

**Recommendations  
for  
National Marine Fisheries Service (NMFS)  
Enterprise Data Management (EDM)**

Version 1.7  
November 2008

James W. Sargent, Information Architect,  
Dr. John Boreman, Director, Office of Science and Technology  
and  
The Fisheries Information Management Committee  
National Marine Fisheries Service (NMFS)  
National Oceanic and Atmospheric Administration (NOAA)

Item	Description
<b>Document Title</b>	Recommendations For NMFS Enterprise Data Management
<b>File Name</b> (auto)	S:\information Architecture\Recommendations\ Recommendations v.1.7.doc
<b>Disposition</b> (Draft, Preliminary, or Final)	Final
<b>Author(s)</b>	Jim Sargent, Dr. John Boreman, and the FIMC
<b>Document Description</b>	This document describes the proposed recommendations for NMFS EDM

Version	Date	Changed By	Description
0.1- 0.4	6/20/2008	Jim Sargent	Initial draft
0.5	7/24	Jim Sargent	Outline revised in response to FIMC comments
0.6	8/25/2008	Tom Fazio	Completed drafts of section1 to 4
0.7	8/27/2008	Tom Fazio	Edits and revisions to draft
0.8	9/6/2008	Jim Sargent	Four Pillars adaptation
0.9	9/17/2008	Jim Sargent	Edits and revisions to draft
0.91	9/28/2008	Jim Sargent	Add pix etc.
.92	9/29/2008	Jim Sargent	Spelling grammar etc.
1.0	9/30/2008	Jim Sargent	Final changes and edits
1.2	10/2/2008	Jim Sargent	Accepted Dr. Boreman's markup; added acronym appx.
1.3	10/7/08	Jim Sargent	Accept Dr. Boreman's additional markups
1.4	10/9/2008	Jim Sargent	Corrected punctuation
1.5	10/14/2008	Jim Sargent	Minor edits and added "Continuing with <i>Status Quo</i> "
1.6	11/2/08	Fran Pfielger	Grammatical edits
1.7	11/7/2008	Jim Sargent	Final edits by FIMC members, et al

## Table of Contents

Executive Summary .....	3
1 Introduction and Background .....	6
1.1 Purpose and Audience.....	6
1.2 Background .....	6
1.3 The Challenge .....	6
1.4 The Response .....	7
1.5 Project Teams.....	7
1.6 References.....	7
2 Current State of NMFS Information Management .....	8
2.1 Interview Results .....	8
2.2 Issues and Concerns.....	9
2.3 Other Drivers and Considerations for NMFS EDM .....	13
2.3.1 Internal and External Data Management Programs and Initiatives .....	13
2.3.2 Data Management Committees and Organizations.....	15
2.3.3 Other Issues.....	16
3 Information Management Business Processes.....	18
3.1 Information Life-Cycle Management .....	18
3.2 Discovery and Access .....	19
3.3 Interfaces with External Systems.....	19
4 The Target State of NMFS EDM.....	20
4.1 Vision.....	20
4.2 Principles.....	20
4.3 Mission, Goals, and Objectives .....	21
4.4 The Four Components of NMFS EDM .....	21
4.4.1 Component One: Governance.....	22
4.4.2 Component Two: Trusted Data.....	27
4.4.3 Component Three: The Authoritative Data Catalog.....	29
4.4.4 Component Four: The NMFS Information Depot .....	31
5 Program Implementation Alternatives.....	37
5.1 Minimal Cost Option .....	38
5.2 Intermediate Cost Option.....	39
5.3 Full Cost Option.....	40
5.4 Recommended Alternative.....	40
5.5 Continuing with Status Quo.....	42
6 Other Recommendations.....	44
6.1 Paperwork Reduction Act (PRA) Business Process .....	44
6.2 Administrative Systems .....	44
7 Risks and Constraints.....	45
Appendix A – Fisheries Information Management Committee Roster .....	47
Appendix B – Data Management Issues Questionnaire .....	48
Appendix C – Objectives and Priorities.....	52
Appendix D – NMFS EDM Components Cross-Reference Tables.....	55
Appendix E – Glossary of Terms.....	56
Appendix F – Acronyms .....	58

## Executive Summary

In January 2008 the Director of the National Marine Fisheries Service (NMFS) Office of Science and Technology (ST) proposed, and the NMFS Science Board endorsed, a project to develop recommendations for improving NMFS enterprise data management (EDM). The Information Architect (IA), whose position was also established in January 2008 in ST, and the existing Fisheries Information Management Committee (FIMC), with representatives from each Regional Office (RO) and Science Center (SC), were charged with the task.

The first step was to determine whether in fact, problems existed and, if so, their nature. The FIMC identified 12 key issues and interviewed management to determine their priorities for addressing them. Issues rated as top priority to address were that NMFS:

- had critical gaps in its data;
- had no authoritative data inventory;
- lacked sufficient metadata for its data;
- experienced data quality and consistency challenges;
- could not effectively integrate its data to answer key questions of the day;
- had administrative systems that did not work well together; and
- had delays in delivery of data to decision makers.

Other issues included:

- data being lost;
- data not being archived for perpetuity;
- historical data in need of rescue;
- inadequate sharing of applications and technologies; and
- lack of buy-in across FMCs to address these issues.

Interestingly, the issue of not having buy-in across FMCs was ranked practically lowest, which seems to indicate that management feels NMFS does have the buy-in necessary to address these enterprise issues collaboratively. Another overall consideration for developing NMFS EDM identified by the FIMC was the need to share data with other NOAA programs and initiatives, such as the Integrated Ocean Observations System (IOOS) Data Management and Communications (DMAC).

The second step was for the FIMC to develop recommendations for the ST Director and ultimately for the Leadership Council for addressing the issues. To this end, they:

- held a workshop and many web meetings;
- worked with a variety of professional data management consultants;
- developed consensus on a vision, proposed principles, mission, goals, and objectives for NMFS EDM;
- developed specific recommendations for a NMFS EDM program;
- developed activities and resources required to carry out the recommendations;
- developed minimal cost, intermediate cost, and full cost options for consideration, and finally;

- recommended a specific practical alternative for moving forward.

The vision of NMFS EDM is that *“NMFS customers can confidently find, access, and use our data.”*

The third (current) step is to present minimal to full cost alternatives for implementing the recommendations. In October 2008, this paper was presented to the ST Director, who will present his findings and recommendations to the Leadership Council at their fall meeting in December 2008. The recommendations focus on establishing the following *Four Components of NMFS EDM*:

- **EDM Governance** to include:
  - A governance structure consisting of:
    - The IA;
    - Regional Information Management Coordinators (RIMCs) in each Regional Office and Science Center;
    - Office Information Management Coordinators (OIMCs) in key HQ Offices;
    - The Fisheries Information Management Advisory Committee (FIMAC), which is composed of the RIMCs and the OIMCs and chaired by the IA. The FIMAC provides advice and recommendations directly to the NMFS Leadership Council (LC).
  - Policies and procedures addressing data stewardship, metadata collection and maintenance, security, confidentiality, and data quality.
  - Policies and operational procedures for managing external requests for information and Headquarters’ requests for information from the FMCs (data calls).
- **An Authoritative Data Catalog** – An inventory of, and appropriate metadata for, all enterprise NMFS information assets, including databases, photograph collections, administrative records, and other digital, analog, and hard copy information assets.
- **Trusted Enterprise Data Assets** – High quality, reliable, and timely data managed by defined IM business processes and continually assessed for sufficiency and improvement.
- **The NMFS Information Depot** - A web portal that provides one-stop shopping for all NMFS enterprise data, based on the Amazon.com portal model, to enable NMFS’ internal and external constituents to effectively discover, access, integrate, and use NMFS information resources to answer key current and future questions that may or may not be related to the original purpose for collecting the data.

The critical foundation for the four components is ongoing management commitment to NMFS EDM.

The recommended alternative for moving forward is a combination of the minimal cost and intermediate cost options that establishes the basics of a sound EDM in FY 2009. It includes:

1. Establishing the Authoritative Data Catalog, and collecting and populating the Catalog with discovery-level metadata for NMFS’ largest and most sharable data assets. Discovery-level metadata includes: data asset title, description, organization code, name and address of data steward, and indicators of data quality and confidentiality.

2. Improving trust in NMFS data by initiating the development of data quality plans for the largest NMFS data assets.
3. Developing a lexicon and taxonomies of NMFS information and increasing corporate understanding of Service Oriented Architecture (SOA) and other emerging technologies as first steps for establishing the Information Depot.
4. Developing a project plan and performance metrics for these tasks and keeping the LC apprised of progress with quarterly reports.

The marginal cost for this alternative consists of increasing the percentage of the regional and HQ FIMC representatives' time dedicated to enterprise data management from 10% to 25%, and an additional \$500,000 for contract support for developing the policies, establishing the data catalog, collecting metadata, and training. Not included are hidden costs of FMC staff working to move the recommendations forward. FMCs will be reimbursed for costs due to the increase their FIMC representative's time from 10% to 25%, as additional funds are available.

Although this alternative will not move things forward as quickly as the FIMC desires, it will, along with management commitment, establish a firm foundation for a sound EDM program in NMFS.

# 1 Introduction and Background

This section identifies the purpose of and audience for this document, presents background information, and describes how this document will be maintained and distributed.

## 1.1 Purpose and Audience

This document specifies recommendations for NMFS Enterprise Data Management (EDM) developed by the Fisheries Information Management Committee (FIMC) to the Director of the Office of Science and Technology (ST), National Marine Fisheries Service (NMFS), for establishing a Program in ST. The intended audience of this document is the ST Director, the NMFS Science Board, and the NMFS Leadership Council, whose collective endorsement, support, and funding are required to establish a successful EDM.

## 1.2 Background

NMFS data and information is the foundation for all the agency's business processes employed to meet its mission of "stewardship of living marine resources through science-based conservation and management, and the protection and restoration of healthy ecosystems." Data and information (which, in this context, include all types of structured and unstructured documents, photographs, sonar images, sound recordings, etc.) are the principal grist for the NMFS mill producing information for regulatory and scientific communities.

We frequently hear concerns about NMFS data and information from practically every sector of our constituent universe: high-level external groups such as the Pew Commission and the National Research Council; managers, scientists, and other users of our data; and programmatic and technical data managers, database administrators, and software developers. The concerns run the gamut—from not having the right data at the right place at the right time, to not knowing what data we have or where it is, to lack of metadata sufficient to integrate and display our data and information to answer pressing questions. Many of these concerns are based on anecdotes of data atrocities (such as the proverbial harpooned scallop discovered in a decades-old research vessel survey, upon which a critical stock assessment was based that led to regulations imposed on the scallop fleet.

Is NMFS' fundamental mission of stewardship of living marine resources at risk? Just how reliable and pervasive is the evidence of NMFS data and information problems? Before making recommendations, the first task was to substantiate that one or more problems exist and determine the nature of the problems. Only after that could we develop a program to address them.

## 1.3 The Challenge

The challenge presented by the ST Director was to determine whether NMFS actually had information management challenges and, if so, to develop recommendations to address them. The ST Director and the NMFS Chief Information Officer jointly established the position of Information Architect in ST to work with the Fisheries Information Management Committee (FIMC). The FIMC was originally chartered by the NMFS Science Board and was made up of

representatives in each of NMFS Science Centers; however, the scope of information to be addressed included all NMFS scientific, regulatory, and administrative data and information.

## **1.4 The Response**

To meet the challenge, the FIMC was expanded to include representatives from all Regional Offices, Science Centers, and key HQ Offices. The FIMC first established baseline understanding of the state of information management at NMFS. This was accomplished through initial 15-minute interviews with top NMFS management to frame the high-level concerns. The FIMC then developed a list of 12 key IM issues and held structured interviews with their managers to determine their priorities for addressing each of the issues. The results of these interviews were compiled and are discussed in Section 2 of this document.

With the baseline issues established, it was clear that a comprehensive approach to NMFS EDM was needed. A workshop was held on July 8–10, 2008, to define a framework and recommendations for a NMFS EDM Program. The results of that workshop are presented in Sections 3 through 6 of this document.

## **1.5 Project Teams**

The following teams collaborated to develop the NMFS EDM recommendations:

- Information Architect (IA) – This position was established within ST to coordinate and lead activities related to NMFS information management. The IA served as the Project Manager for developing the recommendations.
- Fisheries Information Management Committee – A NMFS-wide team, made up of representatives from each FMC, originally designated to answer information management related data calls and to make recommendations to the Science Board.
- Executive Sponsor - The Director, Office of Science and Technology.

## **1.6 References**

Daconta, Michael C., Information As Product: How to Deliver the Right Information, to the Right Person At the Right Time, 2007

Committee on Archiving and Accessing Environmental and Geospatial Data at NOAA, National Research Council, Environmental Data Management at NOAA: Archiving, Stewardship and Access, National Academies Press, 2007

McDonald, K. R., “NOAA's Global Earth Observation - Integrated Data Environment (GEO-IDE), 2007

The Data Reference Model Version 2.0, 2005



## **2 Current State of NMFS Information Management**

This section presents an overview of the current state of information management at NMFS, including interview results, key issues and concerns, and areas of potential risk.

### ***2.1 Interview Results***

Data and information are critical ingredients for all NMFS outputs. Over the years, some internal and external constituents have observed that NMFS faces challenges with its data and data management programs that potentially impact our ability to meet our current and future mission requirements. ST wanted to determine which problems existed, along with their nature and severity, in order to develop steps to address them.

To develop a baseline understanding, the FIMC identified a number of issues regarding data and information management. These issues were captured during interviews with NMFS FMC and HQ management. The respondents were asked to rate each issue based on their perception of its relative priority for NMFS to address.

Figure 1 presents the key information management issues and priorities as identified through interviews with NMFS management. The number to the right of the horizontal bars represents the average priority based on a scale from 0 to 3 (3 being the highest priority).

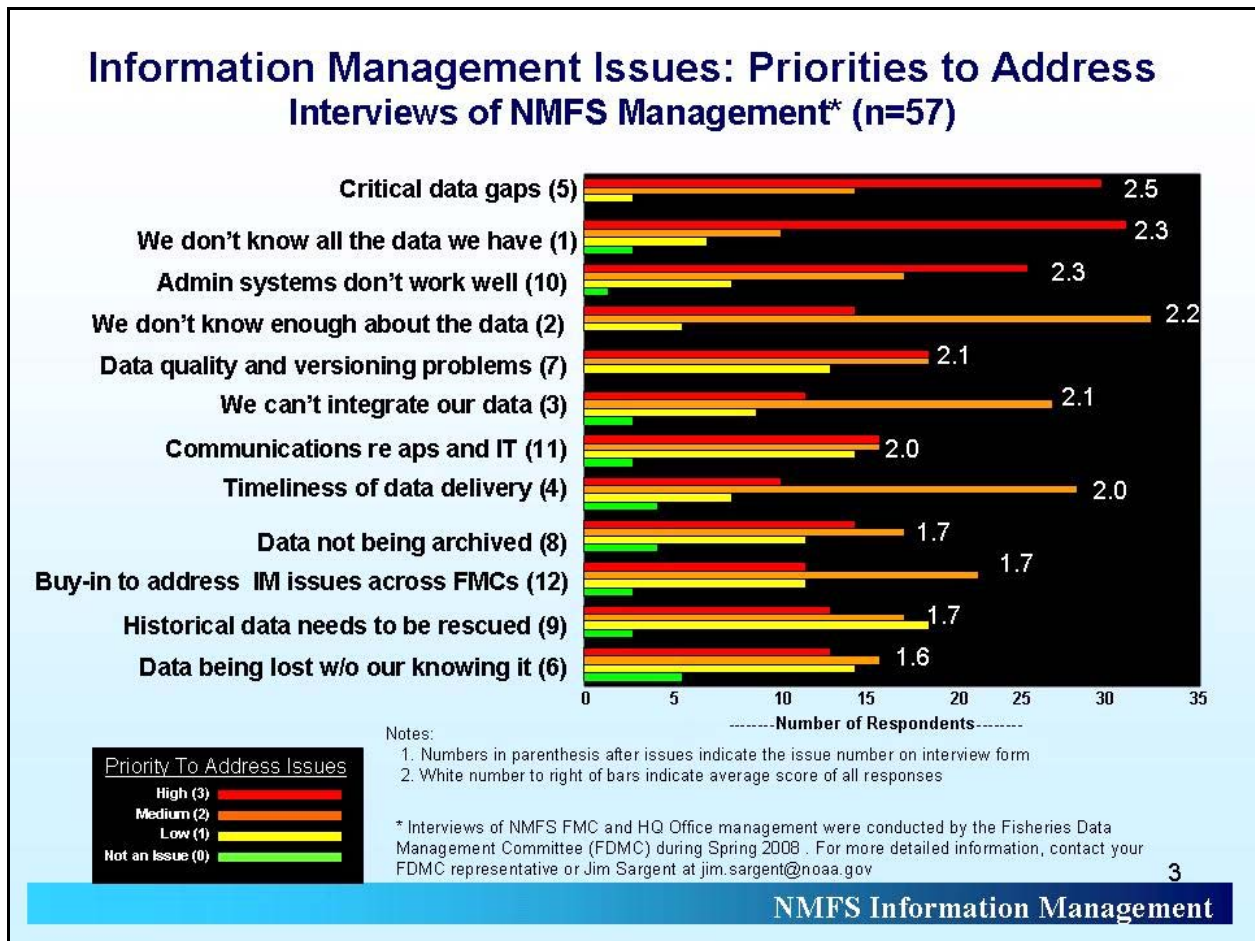


Figure 1. Priorities to Address IM Issues

## 2.2 Issues and Concerns

This section describes the nature and meaning of the information management issues and concerns identified in Figure 1. A summary of comments from the interviews is also provided. The full name of the concern is listed along with a short version (from Figure 1) shown in parentheses.

### Our data has critical gaps that need filling (Critical data gaps)

This issue relates to the concern that current data collections have data gaps that impact management decision-making and scientific assessment. Often, NMFS scientists have said we need more and better data to predict outcomes with greater confidence. As we move toward Integrated Ecosystems Assessment (IEA), is there a need to expand our data collection programs to meet evolving stewardship requirements?

Most interview respondents felt there is definitely a problem with gaps in data collection, especially given the expanding management requirements of the reauthorized Magnuson-Stevens Act. However, it was noted that while gaps certainly exist, we should not expand data collections until we have taken full advantage of what we have. We need to understand what data assets we have across the enterprise and maximize their use. We also should take advantage of cooperative research to augment the data NMFS collects.

### **We don't know all the data sets we have**

This issue relates to the accessibility of NMFS data. NMFS has collected a vast array of data over the past 100-plus years that may be important in answering the questions of today and the future. Interviewees were asked to assess the level of awareness of data across the enterprise and the level of knowledge of how to access that data.

Many interviewees felt there is a lack of coordination between and among offices. While most felt they had a good handle on local data needed to support day-to-day operations, data assets outside their FMC are poorly understood and not readily accessible. There was a feeling that users need a roadmap to find out where data exists, how it is accessed, and who the contacts are to help provide service. In addition, a national-level inventory should be required to enable standardization of data elements and allow data to be shared and combined.

### **The plethora of administrative systems are not connected and don't work well (Admin systems don't work well)**

This issue relates to the perception that administrative systems are not connected and don't work well as support systems. One NMFS senior official noted that we are confronted with a plethora of administrative applications from NMFS, NOAA, DOC, and OMB that overlap, seldom work together, have bugs, and often don't meet NMFS' needs. It is often difficult to determine what applications should be used when, and it is difficult to keep track of passwords. There may also be potential conflicts and redundancy of information in multiple systems.

Most respondents felt there is a significant problem with administrative systems, particularly the way the applications are rolled out with minimal end-user testing. The burden of feeding the applications has caused immense frustration and little benefit due to redundant data entry and lost time. The lack of integration and the expanding need for passwords have added to this frustration.

### **We don't know enough about the data we have to use it effectively and creatively (We don't know enough about the data)**

In order to use data effectively and relate it to other data sets, we need to have adequate information about the data (metadata), such as field names, coding structures, and formats, and how, when, and why the data were collected. It is also important to understand the quality and confidence level of the data so it is used appropriately. Interviewees were asked whether this information exists or whether we need to collect and document metadata to use our data more effectively.

While some organizations are actively developing metadata for their datasets, most interviewees consider lack of metadata a significant problem. In many cases, without proper metadata, data can be misconstrued and thus erroneously applied or be rendered essentially useless. There is a need to define the uses of the data, not just the existing content. We may be collecting data, but missing important components that represent economic and demographic factors essential for policymaking. Some noted that limited metadata, numerous format standards, and lack of exchange standards and common tools to query the data make it difficult to access data across the agency. Data are often in silos and difficult for the average user to access. Data sharing agreements and data quality standards need to be implemented

### **Data quality and versioning problems**

Several studies and reports have indicated that NMFS could improve the quality of its data. The problem of constituents receiving inconsistent data from two different sources is a challenge that

can be traced to not keeping track of different versions of data sets. This contributes to reduced confidence in NMFS data and the decision-making the data support.

All interviewees felt that maintaining data quality is essential for NMFS to adequately perform its mission. There seems to be an even split between those who think data quality is a major problem and those who feel it is under control. Several noted that inconsistent data cause problems for managers, particularly with regard to public perception, thus giving the impression that we don't know what we are doing or that we are making things up. This is also an important issue due to legal or FOIA concerns.

### **We can't integrate our data to determine the effects of climate change on living marine resources (We can't integrate our data)**

Most legacy NMFS information systems were built to meet specific business needs and were not designed with the “hooks” needed to integrate or link systems. The result is siloed data that is difficult to combine with related data sets. In some cases, data sets we would like to relate (e.g., combining research vessel trawl survey data with fisheries catch data, and combining habitat data with marine mammal sightings) cannot be integrated to meet the needs of IEA. Interviewees were asked to assess whether we are able to integrate our data well enough to meet our needs.

Most feel that integration is a problem—systems have grown up in stove pipe manner, each with their own set of codes and data formats. Species codes are a particular problem. There are different names for the same information and similar names for different information. Some noted that integration is difficult because there is no true architectural view of how everything meshes. There needs to be a detailed data architecture map that defines standards for process, data, and systems.

### **NMFS needs to improve communications regarding the development and use of its applications and technologies (Communications re: Apps and IT)**

Does NMFS need to improve communications regarding the development and use of its applications and technologies? NMFS develops and deploys new applications and technologies, such as electronic reporting of fisheries statistics, but often in a stove-piped manner. Interviewees were asked whether we need to share IT development efforts, successes, and challenges more frequently among the NMFS community.

Several interviewees felt there is no shared vision of technology at NMFS. FMCs tend to work almost entirely independently. Collaboration and sharing of ideas, solutions, and technology are increasingly important in a time of limited or shrinking budgets. Science is the driver and technology is often an afterthought. Some felt it would be useful to have a general NMFS/NOAA online clearinghouse for science, technology, and administrative solutions. This should be searchable so someone looking for an electronic fisheries reporting tool, for example, could find available information along with contact data.

### **Data delivery is not timely enough to implement new management strategies (Timeliness of data delivery)**

Is data delivery timely enough to meet the requirements of new management strategies such as individual transferable quotas (ITQs) and dedicated access privileges (DAPs)? In addition, with the advent of web technologies, the demand for timely data has increased. For example, information about marine mammal events is now expected to be available in near real time. Interviewees were asked whether critical data are being audited, analyzed, and delivered to managers and scientists in a timely manner to meet our current and future mission requirements.

Opinions on timeliness were highly variable, dependent on the management need against which the data are being applied. While timeliness is critical for activities such as in-season monitoring, it is not as important when looking at trends (e.g., habitat research). Some managers indicated there is a major time lag (on the order of at least 3 months) between the collection of data and the QA/QC that needs to be done before data can be analyzed and delivered. There is also a major lag between QA/QC and dissemination of that data. Several interviewees thought we are not making adequate investments in electronic data collection and reporting technologies that could improve timeliness of data. For example, the technology exists to collect real-time catch and effort data from U.S. tuna longline vessels by way of the Vessel Monitoring System (VMS).

### **Data is not being archived for perpetuity as per NARA requirements (Data not being archived)**

Is NMFS archiving data for perpetuity as per the National Archives Administration (NARA) requirements? NMFS data represent a corporate resource that needs to be preserved for use in the future. NARA regulations require that all federal records be preserved in archives conforming to their standards. The only archives in NOAA that adhere to these standards are NODC, NCDC, and NGDC. Interviewees were asked whether they thought our data were at risk due to a lack of archive policies and standards.

In general, there was a feeling that staff are not clear on what the requirements are. There is a lack of understanding of NARA requirements, and the issue is not on the radar screen of local management. There are also limited resources to maintain long-term archives. It was noted that some data sets might not warrant preservation for perpetuity. Analysis needs to be conducted to determine which data sets to keep and which to discard.

### **We don't have buy-in for addressing information management across FMCs (Buy-in to address IM issues across FMCs)**

Initiatives that attempt to address information management across FMCs, such as the Fisheries Information Systems (FIS), encounter resistance. Sharing data beyond the jurisdiction where they were collected, and for different purposes, involves risk with little perceived benefit to the originating FMC. Interviewees were asked whether there is a perceived lack of support for establishing NMFS-wide policies and initiatives to address some of these information management issues.

Most felt that everyone recognizes this as a problem but aren't willing to devote the fiscal and human resources needed to solve it. There needs to be a well-communicated campaign to sell the value of good data management and integration across the agency. Everyone benefits from good data management, and getting coordination and cooperation among FMCs is key to successful integration.

### **Historical data need to be rescued to provide a context for the future (Historical data needs to be rescued)**

Historical data provide an important context for the future. There is a concern that data may exist on perishable media and be subject to loss. Is there a need to rescue legacy data and migrate to more modern formats and systems before they are lost?

Interviewees felt that historical data may be at risk of being lost, but not all historical data are necessarily valuable. A review of historical data and a critical evaluation of whether it is needed, complete, and worthwhile need to be undertaken. Obsolete software and hardware are the

biggest risk for loss of data. Some historical data are on floppy disks and may be difficult to retrieve. We need a plan to convert data to the latest storage media and a mechanism by which data would be updated to new media as they become available.

### **Data is being lost without our knowing it**

This concern is based on observations that specific data sets and models on which some NMFS scientific publications were based were not preserved and were eventually lost. In addition, employees may keep data on personal machines, resulting in loss of data when employees leave or retire. Interviewees were asked whether they perceive potential loss of data as a problem that needs to be addressed.

All interviewees agreed that if data is being lost, then it is a significant problem. Many felt this is not a major problem. Some pointed out that data are also lost with full knowledge, in the sense that lack of adequate storage facilities, expanding needs for space, and new safety requirements have prompted the intentional disposal of irreplaceable biological samples from long ago. Consistent standards regarding archival of raw or ancillary data are critical, not only for purposes of scientific publications but also (frequently) as the basis upon which fishery management decisions or policies are based.

## **2.3 Other Drivers and Considerations for NMFS EDM**

In addition to the issues raised by the FIMC and prioritized by NMFS management, a number of other considerations need to be taken into account when developing recommendations for NMFS EDM. This section discusses the more salient ones.

### **2.3.1 Internal and External Data Management Programs and Initiatives**

NMFS information is in very high demand as we address issues such as overfished stocks, global climate change, and integrated ecosystems assessments in addition to our legacy mission requirements. These demands arise from global, international, national, state, and local organizations; from DOC and NOAA; and from within NMFS. Listed below are some prominent initiatives, programs, studies, and reports that require access to NMFS data and metadata, and/or the attention of NMFS information management staff.

#### **2.3.1.1 Global Programs and Initiatives**

##### **Earth Observing Systems (EOS) Global Earth Observation System of Systems (GEOSS) -**

Producing and managing better information about the environment has become a top priority for the United States and nations around the world. In July 2003, the [Earth Observation Summit](#) brought together 33 nations plus the European Commission and many international organizations to adopt a declaration that signified a political commitment toward the development of a comprehensive, coordinated, and sustained Earth Observing System to collect and disseminate improved data, information, and models to stakeholders and decision makers.

#### **2.3.1.2 National Programs and Initiatives**

**Integrated Ocean Observing System (IOOS) Regional Associations (RAs) -** The RAs provide the primary framework for orchestrating the required collaboration within each region and are responsible for the design and coordinated operation of regional coastal ocean observing systems. They consist of:

- Northeast: [GoMOOS](#) (GoMOOS)
- Mid Atlantic : [Mid-Atlantic Coastal Ocean Observing Regional Association](#) (MACOORA)
- Southeast: [Southeast Coastal Ocean Observing Regional Association](#) (SECOORA)
- Caribbean: [Caribbean Regional Association](#) (CaRA)
- Great Lakes: [Great Lakes Observing System](#) (GLOS)
- Gulf of Mexico: [Gulf of Mexico Coastal Ocean Observing System](#) (GCOOS)
- Pacific Northwest: [Northwest Association of Networked Ocean Observing Systems](#) (NANOOS)
- Central and Northern California: [Central and Northern California Ocean Observing System](#) (CeNCOOS)
- Southern California: [Southern California Coastal Observing System](#) (SCCOOS)
- Alaska: [Alaska Ocean Observing System](#) (AOOS)
- Pacific Islands: [Pacific Islands Ocean Observing System](#) (PacIOOS)

### **2.3.1.3 NOAA Programs and Initiatives**

- **Environmental Data Management at NOAA Report** – The National Academies Press was commissioned by the NOAA Science Advisory Board to study NOAA data management, and wrote the book [Environmental Data Management at NOAA: Archiving, Stewardship and Access](#), which is available via the above link. The principles from the book are:
  - *"Environmental data should be archived and made accessible"*
  - *"Data-generating activities should include adequate resources to support end-to-end data management"*
  - *"Environmental data management activities should recognize user needs"*
  - *"Effective interagency and international partnerships are essential"*
  - *"Metadata are essential for data management"*
  - *"Data and metadata require expert stewardship"*
  - *"A formal, ongoing process, with broad community input, is needed to decide what data to archive and what data not to archive"*
  - *"An effective data archive should provide for discovery, access, and integration"*
  - *"Effective data management requires a formal, ongoing planning process"*
- **Global Earth Observation Integrated Data Environment (GEO-IDE) Conops** was developed by the NOAA Observing Systems Council's (NOSC) Data Management Integration Team (DMIT). The GEO-IDE is envisioned as a "system of systems"—a framework that provides effective and efficient integration of NOAA's many quasi-independent systems, which individually address diverse mandates in areas of resource management, weather forecasting, safe navigation, disaster response, and coastal mapping, among others. NOAA Line Offices will retain a high level of independence in many of their data management decisions, encouraging innovation in pursuit of their missions, but will participate in a well-ordered, standards-based data and information infrastructure that will allow users to easily locate, acquire, integrate, and utilize NOAA data and information.

- **NOAA Observing Systems Architecture (NOSA)** – NOSA, developed by the NOAA Observing System Architecture (NOSA) Action Group, directed by the NOSA Senior Steering Group, was established to develop an observational architecture that helps NOAA:
  - Design observing systems that support NOAA's mission and provide maximum value;
  - Avoid duplication of existing systems; and
  - Operate efficiently and in a cost-effective manner.

NOSA includes:

- NOAA's observing systems (and others) required to support NOAA's mission;
  - The relationship among observing systems including how they contribute to support NOAA's mission and associated observing requirements; and
  - The guidelines governing the design of a target architecture and the evolution toward this target architecture.
- **Regional Ecosystem Data Assembly Portals** – The NESDIS National Coastal Data Development Center (NCDDC) provides a coordinated data management system and data discovery mechanism for atmospheric, oceanographic, and terrestrial physical sciences to facilitate sustained economic growth, scientifically sound environmental management, and public safety to the nation and the international community. The NCDDC is a leader in Regional Ecosystem Data Management (REDM) and the establishment of Regional Ecosystem Data Assembly Portals that provide access to data and information through the cataloging, formatting, archiving, and dissemination of the data streams in collaboration with federal and state agencies, regional observing system associations, academic researchers, and non-governmental organizations.
  - **Comprehensive Large Array Data Stewardship System (CLASS)** - CLASS is an electronic library of NOAA environmental data. This website provides capabilities for finding and obtaining those data. CLASS is NOAA's on-line facility for the distribution of NOAA and U.S. Department of Defense (DoD) Polar-orbiting Operational Environmental Satellite (POES) data, NOAA's Geostationary Operational Environmental Satellite (GOES) data, and derived data.
  - **NMFS Fisheries Information Systems (FIS)** – The FIS provides a context for the design, development, and implementation of data collection and data management for fishery-dependent statistics nationwide to improve the timeliness and accuracy of data. The FIS is a portal that identifies the existing federal and state fisheries information systems or databases (data collections) and provides integrated business solutions for effective information sharing. The FIS supports fisheries management decisions by developing a virtual application environment.

### 2.3.2 Data Management Committees and Organizations

As the web matures and access to federal data becomes commonplace, there are many committees that address the needs for agencies to respond. Listed below are some of the more



salient committees and organizations that demand the attention of NMFS IM personnel, and should be considered in the development and administration of NMFS EDM initiatives.

**NOAA Observing Systems Council (NOSC)** - The NOSC serves as the principal advisory body to the NOAA Administrator and the focal point for the agency's observing system activities and interests. The purpose of the Council includes coordinating observational and data management activities across NOAA; proposing priorities and investment strategies for observation-related initiatives; and identifying programs that might benefit most from integration.

**NOAA Data Management Committee (DMC)** - The DMC is the NOSC agent for coordinating the development and implementation of data management policy across NOAA. The DMC has broad latitude and authority for data management decisions for NOAA. It addresses issues and opportunities that require coordination among the Goal Teams, Line Offices, and Data Centers to implement data management responsibilities. The Committee provides guidance and recommendations to the NOSC for managing and integrating data within NOAA.

**Data Management Integration Team (DMIT)** - The DMIT Integration Team provides expertise and advice on the near-term (5-year) actions needed to write and implement a plan for integrating data management within NOAA. The team has representation from all of the NOAA Line Offices and Goals to ensure that all NOAA interests and requirements are represented.

**The Data Archiving and Accessing Requirements Working Group (DAARWG)** - The DAARWG was commissioned by the Science Advisory Board (SAB) to provide recommendations to the SAB regarding accessing and archiving NOAA data. The Group is currently chaired by Dr. Ferris Webster, College of Marine Studies University of Delaware, and meets approximately semi-annually.

**NOAA Integrated Products Team (IPT)** - The IPT, as part of the IOOS Data Management and Communications (DMAC) Subsystem at the NOAA Ocean Service, is charged with developing the Data Interchange Framework (DIF) for IOOS.

**NOAA Enterprise Architecture (EA) Committee Working Group** - The NOAA Enterprise Architecture (EA) Committee serves the CIO Council as a resource to establish Enterprise Architecture (EA) policies and standards, and to develop and promote subject matter expertise in areas/programs assigned to the committee. The EA Committee includes representatives from all NOAA Line Offices and leads NOAA in the implementation and sustained use of EA as a strategic information management and decision-making practice.

**Fisheries Information Systems (FIS) Architecture Committee** - The FIS Architecture Committee is developing the technical architecture for the FIS.

### **2.3.3 Other Issues**

A number of other areas of concern or potential risk must be considered to make NMFS data more accessible.

### **Information Requests from External Sources**

NMFS receives a variety of requests for information from the fishing industry; state, federal, and local governments; the press; and the general public. Some requests are very simple and are responded to in minutes over the phone or e-mail. Others require significant staff time to assemble and compile data in incompatible formats. Some of the requests are formally tracked by Headquarters' correspondence tracking systems, but the majority are informal and may not be tracked at all, which can potentially lead to duplicative calls resulting in different responses to similar questions. As a result, NMFS may be criticized for providing conflicting information and conducting inconsistent science.

### **Information Requests from Internal Sources (Data Calls)**

Another area of potential challenges, often cited by staff and managers, is the many data calls that HQ issues to the regions. The calls are often considered to be poorly defined with unrealistic time schedules, and may duplicate previous requests. Currently, the calls are not tracked by any central system or governed by any policies or best practices.

### **Internet Culture Maturation**

The maturation of the internet, in particular the move to Web 2.0, presents both opportunities and challenges for NMFS. There is unprecedented opportunity for information sharing, networking, and establishing communities of practice. At the same time, there are many challenges related to increased expectations for readily available products and services. There is the potential that NMFS data will be misused or used in ways we never intended. Our data may be shared beyond our control and served up by the use of taxonomies we did not devise. We need to consider these concerns carefully as we develop plans to expose more NMFS data via the web.

### **Legal Issues**

As NMFS data is made more accessible, there will be an increasing need to ensure it is protected and shared appropriately. Personally Identifiable Information (PII) must be secured against intentional or unintentional disclosure. Beyond PII, policies regarding the use and dissemination of confidential data must be followed. However, a balance needs to be struck between the needs of confidentiality and the requirement for sharing within NMFS and partner organizations to ensure access to detailed data by authorized individuals.

There is also a need to maintain records and administrative data in support of litigation and FOIA concerns. This requires that data and information are well-documented, searchable, and easily accessible. Currently, data calls for FOIA or other requests are often extremely time-consuming due to the difficulty in compiling the data needed to complete the request.

### 3 Information Management Business Processes

This section provides a brief overview of the functional components of IM business processes that collectively comprise the scope of NMFS EDM. These components are:

- Information Life-Cycle Management
- Discovery and Access
- Interfaces with External Systems

#### 3.1 Information Life-Cycle Management

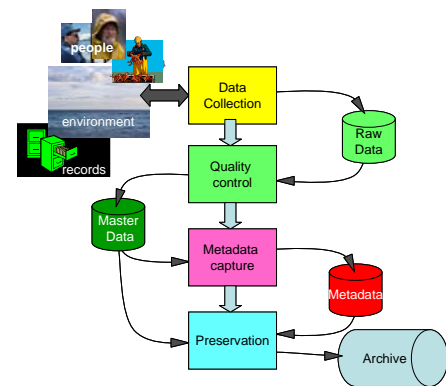
Information life-cycle management involves the end-to-end management of information, from initial collection through long-term preservation and archiving.

It starts with the determination of what data and information to collect, which involves determining the requirements needed to support NMFS business processes. If collection involves requesting information from the public, it requires OMB approval according to the Paperwork Reduction Act (PRA). The PRA requires that the Office of Management and Budget (OMB) approve each collection of information by a federal agency before it can be implemented. The statute defines "collection of information" broadly. It covers any identical questions posed to 10 or more members of the public—whether voluntary or mandatory, and whether written, electronic, or oral. If the collection does not involve PRA, then the survey is designed and developed by the business process responsible for the collection.

Once a data collection has been approved, it is captured by various means (logbooks, VMS, dealer reports, etc.) and stored in electronic form as raw data. Following data collection, the information enters into the transformation and control phase, which is usually performed in conjunction with Information Technology (IT). This involves performing quality control and quality assurance (QA/QC) processes, establishing confidentiality assessments, and ultimately storing the information in a data asset such as a database, Electronic Document Management System (EDMS), web page, or other electronic formats.

Information resources are maintained throughout their useful life. Maintenance activities are governed by best practices for information quality, security, metadata, access, and preservation.

As a corporate resource, NMFS information assets must be preserved for the long term. Information not needed on a regular basis may be archived, but in a manner that still makes it readily accessible. A special case regarding long-term storage is the preservation (and perhaps rescue) of data residing on obsolete or inaccessible media.



**Figure 2. Information Life-Cycle Management**

### 3.2 Discovery and Access

Information discovery and access involves users discovering, accessing, and using NMFS information assets.

Once a data asset is stored, it becomes available for discovery, access, and use. Information is accessible through a variety of NMFS applications, websites and database queries, table lookups, etc. Access to NMFS information is also provided by NMFS staff responding to specific requests through phone calls, e-mails, and other modes of communication. Responding is often a complex process, involving significant staff time to assemble and compile data sets that are sometimes in incompatible formats. Another class of information requests is internal data calls, which usually are generated in an *ad hoc* fashion by NMFS HQ staff for the regions.

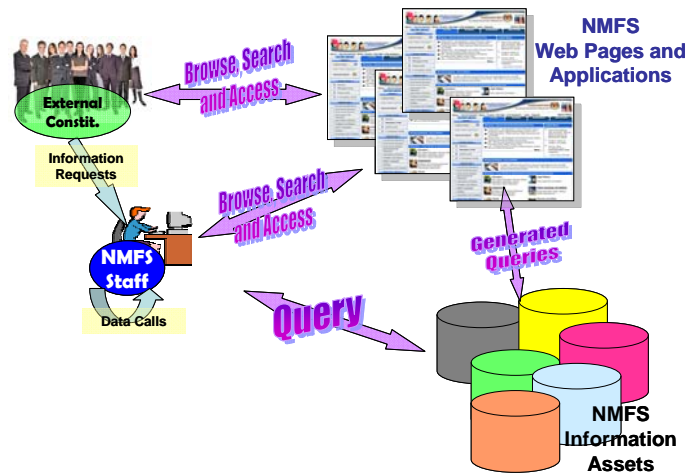


Figure 3. Information Discovery and Access

### 3.3 Interfaces with External Systems

The need to interface with outside systems is increasing as the web culture matures. It involves interfacing systems including those at NOAA; other federal, state, and local governments; and international organizations. Examples include NMFS participation with the Global Earth Observation System of Systems (GEOSS); the Integrated Ocean Observing System (IOOS) and all the related regional observing systems; the Census of Marine Life; and the Ocean Biological Information System (OBIS). Interfacing with these systems involves a high degree of collaboration and the use of standards and common technologies. Interfacing also requires developing agreements on levels of access, availability, response time, and refresh rate for information. Agreements also describe the methods, formats, and services related to enabling information interface and exchange.

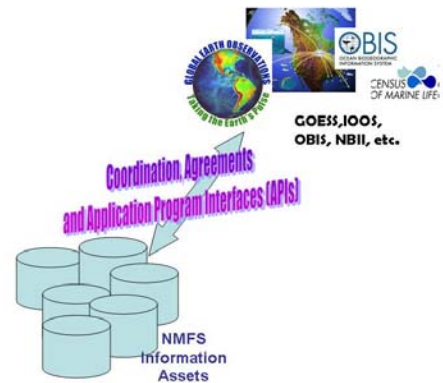


Figure 4. Interfaces with External Systems

## 4 The Target State of NMFS EDM

In reviewing the challenges presented by the current situation at NMFS (see Section 2), it was clear that a comprehensive approach to EDM was needed. The FIMC recommends establishing an ongoing NMFS EDM program. The program will bring focused attention, resources, and coordination to bear in resolving the critical IM issues and concerns facing NMFS. The FIMC developed foundational concepts to help guide the direction and focus of NMFS EDM. This section describes the vision, core principles on which the program is based, the mission statement, and the goals and objectives of NMFS EDM. The four components of the target program are then discussed.

### 4.1 Vision

The vision for NMFS EDM developed by the FIMC:

*NMFS customers can confidently  
find, access, and use our data.*

**NMFS Customers** means all NMFS internal and external customers.

**Can Confidently** means they have confidence (i.e., trust) in the data.

**Find** means that by using an Amazon.com-type portal or portals, they can

- Browse through an ordered hierarchy or taxonomy;
- Search by using discipline-specific key words (taxonomies) (for example, a 9-year-old child uses a different lexicon than a scientist or a Hill staffer) and/or their own user tags (folksonomies); and
- Review metadata, including user feedback and customer experience evaluation (CXE) ratings.

**Access** means they can view data through confidentiality and security filters, by using standard tools and formats.

**Use** means they can download selected data, if appropriate, with sufficient metadata to effectively use the data, including quality indicators, warnings, and information about data elements (coding structures, formats, etc.)

**Our Data** means all NMFS enterprise data we choose to share.

### 4.2 Principles

The FIMC developed and reached consensus on the following guiding principles that form the basic precepts for NMFS EDM:

- Data and Information are an important NMFS corporate resource and need to be managed over their lifecycles (Management, Information Lifecycle Management, and Preservation).
- Our data and information are a global resource and need to be shared appropriately (Discovery and Access).
- Our data and information inspire public trust and confidence (Transparency, Reliability, and Accountability).
- Our data and information are continuously assessed for quality, objectivity, timeliness, and sufficiency to meet the NMFS mission (Sufficiency and Necessity, Integrity, and Completeness).

### 4.3 Mission, Goals, and Objectives

The mission of NMFS EDM is to:

***“Deliver the right data, information, and services when and where they are needed.”***

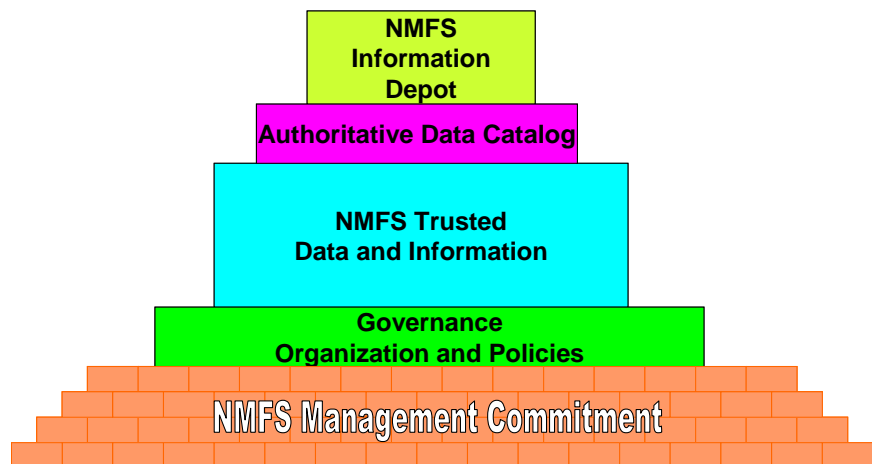
In support of the NMFS information management mission, the following goals were developed by the FIMC:

1. **Manage information with appropriate governance, focus, and accountability.**
2. **Manage information cooperatively to promote sharing of information, solutions, and information management investments.**
3. **Establish and maintain policies, best practices, procedures, and metrics to ensure good stewardship and preservation of NMFS information resources.**
4. **Empower internal and external constituents to effectively discover, access, integrate, and use the information to answer key current and future questions.**
5. **Continually assess and improve the availability, timeliness, and quality of data resources to support NMFS mission.**

The objectives for each of these goals are presented in Appendix C.

### 4.4 The Four Components of NMFS EDM

In the following sections, the four components are detailed in terms of what they are, how they will be implemented, and their expected value and benefits for NMFS. The critical foundation of sound EDM is ongoing management



**Figure 5. The Four Components of NMFS EDM, and Their Foundation**

commitment.

#### 4.4.1 Component One: Governance

The Governance component serves as the main support for NMFS EDM by providing the organization components including the IA, Regional IM coordinators, data stewards for enterprise data assets, and the Fisheries Information Management Advisory Committee (FIMAC). It also involves establishing IM policies, procedures, and best practices, and provides the administration and management to support NMFS through the cultural shift to EDM. Some key sub-components are discussed below.

#### The Governance Model

A governance model describes how organizations align IT strategy with business strategy to ensure they stay on track to achieve their strategies and goals. Proper governance ensures that stakeholder interests are taken into account and that processes provide measurable results. Governance formalizes the decision-making process to help ensure investments align with needs and that the expected benefit of the investment is realized.

To maintain sound information management within NMFS, governance will be an ongoing and evolving process that supports sound project management and mechanisms to communicate progress, success, value, and benefits. The governance effort will be shared between the Regions and Headquarters, with consideration of the need for regional flexibility.

Buy-in and commitment from management will be critical to implementing a successful governance model.

#### Implementation Approach

A key aspect of EDM governance is to define the organizational structure as it relates to management and decision-making for the program. Each organizational element will be designed to ensure appropriate representation between HQ and the Regions, and will be defined in terms of roles and responsibilities. We anticipate the following organizational elements will be needed for NMFS EDM:

- Enterprise Information Architect (IA)
  - Responsible for the overall EDM and administering IM policy at NMFS
  - Serves as the single point of contact for representing NMFS with outside entities such as IOOS, NOAA data management issues and concerns, NESDIS Data Centers, and others
- Regional Information Management Coordinators (RIMCs)
  - Responsible for EDM and administering IM policy in the FMCs
  - Serve as single points of contact in FMC regarding information management issues
- HQ Office Information Management Coordinators (OIMCs)

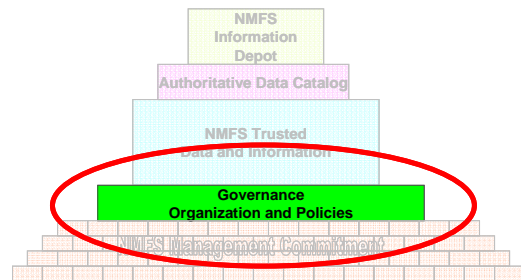


Figure 6. Governance

- Responsible for EDM and administering IM policy in the HQ Offices
- Serve as single points of contact in HQ Offices regarding information management issues
- FIMAC
  - Serves as Configuration Management Board for NMFS IM policies and procedures
  - Composed of RIMCs plus subject matter experts (SMEs) and chaired by the IA
  - Develops recommendations to the Leadership Council for IM policies and procedures
  - Evaluates the applicability of SOA and makes deployment recommendations
- Data Stewards – Individuals who are responsible for establishing and maintaining the quality, integrity, documentation, and preservation of the data asset

Policies and procedures are needed to define how EDM will be managed and executed. At a minimum, policies and procedures will be developed and implemented for:

- Portfolio Management – organizes program projects and initiatives into logical groupings to better facilitate their management and leverage investments.
- Project Selection – a formalized process to review and select project initiatives for funding under the program.
- Project Management – defines a standard, best practices approach to executing program initiatives to ensure that expected outcomes are realized. Included within this is definition of processes for quality, configuration, risk, and communications management.
- Performance Management – defines the methods and measures we will use to assess progress against program goals and objectives. Performance metrics will be put in place, enabling us to continually assess and improve upon the success of the program. Performance plans for information managers and data stewards will be vital in terms of accountability. Periodic reviews of service-level metrics related to the delivery of NMFS products and services will also be critical to measuring the success of the information management effort.

### Value and Benefits

A well-designed governance model will provide the following value and benefits:

- Representative decision-making with input being considered from all interested stakeholders.
- Effective decision-making leading to wise investment of limited human and budget resources.
- Best practices approach to managing NMFS EDM to ensure desired outcomes.

### Stewardship

Effective data stewardship involves the full lifecycle management of information assets, from initial collection through long-term preservation and archiving. It defines the standards, policies, procedures, and best practices that should be used consistently throughout the agency. This consistency is essential to ensure that NMFS information assets use common vocabulary and share formats to allow them to be combined and integrated.



A number of policies and procedures regarding the effective management of data, metadata, and information will be developed, providing for a comprehensive approach to information lifecycle management that will be applied to all assets through creation/receipt, distribution, use, maintenance, and archiving.

### Implementation Approach

As part of the governance model, policies and standards relating to data stewardship throughout the information management lifecycle will be developed and implemented. Policies are needed that address the following aspects of data stewardship:

- Stewardship roles and responsibilities.
- Data inventory and metadata.
- Access and information sharing.
- Security and confidentiality.
- Long-term preservation and archive.

In an effort to make data the center of the business, the following standards will be developed, agreed to, and implemented:

- Metadata at various levels; i.e., discovery, detailed, element levels.
- Minimum required elements for each type of data collection.
- Common definitions, formats, and business rules for data elements.
- Data quality and auditing procedures.

As the information management effort grows and evolves, independent reviews of compliance with policies and standards related to stewardship, as well as revisions of policies and standards, will enable leadership to continually assess the success of the program.

### Value and Benefits

Effective data stewardship will provide the following value and benefits:

- Data and information are managed consistently throughout NMFS ensuring their availability, quality, and integrity.
- Standards are adopted across the agency, thereby allowing related data sets to be easily combined when needed.
- Agreements involving information sharing and appropriate use are in place and followed by all users of NMFS information resources.

### Budgeting Strategy

Appropriate budget resources are needed to develop, implement, and maintain an information management program at NMFS. While there are many potential alternatives and levels of investment for implementing NMFS EDM (see Section 5), each involves trade-offs in terms of the scope of the effort, the time to completion, and the expense. Management will need to assess the level of investment that best meets the ongoing business needs of NMFS.

### Implementation Approach

The budgeting strategy consists of both a short-term and a long-term approach. In the short term (the next 2 to 3 years), the program will need to subsist on existing budget resources. This includes:

- Utilizing existing funds for program activities, e.g., FIS, NIMB funds, IOOS, and REDB.
- Minor reprogramming.
- Potential use of cost recovery strategies.

In the long term, the EDM program should have its own budget line to allow full-scale development and implementation of the program. This entails developing an FY 2012 PPBES budget submission including the following activities:

- Working with EGT programs.
- Including the EDM in most or all EGT programs.
- Developing an IT initiative and driving it through the CPIC process:
  - Creation of an Exhibit 300
  - Submission to the NOAA IT Review Board
  - Submission to the Commerce Review Board
  - Championing of the initiative through the rounds of pass-backs

#### Value and Benefits

Appropriate levels of funding for NMFS EDM will provide the following value and benefits:

- Effective management of NMFS information assets is viewed as a critical priority within the agency.
- NMFS EDM is established and funded at levels that allow steady progress to be made and significant benefits to accrue to the agency.

#### Communication and Technology Sharing

The NMFS information management program will encourage ongoing communication and coordination with the FMCs and related programs such as FIS to leverage efforts, exchange information on common needs and services, and promote sharing of information, solutions, and information management investments.

#### Implementation

The following approach is recommended to promote communication and technology sharing:

- Develop a program communications plan
  - Describes stakeholder needs, methods, and frequency of communications and outreach
  - Create a program website to promote activities
  - Develop and distribute communication materials (newsletters, posters, fact sheets)
- Establish a program collaboration portal
  - Supports management of program activities and schedule
  - Supports review of all program materials

- Provides collaboration space for teams working on individual projects or initiatives
- Establish a central repository for EDM information to support:
  - Communities of excellence
  - Active forums to exchange information and technology solutions
  - Leveraging of knowledge and investments that may be extended from one partner to another

### Value and Benefits

A well-developed communications and technology sharing mechanism will provide the following value and benefits:

- Program activities are transparent to stakeholders promoting discussion and buy-in on initiatives
- Forums are established where NMFS internal and external constituents can exchange knowledge and solutions with practitioners in their areas of interest

### Cultural Shift

The collaboration and coordination of information management efforts may prove to be difficult as regions will want to focus on their own issues. It will be necessary to inspire a cultural shift by selling the EDM paradigm to the regions and making our activities legitimate at a regional level. Achieving a successful cultural shift requires that all stakeholders have a voice in program decisions, that new policies and procedures are properly vetted, and that necessary levels of outreach and training are provided to help institutionalize changes.

### Implementation Approach

Achieving a cultural shift in the business practices of the agency will involve the following activities:

- Integrating information management best practices into all business processes. Information management needs to become a routine part of how NMFS executes its mission. The IA will be established as the enduring champion of EDM in order to provide the focus and consistency needed to achieve change.
- Moving primary interfaces with customers to the web, and defining our business rules for the web and allow customers to self-serve with NMFS information resources and services.
- Harnessing social networking as a means to promote information exchange and foster communities of practice.
- Adopting electronic rulemaking to identify and address e-Discovery issues.
- Incorporating electronic document management systems (EDMS) into day-to-day business, thereby reducing dependence on paper-based records.

### Value and Benefits

Shifting the current culture at NMFS to one more focused on information management will provide the following value and benefits:

- Effective information management is adopted as a standard business practice throughout the agency.

- NMFS personnel recognize the value and importance of common standards, policies, and practices.
- The agency maximizes use of web-based technologies enabling customers to self-serve with NMFS products and services on a 24/7 basis.

## **Information Request and Data Call Management**

Information requests from external constituents and internal data calls are received frequently and need central coordination, tracking, and management (see Section 2.3.3 for a more complete discussion of this sub-component).

### **Implementation Approach**

It is recommended that the following approach be taken to manage information requests and data calls:

- Establish policies and best practices for managing requests and responses
- For information requests
  - Establish the IA (or designee) as the single point of contact for all external information requests
  - Create a web-based tracking system to manage the lifecycle of request and response
  - Establish regular reports for common requests
  - Analyze requests to determine what information should be served up on the web
- For data calls
  - Establish the IA (or designee) as the single point of contact for all internal data calls
  - Create a web-based tracking system to manage the lifecycle of data calls
  - Document the schedule of regular or routine data calls
  - Analyze data calls to determine what information should be centrally retained to avoid redundant calls

### **Value and Benefits**

Central management of information requests and data calls will provide the following value and benefits:

- All information requests and data calls are tracked through their lifecycle to help ensure timely responses are made.
- Central administration will help reduce or eliminate time spent on redundant effort.
- Potential inconsistent responses to constituents will be eliminated.

## **4.4.2 Component Two: Trusted Data**

NMFS' mission is almost entirely dependent on the trust our internal and external constituents have in the quality, reliability, security, timeliness, and sufficiency of our data. By and large, our excellent data management inspires that trust. And yet, occasionally, we question whether we are collecting the right data to answer today's questions, or whether we may be divulging some

confidential or sensitive data, or whether we could increase the quality of our data. This section discusses strategies for institutionalizing trust in NMFS' information resources.

### Analysis of Data Sufficiency and Necessity

Data and information must be continuously assessed to ensure they are sufficient to meet the NMFS mission of providing the right data, information, and services when and where they are needed, as well as meeting the needs of internal and external constituents. Analysis is needed to ensure that the right data is being collected to answer current and future questions. Part of this analysis should include evaluating whether all collected data is truly needed.

### Implementation Approach

The following major activities are recommended to determine the level and degree of data sufficiency and necessity:

- Support assessment efforts with inventories, metadata, and data models of existing data collections
- Map data models to business requirements and determine gaps
- Promote project management strategies that have proven successful in assessing data collection requirements (e.g., with HMS data)
- Promote data assessment workshop formats and facilitators with a proven track record
- Provide direct project management and workshop support
- Develop recommendations on how to close any gaps in data sufficiency or to eliminate unneeded elements from data collections

### Value and Benefits

Analysis of data sufficiency and necessity will provide the following value and benefits:

- Gaps in data collections are identified and recommendations are developed for eliminating those gaps
- Data collection burden is reduced through the elimination of unneeded data elements from current collections

### Security and Confidentiality

Security and confidentiality in the context of information management involves data access controls and the confidentiality rules affecting the ability to share data and with whom they may be shared. Policies must be put in place to ensure that authorized users get access to the data they need while protecting confidential data from inadvertent disclosure.

### Implementation Approach

The program will build on work started by ST to develop policy to account for revisions to data confidentiality under the Magnuson-Stevens Reauthorization Act (MSRA). There also needs to be an enterprise data policy that describes the rules for sharing data both internally and externally. As NMFS information resources are made more available via the web, additional rules regarding the sharing of data, and especially protection of personally identifiable

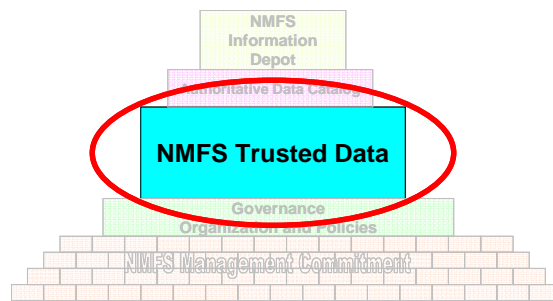


Figure 7. Trusted Data

information (PII), need to be established. The advent of social networking on the web (e.g., chat rooms and blog sites) may present additional challenges in ensuring that NMFS data are not inappropriately disseminated or used.

### Value and Benefits

Effective security and confidentiality policies and standards will provide the following value and benefits:

- Data are shared appropriately, with authorized users getting needed access and confidential data being protected from improper disclosure
- The public has improved confidence in NMFS' ability to safeguard data submitted to the agency

### **4.4.3 Component Three: The Authoritative Data Catalog**

The Authoritative Data Catalog provides the indispensable card catalog for NMFS' vast library of information assets. It is the single definitive source for identifying the information that NMFS collects, analyzes, disseminates, and preserves, and serves as the repository for the data about the data assets (i.e., metadata). The following sections describe the inventory and the metadata documented in the inventory.

### The Data Inventory

For NMFS data and information assets to be shared, they must be discoverable. Too often, access to information is dependent on knowing where to look or with whom to talk. Developing a regularly updated inventory of all data assets will provide for more efficient identification, assessment, and prioritization of our shareable information assets. Making the inventory available to users, combined with appropriate metadata (see next section), will greatly improve access to NMFS' information assets.

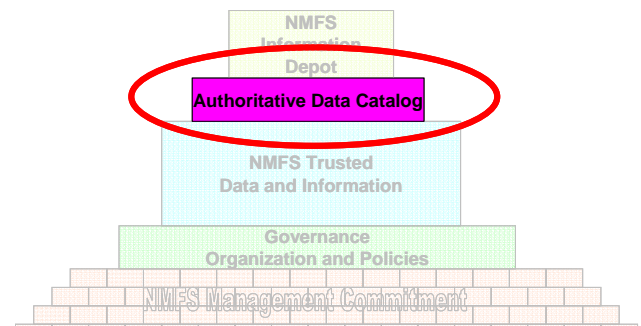
### Implementation Approach

Conducting a complete inventory of all NMFS information assets is a large effort involving a review of all HQ and Regional assets. To do this well requires dedicated resources from across the organization. This can also be an expensive undertaking depending on how quickly we want to complete the inventory. There are trade-offs, and we recommend developing the inventory iteratively in three groups as follows:

- Large size and/or frequently shared data assets
- Medium size and/or occasionally shared data assets
- Small size and/or seldom shared data assets

The inventory and associated metadata needs to be managed and controlled by a professional data librarian to ensure that the information in the inventory stays up-to-date. Policies and best practices relating to the frequency of updating the inventory, versioning, reporting, and performance measures will be developed and implemented by the librarian.

### Value and Benefits



**Figure 8. Authoritative Data Catalog**

Creating and maintaining a data inventory will provide the following value and benefits:

- Information assets across NMFS are discoverable regardless of their physical location
- Better understanding and coordination is promoted between Regions by giving visibility to information assets important to cross-regional initiatives such as ecosystem management
- Potential of redundant data collections is reduced

## Metadata

Metadata is “data about data” and provides the context for data usage. Well-designed metadata helps facilitate the understanding, characteristics, use, and limitations of the data. This is especially important for NMFS, since appropriate understanding of the data is essential for effective decision-making and stewardship of marine resources.

### Implementation Approach

Policies and standards regarding metadata will be developed, as metadata spans the information management lifecycle as part of overall data management and stewardship. It will be important to show the benefits and ease of use of metadata across the regions to achieve buy-in and agreement on the standards.

Once the policies and standards are in place, metadata will need to be developed for all information assets contained in the inventory. There are various levels of metadata to be developed: from high-level “discovery” metadata describing the asset, to detailed metadata describing each data element contained within the asset. It is envisioned that an existing tool, such as InPort, can be utilized as the metadata repository. We recommend the following phased approach to developing metadata for NMFS information assets:

- Phase 1
  - Define the metadata model and policies and standards governing metadata
  - Capture *discovery-level* metadata for *large/often shared* data assets
- Phase 2
  - Capture *detailed-level* metadata for *large/often shared* data assets
  - Capture *discovery-level* metadata for *medium/occasionally shared* data assets
- Phase 3
  - Capture *element-level* metadata for *large/often shared* data assets
  - Capture *detailed-level* metadata for *medium/occasionally shared* data assets
  - Capture *discovery-level* metadata for *small/seldom shared* data assets
  -
- Phase 4
  - Capture *element-level* metadata for *medium/occasionally shared* data assets (as appropriate)

	Metadata Level		
	Discovery	Detailed	Element
Large Assets/ Often shared	1	2	3
Medium Assets/ Occas. Shared	2	3	4
Small Assets/ Seldom Shared	3	4	4

- Capture *detailed-level* and *element-level* metadata for *small/seldom shared* data assets (as appropriate)
- Continually update and maintain all metadata.

*Discovery-level* metadata is defined as that which is required for customers to discover the data asset. It includes information such as data asset title, owner, data steward, description, and information category. *Detailed-level* metadata is defined as enough information for customers to determine whether the data asset is what they want. It includes information such as time, location, data quality, suggestion for use, and formats. *Element-level* metadata is defined as metadata required for the customer to use the data elements. It includes, for each data element, information such as the element name, description, format, and coding structure.

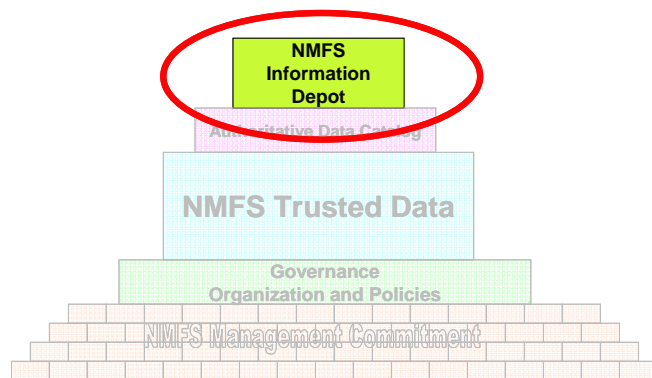
### Value and Benefits

Creating and maintaining metadata for NMFS information assets will provide the following value and benefits:

- Simplifies the discovery of data and helps users better assess whether the information in the asset is what they truly need; this should lessen the time needed to compile information for scientific assessment, data calls, and other uses
- The meaning and nature of information is better understood and is used appropriately

### **4.4.4 Component Four: The NMFS Information Depot**

In order to be useful, NMFS’ information resources must be easily discoverable and appropriately accessible to both internal and external constituents. Component Three: The Authoritative Data Catalog will enable us to understand what data we, as an enterprise, hold. But in order for the data to be useful to NMFS staff and our customers, they have to be able to find it. NMFS has made great strides in serving data up on the web. Too often, however, our customers have to wade through a number of NMFS websites to actually locate the data. For instance, would a student interested in the migratory habits of Beluga whales go to a website associated with the Alaska Regional Office, the Science Center, or NMFS Headquarters Office of Protected Resources to find the information?



**Figure 9. Information Depot**

The following sections discuss the portal and the various sub-components and technologies required to fully implement the concept.

### **The Front Door**

The Information Depot will serve as a front door to all NMFS enterprise data, empowering internal and external constituents to effectively discover access, integrate, and use information to answer key current and future questions. The NMFS Information Depot is envisioned as an



Amazon.com–type website that provides access to services with appropriate security and confidentiality filters. The portal would provide users with the following capabilities:

- One-stop shopping for NMFS information products and services; this may be modeled on the FIS One-Stop
- Provision of multiple taxonomies that allow customers to use their common language to search for and access information
- Discovery engineering tools and techniques to learn about customer search patterns, likes and dislikes, and successes and failures in finding what they want
- A feedback mechanism that accepts customer reviews and assessments of NMFS information and incorporates them into the metadata repository
- The ability to subscribe to specific information assets and receive a notice or alert when new or modified data are available

### Implementation Approach

In developing a NMFS information portal, we recommend that a phased approach be taken, starting with a small pilot to demonstrate feasibility and proof of concept. This approach will involve the following activities:

- Survey a limited set of regional and HQ information resources and document them in the metadata repository
- Develop an initial taxonomy to support discovery
- Design and develop the pilot portal to utilize the regional and HQ information sources documented in the metadata repository
- Validate the pilot through user testing and record suggested changes and enhancements
- Develop a plan for further development and implementation of a fully featured portal
- Eventually adopt a service-oriented architecture (SOA) service model and expand to NMFS-wide implementation, as appropriate

### Value and Benefits

Developing the Information Depot will provide the following value and benefits:

- A single point of entry for access to all NMFS information products and services
- Improved user experience and simplified access to needed information
- The ability to receive direct customer feedback and review of NMFS information products and services

### Discovery Engineering

Discovery engineering provides a means of analyzing customer activities and behaviors on websites in order to better understand their needs and improve service delivery. NMFS data and information are global resources and should be shared appropriately. Through discovery engineering, we will take a systematic approach to the practical application of common terminology and hierarchical taxonomies to develop a logical means of categorizing and organizing data information in ways that are meaningful to our customers.

### Implementation

The following techniques will be applied through discovery engineering:

- Web analysis of visitor activities
  - Identify paths through the portal and how long each page was viewed
  - Identify the keywords used in searching and navigation
  - Identify search patterns and trends
  - Assess the effectiveness in obtaining the desired products and services
- Identify customer groups and link them to their web usage patterns, and assess differences in usage among scientists, management, academia, legislators, children, etc.
- Use the analysis to reshape and improve taxonomies and other features and capabilities
- Allow users to put custom tags on data assets
- Allow users to rate data assets

### Value and Benefits

Discovery engineering will provide the following value and benefits:

- Better understanding of who NMFS' customers are and what their needs are
- Improved user experience and service delivery from NMFS websites

### Information Integration

Information integration refers to the ability to combine data sets of related information that may have been collected via disparate systems. NMFS information systems have typically been developed in a stand-alone or stove-piped manner, which makes integration difficult. True integration will require that systems be constructed using common architectural models of NMFS business processes and data. Before undertaking an effort of this magnitude, we need to determine the business need for integration rather than integrating just for the sake of integration.

### Implementation Approach

The following is a recommended approach to determining integration requirements and establishing the means to achieve information integration:

- Conduct an analysis of integration requirements to support internal and external constituent needs
  - Determine the highest priority integration needs
  - Analyze integration requirements and develop a strategy
- Create architectural approach
  - Define and develop an integrated business model
  - Define and develop an integrated data model
  - Assess the relationship between these models and use the relationship as a guide for developing integrated information systems
- Develop an approach to address data heterogeneity issues including:
  - Metadata requirements
  - Data standards (definitions, formats, etc.)

- Mechanisms such as bridge tables to link existing information resources
- Conversion factors and code translations to allow data to be combined

### Value and Benefits

Appropriate information integration will provide the following value and benefits:

- Eliminates the extensive data manipulation required to compile data for activities such as stock assessments
- Establishes a clear architectural framework to guide future system development, thereby ensuring that new or redeveloped systems can be integrated
- Enables a single version of the truth across the enterprise

### Lexicon and Taxonomies

A standard lexicon and taxonomies are needed to provide the common language and semantics needed to organize, discover, share, and integrate NMFS information assets. *Lexicon* refers to the way the agency uses business terminology. For example, does everyone have the same understanding of “trips,” “landings,” and “weigh outs?”

Taxonomies establish a common framework for describing and organizing information assets. A taxonomy is typically a hierarchical organization. For example, an automobile may be described with a taxonomy of Year/Make/Model/Style. Often, multiple taxonomies are used to describe information assets by using the vocabularies of particular customer or stakeholder groups. These stakeholder groups include internal users such as IEA scientists and stock assessment biologists, and external users such as Congress, fishermen, IOOS customers, and children. Taxonomies must be developed that support these varying needs and perspectives.

### Implementation Approach

We will develop taxonomies for all NMFS data, information, and services, thereby making them more discoverable by both internal and external constituents. The taxonomies will be dynamically built, and will employ common terminology. They will be constructed to enable multiple ways of accessing the same data depending on the inquirer’s starting point of reference. Once the common lexicon and taxonomies are established, they will be incorporated as part of the standard metadata for all information assets.

### Value and Benefits

Building a common lexicon and standard taxonomies will provide the following value and benefits:

- NMFS constituents all use a common lexicon of business terms, enhancing understanding of the meaning of data
- Information searches are simplified through use of standardized taxonomies that facilitate navigation to desired information resources

### Roadblocks Resolution

Roadblocks to data access are an issue across NOAA. Internal and external customers frequently encounter impediments when trying to access or use NMFS data. To manage information

collaboratively, NMFS EDM will try to identify and manage roadblocks to data access and recommend solutions, leveraging current efforts surrounding data access rules and regulations.

### Implementation Approach

Dealing with roadblocks or other impediments to data access will involve the following activities:

- Assess the nature and severity of roadblocks (real or imagined) originating from the following sources:
  - Federal, state, and local statutes
  - Confidentiality issues
  - Data sensitivity issues
  - Personally identifiable information (PII)
  - Other access issues
- Once the roadblocks are known, devise a process to resolve or mitigate the roadblock
  - Identify access restrictions for data assets in the inventory
  - Capture any restrictions or limitations on data access or use in the metadata
  - Develop and propose recommendations to leadership on how to resolve individual roadblocks or access concerns

### Value and Benefits

Finding appropriate solutions to data access roadblocks will provide the following value and benefits:

- Access restrictions are lifted where possible to allow routine access by authorized users
- Restrictions or limitations on data access or use are known and documented through the information asset's metadata

### Service-Oriented Architecture (SOA)

Service-oriented architecture is a system's architectural style for creating and using business processes, packaged as services, throughout their lifecycle. The SOA defines and provisions the IT infrastructure to allow different applications to exchange data and participate in business processes. It will be necessary to educate participants in NMFS EDM on SOA and assess the need and applicability within the agency.

### Implementation Approach

The following steps will be taken to assess the applicability of SOA for NMFS information management:

- Provide education and training on SOA to the FIMC
  - Use Webinars or other means to provide training
  - Establish community of practice Wikis to exchange information on SOA
- Keep abreast of other ongoing initiatives, such as IOOS, NOAA DMIT, and other federal SOA projects
- Develop recommendations on SOA applicability for NMFS to the Leadership Council

## Value and Benefits

Assessing SOA applicability will provide the following value and benefits:

- Better understanding of the role SOA can play within the agency
- Leverage SOA directions and progress made by related initiatives

## Interfaces with Other Systems

NMFS has an ongoing need to coordinate and exchange information with related organizations. Coordination involves interfacing with outside systems including those at NOAA, other federal, state, and local governments, and international organizations.

### Implementation Approach

It is recommended that the following approach be taken to manage interfaces with other systems:

- Provide central coordination
  - Establish a single point of contact for outside systems that develop links to our information resources
  - Handle interface requests like information requests by logging and managing them
  - Keep abreast of NOAA, DOC, federal, and other efforts developing services, standards, and SOAs (e.g., NOAA GEO-IDE, IOOS, EPA, and DHS)
- Establish best practices related to APIs, services, and registry
- Eventually develop APIs and services to support external interfaces; develop a registry of NMFS services (SOA)
- Address issues related to information integration including data heterogeneity, data standards, and conversion factors

## Value and Benefits

Central management of interfaces with other systems will provide the following value and benefits:

- Interface requests are properly coordinated to ensure compliance with NMFS security and confidentiality policies
- Data standards and metadata are put in place to ensure proper understanding and use of NMFS information resources

## 5 Program Implementation Alternatives

The following sections describe alternatives for implementing the NMFS EDM program. The alternatives are intended to serve as a menu of options, describing what could be accomplished given various levels of investment over the next few years.

Sections 5.1 through 5.3 describe the minimal, intermediate, and full cost options for EDM, respectively. Section 5.4 presents a recommended alternative for EDM implementation during FY09.

Each implementation option is organized by components and sub-components (see Section 4 for details on these elements). The right-hand columns depict the activities to be executed over the next few fiscal years. The box to the right presents the color code used to indicate when activities are intended to begin and end. Following each table is an approximation of the annual staff and budget resources needed to support the depicted level of activity. Note that funds quoted do not include the costs associated with existing FTE resources or the increased use of existing staff. Therefore, costs for increasing the percentages of the RIMCs and the OIMCs efforts dedicated to EDP are to be borne by the FMCs and HQ Offices.

<b><u>Color Code</u></b>
<b>Not Started</b>
<b>Started</b>
<b>Partially Done</b>
<b>Done</b>

## 5.1 Minimal Cost Option

The following table presents the schedule of activities under the minimal cost implementation option:

	FY08	FY09	FY10	FY11	FY12+
<b>Component 1. Governance Structure</b>					
Org Structure					
Establish data stewards					
Establish Stewardship Policies					
Establish Data Element Coding standards					
Communication and Technology Sharing					
Budgeting Strategy					
Cultural Shift					
Information Request Management					
Data Call Management					
<b>Component 2. Trusted Data</b>					
<i>Establish Data Quality Plans</i>					
<i>Information Sufficiency and Necessity</i>					
<i>Security and Confidentiality policies and Procedures</i>					
<b>Component 3. Authoritative Data Catalog</b>					
<i>Phase 1 Inventory and Metadata collection</i>					
<i>Phase 2 Inventory and Metadata collection</i>					
<i>Phase 3 Inventory and Metadata collection</i>					
<i>Phase 4 Inventory and Metadata collection</i>					
<b>Component 4. NMFS Information Depot</b>					
Lexicons and Taxonomies					
Roadblocks Resolution					
Depot Portals					
Discovery Engineering					
Interfaces with Other Systems					
SOA					

Implementing the minimal cost option will require the following resources:

- **FTEs**
  - S&T: The Information Architect and administrative support (current level)
  - FIMC Representatives: 10% or less of an FTE per FMC (current level)
- **Funds**
  - S&T: \$300K (current level)

Hidden costs include up to 10% of FIMC members' FTE costs and miscellaneous overhead expenses.

## 5.2 Intermediate Cost Option

The following table presents the schedule of activities under the intermediate cost implementation option:

	FY08	FY09	FY10	FY11	FY12+
<b>Component 1. Governance Structure</b>					
Org Structure					
Establish data stewards					
Establish Stewardship Policies					
Establish Data Element Coding standards					
Communication and Technology Sharing					
Budgeting Strategy					
Cultural Shift					
Information Request Management					
Data Call Management					
<b>Component 2. Trusted Data</b>					
Establish Data Quality Plans					
Information Sufficiency and Necessity					
Security and Confidentiality policies and Procedures					
<b>Component 3. Authoritative Data Catalog</b>					
Phase 1 Inventory and Metadata collection					
Phase 2 Inventory and Metadata collection					
Phase 3 Inventory and Metadata collection					
Phase 4 Inventory and Metadata collection					
<b>Component 4. NMFS Information Depot</b>					
Lexicons and Taxonomies					
Roadblocks Resolution					
Depot Portals					
Discovery Engineering					
Interfaces with Other Systems					
SOA					
Information Integration					

Implementing the intermediate cost option will require the following resources:

- **FTEs**
  - S&T: The Information Architect, data librarian (ZP-4), and administrative support
  - FIMC Representatives: 40% or less of an FTE per FMC
- **Funds**
  - S&T: \$300K
  - Other sources: \$700K

Hidden costs include up to 40% of FIMC members' FTE costs and miscellaneous overhead expenses.



### 5.3 Full Cost Option

The following table presents the schedule of activities under the full cost implementation option:

	FY08	FY09	FY10	FY11	FY12+
<b>Component 1. Governance Structure</b>					
Org Structure					
Establish data stewards					
Establish Stewardship Policies					
Establish Data Element Coding standards					
Communication and Technology Sharing					
Budgeting Strategy					
Cultural Shift					
Information Request Management					
Data Call Management					
<b>Component 2. Trusted Data</b>					
Establish Data Quality Plans					
Information Sufficiency and Necessity					
Security and Confidentiality policies and Procedures					
<b>Component 3. Authoritative Data Catalog</b>					
Phase 1 Inventory and Metadata collection					
Phase 2 Inventory and Metadata collection					
Phase 3 Inventory and Metadata collection					
Phase 4 Inventory and Metadata collection					
<b>Component 4. NMFS Information Depot</b>					
Lexicons and Taxonomies					
Roadblocks Resolution					
Depot Portals					
Discovery Engineering					
Interfaces with Other Systems					
SOA					
Information Integration					

Implementing the full cost option will require the following resources:

- **FTEs**
  - S&T: The Information Architect, data librarian (ZP-4), data technician (ZP-3), and administrative support
  - FIMC Representatives: 80% or less of an FTE per FMC
- **Funds**
  - S&T: \$300K
  - Other sources: \$2.7M

Hidden costs include up to 80% of FIMC members' FTE costs and miscellaneous overhead expenses.

### 5.4 Recommended Alternative

Given the urgency of the data management concerns, it is critical to start executing the full EDM program as soon as possible. However, given limited funding, especially in the early stages of the program, we recommend that a combination of the minimal and intermediate cost options be selected for FY 2009. This option offers a reasonable compromise, providing for significant progress against key activities at a relatively modest cost.

In FY 2009, the following activities would be executed under this option:

- **Establish the EDM governance structure, including:**
  - Adoption of the FIMC as an advisory committee to the Leadership Council
  - Establishment of the roles of the Regional Information Management Coordinators (RIMCs)
  - Establishment of the roles and responsibilities of data stewards for all enterprise information assets
  - Establishment of collaboration tools and methodologies to effectively communicate EDM activities and technologies
  - Establishment of EDM policies and best practices
    - Draft policies for basic data stewardship, including data inventory and metadata maintenance
    - Draft a NMFS glossary of key terminology to define and document terms used in NMFS business processes (e.g., IEAs) and their associated information (e.g., landings)

- **Build the Data Catalog**
  - Use the InPort tool as the authoritative data catalog and conduct **Phase 2** of the inventory, capturing discovery level metadata for all enterprise information assets. The table to the right depicts the phases of metadata development.

	Metadata Level		
	Discovery	Detailed	Element
Large Assets/ Often shared	1	2	3
Medium Assets/ Occas. Shared	2	3	4
Small Assets/ Seldom Shared	3	4	4

- **Create the NMFS Information Depot**
  - Initiate development of a NMFS information taxonomy and our corporate understanding of SOA and other emerging technologies.
- **Develop Performance Metrics**
  - Establish performance measures for the above tasks and keep the Leadership Council apprised of progress with quarterly status reports.

Implementing the recommended alternative will require the following resources:

- **FTEs**
  - S&T: The Information Architect and administrative support (current level)
  - FIMC Representatives: 25% or less of an FTE per FMC
- **Funds**
  - S&T: \$300K (current level)
  - Other sources: \$500K

Implementing this alternative will present NMFS with the following benefits and risks:

- **Benefits**
  - The basic governance structure and procedures for managing the program are in place
  - The authoritative data catalog of NMFS information assets is started
  - All NMFS information is categorized in a draft taxonomy

- The cultural shift regarding stewardship of NMFS corporate information resources is started
- **Risks**
  - The data catalog is not fully populated, so identification of data stewards can only be done for the portion of enterprise data that has been inventoried
  - Key activities will need to be deferred, including:
    - Information request management
    - Data call management
    - Information Depot development
    - Development of alternative constituent-based taxonomies
    - Completion of a confidentiality policy
    - Communications and technology sharing mechanisms

## ***5.5 Continuing with Status Quo***

Having studied the current NMFS EDM environment and analyzed concerns and priorities of NMFS management, the FIMC believes that NMFS IM challenges may put NMFS' reputation as a leader in science and environmental information at increasing risk. In the Internet culture, data and information are thought of as commodities that should be easily discovered and accessed with ever-increasing speed. We should start now to lay the groundwork to understand our data assets from an enterprise-wide perspective. We need to develop transparent data stewardship policies and establish tools to facilitate the public use of our data with an information portal that leverages inventories, data dictionaries, data standards, glossaries, thesauri, and taxonomies. The longer we wait to move beyond the status quo, the more difficult and expensive it will be to catch up.

Unless we start investing in EDM now, NMFS will be increasingly at risk of:

- Being absent or misrepresented in portals for GEOSS, IOOS, Census of Marine Life, and NCDDC and others;
- Continuing to have inconsistent web portals across the Regional Offices and Science Centers with different user interfaces and varying degrees of data sharing;
- Inadvertently sharing inconsistent data, which may reflect poorly on NMFS science and diminish trust in the agency;
- Continuing to make inconsistent and poorly structured data calls to the regions—a continuing source of frustration to Science Center and Regional Office personnel;
- Sharing data and information erratically, with little or no monitoring and without consistent business rules;
- Having no data policies and few data standards and data quality metrics;
- Having no consistent way of tracking and learning from users' satisfaction with NMFS information resources;
- Frustrating constituents with the inability to quickly access the information they need;
- Being accused of hoarding data and being perceived as obstructing NOAA efforts to archive our data;

- Not being much farther ahead than when Dr. Knauss said, 25 years ago, that NOAA information is like a library with no card catalog or bookshelves.

## 6 Other Recommendations

During the development of the discussed recommendations, a number of other opportunities for consolidation or reprogramming in the context of enterprise data management came to light. This section discusses them and makes recommendations to be considered by NMFS management

### 6.1 Paperwork Reduction Act (PRA) Business Process

**Recommendation:** Consider PRA processing an integral part of the EDM business process, but leave it in the OCIO for the time being.

**Rationale:** A significant portion of NMFS information is collected from the public under the provisions of the PRA. Since NMFS EDM includes the full lifecycle of information from collection to long-term preservation, the PRA approval process, which is currently managed by the OCIO, should be considered part of the EDM.

**Background:** The primary purpose of the PRA is to:

- Minimize the paperwork burden for individuals, small businesses, educational and nonprofit institutions, federal contractors, state, local, and tribal governments, and other persons resulting from the collection of information by or for the Federal Government
- Ensure the greatest possible public benefit from and maximize the utility of information created, collected, maintained, used, shared, and disseminated by or for the Federal Government
- Improve the quality and use of federal information to strengthen decision-making, accountability, and openness in government and society

NMFS currently has hundreds of data collection forms, including commercial and recreational landings data and dealer reports, socio-economic surveys, and reports on marine mammal interactions. Aside from resource surveys, information collected under PRA represents the largest set of NMFS data.

Approval must be obtained from OMB for each data collection system. The NMFS process for obtaining this approval is highly structured and is carried out currently by the NMFS CIO—a business practice that was developed in 2001 when it was excised from F/MB.

### 6.2 Administrative Systems

**Recommendation:** Have the NMFS CIO develop an enterprise-wide portal to all of the administrative systems NMFS managers are required to use, that is organized by administrative function and contains password reminders.

**Rationale:** During interviews conducted by the FIMC, one of the highest priorities cited by NMFS management was to address the plethora of administrative systems that are not connected and don't work well together.

**Background:** During the initial 15-minute discussion with NMFS top management, one senior official noted that we are confronted with a unwieldy variety of administrative applications from NMFS, NOAA, DOC, and OMB that overlap, seldom work together, have bugs, and often don't meet NMFS' needs. It's often difficult to determine what applications should be used when, and it's difficult to keep track of passwords. Also, there may be potential conflicts and redundancy of information in multiple systems. He suggested developing an enterprise-wide portal.

## 7 Risks and Constraints

In this section we explore the risks and constraints associated with establishing and administering the EDM Program. Implementing a new program, like any business or organizational change, presents challenges. However, proper planning, including effective monitoring of progress and a corrective action process, can minimize the impact of potential problems and keep the implementation process moving efficiently toward its intended results.

The following describes possible areas of risk or constraint and presents a mitigation strategy for addressing them:

- **Buy-in from management** – To be successful, the NMFS EDM needs the support and buy-in from management, both at HQ and in the Regions.

*Mitigation:* To achieve management buy-in and ongoing support, plans will be vetted and progress will be reviewed with management stakeholders on a continuing basis. In addition, a communications plan will be established for the Program describing the mechanisms for keeping managers and other stakeholders informed.

- **Regional support** – Similar to above, the Program needs to have the support of regional leadership and data stewards to be successful.

*Mitigation:* The RIMCs, supported by a communications plan, will play a key role in ensuring buy-in by the regions through regular communications and vetting with regional stakeholders.

- **Regional concerns regarding centralization** – There is a risk that the regions may not fully cooperate if they view the Program as a HQ initiative that doesn't provide value to the Regions.

*Mitigation:* The RIMCs will play a key role by acting as liaison between the Regions and HQ. Part of their responsibility will be to ensure that the needs of individual Regions are being addressed throughout Program implementation.

- **Funding / staffing** – Lack of adequate funding and staffing levels could severely hinder the ability of the Program to achieve its objectives. At the very least, timelines to complete certain activities may become unacceptably extended.

*Mitigation:* The Program will ensure that management is kept informed of the trade-offs at various levels of funding and staffing, so they can make informed decisions regarding program implementation. In addition, regular progress reporting will convey the accomplishments and benefits being derived from investments in the Program.

- **PII issues** – As NMFS data become more exposed and accessible, there is a risk that PII or other confidential data may be used or disseminated through unauthorized channels.

*Mitigation:* As part of Program implementation, policies and procedures for the management and dissemination of data need to be developed. This includes establishing policies and procedures for the security and handling of PII and confidential data.

- **Cultural Shift** – The EDM Program will present new ways of doing things and present new tools for users that represent a significant cultural shift from current practices. There is a risk that people will resist this change if the benefit to them is not immediately obvious.

*Mitigation:* Culture change challenges can be successfully minimized by effective planning and orientation. The implementation plan and the EDM tools and services must be designed to ensure successful application—one cannot make a good first impression the second time. Therefore, it is important to choose a scale and setting for initial pilots that will lead to visible accomplishment. In order to be adopted, tools and services must provide benefits such as saving time, making people’s jobs easier, and making their tasks more rewarding. It is crucial to take the time to design in advance the robust management systems that, when coupled with the proper level of automation, deliver positive and lasting results.

## Appendix A – Fisheries Information Management Committee Roster

F/NER-Martin Colby, Anthony Conigliari  
F/NEC-Joan Palmer

F/SER-John Reed  
F/SEC-Brion Cook

F/SWR-Jerry Hornof  
F/SWC-

F/NWR-April Wolstencroft  
F/NWC-Richard Kang

F/AKR-Corinne Brown  
F/AKC-Mike Brown

F/PIR-Eric Forney  
F/PIC-Brent Miyamoto

F/CIO-Susan Molina  
F/PR-Angela Collins-Payne, Laura Gutierrez  
F/HC-Joan Moubleaux  
F/SF-Regina Spallone  
F/ST-Jihong Dai, Tina Chang, Jim Sargent, Chair

### Subject Matter Experts

Karen Sender-F/PIC (Metadata)  
Larry Talley-F/AKR (Information Management)  
John Witzig-F/ST (FIS)  
Ajith Abraham-F/CIO (Applications Development)



# Appendix B – Data Management Issues Questionnaire

## NMFS Information Management Issues Interview Form

### Overview

Data and information are the critical ingredients for all NMFS outputs. Some internal and external constituents, over the years, have observed that NMFS faces challenges with its data and data management programs that potentially impact our ability to meet our current and future mission requirements. The Office of Science and Technology wants to determine if we do, indeed, have problems, and if so, what they are so we can take steps to address them.

The purpose of these interviews is to help the Fisheries Data Management Committee (FDMC) gain an understanding of the information management challenges facing NMFS today. In response to challenges identified, we plan to brief the Leadership Council on proposed solutions. If there are no critically perceived problems, then we have completed our mission and can invest our resources elsewhere. If, however, as has often been observed by internal and external constituents and studies, there are significant problems, then we need to do something about them.

As follow-up to the interview, you will have the opportunity to review the notes on your responses before proceeding further. Presentations to the Leadership Council will be summarized and will not disclose individual responses without approval.

The FDMC has identified the following issues regarding data and information management. Please rate each of them, on your perception on the priority for NMFS to address, on a scale of 0 - 3, where:

- 0 = No priority, This is not an issue;
- 1 = Low Priority;
- 2 = Moderate Priority; or
- 3 = High Priority.

### Guidance

**Scope:** Consider all data and information collected, developed or obtained by NMFS including environmental, administrative, legal, socio-economic, enforcement, etc.

**Data vs. information:** In the following, consider "data" to mean "data and/or information"

**Context:** All issues are to be considered in the NMFS-wide context. Please provide scoring from this perspective. If the priority of address an issue would be different within your Office or Center, please note it in the comments.

**Please Note:** The purpose of this interview form is to aid FDMC members in discussing the following issues with their management and to record the results of those discussions. It is not intended to be a form filled in by interviews or to be a statistically reliable survey. Also feel free to add any additional issue you see fit.

## **The Issues**

### **1. We don't know all the data we have.**

**Rating:** \_\_\_\_\_

Discussion: Over the past 30+ years, NMFS has collected a vast array of data that may be useful in answering questions of today and the future. Are you confident that we are aware of all these data across the enterprise and/or know where to go to determine what data we have?

Comments:

### **2. We don't know enough about the data we have to use it effectively and creatively.**

**Rating:** \_\_\_\_\_

Discussion: In order to effectively use data and relate them to other data sets, we need to have adequate information about the data (metadata), such as field names; coding structures and formats; and how, when, and why the data were collected. Do we have this information or do we need to collect and document metadata to use our data effectively?

Comments:

### **3. We can't integrate our data to determine effects of climate change on living marine resources.**

**Rating:** \_\_\_\_\_

Discussion: In some cases, data sets that we would like to relate, such as research vessel trawl survey data with fisheries catch data and habitat data with marine mammal sightings, can not be integrated to meet the needs of integrated ecosystems assessments. Are we able to integrate our data well enough to meet our needs?

Comments:

### **4. Data delivery is not timely enough to implement new management strategies.**

**Rating:** \_\_\_\_\_

Discussion: New management strategies such as ITQs and DAPs require timely data. With the advent of web technologies, the demand for timely data has increased. For example, information about marine mammal events are now expected near real time. Are critical data being audited, analyzed and delivered to managers and scientists timely enough to meet our current and future mission requirements?

Comments:

### **5. Our data has critical gaps that need filling.**

**Rating:** \_\_\_\_\_

Discussion: Often NMFS scientists have said that we need more and better data to confidently predict outcomes. As we move toward IEA, do we need to expand our data collection programs?

Comments:

### **6. Data are being lost without our knowing about it.**

**Rating:** \_\_\_\_\_

Discussion: Observations have been made that specific data sets and models on which some NMFS scientific publications were based on were not reserved and were eventually lost. Do you perceive potential loss of data as problem that needs to be addressed?

Comments:

**7. Data quality and versioning problems.**

**Rating:**

\_\_\_\_\_

Discussion: Several studies and reports have indicated that NMFS could improve the quality of its data. Also, constituents receiving inconsistent data from two different sources is a challenge that can be traced to not keeping track of different versions of data sets.

Comments:

**8. Data not being archived for perpetuity (as per NARA requirements).**

**Rating:**

\_\_\_\_\_

Discussion: We obviously need to preserve our data for use in the future. NARA regulations require that all federal records be preserved in archives conforming to their standards. The only archives in NOAA that adhere to these standards are NODC, NCDC, and NGDC. Is it a concern that, because we have no archive policies and standards, that our data is at risk?

Comments:

**9. Historical data needs to be rescued to provide a context for the future.**

**Rating:**

\_\_\_\_\_

Discussion: Are there data that may be on perishable media that need to be rescued and put in digital formats for our use?

Comments:

**10. Plethora of administrative systems are not connected and don't work well.**

**Rating:**

\_\_\_\_\_

Discussion: One NMFS senior official noted that we are confronted with a plethora of administrative applications from NMFS, NOAA, DOC and OMB that overlap, seldom work together, have bugs, and often don't meet NMFS needs. It's often difficult to determine what applications should be used when and it's difficult to keep track of passwords. Also, there may be potential conflicts and redundancy of information in multiple systems.

Comments:

**11. NMFS needs to improve communications regarding the development and use of its applications and technologies.**

**Rating:**

\_\_\_\_\_

Discussion: NMFS develops and deploys new applications and technologies, such as electronic reporting of fisheries statistics, but often in a stove-piped manner. Do we need to share IT development efforts, successes, and challenges among ourselves more?

Comments:

**12. We don't have buy in for addressing information management across FMCs.**

**Rating:**

\_\_\_\_\_

Discussion: Initiatives that attempt to address information management across FMCs (such as FIS) encounter resistance. Sharing data beyond the jurisdiction where it was collected, and for different purposes, involves risk, with little perceived benefit to the originating FMC. Is the lack of support for establishing NMFS-wide policies and initiatives to address some of these issues a challenge?

Comments:

**13. Are there additional issues? Please list your top three information management related issues not found on this list.**

Discussion: Are there other information management related issues that NMFS should be addressing?\_\_

Comments:

**ACRONYMS**

DAP - Dedicated Access Privilege

FMC - Financial Management Center; HQ and each Regional Office and Science Center is a unique FMC

IEA - Integrated Ecosystem Assessment

ITQ - Individual Transferable Quota.

NARA- National Archives and Records Administration

NCDC -National Climatic Data Center

NGDC - National Geophysical Data Center

NODC - National Oceanographic Data Center

OMB - Office of Management and Budget

## Appendix C – Objectives and Priorities

During the information management workshop conducted July 8 – 10, 2008, the FIMC identified objectives for each EDM goal area (see Section 3) and established relative priorities as to the importance of each objective. The following table presents the objectives by goal area, indicating the desired time frame for completion (short, mid, or long-term) as well as the priority level (low, medium, or high). These timeframes and priorities were used to help inform the analysis that went into development of the program implementation alternatives presented in Section 5.

<b><i>Goal 1. Manage information with appropriate governance, focus and accountability</i></b>		
<b>Objectives</b>	<b>Time Frame</b>	<b>Priority</b>
Establish role of information coordinator in each FMC to work with local data managers and act as point of focus for IM in the FMC	Short-Term	Medium
Establish a governance structure for IM (project/portfolio mgmt)	Short-Term	Medium
Develop elements of performance plans for information managers and data stewards within the agency	Mid-Term	Low
Establish an information architect position	Short-Term	N/A
Achieve buy-in and secure commitment from management	Short-Term	Medium
Develop mechanisms for communicating progress, success, value, and benefits	Short-Term	Low
Develop recommendations for funding and budget initiatives needed to support IM efforts	Short-Term	Medium
<b><i>Goal 2. Manage information collaboratively to promote sharing of information, solutions and IM investments</i></b>		
<b>Objectives</b>	<b>Time Frame</b>	<b>Priority</b>
Establish mechanisms for ongoing communication and coordination with FMCs and related programs (e.g., FIS)	Short-Term	Low
Create mechanism for collaboration to allow exchange of information on common needs and services	Mid-Term	Low
Develop a service oriented architecture, enterprise service bus	Long-Term	Low

Identify and manage roadblocks to data access and recommend solutions to management and data stewards	Short-Term	High
Define information terms	Short-Term	Low
<b><i>Goal 3. Establish and maintain policies, best practices, procedures and metrics to ensure good stewardship and preservation of NMFS information resources</i></b>		
<b>Objectives</b>	<b>Time Frame</b>	<b>Priority</b>
Develop metadata policy and standards for data access and use	Short-Term	High
Identify, assess, and prioritize shareable information assets	Short-Term	High
Publish and maintain metadata on all shareable information assets	Mid-Term	Low
Establish quality plans for all data collections including minimum elements, formats, and values	Mid-Term	Low
Define policy and standards related to stewardship of data through the life-cycle including metadata, security, access, preservation	Short-Term	High
Define concept of operations for NMFS IM including high-level view of functions, data subjects, tools, and technology	Short-Term	Low
<b><i>Goal 4. Empower internal and external constituents to effectively discover, access, integrate and use the information to answer key current and future questions</i></b>		
<b>Objectives</b>	<b>Time Frame</b>	<b>Priority</b>
Develop taxonomy for all NMFS data, information, and services	Mid-Term	Low
Educate ourselves on SOA and assess the need and applicability of SOA within the agency	Short-Term	Low
Develop policies, guidelines, MOUs, etc. needed to facilitate access to data	Mid-Term	Low
Define business glossary for NMFS information assets	Short-Term	Low
Define and implement a "front door" to enable access to NMFS information access by both internal and external constituents	Mid-Term	Low

***Goal 5. Continually assess and improve the availability, timeliness and quality of data resources to support NMFS mission***

<b>Objectives</b>	<b>Time Frame</b>	<b>Priority</b>
Provide training, education, and outreach to data stewards, managers, and users of NMFS data	Mid-Term	Low
Establish mechanisms to receive feedback from users on satisfaction with NMFS information and services	Mid-Term	Low
Establish service level and metrics related to delivery of NMFS products and services. Enable periodic review of service level	Mid-Term	Low
Establish independent review of compliance with policy and standards related to data stewardship and evolve as needed	Mid-Term	Low

# Appendix D – NMFS EDM Components Cross-Reference Tables

Table D.1. EDM Components x IM Issues

Enterprise Data Management Components	Information Management Issues										
	Critical Data Gaps	We don't know all the data we have enough about the data	Admin systems don't work well	Data Quality and versioning problems	We can't integrate our data	Communications re apps and IT	Timeliness of data delivery	Data not being archived	IM issues across FMCS	Historical data needs to be rescued	Data being lost w/o our knowing about it
<b>Governance Structure</b>											
Governance Model	Red	Red	Red	Red	Red	Yellow	Yellow	Red	Red	Red	Red
Stewardship	Red	Red	Red	Red	Red	Yellow	Yellow	Red	Red	Red	Red
Budgeting Strategy	Red	Red	Red	Red	Red	Yellow	Yellow	Red	Yellow	Red	Red
Communication and Technology Sharing						Red			Yellow		
Cultural Shift	Yellow	Yellow		Yellow	Yellow			Orange	Orange		Orange
Inf. Requests & Data Calls				Orange							
<b>Authoritative Data Catalog</b>											
Data Inventory	Red	Red	Red		Red			Orange		Yellow	Yellow
Metadata	Red	Red	Red		Red			Orange		Yellow	Yellow
<b>Trusted Data</b>											
Analysis of data sufficiency and necessity	Red										
Security and Confidentiality				Yellow							
<b>NMFS Information Front Door</b>											
The Front Door				Yellow			Yellow				
Discovery Engineering											
Information Integration					Red						
Lexicon and Taxonomies	Yellow										
Roadblocks Resolution											
Interfaces with Other Systems											
SOA					Orange						

Table D.2. EDM Components x IM Issues

Enterprise Data Management Components	Enterprise Data Management Goals				
	Goal 1. Manage information with appropriate governance, focus and accountability	Goal 2. Manage information cooperatively to promote sharing of information, solutions and information management investments	Goal 3. Establish and maintain policies, best practices, procedures and metrics to ensure good stewardship and preservation of NMFS information resources	Goal 4. Empower internal and external constituents to effectively discover, access, integrate and use the information to answer key current and future questions	Goal 5. Continually assess and improve the availability, timeliness and quality of data resources to support NMFS mission
<b>Governance Structure</b>					
Governance Model	Red	Red	Red	Red	Red
Stewardship	Red	Red	Red	Red	Red
Budgeting Strategy	Red	Red	Red	Red	Red
Communication and Technology Sharing			Yellow	Yellow	Yellow
Cultural Shift	Yellow	Yellow			
Inf. Requests & Data Calls	Orange		Yellow	Orange	
<b>Authoritative Data Catalog</b>					
Data Inventory	Red	Red	Red	Red	Red
Metadata	Red	Red	Red	Red	Red
<b>Trusted Data</b>					
Analysis of data sufficiency and necessity				Orange	Red
Security and Confidentiality	Red	Red	Red	Orange	
<b>NMFS Information Front Door</b>					
The Front Door		Yellow		Red	
Discovery Engineering				Red	
Information Integration	Orange	Orange		Red	
Lexicon and Taxonomies	Orange	Orange		Red	
Roadblocks Resolution				Red	
Interfaces with Other Systems	Yellow	Yellow		Red	
SOA	Orange	Orange		Red	



## Appendix E – Glossary of Terms

This appendix provides definitions for terminology used in this document

**Data** - A value, or set of values, representing a specific concept or concepts. Data becomes "information" when analyzed and possibly combined with other data in order to extract meaning, and to provide context. The meaning of data can vary according to its context. (DRM usage). Note: In the NMFS EDM program, the terms “data” and “information” are treated and used synonymously, principally because whether a data has meaning is in the mind of the beholder.

**Data asset** – A managed container for data; examples include a relational database, Web site, document repository, directory or data service (FEA DRM)

**Data steward** - The individual who is responsible for establishing and maintaining the quality, integrity, documentation, and preservation of the data asset

**Enterprise Data (Information)** - Data and information that is:

- Already shared with some entity outside of the FMC of origin (for example, observer data that is already shared between a science center and a region);
- About entities or issues that cross FMC boundaries (for example, data about vessels that fish in multiple regions); or
- Routinely aggregated to support decisions at a higher organizational level (i.e., data that is rolled-up to support national decisions, such as catch and value data that rolls-up to Fisheries of the United States).

**Information Management (IM)** - IM is the management of the information that resides on the IT Infrastructure, and addresses issues such as collection, data stewardship, IM tools and applications development, preservation, and confidentiality.

**Information Technology (IT)** - IT is the infrastructure and the applications running on the infrastructure, and addresses related issues such as operations, security, capital planning and investment control, and IT tools and applications development (i.e., what are primarily the responsibility of the NMFS CIO).

**Metadata** – An external description of a distinct data resource. Common usages for metadata include providing the context of the data resource, managing its lifecycle, and extending it to new uses. An example of metadata is the external description of an audio file specifying the artist that created it, when it was created, the length of playtime, and its genre of music. The purpose of metadata is to manage and improve the use of data and thereby turn it into a strategic asset. (Daconta) TBD

### Metadata levels

- *Discovery-level* metadata is defined as that which is required for customers to discover the data asset. It includes metadata such as data asset title, owner, data steward, description, and information category.

- *Detail-level* metadata is defined as enough information for customers to determine whether the data assets are what they want. It includes information such as time, location, data quality, suggestion for use, and formats.
  - *Element-level* metadata is defined as information required for the customer to use the data elements. It includes, for each data element, information such as element name, description, format, and coding structure.
- (NMFS EDP)

## Appendix F – Acronyms

AOOS	Alaska Ocean Observing System
API	Application Program Interface
CaRA	Caribbean Regional Association
CeNOOS	Central and Northern California Ocean Observing System
CLASS	Comprehensive Large Array Data Stewardship System
CXE	Customer Experience Evaluation
DMAC	Data Management and Communications
DMIT	Data Management Integration Team
DAP	Dedicated Access Privileges
DOC	Department of Commerce
EOS	Earth Observing Systems
EDMS	Electronic Document Management System
EDM	Enterprise Data Management
FMC	Financial Management Center
FIMAC	Fisheries Information Management Advisory Committee
FIMC	Fisheries Information Management Committee
GEO-IDE	Global Earth Observation Integrated Data Environment
GEOSS	Global Earth Observation System of Systems
GCOOS	Gulf of Mexico Coastal Ocean Observing System
ITQ	Individual Transferable Quota
IA	Information Architect
IT	Information Technology
IEA	Integrated Ecosystems Assessment
IOOS	Integrated Ocean Observation System
MSRA	Magnuson-Stevens Reauthorization Act
MACOORA	Mid Atlantic Coastal Ocean Observing Regional Association
NARA	National Archives Administration
NCDC	National Climate Data Center
NGDC	National Geophysical Data Center
NMFS	National Marine Fisheries Service
NOAA	National Oceanic Atmospheric Administration
NODC	National Oceanographic Data Center
FIS	NFMS Fisheries Information System
DMC	NOAA Data Management Committee
IPT	NOAA Integrated Products Team
NOSA	NOAA Observing Systems Architecture
GoMOOS	Northeast
ANOOS	Northwest Association of Networked Ocean Observing Systems

OBIS	Ocean Biological Information System
OIMC	Office Information Management Coordinators
OMB	Office of Management and Budget
ST	Office of Science and Technology
PacIOOS	Pacific Islands Ocean Observing System
PRA	Paperwork Reduction Act
PII	Personally Identifiable Information
RIMC	Regional Information Management Coordinators
RO	Regional Office
SC	Science Center
SOA	Service Oriented Architecture
SECOORA	Southeast Coastal Ocean Observing Regional Association
SCCOOS	Southern California Coastal Observing System
SMEs	Subject Matter Experts
DAARWG	The Data Archiving and Accessing Requirements Working Group
VMS	Vessel Management System