve visibility and discoverability of documents associated with datasets. 4) Networks will be required to manage and audit all metadata documentation (ongoing).
Cumulative (all years) data & consistent structure	1) IMD will develop a policy on providing tabular and non-tabular data in a consistent structure. 2) Raw data are submitted to pre-defined tables in EQUIS 3) Data will be stored in a cumulative format except in rare cases where incremental, non-overlapping datasets are necessary. 4) Data Store will manage relationships to associated non-tabular items	1) Central Office will develop a policy to ensure authoritative tabular dataset are stored in a structure consistent across time. 2) Data Store enhanced to handle nontabular relationships.
Data quality flagging	1) Data flags are pre-defined in EQUIS table schemas. 2) Networks incorporate Data flagging values prior to uploading datasets. 3) Specific quality assurance guidelines are developed at the protocol level. 4) For Networks to use the EQUIS flagging, a single IMD wide flagging process would need to be established. 5) Networks will use the most current IMD Certification definitions	1) Central Office finalizes, with network input, guidelines for row/field level flagging. 2) Networks will be required to implement and audit flagging of row-level data (ongoing). 3) Central office finalizes Certification/Data Processing Level Guidelines (in particular ensure alignment with the draft QAP guidance)

Table C-7 (continued). Assumptions and activities for EQuIS.

Aspect	Assumptions	Activities
Customizable querying/subsetting of data	1) EQuIS provides extensive query, subsetting, and visualization capabilities 2) EQuIS Services could be modified for IMD specific needs	1) Create custom output using built-in EQuIS tools 2) Create custom tools if needed (e.g. visualization)
Connectedness to SOPs/protocols	1) Central office establishes policy to define minimum requirements for Line-of-Sight. 2) Networks will maintain necessary documents to complete Line-of-Sight in Data Store. 3) IMD will establish a consistent model for managing protocol versioning and SOP's. 4) Data in EQuIS will be linked in some way with associated items in Data Store, and vice versa	1) Workgroup formed to establish Line-of-Sight policy and requirements. 2) Data Store would be augmented to improve visibility between data products and documentation. 3) Networks maintain documents in Data Store (ongoing). 4) Data Store will be updated to manage versioning of Protocols and SOPs. 5) Develop a link between data sets/results in EQuIS and associated items in Data Store.
Capability to interact with unique (external) solutions	1) EQuIS includes built-in interfaces to several 3rd Party software applications including ESRI ArcGIS 2) From EQuIS Documentation: "Many organizations are using EQuIS as the backbone of a corporate or enterprise data exchange platform. EQuIS is an open system with available source code. You can customize and extend the kernel schema. You can modify the schema. You can share data with consultants that have EQuIS, or consultants that use other systems. You can export and import data from any other electronic source or other data systems."	1) Central Office and/or networks identify and develop data exchange solutions within EQuIS 2) Establish an ongoing work-group to determine which unique external solutions to pursue (these will evolve over time in # & functionality) and to what level integration/interaction should occur.
Continuous connection to network data	1) This applies to Physical Network Connectivity and is integral to the Technical solution and implementation 2) Ability to have local mirroring is currently restricted by DOI policy	1) Examine network bandwidth program wide and compare to solution needs 2) IMD will to invest time and effort into improving Physical Network Connectivity 3) Increase failover capacity at Data Center level
Degree/ease with which data is made available to external systems (e.g. ability to push data to external systems)	1) EQuIS reportedly supports open data exchange to meet client requirements 2) Metadata can be shared with anyone (internal and external).	1) Identify and develop support for appropriate exchange formats to facilitate tabular data pushes 2) Networks ensure that nonshareable data are flagged 3) Develop Data Store to harvest metadata from EQuIS

Table C-7 (continued). Assumptions and activities for EQuIS.

Aspect	Assumptions	Activities
Degree/ease with which data is made available to external systems (e.g. ability to push data to external systems) (continued)	3) Survey design (EDD's) in EQuIS flag sharable and nonsharable data. 4) Data Store will be enhanced to harvest complete metadata (see Metadata performance metrics above)	–
Degree of efficiency for providing IMD operational information (who, what, when, & where of the program)	***Same as CIA*** 1) IMD will develop policy on elements of dataset publication that need to be tracked 2) Enhancements have been made to Data Store (see Metadata performance metrics above) to harvest complete metadata 3) These reports can be extracted directly from protocol databases or from metadata in Data Store	1) Include operational information in EQuIS Survey designs according to IMD policy 2) Enhance Data Store to include elements defined in IMD protocol tracking policy. 3) Enhance interaction between Data Store and Protocol Tracking and EQuIS 4) IMD Wide Governance team established to manage operational norms for the use of EQuIS
Compliance facilitation (e.g. record management)	1) IMD develops and issues policies and guidance on compliance standards. 2) Data Store is enhanced to meet all compliance standards for records management. 3) Authoritative data sets in EQuIS meet records management mandates, or the data are harvested/pushed to a compliant records management system 4) Networks are not responsible for meeting compliance (they are responsible for meeting IMD reporting requirements which should automatically be harvested and used to meet compliance).	1) IMD will develop compliance standards 2) Enhance Data Store to support compliance requirements 3) Obtain NARA-certification for EQuIS content or develop data exchange system to a NARA-approved records management system

Table C-7 (continued). Assumptions and activities for EQulS.

Aspect	Assumptions	Activities
QA/QC performance reporting	1) Whatever metadata format is chosen must support QA flags for auditing 2) EQulS will be enhanced to comply with IMD QAQC requirements. 3) Networks may need to separately report additional/other QAQC metrics (e.g. nontabular) 4) For Networks to use the EQulS flagging, a single IMD wide flagging process would need to be established. 5) Automatic comparison of data to "standards from documentation" is possible but unrealistic	1) Create a clear policy on which QA metrics are auditable 2) Include required QAQC fields and metrics in EQulS Survey Designs and audit reports 3) Networks report other required QAQC elements, e.g. QA built in to data collection and processing procedures
Versioning (dataset-level) capabilities	***Modified from Data Store*** 1) This versioning deals with the entire dataset 2) IMD will develop a consistent policy for all networks 3) Versioning applies to both cumulative and non-overlapping incremental datasets. 4) Whatever metadata format is chosen must support versioning. 5) EQulS Survey Designs will track dataset level versioning 6) Versioning of cumulative datasets has been accounted for	***Modified from Data Store*** 1) Create a subteam to establish how versioning will be defined and work (including enforcement) in IMD 2) Design EQulS Surveys to include dataset level versioning
Change Tracking (record-level modifications to data) capabilities	1) Change-tracking deals with the record-level modifications to data. 2) Change-tracking only required for changes made to data with a Data Processing Level (DPL) of "Provisional" or above. 3) IMD will develop a consistent policy for all networks 4) Change-tracking applies to both cumulative and non-overlapping incremental datasets. 5) The audit features in EQulS will be enabled in order to track changes to EDDs based on appropriate user permissions 6) All networks will perform record level auditing in EQulS	1) Create a subteam to establish how change tracking will be defined and work in IMD 2) Networks will audit data in EQulS as needed

Centralized Interface and Administration

Table C-8. Description, policy, process, technology, and training for centralized interface and administration.

Aspect	Description
Brief Description	Network data associated with monitoring protocols are stored in data center databases that are collaboratively designed/maintained with Central Office staff. Field databases/apps may be used, but networks must connect to centralized databases to load data (i.e. network located data entry databases would be for temporary storage only). Networks would use NPS Data Store to catalog the centralized databases, and Data Store would hold and manage relationships between associated non-tabular files, protocol documents, and products (e.g. published reports).
Policy	• no data entry databases in offices • all data must be entered into centralized SQL Server databases (one for each protocol) • define metadata standard • metadata must be loaded to catalog • metadata must be updated in conjunction with modifications to database • finalize QA/QC policy • require posting data in certain timeframe • assume this option is getting ahead of IT restrictions about "applications" and works to centralize databases in data centers.
Process	• limitations on where data may be entered by permissions for appropriate staff/network • network submits database model/database or meets with central office to design database in data center SQL Server • central office works with network to design user interface for database • changes to schema and batch edit scripts get queued into help system • allow networks to connect to databases to read data • networks must work with central office to design databases • networks can build reports on databases using SSRS • web services can be created on databases • central office works on converting legacy databases to SQL Server • could potentially do some compliance checking • central office works with networks to design QC checks into databases

Table C-8 (continued). Description, policy, process, technology, and training for centralized interface and administration.

Aspect	Description
Technology	• SQL Server at data centers • web-based interfaces for user interface at networks • SSRS for reporting • performance tweaks to ensure adequate speed at networks • data upload and download via mobile devices (that can connect to centralized SQL Server) (aka support digital data collection) • Data Store catalogs the SQL Server datasets • stand up web services on SQL Server databases
Training	• SSRS training for getting data back out • consuming web services • networks train users on entering data into web-based UIs • how to post data from mobile devices

Table C-9. Assumptions and activities for centralized interface and administration.

Aspect	Assumptions	Activities
Metadata Format Consistency	***Same as Data Store*** 1) Metadata standard will be required. 2) IMD will develop policy on metadata formats with science input (FGDC or EML) 3) IMD will develop minimum set of required metadata fields 4) Solution will include non-tabular data formats 5) Metadata are managed in Data Store for posted items; management of metadata for items not posted to Data Store is ad hoc 6) Metadata (including linked protocols) are available at time of download. 7) Metadata manually updated in conjunction with modifications to database.	***Modified from Data Store*** 1) Choose/develop IMD-wide metadata standard(s) to be applied to all tabular and non-tabular (ex. GIS, photos) data. 2) Central office modifies or develops tools to ease metadata generation & maintenance 3) Data Store updated to accommodate properties of metadata standard(s) for tabular and non-tabular data for both upload and download. 4) Develop a group to define metadata format (one-time) 5) Maintaining and enforcing (ongoing) 6) Develop local protocol database/Data Store link for metadata interchange 7) Update metadata records on existing Data Store records to match chosen standard

Table C-9 (continued). Assumptions and activities for centralized interface and administration.

Aspect	Assumptions	Activities
Metadata Format Consistency (continued)	***Same as Data Store (continued)*** 8) Data Store is essentially the "one-at-a-time" (i.e. not batch) metadata creation tool by filling in the fields for a citation 9) Morpho or other existing tools could be used to generate EML metadata locally; entering records into Data Store is another way to generate metadata 10) Technology is not the limiting factor in metadata handling for Data Store.	–
Metadata / documentation auditing	***Modified from Data Store*** 1) Metadata auditing will be implemented over time. 2) Enforcement is achieved for protocol-level metadata by including more required fields in Data Store; automated auditing becomes a possibility because metadata are entered into Data Store 3) Line-of-Sight supported/easily discovered (Protocol, SOPs, QAPs, etc.)	***Modified from Data Store*** 1) Enhancement of Data Store to support meta data schema check against data provided. 2) Development of data auditing tool. 3) Network to provide additional Metadata to describe Schema. 4) Build tool to read record-level metadata directly from protocol database 5) Enhancement to improve visibility and discoverability of Line-of-Sight documents associated with datasets. 6) Networks will be required to manage and audit all metadata documentation (ongoing).
Cumulative (all years) data & consistent structure	***Modified from Data Store*** 1) IMD will develop a policy on providing tabular and non-tabular data in a consistent structure. 2) Raw data will be provided using one of two methods a) Entered directly into a CIA database through a UI b) Uploaded into a CIA database 3) Data will be stored in a cumulative format except in rare cases where discrete, non-incremental datasets are necessary.	***Modified from Data Store*** 1) Central Office will develop a policy to ensure authoritative tabular dataset are stored in a structure consistent across time. 2) Data Store enhanced to handle nontabular relationships.

Table C-9 (continued). Assumptions and activities for centralized interface and administration.

Aspect	Assumptions	Activities
Cumulative (all years) data & consistent structure (continued)	***Modified from Data Store (continued)*** 4) Data Store will manage relationships to associated non-tabular items 5) Non-protocol databases could be converted as time/resources permit. They would live in Data Store until that point.	–
Data quality flagging	***Modified from Data Store*** 1) It is expected that each database would support row-level flagging 2) There is one single standard (from IMD) for row-level flagging values 3) Specific quality assurance guidelines are developed at the protocol level. 4) Networks will use the most current IMD Certification definitions	***Modified from Data Store*** 1) Central Office finalizes, with network input, guidelines for row/field level flagging. 2) Networks will be required to implement and audit flagging of row-level data (ongoing). 3) Central office finalizes Certification/Data Processing Level Guidelines (in particular ensure alignment with the draft QAP guidance)
Customizable querying/subsetting of data	***Modified from Data Store*** 1) Coarse queries based on values stored in metadata. 2) Dynamic subsetting and/or visualization is available for all datasets.	***Modified from Data Store*** 1) All IMD databases would need to be converted with Central Office assistance 2) Data owners would need to provide usage metadata to assist in using data from queryable source. 3) Central office and networks will develop visualization tools
Connectedness to SOPs/protocols	***Same as Data Store *** 1) Central office establishes policy to define minimum requirements for Line-of-Sight. 2) Networks will maintain necessary documents to complete Line-of-Sight in Data Store. 3) IMD will establish a consistent model for managing protocol versioning and SOP's.	***Same as Data Store *** 1) Workgroup formed to establish Line-of-Sight policy and requirements. 2) Data Store would be augmented to improve visibility between data products and documentation. 3) Networks maintain documents in Data Store (ongoing). 4) Data Store will be updated to manage versioning of Protocols and SOPs.

Table C-9 (continued). Assumptions and activities for centralized interface and administration.

Aspect	Assumptions	Activities
Capability to interact with unique (external) solutions	***Same as Data Store *** 1) Capability to interact with unique/custom solutions linked to ability to support customizable querying and subsetting of data.	***Same as Data Store *** 1) Additional network level training to facilitate making connections to data. 2) Establish an ongoing work-group to determine which unique external solutions to pursue (these will evolve over time in # & functionality) and to what level integration/interaction should occur.
Continuous connection to network data	***Modified from Data Store*** 1) WAN at network is suitable for use of CIA 2) Ability to have local mirroring is currently restricted by DOI policy	***Modified from Data Store*** 1) Examine network bandwidth program wide and compare to solution needs 2) IMD will be required to invest time and effort into improving Physical Network Connectivity 3) Increase failover capacity at Data Center level
Degree/ease with which data is made available to external systems (e.g. ability to push data to external systems)	***Modified from Data Store*** 1) Degree/ease with which data is made available to external systems is linked to ability to support customizable querying and subsetting of data. 2) Metadata can be shared with anyone (internal and external). 3) Tabular data can only be shared externally if it has flags identifying nonsharable data. 4) Enhancements have been made to Data Store (see Metadata performance metrics above) to harvest complete metadata.	***Same as Data Store *** 1) Identify and develop support for appropriate exchange formats to facilitate tabular data pushes 2) Networks ensure that nonshareable data are flagged
Degree of efficiency for providing IMD operational information (who, what, when, & where of the program)	***Modified from Data Store*** 1) IMD will develop policy on elements of dataset publication that need to be tracked 2) Enhancements have been made to Data Store (see Metadata performance metrics above) to harvest complete metadata 3) These reports can be extracted directly from protocol databases or from metadata in Data Store	***Modified from Data Store*** 1) Enhance Data Store to include elements defined in IMD protocol tracking policy. 2) Enhance interaction between Data Store, Protocol Tracking, and Protocol Related Databases

Table C-9 (continued). Assumptions and activities for centralized interface and administration.

Aspect	Assumptions	Activities
Compliance facilitation (e.g. record management)	***Same as Data Store *** 1) IMD develops and issues policies and guidance on compliance standards. 2) Data Store is enhanced to meet all compliance standards for records management. 3) Networks are not responsible for meeting compliance (they are responsible for meeting IMD reporting requirements which should automatically be harvested and used to meet compliance).	***Same as Data Store *** 1) IMD will develop compliance standards 2) Enhance Data Store to support compliance requirements
QA/QC performance reporting	***Modified from Data Store*** 1) Whatever metadata format is chosen must support QA flags for auditing 2) Data Store will be enhanced to comply with IMD QAQC requirements. 3) That each individual collection database manages QA flags appropriately 4) Automatic comparison of data to "standards from documentation" is possible but unrealistic	***Same as Data Store *** 1) Create a clear policy on which QA metrics are auditable
Versioning (dataset-level) capabilities	***Modified from Data Store*** 1) This versioning deals with the entire dataset 2) IMD will develop a consistent policy for all networks 3) Versioning applies to both cumulative and non-overlapping incremental datasets. 4) Whatever metadata format is chosen must support versioning. 5) There is an enhanced connection between queryable dataset and related protocol documents in Data Store 6) Versioning of cumulative datasets has been accounted for	***Modified from Data Store*** 1) Create a subteam to establish how versioning will be defined and work (including enforcement) in IMD 2) Data Store would be enhanced to connect queryable dataset and related protocol documents

Table C-9 (continued). Assumptions and activities for centralized interface and administration.

Aspect	Assumptions	Activities
Change Tracking (record-level modifications to data) capabilities	***Modified from Data Store*** 1) Change-tracking deals with the record-level modifications to data. 2) Change-tracking only required for changes made to data with a Data Processing Level (DPL) of "Provisional" or above. 3) IMD will develop a consistent policy for all networks 4) Change-tracking applies to both cumulative and non-overlapping incremental datasets. 5) Change-tracking occurs within each database	***Modified from Data Store*** 1) Create a subteam to establish how change tracking will be defined and work in IMD 2) Add change tracking to each protocol database

Status Quo

Table C-10. Description, policy, process, technology, and training to maintain status quo.

Aspect	Description
Brief Description	Don't change anything - every network does what they want and central office doesn't issue much guidance, but just provides assistance when needed.
Policy	• no central QA/QC policy (draft state) • no metadata standard • no policy to post to Data Store • peer review of protocols? • publish protocols through NRR series
Process	• minimum implementation standards • db standards • NRDT • certification guidelines

Table C-10 (continued). Description, policy, process, technology, and training to maintain status quo.

Aspect	Description
Technology	• Collection tools ○ Access ○ Filemaker ○ SQL Express ○ paper datasheets ○ ... • Data entry tools ○ Access ○ Lightswitch • Data storage tools ○ network share/server ○ Access ○ Data Store ○ SQL Server (@network) ○ SQL Server (@data center) ○ Aquarius • Reporting tools ○ SSRS ○ Access reports ○ documents (Word/PDFs) ○ R Shiny ○ WRD EQuIS • Delivery tools ○ Website ○ SharePoint ○ Data Store ○ share/server
Training	–

Minimally Compliant

Table C-11. Description, policy, process, technology, and training to be minimally compliant.

Aspect	Description
Brief Description	Satisfy compliance (meets legal regulations/directives from Fed. Gov., DOI, NPS, NRSS, IMD), but with minimal impact (change) to network operations
Policy	• choose metadata standard(s) • finalize QA/QC policy • define authoritative data and make Data Store official repository for authoritative data • require posting files to Data Store in certain timeframe
Process	• Protected records • Guidance on redacting data from datasets • Network provides redacted + full dataset, each with appropriate metadata • Options 1) Redacted data and full data on same DS citation with different access levels (mod to DS) 2) Separate citations for redacted data vs. full data (policy, no mod. to DS) • Start to upload data and metadata annually with DS mod • Versioning ○ Year 1: Just post it ○ Year 2: Narrative of changes in text file goes with data ○ Year 3: Details of changes in text file goes with data ○ Year 4-5: Change tracking in database table • Visualization on basic protocol metadata rather than analysis
Technology	• Data Store ○ Implement DOIs ○ Protected Records ▪ Modify DS “Sensitivity” to reflect Protected Records terminology ▪ Allow different Protected Records status for redacted vs. full datasets ○ Attribution of derivative or ancillary data • Line of sight/protocol package • Better tools for association of related products (protocols, SOPs, datasets, etc.) • finish allowing links to SQL databases and web services

Table C-11 (continued). Description, policy, process, technology, and training to be minimally compliant.

Aspect	Description
Training	need documentation for allowing links to SQL databases and web services

Park Focused

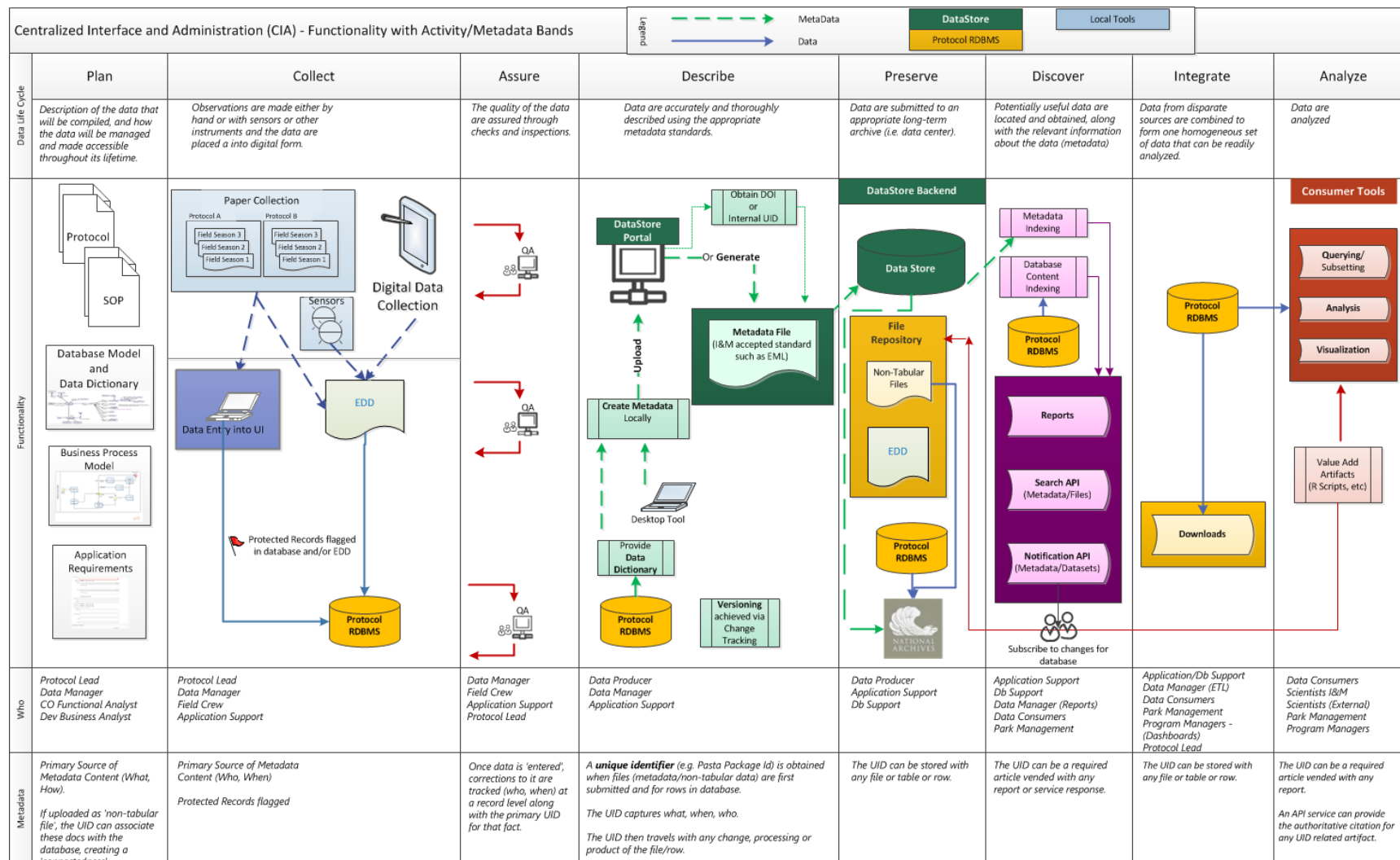
Table C-12. Description, policy, process, technology, and training for park focused activities.

Aspect	Description
Brief Description	- Meets minimally compliant requirements, while maximizing delivery to parks and de-emphasizing WASO overhead. - Implementation may include lightweight enhancements to the existing Data Store application.
Policy	- each network develops policy/agreement to provide info. to parks using standards that best satisfy each park - WASO defines what they need for authoritative data lineage - authoritative data lineage designated (source over time) by each network - metadata per park need - QA/QC per park need as determined by network - data are provided to park per agreement with park steering committee/board
Process	- Protected records - just need to ensure delivery of data/info. to appropriate park personnel; label products as containing sensitive information - Data Delivery - minimal would be post to Data Store - Data visualization (live) - process data for visualization tool; aggregate/filter/mask/analyze; include derived products - WASO migrates files to archival/newer format
Technology	- Data visualization - tool that consumes data and provides interactive visualization snapshot - Data Store for storage, including copies of data optimized for vis. tool - non-tabular data managed locally by network, but could ask WASO to build tools to facilitate
Training	- Bulk of training done by networks for their own staff - Data Store training offered by WASO

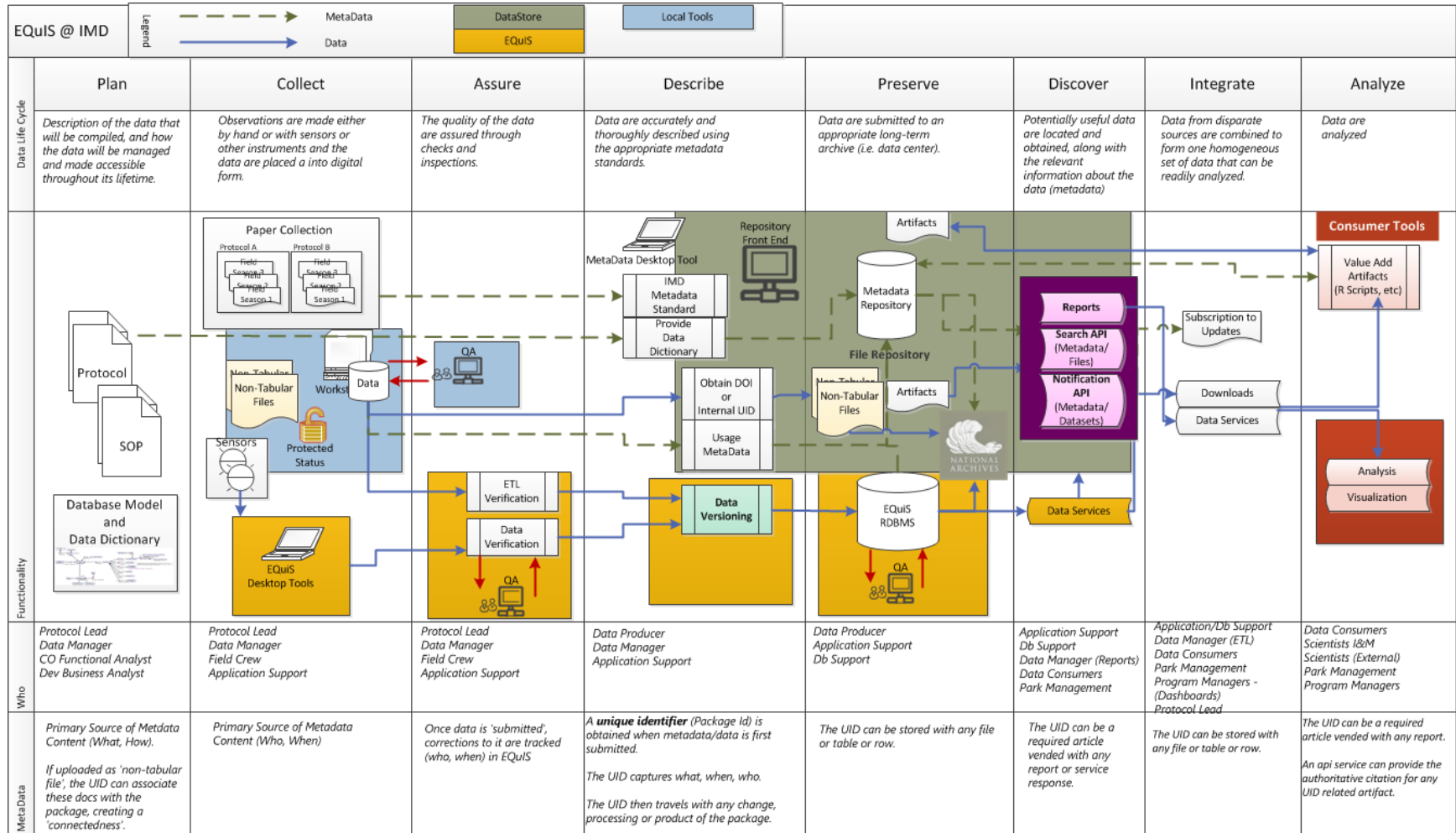
Appendix D. Workflow Diagrams for Alternatives

Centralized Interface & Administration

D-1

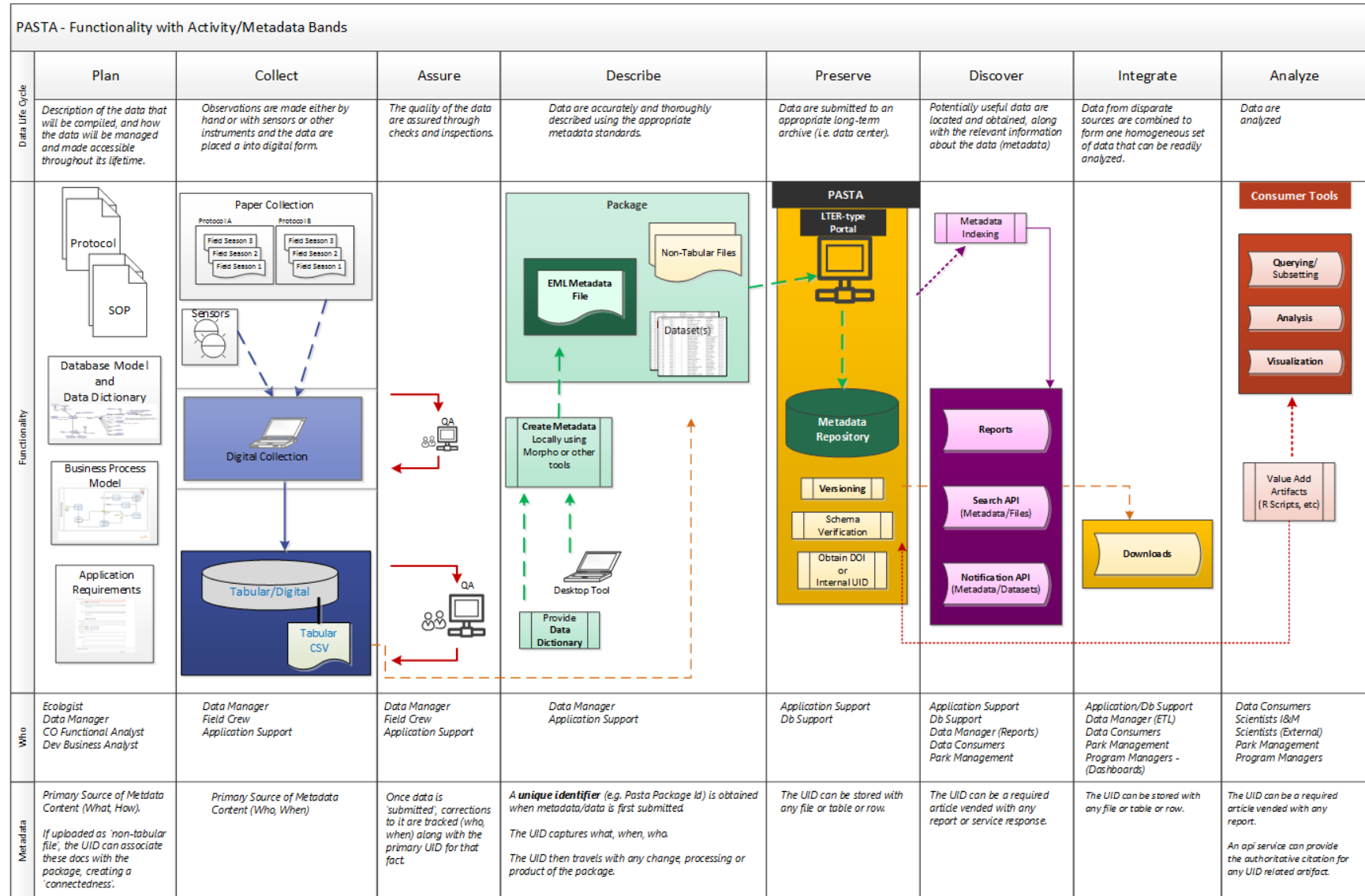


EQulS



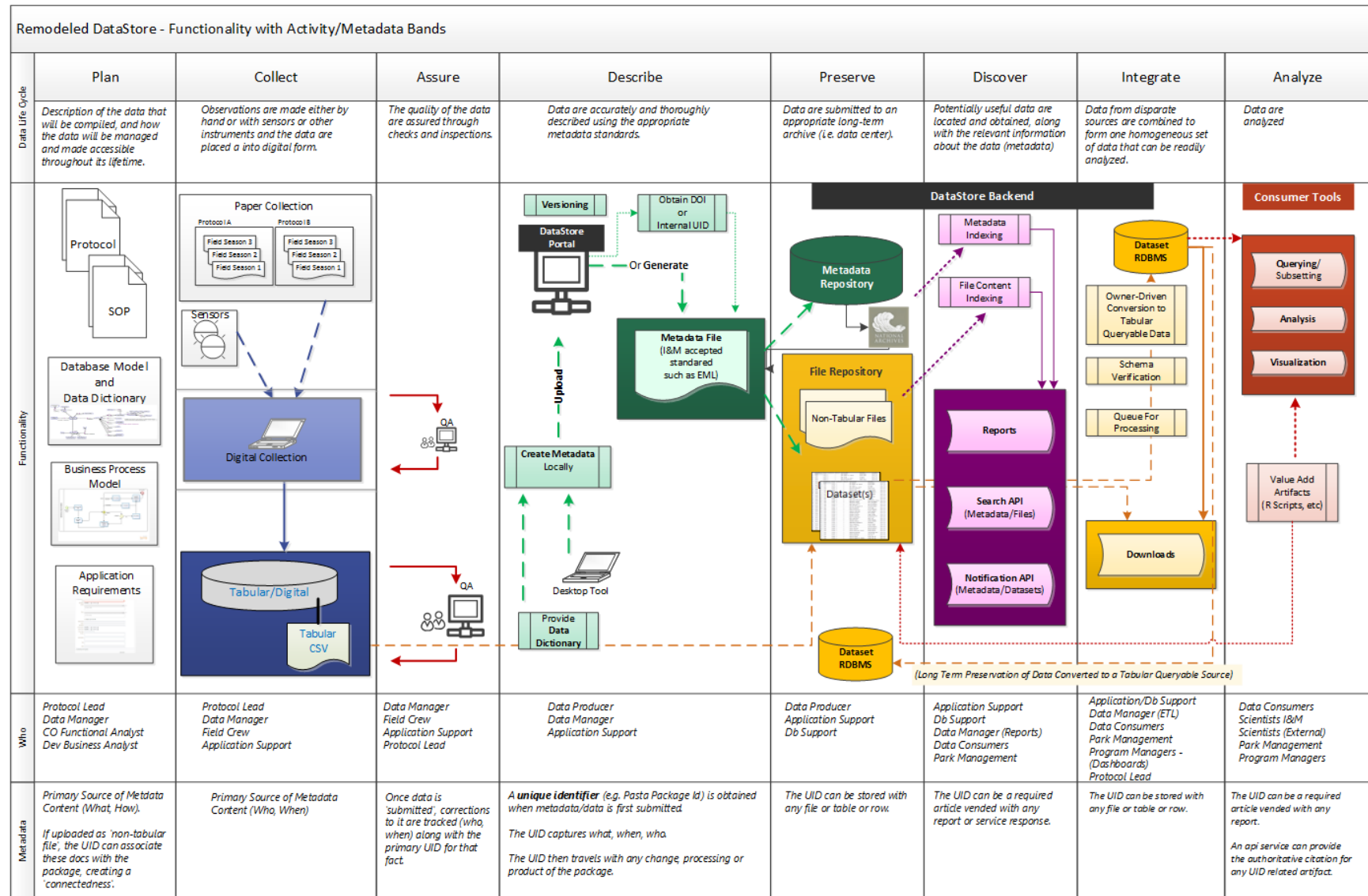
LTER/PASTA

D-3



Remodeled Data Store

D-4



Appendix E. Long-Term Data Management Task Force Membership

Task Force Lead: Marianne Tucker – IMD Central Office, Deputy Division Chief

Participants:

- 1) Paula Capece – Structured Decision Making Advisor
- 2) Nicole Athern – Structured Decision Making Advisor
- 3) Judd Patterson – SER
- 4) John Boetsch - PWR
- 5) Scott Miller- AKN
- 6) Adam Kozlowski - NER
- 7) Angela Jarding - MWR
- 8) Rob Daley - IMR
- 9) Bonnie Campbell – IMR DM, Biologist
- 10) Geoffrey Sanders - NCRN NPM, RPM, DM
- 11) Kristin Legg - NPM and IMR Ecologist
- 12) Isabel Ashton - MWR Veg Ecologist
- 13) Tom Philippi – Quantitative Ecologist
- 14) Lisa Nelson – Central Office
- 15) Brent Frakes – Central Office/FWS
- 16) Simon Kingston – Central Office
- 17) Peter Budde –NRSS Data/Systems Officer and Meeting Facilitator
- 18) Dan Kocol – MBS Program Manager
- 19) Margaret Beer – Communication Specialist
- 20) Mike DiSibio – MBS Lead Architect
- 21) Bill Johnson* - acting National Data Manager
- 22) Mark Hart* - acting National Data Manager

The Department of the Interior protects and manages the nation's natural resources and cultural heritage; provides scientific and other information about those resources; and honors its special responsibilities to American Indians, Alaska Natives, and affiliated Island Communities.

NPS 909/140574, November 2017

National Park Service
U.S. Department of the Interior



Natural Resource Stewardship and Science

Natural Resource Stewardship and Science

1201 Oakridge Drive, Suite 150
Fort Collins, CO 80525

www.nature.nps.gov