DFO National Science Data Management Strategy

R. Keeley[1], L. Barton[3], R. Eisner[5], J. Goodman[2], J. Holmes[3], S. Hurtubise[4], G. MacDonald[2], D. Nicholson[5], R. Nowlan[6], J. O'Neill[5], D. Senciall[7], T. Trivedi[8]

[1] Integrated Science Data Management Branch
Department of Fisheries and Oceans
S1202-200 Kent St.
Ottawa, ON, K1A 0E6

[2] Information Management and Technology Services
Department of Fisheries and Oceans
200 Kent St.
Ottawa, ON, K1A 0E6

[3] Regional Science Branch
Department of Fisheries and Oceans
3190 Hammond Bay Road
Nanaimo, BC, V9T 6N7

[4] Regional Science Branch
Department of Fisheries and Oceans
Institut Maurice-Lamontagne
850 route de la Mer, PO Box 1000
Mont-Joli, QC, G5H 3Z4

[5] Regional Science Branch
Department of Fisheries and Oceans
Bedford Institute of Oceanography
1 Challenger Drive PO Box 1006
Dartmouth, NS, B2Y 4A2

[6] Regional Science Branch
Department of Fisheries and Oceans
PO Box 5030
Moncton, NB, E1C 9B6

[7] Regional Science Branch
Department of Fisheries and Oceans
Northwest Atlantic Fisheries Centre
80 East White Hills Road PO Box 5667
St. John's, NL, A1C 5X1

[8] Regional Science Branch
Department of Fisheries and Oceans
Great Lakes Laboratory for Fisheries and Aquatic Sciences
867, Lakeshore Road
Burlington, ON, L7R 4A6

2006

Canadian Technical Report of Hydrography and Ocean Sciences 251

Fisheries and Oceans Pêches et Océans

Canada

Canadian Technical Report of
Hydrography and Ocean Sciences

Technical reports contain scientific and technical information that contribute to existing knowledge but that are not normally appropriate for primary literature. The subject matter is related generally to programs and interests of the Department of Fisheries and Oceans, namely, hydrography and ocean sciences.

Technical reports may be cited as full publications. The correct citation appears above the abstract of each report. Each report is indexed in the data base *Aquatic Sciences and Fisheries Abstracts*.

Technical reports are produced regionally but are numbered nationally. Requests for individual reports will be filled by the issuing establishment listed on the front cover and title page. Out-of-stock reports will be supplied for a fee by commercial agents.

Regional and headquarters establishments of Ocean Science and Surveys ceased publication of their various report series as of December 1981. A complete listing of these publications is published in the *Canadian Journal of Fisheries and Aquatic Sciences*, Volume 39: Index to Publications 1982. The current series, which begins with report number 1, was initiated in January 1982.

Rapport technique canadien sur
l'hydrographie et les sciences océaniques

Les rapports techniques contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles, mais qui ne sont pas normalement appropriés pour la publication dans un journal scientifique. Le sujet est généralement lié aux programmes et intérêts du ministère des Pêches et des Océans, c'est-à-dire l'hydrographie et les sciences océaniques.

Les rapports techniques peuvent être cités comme des publications intégrales. Le titre exact paraît au-dessus du résumé de chaque rapport. Les rapports techniques sont indexés dans la base de données *Aquatic Sciences and Fisheries Abstracts*.

Les rapports techniques sont produits à l'échelon régional, mais numérotés à l'échelon national. Les demandes de rapports seront satisfaites par l'établissement d'origine dont le nom figure sur la couverture et la page du titre. Les rapports épuisés seront fournis contre rétribution par des agents commerciaux.

Les établissements des Sciences et levés océaniques dans les régions et à l'administration centrale ont cessé de publier leurs diverses séries de rapports en décembre 1981. Une liste complète de ces publications figure dans le volume 39, Index des publications 1982 du *Journal canadien des sciences halieutiques et aquatiques*. La série actuelle a commencé avec la publication du rapport numéro 1 en janvier 1982.

Canadian Technical Report of Hydrography
And Ocean Sciences 251

2006

DFO NATIONAL SCIENCE DATA MANAGEMENT STRATEGY

R. Keeley[1], L. Barton[3], R. Eisner[5], J. Goodman[2], J. Holmes[3], S. Hurtubise[4], G. MacDonald[2], D. Nicholson[5], R. Nowlan[6], J. O'Neill[5], D. Senciall[7], T. Trivedi[8]

[1] Integrated Science Data Management Branch
Department of Fisheries and Oceans
S1202-200 Kent St.
Ottawa, ON, K1A 0E6
Dartmouth, NS, B2Y 4A2

[2] Information Management and Technology Services
Department of Fisheries and Oceans
200 Kent St.
Ottawa, ON, K1A 0E6

[3] Regional Science Branch
Department of Fisheries and Oceans
3190 Hammond Bay Road
Nanaimo, BC, V9T 6N7
St. John's, NL, A1C 5X1

[4] Regional Science Branch
Department of Fisheries and Oceans
Institut Maurice-Lamontagne
850 route de la Mer, PO Box 1000
Mont-Joli, QC, G5H 3Z4

[5] Regional Science Branch
Department of Fisheries and Oceans
Bedford Institute of Oceanography
1 Challenger Drive PO Box 1006

[6] Regional Science Branch
Department of Fisheries and Oceans
PO Box 5030
Moncton, NB, E1C 9B6

[7] Regional Science Branch
Department of Fisheries and Oceans
Northwest Atlantic Fisheries Centre
80 East White Hills Road PO Box 5667

[8] Regional Science Branch
Department of Fisheries and Oceans
Great Lakes Laboratory for Fisheries and Aquatic Sciences
867, Lakeshore Road
Burlington, ON, L7R 4A6

Cat. No. Fs 97-18/251E ISSN 0711-6764

Correct citation for this publication:

Keeley R., L. Barton, R. Eisner, J. Goodman, J. Holmes, S. Hurtubise, G. MacDonald, D. Nicholson, R. Nowlan, J. O'Neill, D. Senciall, and T. Trivedi. 2006. DFO National Science Data Management Strategy. Can. Tech. Rep. Hydrogr. Ocean Sci. 251: viii+62 p.

TABLE OF CONTENTS

LIST OF ANNEXES

ABSTRACT

Keeley R., L. Barton, R. Eisner, J. Goodman, J. Holmes, S. Hurtubise, G. MacDonald, D. Nicholson, R. Nowlan, J. O'Neill, D. Senciall, and T. Trivedi. 2006. DFO National Science Data Management Strategy. Can. Tech. Rep. Hydrogr. Ocean Sci. 251: viii+62 p.

Fisheries and Oceans Canada, through its own programs and through exchanges with national and international organisations, has acquired a large volume of scientific data and information over the years. Over the same time, the management systems for these data have grown up largely in an uncoordinated way. These historical data, augmented with on-going data collections, are an extremely valuable and irreplaceable resource for the Department. In 2001, Science Sector adopted a Data Policy, a high level statement of how data collected within Science will be managed. This Strategic Plan represents the next step towards implementing the DFO Science Policy for the management of scientific data. As such, it addresses the archiving of data, the access to data and information, the standards and their application to managing data and finally how the data management work plan should be organized. Turning the recommendations and actions noted in this strategy into concrete activities across Science will require the support of all Science staff and help from other Sectors, notably IMTS. As the detailed plans develop and are implemented, and as technology changes, there may be some changes in strategy. This plan needs to remain flexible in its implementation but of sufficient vision to give a solid target.

RÉSUMÉ

Keeley R., L. Barton, R. Eisner, J. Goodman, J. Holmes, S. Hurtubise, G. MacDonald, D. Nicholson, R. Nowlan, J. O'Neill, D. Senciall, et T. Trivedi. 2006. DFO National Science Data Management Strategy. Can. Tech. Rep. Hydrogr. Ocean Sci. 251: viii+62 p.

Par le biais de ses propres programmes et d’échanges avec des organisations nationales et internationales, Pêches et Océans Canada a acquis un grand volume de données scientifiques et d’informations au cours des ans qu’il gère à l’aide de procédures développées au fil des ans. Durant la même période, les systèmes pour gérer ces données ont pris beaucoup d'ampleur, mais de façon non coordonnée. Ces données historiques, et celles encore récoltées aujourd'hui, constituent une précieuse ressource irremplaçable pour le Ministère. En 2001, le secteur des Sciences a adopté une Politique des Données, un énoncé de haut niveau sur la façon dont les données récoltées par le secteur des Sciences seront gérées. Ce plan stratégique représente un pas vers la mise en œuvre de la Politique du Secteur des sciences du MPO sur la gestion des données scientifiques. À cet effet, il aborde les questions de l'archivage des données, de l'accès aux données et à l'information, des normes et de leur application pour gérer les données et finalement l'organisation du plan de travail pour la gestion de données. La transformation des recommandations et des mesures signalées dans la présente stratégie en activités concrètes dans le Secteur des sciences nécessitera l’appui de tout le personnel du Secteur et l’aide d’autres secteurs, notamment de la Direction générale de la gestion de l’information et des services de la technologie (GIST). Lorsque des plans détaillés seront élaborés et mis en œuvre et que la technologie évoluera, il est possible que des changements soient apportés à la stratégie. Il est essentiel de maintenir une certaine souplesse dans la mise en œuvre du présent plan mais il doit être animé par une vision suffisante pour lui donner un objectif solide.

EXECUTIVE SUMMARY

In 2001, Science Managers in DFO agreed to a national Data Policy (Annex 1). The Policy stated 5 principles.

1) Fisheries and Oceans Canada (DFO) scientific data sets are a valuable national resource that have been acquired through decades of investment, enabling the Department to maintain world leadership in aquatic sciences and aquatic management. These data are irreplaceable and must be properly managed to ensure long-term availability.
2) Because of the complex and often unique nature of scientific data, it is essential that DFO Science/Oceans maintain responsibility for their quality control, management, archiving and dissemination.
3) To ensure their proper management, all scientific data collected by the Department must be migrated to a 'managed' archive immediately after they have been processed.
4) To obtain maximum benefit to the Department and to the user community at large, scientific data must be made available in a timely manner with full and open access, consistent with any obligation with respect to DFO's data holdings.
5) To obtain access to international data and information that are pertinent to Canadian needs, Canada must be able to exchange its data with other world data centres, subject to the 'Exceptions' listed in the section 'Availability of Access' of the Policy.

This Strategic Plan represents a step towards implementing the DFO Science Policy for the management of scientific data. It is organized under four topics concerned with

- Archiving of data
- Access to data and information
- Standards and their application to managing data
- Governance (i.e. organization of the work plan)

Each section presents a discussion of the issues, states recommendations, and then proposes actions to meet the recommendations that will have implications on DFO staff at all levels For each of the topics, the highlights are as follows.

<u>Archives:</u>

- All Science data are acknowledged as an "extremely valuable and irreplaceable resource", they must be "managed as part of an integrated system accessible through regional, zonal and national data centres.
- All archives must faithfully maintain data over the long term and meet the test of accessibility to both the original providers of the data and to other users, regardless of their form (numerical or not).
- Guidance on the appropriate archive strategies for different kinds of data will be provided following the determination of the diversity and kinds of data as well as the present archives of Science data.
- Designating a primary archive for each kind of data that is being managed will assist in implementing the coordination required to manage a distributed archive system.
- Each region will provide adequate funding to manage project data properly and will designate a coordinator to review the project plans and work with project leaders to determine the level of

effort to move the collected data to the designated archives.

- The manager of the archive and the data collector will work together to move the data as quickly as possible to the archive.
- Procedures will be employed to manage the quality of data in the archives and to identify all near or exact duplications of archived data in newly provided data.
- Ways of standardizing use and sharing of software applications within Science will continually be explored.
- Data archives must be robust to changing attributes, to accommodating new variables and to supporting data versions, when appropriate.
- Mechanisms for the documentation of current contents and management of new samples and tracking information of non-numerical archives will be implemented.
- A list of data at risk will be re-established to set priorities on rescuing data according to determined criteria and depending on available resources.

Access:

- Detailed descriptions of the contents of primary archives, including the appropriate contact names, must be provided as part of readily accessible inventories.
- Every archive will provide on-line browse facilities to users, and employ common tools promoted by NSDMC.
- Each archive centre will provide flexible delivery mechanisms and evaluate the feasibility of placing its data on-line with immediate processing of requests.
- Archive centres will provide a common suite of output formats and ensure sufficient documentation comes with the data.
- Accessible and well documented products and performance metrics will be provided by archive centres and provision will be made to accept user feedback and suggestions.

Standards:

- The adoption of standards is the first step towards developing interoperability between the distributed components of Science's data system.
- The adoption of standards will require changes to existing technologies to implement their usage.
- The adoption of standards as close to data collection as possible will pay more dividends.
- Agreement will be sought on adoption of standard practices as broadly as possible.
- Before adopting any standard, an analysis of impacts will be made from the national perspective.

Governance:

- Science projects must demonstrate the existence of a comprehensive and costed data management plan that follows the guidance provided by this Strategy.
- NSDMC was formed by NSDC with representation from each region and other sectors to foster national cooperation on data management issues.
- There needs to be staff in every region who are available to work within this national structure and resources (~ 5 to 10 % of total budget) must be made available on an ongoing basis and independent of project funding to support its work.
- It is necessary to form a regional data management coordination group, chaired by the NSDMC representative, with the responsibility to ensure better coordination among the business lines

and to support the national approach to managing valuable data resources.

- The responsibilities and reporting lines will be clear and readily connected to the activities promoted by NSDMC by designating consolidated regional organizational units dedicated to data management.
- Data management proposals will be solicited from regional staff, in collaboration with their NSDMC member and colleagues across Science.
- Accountability mechanisms and performance indicators to measure improvements made in data management will be developed and implemented and the work and results of the NSDMC will be shared with Science staff by the building of a Web site containing relevant documents and links to NSDMC activities.

DFO Science acquires and manages a wide variety of data. There is much work to be done of varying degrees depending on the kind of data considered. By taking a broad approach focused on functions we are striving to identify and build on the commonalities that exist.

Turning the recommendations and actions noted in this strategy into concrete activities across Science will require the support of all Science staff and help from other Sectors, notably IMTS. As the detailed plans develop and are implemented, and as technology changes, there may be some changes in strategy. This plan needs to remain flexible in its implementation but of sufficient vision to give a solid target.

1. INTRODUCTION

In 2001, DFO Science Managers agreed to a Science Data Policy (Annex 1). The policy is a high level statement of how data collected within Science will be managed. A strong impetus for the creation of this policy is the wide recognition that observations made in the past are useful in a variety of ways. They can provide baseline conditions before natural or artificial disturbances changed the system. They provide the ability to look for trends or variability that helps to explain current or future conditions. They provide data that can be used to test hypotheses in different conditions.

This document describes the overall strategy that will guide the development of a national data and information management system for Science. In most of this document the word "data" is used to describe measurements made. However, to support the interpretation of the measurements, it is necessary to hold other information such as data origins, sampling methods, instrumentation, variable names, identifiers, etc. These are referred to as "metadata" in the text. In other areas, the document refers to "information". In these cases, the term is used to describe reports, documentation and similar items that contribute to turning data and metadata into knowledge.

Science holds a wide variety of data, some of which are not collected by Science staff. For pragmatic reasons, we sometimes copy data held or collected by others. But, data in archives are never static. They are processed and reprocessed, problems found and fixed, additional information added, etc. Generally, copying data is not the best choice of operations since any copy may quickly get out of date. Rather, it is better to negotiate acceptable access to these archives, and to query them when data are required.

Preliminary work on analyzing the state of data archives in Science was carried out by the National Data Management Working Group in 2002. Data were grouped into the broad categories of physical oceanography, fisheries, and environmental sciences. The first two reports were completed and are included as Annexes 2 and 3. The third report was not completed. These studies surveyed the data being collected by Science and evaluated how they were being brought together into well managed archives and made available to others. They were carried out at a high level (no information on specific data files) but still are very useful for indicating areas of weakness in the data management practices in Science.

In preparation for writing this strategy document, a series of reports were prepared that identify important concepts and issues that must be dealt with to succeed in building a national data management system for Science. These are highly detailed and were used as the basis for the more general discussions of this strategic plan. These scoping documents are continuing to evolve and will be made available as they mature.

This strategy partitions the key data management activities into seven categories. These are:

- Archives – includes all of the issues about acquiring the data from researchers, quality assurance, managing data versions, safe guarding from loss, etc.
- Non-numerical assets – addresses the issues surrounding assets (such as physical samples of fish maintained in freezers) that are not now or cannot in future be handled in

numerical ways (such as in data bases).

- Data rescue – includes identification and recovery of data considered to be at risk. This is closely related to Archives, but because of the aging of science staff and staff reductions, it is singled out for special attention.
- Access – includes the issues surrounding providing information to users about the data holdings of Science, how they are maintained and providing access to the data and information.
- Inventories – essentially a sub-component of Access, the adoption and development of a data cataloguing system is a specific recommendation of the national data policy and a key enabler for the other themes.
- Standards – includes the identification and adoption of existing standards, development of new standards where none exist and national implementation of these standards to improve efficiency.
- Governance – includes the issues relevant to coordination of national data management activities.

2. DATA SYSTEM OBJECTIVES

The broad objectives of the data management system for Science are:

- to safe guard those data collected by Science or hosted by Science on behalf of others.
- to provide easy access to these data for a broad user community.

Managing the data and information assets of Science is a significant endeavour that encompasses a wide range of activities. They include everything from data acquisition systems to generating products from existing archives. To achieve this, Science needs to develop a comprehensive view of the components and processes that support these objectives.

It is not possible to start to build a national data system from a “clean slate”. First, the resources to do this are not available. Second, there is not the organizational structure that would allow this. Third, because of the nature of scientific data acquisition and use, there will always be a need for systems to evolve. The strategy, therefore, advocates a pragmatic approach that will use common themes to knit together disparate views and programs, will promote adoption of regional initiatives on a national scale, will exploit the use of standards as possible and generally try to bring about a convergence of systems and approaches to achieve interoperability between existing systems.

The desire of Science to improve its data systems does not stand in isolation. There are similar initiatives in countries such as Australia (BLUElink), the United States (DMAC), and in the European Union (SeaDataNet). Each of these has similar goals, and experiences that Science can learn from. It is important for Science to be connected to these endeavours.

3. CONCEPT OF OPERATIONS

Data management does not and cannot operate in a silo. There must be a close working

relationship with researchers who plan data acquisition activities and the IT specialists responsible for the 24/7 operations of the network and hardware. The work should start in the project definition, and continue with researchers, data managers, and IT specialists contributing at the appropriate stages in the flow of data from acquisition, to archives, to users.

The project definition stage should be the first point of contact with the data system. It is here that the researcher/data collector decides such issues as the types of data contemplated for collection, and the supporting metadata that need to be collected. In combination with data managers, they decide the data structures to be used, where the data should go once acquired, and the time frames for all this to happen. Close cooperation at this stage between data collectors and data managers is essential in streamlining the data management operations and ultimately providing efficient processing and timely access to the data.

At data collection, ideally, all data of a single type will present to the data system in a single format or data structure. Even if this can be accomplished by Science, data arriving from external sources are unlikely to conform to Science Sector's internal standards. This cannot be ignored since external data sources provide significant quantities of data to DFO. It is crucial, therefore, for the data system to be capable of accommodating new input data structures and content.

The data transport and processing function must be robust and error free. It supports getting the data and metadata from the platform or location where the measurements are made to the location of the archive. Some of the data processing may be done before the data are presented to the archives and some is done at the archive centre. There are many variants on how data are collected and each of these must be managed taking into consideration the unique characteristic of each. Where possible, common procedures should be used, and all of the procedures must be documented.

Archive structures must be flexible enough to accommodate new measurements without requiring major alterations. New kinds of measurements are produced from new instrumentation and these must have a designated archive.

Metadata are a key component of the archives and must be carefully collected and maintained with the measurements. The metadata include such information as where and from whom the data came, how the measurements were collected and analysed, and what additional processing they have been subjected to. These and other attributes are very helpful in resolving questions that inevitably arise when problems in the data are found. Much of these metadata can be recorded using controlled vocabularies. This supports reliable queries of the archives, as well as resolving any ambiguities that may be associated with the data as they enter the data system.

The data system must provide knowledge about where data reside. This can be accomplished by building a catalogue of holdings. This catalogue will tell what data are held, where they are located and provide links to get additional information. The production of this catalogue will be closely tied to the archives and, therefore, is a crucial product of the data system and the first point of access for many users.

Once a user has located the archives of interest, they may want more detailed information to

determine more precisely if what they want actually exists in the archive. This will require more sophisticated tools than those needed for data discovery and will be provided by the data system. Having found data of interest, delivery tools will be used for data selection and to provide the results to users.

Provision of analyses and products from the available data can be carried out by both data managers and users. The nature of the products will determine the degree of cooperation needed. Generally, data managers will produce analyses that show the state of the archives, and metrics to describe how well the data system works. Where problems are found, steps will be taken to improve operations. Users may wish to prepare more scientifically oriented products, such as climatologies. Data managers may contribute to this in performing the necessary computations or preparation of the data to ease the task.

4. ARCHIVES

The Data Policy declares that all Science data are an "extremely valuable and irreplaceable resource" that must be "managed as part of an integrated system accessible through regional, zonal and national data centres". This means that the archiving strategy for different kinds of data may be different. But all archives must faithfully maintain data over the long term and meet the test of accessibility to both the original providers of the data and to other users.

The first question for any type of data is to decide the appropriate archive strategy. Data managers should consider such questions as:

- Do the data exist in a more usable form elsewhere?
- Given the right circumstances, are the data reproducible?
- Are the data of wide or narrow interest?
- Should the archive be maintained by Science or is there a more suitable and available agency?
- Do the data need to be maintained in the long term?

Once the decision is made on creating an archive, the form, whether numerical or not, needs to be decided. Numerical archives are those which hold data that are collected digitally and may be stored in computer databases or files. The creation of these archives is not simply the process of acquiring data from providers and inserting them into some data basing scheme. The following items are necessary to consider in creating and maintaining archives.

- They must contain not only the numerical values, but also sufficient metadata such that the values are interpretable by users many years in the future.
- The content needs to be verified to be correct.
- Procedures need to be developed to guard against the introduction of duplications.
- Because the same data are sometimes presented to archives from different sources, some degree of version control may be required.
- Technology changes and there is the requirement to migrate archives from older to newer media, and to change archive systems to use newer technologies.
- Impacts on users of changing technology should be taken into consideration.

- A regime of routine backups of the archives must be implemented to guard against accidental loss.
- As appropriate, incoming data may be converted to the archive format to provide consistency to users. Any content conversion must be carried out without corruption of information.
- Procedures carried out on data entering an archive must be well documented and the documentation readily available.
- Appropriate ease of access to the archive need to be determined.

There is a wide variety of data managed by Science and there is no one location where all of the resources exist to manage them all. The Data Policy states that the data system in Science is distributed with a coordination role being played by Ottawa.

There are 3 roles to be performed in the archiving process. Any particular data management group may deliver one or more of these roles. They are as follows.

- The Data Assembly Centre (DAC) archives source data generated by research and observational programs. It provides the initial processing of the data (quality checking and corrections, navigation, smoothing, etc.) and provides a processed dataset to the Data Distribution Centre.
- The Data Distribution Centre (DDC) archives data, often based on type, from the variety of sources (DFO DACs, international sources, originators) into a consolidated collection. It also provides a second level of data verification. The DDC has the primary responsibility for providing access (Discovery, Browse, and Delivery) for those data types for which it is a designated DDC. Most users will acquire data through the appropriate DDC.
- The Product Generation Centre (PGC) uses available data to construct various types of products. These may be scientific products such as climatologies or products that describe the contents of the archives.

The data assets of Science can be classified as digital or analogue. In the former category are the numerical data returned from instruments such as CTDs or counts of plankton from a net tow. From a data management view, these are already in digital form and so are readily managed in computer files.

The analogue category includes such media as video, still imagery and audio files. An important attribute is that the files tend to be very large. Another important feature is that it is common that there has not been any classification of the contents or interpretation of what can be seen or heard. A very simplistic approach would be to build a database with file names and download the data file by file. This would require users to spend a tremendous amount of time reviewing the files and looking for content of interest.

A more useful approach is to index individual subsets of the files by time and/or location (even content if possible). Such procedures are now technologically possible, but challenging. Having done so, it would be possible to search the archives at least by these criteria and so provide a more targeted result to users.

Finally, Science holds a significant volume of physical samples (later referred to as non-numerical assets). In this case, the archives are preserved in freezers, glass jars or envelopes containing the samples. These require a different strategy for management.

4.1 ARCHIVE STRATEGIES

Data of interest or collected by Science are varied. Some data result from experimental programs in laboratories. Other data are collected in the environment whether in a natural system, such as the open ocean, or in a manipulated system, such as the Experimental Lakes Area. Data may come to Science from its own researchers, from researchers in other government departments, from universities or from colleagues or organizations in other countries.

Recommendation 4.1: Each type of data whether collected by Science staff or that is managed by Science on behalf of other providers, must have a managed archive.

Action 4.1a: The data assets of Science will be surveyed appropriately to determine the variety and kinds of data and their present archives. Based on this information, the National Science Data Management Committee (NSDMC) will assess the weaknesses identified and take actions as appropriate to correct the problems.

Action 4.1b: NSDMC will develop a document that provides guidance on what are considered to be the appropriate archive strategies for different kinds of data. This document will help to ensure a consistent approach to archive strategies in Science.

A distributed archive system calls for a level of coordination that is more complicated than if all data were held in a single location. Part of the coordination is to designate a primary archive for each kind of data that is being managed. The role of a primary archive is:

- To accept, process and maintain data of a designated type.
- To provide the first point of contact to the data for anyone wishing to access them (including providers).
- To coordinate moving data from providers to the archive and controlling versions of data.

Action 4.1c: NSDMC will further develop the ideas of DACs, DDCs, and PGCs to clarify roles and responsibilities. Where there are no currently designated archives, NSDMC will take appropriate actions to create them.

Accessibility to archives is a key element. While it does not guarantee ease of access, maintaining archives on-line is one step in improving accessibility. Archive size, complexity, technical and security issues, telecommunications capabilities, etc. all play a role in determining if archives may be placed on-line. To the extent possible, it is the intention of the Science data system to have all archives available on-line.

Action 4.1d: Each primary archive will examine the resource requirements and capabilities to place their archives on-line and formulate a plan for doing so.

No matter the type of data being handled, it is certain that new instrumentation, new variables, new techniques, etc. will provide unanticipated data being available to an archive. Continual changing of archives puts a demand on resources that is not likely to be supportable. On the other hand, it must be expected that archive systems will need to be rebuilt periodically. The challenge is to find the right middle ground such that an archive system is sufficiently robust to be able to adapt to most new demands in a way that does not require significant redesign and rebuilding of the archives every few years.

Action 4.1e: Archive systems must be built in such a way that they are extensible and can adapt to new variations of data. This may mean use of indirect referencing (using code tables, for example, rather than names in the archive), modular designs for processing systems, software coding strategies that allow for easy reuse of code, etc.

4.2 PROJECT INFORMATION

Projects are conceived to meet specific goals for DFO. They may be targeted to answer one time scientific questions or they may be projects that intend to conduct measurement programs over a long period of time and on a regular basis. In all cases, contact between the data manager and the project leader is very important. New projects may include new variables not yet seen by the data system and the data system must ensure an efficient processing stream that can handle regular and ongoing input.

In formulating the project, there should be collaboration between the data manager and scientific staff to discuss the nature of the data to be collected, and to formulate a plan for handling the resulting data. A key part of this collaboration is to determine what data processing and management will be handled by the project, and what will be handled by data managers using DFO data systems. Even if all data are to be handled by the project, there must be a plan to turn over the data to the DFO system. This plan must indicate the time that the turn over will start and must also provide sufficient funding so that appropriate preparations can be made to manage the data within the DFO data systems.

Recommendation 4.2: All Project Plans will contain a section on data management and this section will be used by data management personnel to plan for appropriate archiving of the resulting data.

Action 4.2a: Each region will designate a coordinator who will ensure the review of project plans and work with Project leaders to determine the level of effort that will be needed to take data collected by the project and move them into designated archives.

Action 4.2b: Each region will, in concert with the NSDMC, provide adequate funding to ensure the resulting project data are managed in a way that is consistent with the Data Policy

4.3 DATA COLLECTION

Data management personnel may, and should be encouraged to participate in data acquisition activities. This is important since it builds a trust between researchers, technicians, and data managers. It is important for data managers to be knowledgeable in the technical issues and operational problems associated with data acquisition at sea. Likewise, it is important for all DFO staff to understand how the data system works and what role they play in successfully managing the acquired data.

Recommendation 4.31: Data management staff should receive experience in data collection procedures.

Action 4.31: Data managers will have reasonable opportunities to take part in data acquisition activities. These should be considered as both training activities and a way to develop closer collaboration between scientists and data managers.

Data may be presented to the data system very soon after data collection or much later. It is normal for data acquired electronically, such as from a CTD, to be provided quickly (hours to days). Other data must be measured on shore, for example from chemical analyses of water samples. In either case, appropriate metadata must accompany the data.

When data are presented to the data system, managers must ensure that the appropriate identifiers (cruise, samples, etc.) are attached to the data to allow association of these data to other data acquired at the same time and place and with known sampling characteristics. Depending on the type of data being presented, there will be the need for other kinds of metadata that describe such attributes as the instrumentation, sample collection or storage methods, analysis methods, etc. The requirements for these additional metadata will differ based on the kind of data collected. Without appropriate metadata, the data provided to the archive may be unusable by another user either immediately or at some time in the future.

The IODE (Intergovernmental Oceanographic Data and information Exchange) Committee and JCOMM (Joint Commission on Oceanography and Marine Meteorology) are international organizations that cooperate in exchanging data. Through these groups, Science acquires a significant volume of data collected by foreign platforms in waters around Canada or of interest to DFO researchers. The data from these sources are treated the same as data from domestic sources. They often have more limited metadata and there may not be a scientist associated with the data collection. Despite these characteristics, they are generally valuable and warrant the effort to link to appropriate external systems to gain access to them.

Recommendation 4.32: Data presented to the data system will have all of the required metadata to ensure they are useable by others.

Action 4.32a: It is necessary to ensure that the basic information about where, when and what have been collected are present with the observations. If this information is missing, action must be taken to acquire it.

Action 4.32b: The data manager will confer with appropriate staff providing data to determine what additional metadata are required to ensure future users of the data will understand the constraints on correct interpretation of the observed values.

The Data Policy provides for some time periods when distribution of data may be restricted. The collector may hold the data until the restriction period has elapsed, or provide the data to the archives with the understanding that the data will be held with restricted access until the restriction period has ended. Providing data to the archive before the restriction period is over ensures that data get to the archives and allows for preprocessing before insertion into the archive. However, it puts onus on the archive to control access to the data.

Recommendation 4.33: The data collector and data manager of the archive that will hold the data will consult to move the data to the archive with minimum delay while respecting any distribution restrictions that may apply.

Action 4.33a: The manager of the archive must satisfy the data collector that data will not be distributed before any restriction period elapses.

Action 4.33b: The data collector will work with the manager of the archive to move the data as quickly as possible to the archive.

4.4 Data Transfer and Processing

Data provided to the data system will be in a variety of formats. It is the responsibility of data managers to work with the providers to ensure that the contents are complete and understandable and that if any format conversion is necessary, that no important information is lost. Such activities may be undertaken as part of the project data management or may be done when the data are presented to the data system.

Recommendation 4.41: The contents of data presented to the data system will be documented and this documentation is readily available.

Action 4.41a: Data managers will ensure they have documentation that describes the format or data structure and content of accepted data. This documentation must be maintained in a formal repository meeting the requirements established by the NSDMC.

Action 4.41b: Data managers will ensure that metadata provided with the data are consistent with archive contents. As necessary, they will consult with data providers to ensure metadata content meets archive requirements.

Action 4.41c: Data managers will undertake necessary transformations of the data and metadata into archive standards while ensuring that none of the information is corrupted.

It is the intention of this strategy to reduce the number of different formats presented to the data system. To this end, any processing that takes place before the data are provided to the data system should take this intention into account. This is where early collaboration during project

definition can ease the work load to manage the data.

Recommendation 4.42: The number of different data formats delivered to the data system will be reduced without endangering the flow of data to the system.

Action 4.42: Data managers must have well documented descriptions of the preferred formats for all types of data. These must be maintained in a formal repository meeting the requirements established by the NSDMC and be readily available to all Science staff. Data providers will be encouraged to use these formats as much as possible.

Before data enter Science archives, they will be subjected to tests of their quality. These tests may happen in the region in which the staff collected the data or in the region hosting the archive. To ensure consistency of treatment, testing procedures will be standardized. Work may be shared between the region assembling the data and the region hosting the archive but no matter how the work is done, the testing will be consistent.

Because the volume of data is sufficiently high, it will be necessary to develop computer implemented test procedures that exploit characteristics of the type of data under consideration. These characteristics include the precision and accuracy of the measurements, the type of instrument employed, known failure modes of the instrument or sensor, differences from climatologies, etc. The tests will produce results that are recorded with the data and are readily understood.

The tests will change as experience is gained and there must be mechanisms to identify the tests and versions deployed against the data.

In every case, if there are questions about the data, the first consideration will be to go back to the collector or knowledgeable person to resolve the questions. If no such person is available, the data will be corrected if possible or accepted as is with appropriate quality indicators attached to the data.

Recommendation 4.43: Procedures will be employed to assess, correct where possible, and document the quality of data in the archives.

Action 4.43a: Procedures will be set for determining the quality of data, preserving the test results, and documenting the specific tests applied. Corrections of detected errors will be made in collaboration with data providers when possible. The procedures will be well documented with descriptions available from a formal repository established by NSDMC.

Action 4.43b: Procedures will be developed to ensure that data of the same type coming from different providers are processed in the same way. Documentation of the processing steps will be managed in a formal repository established by NSDMC and be readily available to anyone.

Action 4.43c: As test procedures improve or are added, the new procedures will be used against newly acquired data. Consideration will need to be given for reprocessing existing archives to bring all data to the same level of quality assurance or determine an effective way to

communicate the differing levels that exist in an archive.

Periodically, data that have been moved to an archive will undergo reprocessing. This sometimes is required to add more variables than originally available or to fix newly discovered problems. Sometimes, the same or associated data arrive from two different providers and on different time scales. Before data are placed in archives it is important to know if they are replacements or additions. This requires a way to identify different versions of the data or a scheme for determining if newly arrived data are already represented in some form in the archives.

Recommendation 4.44: Procedures will be employed to identify all near or exact duplications of archived data in newly provided data.

Action 4.44: Actions will be developed to identify if data arriving at an archive are new or not. These actions may include standardizing on cruise identifiers, building unique tags for data, devising algorithms and applications to find duplications and near duplications between incoming and archived data, or any appropriate combinations of these and other procedures.

Some software applications that carry out routine processing of Science data can be purchased from commercial vendors or can be found as open source software. Use of such applications provides advantages of consistency. But there are many other cases where custom software is produced by data managers themselves. These applications are written in a variety of software languages to run on different operating systems and to access data structures that are often unique. There would be a large gain in productivity and consistency if software written by staff in one group in Science could be easily used by another needing the same functionality.

Recommendation 4.45: Software developed for data management purposes will strive to be platform independent and exploit commonalities of data structures.

Action 4.45: Data managers will continually explore ways to standardize use of software applications within Science so that the same application may be used on a variety of platforms and computer operating systems, or so that processing may be shared on a regional or national basis.

4.5 DIGITAL ARCHIVES

When data are turned over to the data system by the provider, there may be a lot or a little work required before the data can enter the archives. The time delays between receipt of the data and incorporation into archives is dependent on many factors including the completeness of the data, the format, content, and quality. No matter how much work is required, it is important that the data provider be informed when the data actually enter the archives. This provides confidence to the data provider in the operations of the data system and is a measure of how well the data system is coping with submissions.

Recommendation 4.51: All data providers will be informed when the data they provide are placed in the archives.

Action 4.51: Data managers will set up mechanisms to ensure that notification will be sent to originators when the data they provide enter the archive. Data managers will also ensure that they can measure the time delay between data receipt and insertion into the archive as a way to gauge the efficiencies of the data system and to find weaknesses.

It is a common practice to make use of lookup tables to describe all kinds of attributes in archives. This provides a consistency in description that is particularly important when the attribute is used to support queries. But although the use of lists is common, different archives in Science (and elsewhere) use different lists for the same or similar attributes. This is a source of confusion for a user especially if they are trying to combine data from different archives.

Recommendation 4.52: All Science archives will share common lookup tables and controlled vocabularies.

Action 4.52: Data managers will cooperate to consolidate controlled lists used across Science into common lists used nationally. Mechanisms will be put in place to share the lists and to ensure they are kept current and easily accessible to all.

Different kinds of data have different characteristics and this is reflected in the structures of established archive schemes. But, there are commonalities across these data structures that generally have not been exploited. For example, almost every archive retains details about the source of the data. In addition, archives often store information that describes the instrumentation or procedures used to acquire the stored measurements. In most cases, the content and structure for even these subsets of information are unique. This causes at least two problems. First, the information (metadata) content is variable and therefore it is more difficult to compare when combining data from two different archives. Second, the differing data structures themselves complicate combining measurements stored in different archive structures.

Recommendation 4.53: Archive data structures will converge so that a wider variety of types of data will be held in fewer archive data structures.

Action 4.53a: NSDMC will undertake work to examine the various archives holding data in Science and look for common features. As possible, these commonalities will be incorporated into existing or developing archives so that in time the same metadata held in different archives will be held in similar data structures.

Action 4.53b: Data managers will continually keep abreast of developments that offer greater standardization of data structures across differing types of data. As these are encountered, they will be brought to the attention of NSDMC members for consideration.

On rare occasions, it becomes necessary to remove data from archives. In this case, deletion of the records is a straightforward issue. What is more complicated is maintaining the knowledge of when and why the deletion occurred. Without this knowledge, it is possible for the data to re-appear through data exchanges with partners and so once again be placed in archives.

Recommendation 4.54: Information will be retained about all data removed from the archives,

including the date, reasons and whatever other information is needed to be sure they do not re-appear.

Action 4.54: NSDMC will develop a mechanism to follow to delete data from archives, including how to record when and why data are removed from archives.

All of the data and metadata collected and managed by Science represent a significant monetary investment by the Government of Canada. It is crucial that these assets be guarded against loss caused by persons with malicious intent or through accidents. An important safeguard to this end is the regular creation of backups and the safekeeping of these for a suitable period of time.

Backups evidently apply to archive files. But they also apply to the originals of data as they are received at the archive. These will have all of the original content as was provided to the data system, and therefore may be used to answer questions about the data or subsequent processing.

Recommendation 4.55: Backups of data will be created and stored on a regular basis.

Action 4.55a: Data managers will decide on a suitable schedule for backing up archives and make these schedules known. They will also decide what copies of these backups will be stored off site to safeguard against catastrophic damage to local computing facilities.

Action 4.55b: Data managers will determine the most appropriate way to retain originals of the data provided to the archive. Appropriate backup strategies will be implemented for these data.

Action 4.55c: Data managers will decide the appropriate retention schedules for the data that are managed in their archives. These schedules will conform to those mandated by Government of Canada policies. Science may choose to retain data longer than maximum requirements.

Measurement technology is continually evolving and improving. Measurements of variables already archived, for example, become more precise, and measurements of new variables become available. Information that is needed for a proper interpretation of results from new technologies may be different from that which was required previously. A data system and an archive must be robust enough to be able to accommodate these changes without breaking down.

Recommendation 4.56: Archives will permit the ready inclusion of new variables and changes in measurement attributes of existing variables.

Action 4.56: Data managers will endeavour to ensure that existing and future archives are robust to changing attributes or accommodating new variables.

In the domain of satellite data management, it is common to speak about the “processing level” of the imagery. Level 0 represents the raw data as it streams from the satellite sensors and Level 3 represents interpreted and georeferenced values. Similar ideas have already been encountered when comparing the raw data delivered from a CTD sensor (more like level 0) to archives of measurements of from various instruments (more like level 3). This idea can be expanded to include the degree of processing through which data have gone. Having a concise index for

each archive may be helpful to a user to quickly identify the appropriate archive holding the data of interest.

There are other ways to look at what constitute versions of data. Data received at an archive within a few hours of collection may undergo simple verification procedures but let through more subtle errors. With more time available and greater scrutiny, subtle errors will be identified in the data and this version will be the next to appear at the archive. Subsequent processing either by the originator or by others may produce yet another and perhaps different version of the data.

There are many variants of how data versions may be created. From the point of view of the archive user, it is likely that in most cases they will want the "best" version available.

Recommendation 4.57: NSDMC will use as appropriate, an indicator of the version of the data.

Action 4.57: NSDMC will investigate data version control issues and determine the most appropriate strategies to be used.

4.6 NUMERICAL MODEL OUTPUTS

Computer modeling is an activity carried out in many parts of Science, for uses ranging from research to product development. The outputs from these models are generally only available to a limited few. A model could be one that computes global ocean circulation or one that simulates interactions within ecosystems. The results are closely linked to how the observational data are assimilated into the model and how the computations are carried out by the software. Numerical models can produce large volumes of data since they can provide a continuous, quantitative representation of ocean or ecosystem variability in four dimensions (space and time).

The results of models are valuable to others because they can take limited observational data and perform a kind of interpolation/extrapolation to provide results where there are few observations or when observations are poorly sampled in space and/or time. Models can be used to hindcast or reconstruct past variability; nowcast or provide the state of the ocean/ecosystem by combining observations, dynamics and empirical information; and forecast conditions in the future. The resulting value-added fields and products are used by others directly or as inputs to other kinds of models.

There are many classes of models with the majority in past times falling within the domain of research. These are run to explore scientific issues, are constantly being improved or changed, and have results that are generally of immediate use to only a small audience. There are an increasing number of models that run in a more "operational" mode, providing outputs at time intervals ranging from hours to seasons. These models are run on a routine schedule, have characteristics that are fixed for considerable periods of time and hence can be readily documented, usually have undergone some degree of observational validation, and provide products that are of wider use and distributed to clients on a routine basis. These latter characteristics define the key attributes that determine if model results have value to archive. Science will consider the results of such operational models as data assets and consequently they should be managed appropriately.

At present, there is no clear definition in Science that can be used to decide if a model is operational or not. A more precise definition is needed so that a consistent view is obtained of what model results should be archived. Collaboration with the Science national Centre of Expertise for Ocean Model Development and Application (COMDA), which in part is focused on advancing operational oceanography in Canada, provides one option for addressing this evolving area of data management.

Recommendation 4.61: Science will consider certain outputs from operational models as data assets to be archived.

Action 4.61: NSDMC will work with COMDA and the modeling community in Science to define the characteristics that indicate whether a model is operational and consequently which of its outputs should be archived.

The volume of data produced may be an issue. In addition, it will be important to devise an appropriate indexing scheme so that subsets of the outputs can be quickly identified and accessed.

Recommendation 4.62: NSDMC will work with relevant modeling groups to develop cost-effective strategies for the storage and archival of operational model outputs and products.

Action 4.62: NSDMC will work with COMDA and the modeling community to develop such strategies.

Recommendation 4.63: NSDMC will develop an efficient indexing system so that reasonably sized subsets of the results can be readily identified.

Action 4.63: NSDMC will work with the modeling community to develop the indexing required.

The model characteristics are of great importance as they impact what data and information to archive, and how long it should be archived for. In addition to this, the data assimilation schemes, observational inputs used, computational algorithms and generally the important internal operations of the model are necessary to document so that comparisons may be made between models and observations, and reliability can be assessed.

Recommendation 4.64: Appropriate model characteristics will be archived with model results.

Action 4.64: NSDMC will work with COMDA and the modeling community to define what are the important characteristics of model operations that should be archived with the model results.

Being considered a data asset means that archived model results need to conform to accessibility requirements and to standards adopted for national use. Because of data volumes, it may be necessary to put some limitations on archival and accessibility functionality.

Models change and improve so that older versions of models are retired and newer versions come into operation. Each time there is a change to an operational model, the value of retaining

output from the earlier version should be assessed.

Recommendation 4.65: Operational model outputs will be archived in a manner that is compatible with national accessibility requirements and standards for other digital data archives.

Action 4.65a: Data managers will work with the modeling community to define the archival requirements for model outputs.

Action 4.65b: NSDMC will collaborate with COMDA and model developers to decide the long-term value of preserving outputs of retired versions of models.

4.7 NON-NUMERICAL ASSETS

Non-numerical archives are those which have not or cannot be converted to digital archives. They may be samples in the form of whole or partial organisms either frozen or pickled or they may be extracted portions such as otoliths or fish scales. They may also be in other forms, such as hard copy photographs or analogue imagery. In some cases, it may be possible to convert these to digital archives by using imaging or scanning techniques. In general, non-numerical assets represent the source samples from which digital data (e.g. fish age, number of organisms, etc.) may be derived.

These assets are retained as references for further or different analyses in future. They represent an archive every bit as important to Science as its numerical archives. Maintaining these assets presents a set of different problems that need to be faced. The problems include the following.

- How to maintain the facilities that contain the samples such as freezers, jars or temperature and humidity controlled rooms.
- Who is responsible for maintaining both the non-numerical assets themselves, and the infrastructure (such as climate control).
- An indexing system must be built and updated to record what samples exist in the archive.
- If samples are extracted for further analysis, information about by whom and when the sample was removed need to be retained.
- It is important to keep records up to date as to when new samples are provided.
- A record of actions taken to maintain the integrity of the samples must be preserved (such as if the preservative was refreshed or changed) so that any changes in the samples can be accounted for.

In the case of non-numerical assets, the processing and archiving procedures have different characteristics than for digital archives. The procedures are more akin to museum collection preservation and it is likely that much can be learned from that discipline. If these samples are to be of use in future, appropriate steps must be taken to ensure they are properly managed.

Recommendation 4.7: Science will take steps to ensure the future integrity of its non-numerical assets.

Action 4.7a: NSDMC will determine how its non-numerical assets may be properly preserved. This will consider the required infrastructure, costs, and possible partnering arrangements

such as with museums. Sample tracking will be one of the considerations. Roles and responsibilities will be set for those maintaining the non-numerical assets.

Action 4.7b: Data managers will implement mechanisms to document the current contents of non-numerical archives and links to measurements contained in digital archives. The mechanisms must be able to manage the inclusion of new samples or content and track information about samples removed or altered among other information.

4.8 DATA RESCUE

If all data and non-numerical assets collected in the past had had a managed destination archive, there would be no need to worry about data rescue activities. But such an ideal has not occurred in the past, and is unlikely to be fully realized in the future. What makes this issue more pressing at this time is the aging population of research scientists in DFO and the data that they hold outside of managed archives. And, whereas the actual measurements may be in less danger of loss, it is the metadata that describe the measurements and how they were made that is at greater risk because this information is often not written down with the measurements. For some kinds of data, the lack of metadata may be inconvenient, but for other kinds, this lack may completely nullify the usability of the measurements.

It is a requirement to establish a reliable list of data that are not maintained in managed archives and therefore are at some risk of loss. The risk level will vary depending on the media on which the data reside, whether backups exist, how well the data structures are documented, how much metadata reside with the data, how close the person holding the data is to retirement and so on. A few years ago, a data base was built, called SCIDAT, whose purpose was to construct this list of data at risk and to make a preliminary assessment of risk for each entry. The data base was subsequently used to catalogue other files and information that were considered of importance, and so went beyond the original purpose. The result is that there are many entries in SCIDAT that do not reference data as such and there is no simple way to separate out these entries from records that do describe data.

Recommendation 4.8: NSDMC will re-establish a list of data at risk, use this to set priorities on rescuing data and each year show progress in bringing listed data into managed archives.

Action 4.8a: NSDMC will prepare a document to provide guidance on assessing if data are at risk.

Action 4.8b: NSDMC will determine the best way to re-establish a list of data at risk of loss. It will take into consideration the present version of SCIDAT and establish mechanisms for keeping information current. It will consider how SCIDAT assessed the vulnerability of the data and, as appropriate, reconfirm the assessment.

Action 4.8c: Determining which data should be brought into established archives will depend on a number of factors including cost, importance of the data, and vulnerability to loss. NSDMC should establish guidelines to set the priority of which data collections should be treated first.

Action 4.8d: Each year, NSDMC will review the list of data at risk and as resources permit, undertake to secure data in archives. Resources external to Science may be available to assist (such as the IODE GODAR Project) and these should be investigated.

5. ACCESS

Archives are built so that historical data may be used by future researchers and the public. Providing access can be considered to take place in three stages. They are as follows.

- The Discovery stage allows users to identify archives that are likely to contain the data or information of interest to them.
- The Browse stage allows users to look into archives to determine if the specific kind of data is present at the locations, times, in the quantities and with the other attributes that they require.
- The Delivery stage allows users to formulate a query of the archive and to have the data delivered to them in a form that is readily usable.

The technology supporting access to data and information is changing at a rapid rate. The ubiquity of the Internet and World Wide Web services is a clear example. Predicting which of these technologies is most suitable to Science requirements is not easy. However, Science data managers need to be conversant with these developments, to experiment with ones that look promising and to recommend for national adoption, those that provide significant benefits.

Of equal importance is the requirement to protect Science data assets from damage from malicious individuals operating on computer networks. In this aspect, Science must work closely with IMTS to find the correct balance between accessibility and protection.

Recommendation 5.0: Science data managers will recommend for adoption those technologies that have significant advantages for a national data system.

Action 5.0a: Data managers in Science will monitor and experiment with developing technologies that appear to have potential use to national data systems. As appropriate, those technologies with significant benefits will be recommended for national adoption after appropriate discussions with IMTS and computer security experts.

The Science Data Policy allows for some restrictions on the access to data collected by DFO researchers. These restrictions would include such factors as a researcher's first right to publish, certain kinds of data that are sensitive due to their interpretation in health and safety issues, and commercial data. These are the exceptions; the vast majority of the data held in Science archives are in the public domain. Access restrictions must be supported in the Science data system, but must operate within the scope of the Science Data Policy. These restrictions must be built into the archiving process, and operate when users seek access to the data.

5.1 INVENTORIES (DISCOVERY)

DFO Science archives are distributed across the regions and in many different forms. In many

cases, archives are distinguished by the kind of data they hold. In other cases, the same type of data is held in two separate archives with the division being based on geographical boundaries. Whatever the historical reasons, or the internal (to Science) management of the archives, users must be able to find data of interest and in a way that does not rely on collegial arrangements.

To provide knowledge about where data reside, the data system needs to build a catalogue of its holdings. This catalogue tells generally what data are held, in what location and provides links to get additional information. The production of this catalogue must be closely tied to the archives and, therefore, is a crucial product of the data system. This catalogue must have the following attributes.

- It is easily located
- It describes only those data collections that Science manages
- It is clear what are the source archives and what are other versions
- Its contents can be delivered to broader cataloguing systems, such as GeoConnections
- It is readily updated

Recommendation 5.1: The national data system will maintain an on-line, publicly searchable inventory of archives (digital and non-numerical) maintained by Science.

Action 5.1a: Data managers of primary archives will construct descriptions of their archive contents in sufficient detail to permit providers to identify the proper location to which they should send their data and so that users can determine the most likely sources of data meeting their interests. These descriptions must identify any access restrictions that may apply.

Action 5.1b: Data managers will ensure that the inventory is readily accessible and records are maintained up-to-date. They will employ appropriate technologies and standards to build and maintain the inventory.

Action 5.1c: Data managers in other locations with similar data as in primary archives will indicate to users who contact them, where the primary archive resides and should be first consulted.

5.2 BROWSE

Once a user has located the data of interest, they will want more detailed information to determine more precisely if what they want actually exists in the archive. This will require more sophisticated tools that support such functions as:

- Data location mapping tools showing where data were collected
- Other tools that can show the complete spatial and temporal distributions of the data as well as more details about the measurements collected.
- Tools that allow for some degree of visualization of the measurements or contents.

These facilities may be delivered through web based tools.

Recommendation 5.2: All archives will provide browse facilities.

Action 5.2a: Digital archives will provide on-line browse facilities to users. They should show what was collected, where and when, and perhaps even the measurement values. Common tools for browsing will be promoted by NSDMC.

Action 5.2b: Archives of non-numerical assets will provide on-line browse facilities to users. They should show the nature of the assets, where and when they were collected, and what access restrictions may apply.

5.3 DELIVERY

Data delivery is carried out once a user has determined that they are reasonably certain the archive examined has data of interest. It begins with the user specifying the criteria to use in searching the archive. For on-line archives, the specification of criteria may be done interactively. For archives that do not have on-line facilities, there will need to be some human intervention to set criteria and carry out the request. Having a personal contact with users can be valuable as this may streamline the request and improve the satisfaction of users.

Users will request data from an archive either on an ad-hoc basis or as a "subscription" service. Ad-hoc requests are those that are posed as the need arises. They usually are to meet a one-off purpose. Subscriptions are those where a user requires a regular delivery of data meeting their criteria.

For either kind of request, data may delivered through "push" technology, such as uploading files to a user's ftp server, or through a "pull" process where the files for a user are placed on the archives ftp server and the user must come to get them.

Just as Science receives significant data from foreign sources, we also contribute to international data centres and researchers. This is done through the auspices of IODE and JCOMM. The data exchange can take place through a subscription type service, or through ad hoc data requests.

Some of the archives in Science are quite large and it is possible that a single request may ask for very large portions of such archives. Immediate processing of these requests may put unacceptable demands on computing resources and result in degraded performance for all users. It will be necessary for archive centres to take this into consideration and configure their request processing systems accordingly.

Archive centres will be under constant demand to provide data in a wide variety of structures and formats. It is unrealistic to expect that any and all output formats can be delivered. Rather, archive centres will need to consider the more frequently requested outputs and provide these. It is through the provision of standard formats that the first stage of data integration across archives will be accomplished. As resources permit and demands change, archives may provide other outputs.

Recommendation 5.3: Archives will provide easy access to their contents.

Action 5.3a: Archives will provide appropriate facilities for users to specify search criteria and sub-setting of data.

Action 5.3b: Each archive centre will need to examine the feasibility of placing its archives on-line and provide immediate processing of requests. Though the recommendation is to have all archives available on-line with immediate processing, there may be good reasons why this is not possible. Each centre will develop a plan for putting archives on-line and designating conditions for immediate processing of requests.

Action 5.3c: Archives will provide delivery mechanisms that allow for ad-hoc requests and subscription services. The provision of these services will consider the needs of users and will be provided as resources permit.

Action 5.3d: Archive centres will provide a common suite of output formats and additional ones as needed for their particular archives. The common suite of formats will need to be decided based on experience of user needs and in consultation with users. Archives will be free to provide other outputs that cater to the particular kind of data they hold.

Action 5.3e: Archives will support both push and pull delivery technologies.

Action 5.3f: Archive centres must ensure that sufficient documentation is provided with the data and information so that a user can be reasonably expected to judge correctly how data should and should not be used.

5.4 PRODUCTS

The wealth of Science archives it not simply in the measurements they contain but also in the information and knowledge that can be derived from them. These derived results can be generally categorized as products.

There are two classes of products. They are:

- Information derived from the measurements that are used for scientific purposes or for the public.
- Data management information used to monitor the performance of the data system itself.

Information products that archives may generate include such items as climatologies, fields or vertical sections of different variables, and any of the other possible analyses. These products will be made in collaboration with appropriate scientific staff and descriptions of how they are produced should be provided to ensure that they are used appropriately.

Some products may be prepared in advance and routinely updated, such as climatologies. Typically, these will be described on web pages prepared by the archive centre. Other products may be produced on-line and on demand in which case they will be distributed using the same delivery mechanisms as for data.

Data management products usually are for the use of managers who need to know how well the data system is performing. These include such information as the size and growth of archives, how quickly data enter archives after receipt at the archive centre, the frequency of error detection in received data, the volume and type of data received and distributed by the archives, etc.

Recommendation 5.41: Information products derived from archives will be well described and readily available.

Action 5.41a: Data management and scientific staff will collaborate to define and generate appropriate products that can assist users seeking information from archives. Users must be able to find these products easily. Documents that describe the production process must be available and clearly associated with each product.

Action 5.41b: Data management products will be generated as required by archive centres to gauge their level of performance. These should be produced on appropriate schedules and used to monitor the performance of and to correct problems found in the data system.

An important part of delivery of data or products to users is to provide them with a mechanism for feedback. It is through this feedback that archive centres will identify weaknesses and strengths, and hear about other needs of users.

Recommendation 5.42: Archive centres will provide mechanisms to acquire user feedback.

Action 5.42a: Archive centres will examine their data delivery mechanisms and make provision for accepting user feedback and suggestions for new products or services. There are many ways this can be done such as surveys, providing on-line facilities, etc. Regular analyses of the feedback will provide guidance on what additions or changes should be considered.

6. STANDARDS

The adoption of standards provides the means to develop interoperability between the distributed components of Science's data system. Without them, users will be frustrated by inconsistencies of treatment, will be unable to find data or information, and will encounter significant obstacles to their work. Standards are a cross cutting issue that have application across all components of the data system as described here. Because of the broad ranging nature of applicability, this section will not deal in specifics. Rather, these details are left to be described in the scoping document on standards. However, this section will categorize the standards issues in the same functional components as used to discuss archives.

Standards are taking a greater prominence in technology solutions driven by the high degree of connectivity permitted by the Internet. To exchange information widely in an efficient way requires the adoption of standards. The W3C (World Wide Web Consortium) is a collection of interested individuals and groups that actively pursue the development of standards of all kinds. The OGC (Open Geospatial Consortium) is another group whose main focus is on

geospatial referencing and representation of data. Many of the proposals from these groups are appearing as ISO (International Standards Organization) standards and once that happens, businesses can confidently build technologies that conform. From this broad spectrum of existing and developing standards, there can be found some that have direct application to the tasks of managing data. Where useful standards exist, Science should adopt them.

There are other areas in data management where there are no existing standards. In some cases, there exist practices that are fairly broadly accepted but have not yet achieved formal status. Science needs to choose from these, to decide which meet our requirements and will provide positive benefits if adopted for our use.

Finally, there are other areas where there exist no standards and no widely accepted “best practices”. In these areas Science is on its own to decide what works optimally for our purposes. But, the problems we face in Science are also faced by most other scientific organizations and so it makes sense for us to work closely with partners to develop practices with wide support. These will form the basis of proposals of standards at wider national or international scales.

The adoption of standards will require changes to existing technologies to implement their usage. This requires resources to make the required modifications in software or to purchase appropriate hardware or software solutions. Because of this investment, it is likely that the national data system will be in varying states of implementation of standards. It is a challenge to build and maintain an effective system where different components have different capacities.

Standards also have a role in helping to gauge the performance of a national data system. In this case, the standard is set by ourselves and is a target to achieve. These performance metrics should be generated throughout the data system, and must be standardized in form so that they can be compared.

6.1 Data Collection

Within this component there are opportunities to apply standards that will assist in the identification of duplicates and data versions, in streamlining processing by reducing the number of formats that need to be handled, and in clarifying data and information delivered to the archives. Standards adopted early in the data collection process pay dividends in downstream functions. As possible, the adoption of standards as close to data collection as possible is desirable.

6.2 Data Transfer and Processing

The application of standards here can ensure that minimum requirements of metadata accompany data, that data formats are well described and therefore easy to handle, and that all data of a type are handled in equivalent ways no matter where the processing takes place.

6.3 ARCHIVES

Archives are built on computer management systems and applications that run on a variety of operating systems. A data provider and a data user should not need to know or care about these technologies. Presently, differences in the content of archives, even for the same instrumentation, causes differences in what users see when they request data. This can be remedied to some degree by standardizing outputs. At a deeper level, there is the opportunity to standardize archive contents by developing common views across archives. Part of this process will be the adoption of common, controlled vocabularies.

Standards play a different role in managing non-numerical assets. Here, it is important to look at what are the recommended ways to manage physical samples (substantial knowledge of this kind is to be found in museum staff). It is also important to keep aware of new ways to handle these assets that may help convert them into digital records even if they must be handled in different ways compared to measurements of the environment. So, image formats for scanned material, indexing and attaching attributes to audio or visual records, among others, must be under continual review.

6.4 ACCESS

Presently, there is a heightened interest internationally to develop catalogues of data holdings. Much effort is being put into standardizing catalogue content and there is an ISO standard recently published for this. Adoption of this standard will aid the production of catalogue material and will simplify the promulgation of this information around the world.

Preliminary work is also underway in defining standards for browse capabilities. In certain aspects, Web Map Services (WMS) represent a well defined standard for displaying georeferenced information. Adoption of this standard will allow purchase of commercial software, rather than having to write the software ourselves, and a degree of interoperability between maps produced at one archive or another.

Standards related to data delivery are not available but are under development. Different groups are proposing different solutions and so it will be necessary to watch or participate in these discussions.

Standards also apply in the presentation of on-line tools to users. It is a good idea to have a common look across all archives so that no matter where the user touches the data system, the presentation is familiar and consistent. There is no broadly accepted standard for such presentations, but there are some candidate technologies that bear watching.

Recommendation 6.4: NSDMC will analyze its requirements for standards and adopt existing standards where possible, accept “best practices” where applicable and develop local standard solutions when required.

Action 6.4a: NSDMC will continue to review and revise the scoping document on standards to stay abreast of the standards issue. Agreement will be sought on adoption of standard

practices as broadly as possible.

Action 6.4b: NSDMC members and others will participate in discussions to develop new standards. As these evolve, the information will be brought back to Science for incorporation as and when appropriate.

Action 6.4c: Before adopting any standard, an analysis of impacts will be made from the national perspective. This analysis will be used to plan how to accommodate the uneven implementation in the national system so as to minimize impacts on providers and users.

7. GOVERNANCE

7.1 PROJECT DATA MANAGEMENT

The Science Data Policy states the requirement for science projects to "demonstrate the existence of a comprehensive data management plan". In the past, even when this was specified as a requirement, it was unclear what such a plan should discuss and to what level of detail. To assist in this task, this strategy provides guidance (Annex 4) on the important elements to address and some of the inherent costs. When the project is approved and funded, the funding identified for project data management will be used to support activities to ensure the data are incorporated into the national data system.

For some types of data, a well developed and functioning data system may already exist. In this case, it is enough for the project to identify the system into which the data may flow and to coordinate data preparation activities to streamline this.

In other instances, no data system will exist for data collected by a project. In this case, work will be required to define the new components needed for the national system, and a project may need to be funded to assist in building these.

Recommendation 7.1: All science projects must have a section that discusses data management and its costs and the contents of this section should follow the guidance provided by this Strategy.

7.2 NATIONAL COORDINATION AND ORGANIZATION

The National Science Directors Committee (NSDC) recognized the importance of data management for meeting the DFO Science mandate. It acknowledged that data management is one of five key functions requiring particular attention under Science Renewal. Consequently, it formed the NSDMC with a mandate to coordinate and forge a national data system.

Maintaining a national data system requires

- staff in every region to work within this national structure.
- resources on an ongoing basis to support the work of the committee.
- salary and operational resources to support activities from A-base funding.

The terms of reference of the NSDMC are found in Annex 5.

At a national level, the NSDMC is composed of a chair reporting to a Director General in Ottawa and members from each region. As appropriate, the NSDMC will have other members to ensure strong connections to other sectors and groups. NSDMC administers allocated funding from NSDC. Members of NSDMC report progress on data management activities to the chair.

Managing a distributed system such as Science has requires strong coordination activities. One of the ways this is accomplished is through regular meetings of NSDMC. At these meetings, members discuss progress on projects, relate problems or new concerns, and set future directions. These meetings build a national view for data management that members take back to their regions.

National cooperation must occur at many levels. Clearly there is a need to establish common practices and procedures that employ standards so that a greater degree of interoperability is achieved by the distributed system. Cooperation also extends to developing or using common software whether produced in-house, or obtained from open or commercial sources. If there is common use of commercial products, there may be opportunities to reduce the unit cost of licenses by consolidating needs into a single license purchase. There may also be the opportunity to share hardware resources, either computing resources or disk storage. Shared use may require shared cost, so procedures will need to be developed to support this.

The structure of the data management organization is one important factor influencing its effectiveness. At present, data management activities, with a few exceptions, are dispersed in all of the various Science programs. Many staff devote some fraction of their time to data management tasks. But this dispersed nature makes it difficult to forge even local approaches, much less national ones. Although the summing of all of these resources may amount to what is considered an acceptable funding level, the focus is diffused and therefore less efficient and effective Some consolidation of data management activities helps in recognizing commonalities, in identifying opportunities for re-engineering systems, in discovering better practices, and achieving better results. At the same time, the organizational structure must still ensure that the data management activities undertaken are strongly connected to the business lines.

There are national costs to a national data system. Where new capabilities are required, increased costs are evident. Where there are existing facilities to manage the data coming from a project, there will still be incremental costs to be considered. These costs are related to the processing of data into archives, maintenance of the archives including migration of archives from older to newer media, and upgrading or extending existing archives, processing and access systems to exploit newer technologies or to extend the range of services provided by the data system. These activities operate outside of individual project data management activities but are essential to ensure that the data from the many sources in Science and outside can be integrated and made available into the future. The NSDMC operates to identify opportunities where regions can cooperate to extend regional systems to national ones, to coordinate building new components as needed, and to encourage standardization in the data system.

Establishing what is an appropriate level of funding is not trivial. One could examine the minutia

of defining what is and is not included in data management and from this build a case for an appropriate funding level. A simple but effective strategy that does not get embroiled in the details of data management is to set a funding level at some fraction of the total budget. A common figure is 5-10%.

Recommendation 7.21: There needs to be staff with ongoing salary and operational resources in every region and who are tasked to work for a national data management system.

Recommendation 7.22: The NSDMC requires a member from every region who will coordinate the data management activities in their region.

Recommendation 7.23: Building a national data system requires resources to hold regular meetings in regions.

7.3 Regional Coordination and Organization

Management of data in regions varies widely. Within certain business lines, data management practices are more consistent with the data policy while in others much improvement is required. This was established in the reviews done by the National Data Management Working Group (as reported in annexes 2 and 3) and in the Science Review.

The NSDMC representative in each region cannot be expected to know all of the relevant data management issues for the region. It is, therefore, necessary to form a regional data management coordination group, chaired by the NSDMC representative. The group will have the responsibility to ensure better coordination among the business lines and to support the national approach to managing valuable data resources. Members of this group should consist of staff whose sole or main activities are in data management with the group having expertise that encompasses all business lines. In regions where facilities are situated in different physical locations, it is desirable to have a member from each location. The group should also have representation from other sectors that have strong interests in how data are managed by Science.

Responsibilities and authorities for data management activities are defined most simply through a line organization. With designated divisions and sections, the responsibilities and reporting lines are clear and readily connected to the activities promoted by NSDMC. Alternative organizations and reporting structures will complicate matters.

Members of NSDMC are the leaders of the regional group and consequently are responsible for coordinating the data management activities of their region. They need to report to regional managers on national activities, and to use their regional knowledge to provide ideas to NSDMC on how national data systems should be built and operated. In addition, they act as regional points of contact when Science is asked for advice on matters.

Recommendation 7.3: Regions should establish data management coordination groups, chaired by the NSDMC member and organize data management activities in the region to support a national, integrated data system.

7.4 DATA MANAGEMENT ACTIVITIES

As of 2006, the NSDMC has operated for a single year. At its first meeting, a series of national projects were developed in an ad hoc process. During the course of the year, procedures were developed to ensure the initiation and funding of projects is transparent and carried out with full national cooperation.

Regional staff is encouraged to work with their NSDMC member and colleagues across Science to prepare proposals for data management activities to be undertaken each year. The template for the proposals and instructions about how it should be completed are included in Annex 6.

NSDMC will meet as early as possible after funding levels are known to review submitted projects, to look for opportunities to combine projects into ones with national scope, and to allocate funding. The process of project approval will be guided by the national strategy and objectives, take into consideration how proposals fit together, be based on available resources and respect funding or other pressures. At the end of the meeting, NSDMC will provide a summary of how the funding levels for proposals were set.

During the life of the project, NSDMC will review the progress of each and make corrections as considered necessary.

Recommendation 7.4: NSDMC will solicit data management projects with clear instructions on how the projects will be judged, will determine the funding to be provided to the projects and will oversee the work carried out.

7.5 LINKS TO COMMUNITIES

The NSDMC has contacts with other groups who have an interest in Science data systems. Some of these groups provide staff to take part in NSDMC meetings while others allow NSDMC members to sit in on their meetings.

Because of the very central role of computing infrastructure in managing and protecting data, Information Management and Technical Services (IMTS) has designated the Science Portfolio Manager as a member of the committee. NSDMC meetings will invite other representatives from IMTS to attend as possible or needed. It is important to Science that staff from IMTS understand how existing resources are being used and be aware of what computing resources are required in future by Science. Although NSDMC meetings will not cover the full range of Science Sector activities, what is discussed should provide a starting point for further contacts to be made between IMTS and Science. Likewise the Portfolio Manager can communicate the IMTS policies and issues that have to be considered by Science.

Science manages data on behalf of other sectors or uses data collected by groups in other sectors of DFO. The NSDMC should have contacts in these sectors. Depending on the degree of knowledge resident in NSDMC members, the committee may need a permanent representative from other groups, or simply invite attendance from someone in the region in which the committee meeting is being held. The NSDMC will take advice and consider budget constraints

to decide on such membership.

Contacts with these other Sectors is not simply for informational purposes. Science has an interest in the data managed by these others, and so it is in everyone's best interest to not only exchange ideas, but build working relationships in cooperative projects. In the development of cooperative projects, each partner must indicate his role and responsibility in managing the data or information.

Science data managers collaborate with colleagues in other government departments, such as Environment and National Defense. At present, the NSDMC has not yet tried to make formal connections into these other departments. Effective cooperation will take place in the context of sharing data and information of mutual interest. For example, the more recently developing coupled modeling initiative will encourage a closer working relationship between the data management systems in the partner departments. Cooperation with other departments will develop as such projects mature and as NSDMC consolidates a national strategy of data management.

DFO Science data managers have strong links to international partners. These collaborations are very useful to exchange new ideas for managing data, developing or adopting standards, and maintaining and developing contacts for data of interest. Members of NSDMC need to be cognizant of existing collaborations and to cultivate new ones as appropriate.

Recommendation 7.5: NSDMC will invite additional members beyond regional Science staff as deemed appropriate and will provide members to attend meetings of other Sectors, or organizations engaged in data management activities of interest.

7.6 REPORTING

At the end of each year, the status section of the project proposal will be updated by the lead project manager to describe the work completed. This will be submitted to the chair of NSDMC as a record of work done.

Members of NSDMC will prepare a Regional Report of work carried out and distribute this to all NSDMC members. The template for this report is included in Annex 7. NSMDC members can use this regional summary to explain to their Directors what has been accomplished in their region in data management activities.

The Regional Report also includes a section where information about other data management projects carried on outside of the national context may be reported. Members are encouraged to use this section as a way to inform others on the NSDMC of work undertaken in their region. This will stimulate inter-regional cooperation and further the goals of greater national coordination.

The chair of NSDMC will use the regional summaries to compile an annual report. This report will be provided to the NSDC on progress achieved. The report will provide details on the work carried out that year and a description of how the work fits into the longer term goals. The report

will also contain indications as possible of proposed work for the upcoming year for their approval.

Recommendation 7.61: The lead project manager will prepare a report annually on what was accomplished.

Recommendation 7.62: Members of NSDMC will prepare annual status reports on each national project in their region and regional data management projects undertaken.

Recommendation 7.63: The chair of NSDMC will provide an annual report to NSDC to explain what was accomplished in the previous year, and to propose the activities in the coming year.

Communicating the work and results of the NSDMC to Science staff is an important task. One way this can be accomplished is through a web site. This will be built and maintained by the chair of NSDMC. The web site will contain the various reports described above, policy and strategy documents, information about proposed and adopted standards, upcoming events and milestones, studies or reports of interest and links to appropriate inventories and catalogues.

Other communication methods should also be employed. This could include notices in the "In The Loop", seminars on project activities, initiation of requirements surveys, and others.

Recommendation 7.64: NSDMC will maintain a web site containing relevant documents and links related to its activities.

Recommendation 7.65: NSDMC will undertake appropriate activities to communicate the activities of the committee and supported data management activities.

8. CONCLUSION

Science is an ever changing endeavour that continually explores the limits of our knowledge and capabilities. The data and information resulting from this exploration is widely varied and evolves along with science.

DFO has a substantial financial investment in the data it acquires. Managing the data so that they are available to future users is not only financially prudent, but is the only way that long term trends in our environment can be assessed.

In adopting a data policy, Science began the process of building a formalized data management system for itself. The next step in this process is the development of a strategy to take the intentions of the data policy and to turn these into a reality. This Strategy lays out the general approaches to address the numerous issues of data management that when completed will forge a national data system.

Annex I. Science Data Management Policy.

June 12, 2001

Management Policy for Scientific Data

Preamble

Fisheries and Oceans Canada, through its own programs and through exchanges with national and international organisations, has acquired a large volume of scientific data and information over the years, and manages these through a set of practices evolved over the years. Since these historical data sets are an extremely valuable and irreplaceable resource of the Department, it is essential to develop and implement a Science and Oceans data management policy to ensure the preservation and enhancement of the data, while facilitating efficient and appropriate utilisation. It is recognised that this policy has to be consistent with the many data sharing arrangements the Department has with external agencies in Canada and international organisations and with the obligations associated with these arrangements. The policy will have to be flexible enough to permit effective new partnerships and to be responsive to new priorities. The intent of this policy is to safeguard the present and future holdings of scientific data, to strengthen the promotion of data interconnectivity, to maximise the usefulness of existing data through standards, and to determine cost-effective ways to manage data holdings. The implementation of such a policy is consistent with the Government of Canada's initiative to rationalise and improve the overall cost-effectiveness of its data holdings.

Priorities Influencing the Policy

This policy is based on current Departmental priorities, which include:

- Support scientific research projects and resource assessments at a regional, zonal, national, or international level;
- Provide scientific information and data on ocean, coastal and inland waters and ecosystems in support of integrated resource management, conservation of marine, anadromous and freshwater fishery resources, and the sustainable development of aquaculture;
- Provide scientific information and data for the achievement of marine and freshwater environmental and fish habitat protection and conservation through an integrated approach;
- Support the information and data requirements for marine services, transportation, and navigation;
- Support the Departmental responsibility to review environmental impact assessments for approval of environmental design parameters associated with offshore, coastal zone and inland waters development;
- Collaborate with other federal and provincial governmental departments to ensure greater flexibility in timely and cost-effective access to data and information;
- Provide scientific information in support of policy development in the department;
- Support Canada's commitment to international organizations.

Basic Principles

1. Fisheries and Oceans Canada (DFO) scientific data sets are a valuable national resource that have been acquired through decades of investment, enabling the Department to maintain world leadership in aquatic sciences and aquatic management. These data are irreplaceable, and must be protected and managed to ensure long-term availability.
2. Because of the complex and often unique nature of scientific data, it is essential that DFO Science/Oceans maintain responsibility for their quality control, management, archiving and dissemination.
3. To ensure proper management and archival of data, all scientific data collected by the Department must be migrated to a 'managed' archive immediately after the data have been processed.
4. To obtain maximum benefit to the Department and to the user community at large, scientific data must be made available in a timely manner with full and open access, consistent with Departmental, national and international obligations with respect to its data holdings.
5. To obtain access to international data and information that are pertinent to Canadian needs, Canada must be able to exchange its data with other world data centres, subject to the 'Exceptions' listed in the section 'Availability of Access' below.

Data Management Policies

Data Archiving

All DFO scientific data must be managed as part of an integrated system accessible through regional, zonal and national data centres. The Marine Environmental Data Service, Science Sector, (MEDS) will provide co-ordination among regional, zonal and national centres as appropriate, to ensure that all data are properly managed. Where no data management centre exists in a Region, Science and Oceans managers will be required to designate and support indeterminate A-base staff positions that include data management responsibilities.

MEDS will continue to function as a national data centre for Departmental data with archiving functions shared as appropriate with existing Regional data centres, and will serve as the primary point of contact for international data exchanges except in cases where the ADM Science or the ADM Oceans has designated in writing an alternate data centre as the primary contact.

The responsibilities of the integrated system of data centres will be to:

- Respond to internal and external data requests, in accordance with 'Availability of Access' Section below.
- Maintain inventories and documentation for all data holdings for which they have designated responsibility, including references to data sets not stored at the data centre.
- Provide basic data retrieval, integration and summarization capabilities to satisfy common requests.

- Provide or authorize computerized networking linkages.
- Perform, in concert with the data providers, data quality control, verification and removal of duplicate data.
- Ensure long term accessibility and documentation in the event of organizational changes, retirements, etc.
- Protect data against loss resulting from error, accident, technological change, degradation of media, etc.

In cooperation with Regional staff, MEDS may provide any or all of the above services on behalf of a Region, if so requested by that Region.

Data Submission

It is the responsibility of Science and Oceans managers to ensure that data collectors under their management submit their data as well as data collected under contract to or partnership with other agencies, to the appropriate data centre in a timely fashion. This is important to ensure that data are quickly migrated into a 'managed' environment where they are properly backed up and secured from accidental or circumstantial loss, and where the supporting metadata are integrated with the data to preserve the long-term usefulness of a data set.

Timely fashion will be taken to mean that: (a) data sets will be submitted immediately after the data are processed (b) submission will not be delayed while data analysis, statistical treatment, interpretation and publication occur, and (c) submission will include metadata prepared by the data collector to accompany the data set and document the methodologies and other details needed so that others are aware of the potential limitations of the data.

Data encompassed by this policy include data identified in Annex 1, and any other scientific data that may be created or otherwise acquired by DFO.

Exceptions to this policy are possible if: (a) the responsible manager and the responsible data centre have agreed that the data in question are not appropriate for submission, or (b) it can be demonstrated that there is a legal imperative (e.g. legal chain of custody requirements) that categorically prohibits submission of the excluded data, or (c) an extension or exemption from the policy is sought for other reasons and granted in writing by the Regional Science/Oceans Director.

Data submission to the responsible data centre does not mean that the data will be openly accessible. Thus concerns about access shall not be seen as a valid reason for not submitting data. It is the responsibility of the Regional Science/Oceans Director to designate data as classified for the purpose of preventing access to data which may not and must not be openly accessible.

Availability of Access

DFO scientific data are a public resource and subject to full and open access within two years of being acquired. In cases where, in the opinion of the Regional Science/Oceans Director, there may be a danger of improper or incorrect interpretation of the data, steps shall be taken to ensure that potential users are fully apprised of this possibility and a contact person should be identified who can provide assistance in proper use and interpretation.

Exceptions will be made to this policy in the event that one or more of the conditions below are met:

- DFO investigators have written approval from the Regional Science/Oceans Director to delay access to the data; in such cases, the letter of approval will include the rationale for the delay, and an agreed-upon date for the release of the data;
- There are third party agreements, privacy concerns, or legal restrictions;
- The data are of commercial benefit to DFO, in which case they will be managed according to Departmental intellectual property management regimes and prevailing policy. The data would be protected under s.18 of the Access to Information and Privacy Act.

Where there is uncertainty or dispute over whether a data set meets the second or third condition, legal advice shall be sought and followed.

Future third party agreements for the provision or exchange of data will certainly have an impact on data management in DFO and must therefore be approved by NSDC to ensure consistency with this Policy.

Inclusion of a Data Management Component in Science Project Plans

All science project proposals and plans must demonstrate the existence of a comprehensive data management plan, or must develop one if the existing infrastructure cannot adequately respond to the requirements of the project, to address the management of scientific data collected during the life of the underlying project. This plan must include strategies and schedules for the transfer of the data to the responsible data centre. The project budget must clearly indicate the allocation of resources for data management and how these resources will be used. The Regional Science/Oceans Director or their

designate will be responsible for conducting periodic reviews of data management activities to ensure that they are consistent with the plan.

National Inventory

A national inventory of DFO scientific data holdings will be maintained. It will be the responsibility of each designated data centre to maintain and update the inventories of its holdings. MEDS will be responsible for maintaining national links to all data inventories

and the infrastructure to ensure the inventories are nationally accessible.

Acquisition of Data from Third Party Sources

DFO Science and Oceans sectors should pursue the acquisition of relevant scientific data from other national and international sources where these data contribute to the goals of the Department. This must be done in an open and transparent manner and DFO's rights and duties must be agreed upon by all concerned parties and approved by NSDC.

Data Submitted under Regulations or Having Legal Aspects

Scientific data that have legal aspects constraining their distribution, whether collected by DFO or submitted by third parties, will be kept in their original form, and appropriately secured. If confidential data are submitted by third parties, a letter from the third party will be obtained indicating that the data are confidential. As well, the data manager responsible for that data set should designate the data as "Protected - Third Party Information".

Data Rescue

DFO Science and Oceans sectors will develop a national data rescue program to locate and preserve scientific data that are of value to departmental programs and may be in danger of being lost.

Application of Technology

Science and Oceans data centres will manage their data and will service users in an efficient manner by taking full advantage of current technology within the existing Informatics framework where appropriate.

Access to Information and Privacy Act Considerations

DFO Science and Oceans sectors will manage their data in a manner consistent with the Access to Information and Privacy Act (ATIP) and the requirement to document the location, status, and availability of the data consistent with good data management

practices. When scientific data are requested under the Act, MEDS officials or the responsible Science/Oceans Regional Director should provide the data to the ATIP Secretariat in HQ and inform ATIP as to whether the data are confidential (along with supporting rationale for confidentiality) or inform ATIP that the data can be disclosed.

Working Mechanisms

A permanent National Data Management Working Group (NDMWG), with representation from Regions and Sectors and a chairperson from MEDS, will be established, reporting to the ADM Science and ADM Oceans. MEDS will carry the secretariat function for the group. Annually, the group will review the data management activities, assess last year's performance against

plans and define the tasks and milestones for the coming year. MEDS will have the responsibility of presenting a report on the status of scientific data management to the ADM Science and ADM Oceans, and to make recommendations to correct any deficiencies that prevent the policy from meeting its objectives.

Implementation

It will be the responsibility of the Regional Directors of Science/Oceans to implement and ensure adherence to this policy. Inter-regional and inter-sectorial issues and concerns will be addressed by the ADM Science and ADM Oceans, as appropriate.

Contacts

For further information on this policy or on accessing the scientific data please contact:

Director, Marine Environmental Data Service
W082, 12th Floor, 200 Kent St.
Ottawa, Ontario, Canada, K1A 0E6
(Tel) 613-990-0265
(FAX) 613-993-4658
e-mail: services@meds-sdmm.dfo-mpo.gc.ca

Data Policy Annex 1:

Some Data Types Covered Under the Management Policy of DFO Scientific Data

A. Physical oceanographic data

B. Hydrological data (e.g. Flow volumes of streams and rivers)

C. Meteorological data

D. Biological oceanographic data

E. Marine chemistry data

F. Contaminants data

G. Fisheries data

- Biological data (from catch sampling, trawl and acoustic surveys, sentinel fisheries and industry surveys, science logbooks, etc.)
- Field and lab data in support to stocks' assessment process
- Fish health data

H. Freshwater and marine habitat data

I. Freshwater biological data

J. Experimental Lakes Area (ELA) data

K. Data collected by the Canadian Hydrographic Service, subject to CHS agreements and operational practices.

Annex II. The State of Physical Oceanographic Archives in 2002

Data Management Policy Implementation – Oceanography

I. Introduction

One of the tasks assigned during the National Science Data Management Workshop - September 2002 was to develop a data policy implementation strategy. A number of sub-groups were formed to report on Oceanography, Fisheries, Aquaculture, Hydrography and Environmental Science. This report is for Oceanography.

The terms of reference (Annex I) were established after the group began its work and are too ambitious to be fulfilled within the end of December time frame. The emphasis for this report has been to focus on a set of problems and priorities that are common to a number of regions, and propose specific initiatives to address these problems. Many of the objectives in the initial TOR will be addressed in these initiatives.

In order to keep the main body of the report as brief as possible, much of the detail is contained in a series of annexes at the end of the report.

II. Membership

There was representation for all regions and HQ. The names of the committee members are listed below.

Doug Gregory Maritimes - Lead	Dave Senciall Newfoundland	Robert Nowlan Gulf
Bernard Pelchat Quebec	Bob Keeley MEDS	Aaron Carswell C&A (Burlington)
Joe Linguanti Pacific	Christine Michel, Bernard LeBlanc	C&A (Winnipeg)

III. Process

The initial task was to first survey the regions and HQ to prepare an assessment of the current situation. A parallel exercise established objectives to allow us to assess the current situation against rated criteria. Regions then prepared a set of priorities based on the ratings. Ratings were collated to establish national or zonal priorities. The priorities were used to generate “proposals of intent” which are targeted at specific problem areas. These proposals, designed to more closely integrate to data management activities across regions, form the basis of the implementation plan for oceanography.

IV. Principles and Objectives

In carrying out the assessment, we developed criteria on which we can measure our performance in adhering to the principles of the data policy.

Principles

The data policy establishes the twin pillars of **Data Archival** and **Availability of Access.**

The objectives define the goals to be achieved to satisfy the principles of the policy. An over-riding aim is to provide some convergence in the way we handle and distribute our data across the regions. Objectives 1-5 are primarily concerned with Archival. Objectives 5-8 focus on Access. Note that conforming to international standards (#5), is common to both activities.

Objectives

1. Data are maintained by a designated data center (not an individual)
2. Data are maintained in a managed environment with formal backup and archival procedures
3. All data subjected to standard processing procedures
4. Data versions are controlled
5. Metadata conform to international standards
6. Data access is provided by a designated data center (not an individual)
7. Data access is co-coordinated nationally, or at a minimum, zonally
8. Data are accessible through a web/ftp portal to the public, or minimum, DFO.

V. Assessment

The first phase of the project was to establish baseline information on the current situation within each region. Because of the wide range of data encompassing the oceanography sub-group, the focus was data management on a parameter basis. The information compiled is described in Annex II Assessment Process. The detailed survey results are available from Gregory as an Access table or Excel spreadsheet.

A total of 92 entries from the six regions and MEDS were returned. Due to variations in reporting, these entries were combined into “parameter groups” which were based on the traditional view of the management of oceanographic data within the Department. These parameter groups are described in Annex II.

The individual entries were all rated against the objectives in section IV. A simple satisfies / does not satisfy rating was used. The results are tabulated in Annex III Assessment Results.

V.1 By Objective

It is obvious from looking at the results in Annex III that we have a long way to go to meet the overall objectives. It is also equally apparent that we have made considerable progress in identifying designated data managers in all of the regions. The concept of a designated data centre or data manager was new in the policy, and with a 72% rating it is clear that management has taken some positive steps to ensure this happens.

On all the other criteria, the results are poor with failing grades in every category.

V.2 By Parameter Group (ordered by “Badness”)

Biology – This was the largest single category representing a quarter of the entries and encompassing all regions. It is also the least well managed. The BioChem project was designed

to address this shortcoming and implementation in the Maritime region has been relatively successful.

Drifters – Newfoundland, Maritimes and Pacific reported these data. MEDS is a WDC for drifting buoy data. The older legacy data has no clear data manager ownership in any of the regions. There is some cause for optimism as the new Argo Palace float project evolves. These data are managed nationally and internationally from the outset.

Water Chemistry – All regions reported water chemistry data and is being managed only marginally better than biological data. The BioChem project also addresses these data.

TS Profiles – This is our best managed dataset with a long history of a national archival program. Major areas for improvement would be standardization of processing protocols and software. Open access is readily available, but in a wide variety of forms and formats. Standardized access would be a major improvement.

Underway Currents – These data are in a similar state as drifting buoy data but lacking the benefit of a national focus. The regions involved are Newfoundland, Maritimes, Pacific and MEDS.

TS Series – Moored thermograph data are reported from all regions except MEDS and C&A with archival centred in Maritimes, Quebec and Pacific regions. Improvements could be made to co-ordination of both archival and access to these data.

Underway TS – Data were reported from Maritimes, Quebec, MEDS and Pacific. There is little in way of co-ordination or even similarity of the programs. An international project, the Global Ocean Surface Underway Data Project, is forming around underway data. This may be used as a vehicle to organize our national holdings.

Moored Currents – This is a reasonably well managed parameter group with archival and access concentrated within Maritimes and Pacific regions. Improvements could be made in standardization of processing and access.

Remote Sensing – Data were reported by Newfoundland, Maritimes, Quebec and Pacific regions, although Newfoundland is no longer actively collecting data. The data appear to be reasonably well managed with Maritimes, Quebec and Pacific although all three regions are functioning autonomously with limited coordination.

Optics and Acoustics – These are grouped together because they are single-issue parameter groups with Maritimes Region.

Waves and Water Levels – These are both national parameter groups managed by MEDS. No other region reported these data.

VI. Priorities

Each region was asked to assign a **H**igh, **M**edium, or **L**ow priority to each objective within a parameter group. To assist in the collating, a High response was re-assigned a rating of 2; Medium was given a value of 1. Low or no response was assigned a zero. This is a somewhat arbitrary rating, but the overall results were so striking it is unlikely a different rating scheme would make any difference.

The collated priorities are reported in Annex IV Priorities. The individual responses on a regional basis are reported in Annex V.

Within the objectives, the requirement for standardized processing was the main priority. Data versioning, National or zonal co-ordinated access, and common access through an ftp or web portal were all close together in a second group.

The overall priorities look much like the assessment with two exceptions. Drifters and Underway currents (ADCP) were given a lower priority than their “assessment health” would suggest. Neither result is surprising. Much of the drifter data is of the legacy category and is a data rescue problem. The low priority for underway currents is due to the fact we are not sure what to do about the problem.

Biological data (Plankton) and water chemistry data were the #1 and #2 priorities. Both groups associated a high priority to all objectives with the exception of a much stronger response to objective 2 for biological data. This suggests data rescue is a bigger issue for biological data than for water chemistry.

Temperature / salinity profile data was the #3 priority with the primary issues being standardized processing and portal access. Controlled versioning and national or zonal co-ordination were also strong priorities.

The fourth and fifth priorities (T/S series and Moored current meters) could readily be combined, as the archival, processing and access procedures are very similar for both. The priority objectives, data versioning, co-ordination and standardized processing were also the same.

Physical Oceanographic Archives: Annex I – Terms of Reference

- Describe the data processing/management and dissemination system, and the roles & responsibilities of each region, for each subset of the data in that group.
- Document the QC procedures for each, and assess consistency among regions.
- Develop the strategy for respecting the designated authority/ownership of the region collecting data while managing databases as a national asset with access / delivery through national or zonal portal.
- Address the life-cycle management issues for the system(s) that is (are) selected for each data type.
- Objectively assess the best system for management of each data type, designated archival centre, designated data manager and the associated responsibilities.

Physical Oceanographic Archives: Annex II – Assessment Process

Survey Information - Regional representatives were asked to report on a number of topics consisting of:

- Parameter name
- Data Manager – name of primary contact
- Primary Archival location – i.e. national, zonal, regional, individual
- Secondary Archival Location - i.e. national, zonal, regional, individual
- Archival Format System – brief description of software / formats
- Archival Protocol – description of archival process
- Primary Access – description of who can access and how
- Secondary Access – description of who can access and how
- Access Format System – brief description of software / formats
- Access Protocol – description of how to access the data

Parameter Grouping

Parameter Group	Parameter Description
TS Underway	Batfish, moving vessel samples
TS Series	moored thermographs, salinographs, lighthouse time series
UnderwayCurrents	Moving Vessel ADCP
Remote Sensing	SST, ocean colour, CODAR, HF radar
Waterchemistry	discrete water chemistry
Biology	Phytoplankton, Zooplankton, Benthos, pigments, bacteria
Moored Currents	Moored current measurements, conventional and ADCP
Waves	various types of wave measurements
Water Levels	Tides and Water levels
Optics	Optical data including VOPC
Acoustics	Acoustic ADCP, multi-frequency
Drifters	Drifting Buoys, Palace floats, profiling floats
TS Profiles	CTD, XBT, bottle, BATHY, TESAC

Physical Oceanographic Archives: Annex III – Assessment Results

Objectives 4 and 5 were not rated due to insufficient information in the initial survey. The number in each column refers to the number of times an individual met an objective.

Parameter Group	No. of Entries	#1 Designated Data Center	#2 Formal Backups	#3 Standard Processing	#6 Designated Access	#7 Nat / zonal co-ordination	#8 Web / ftp portal	Compliance
Biology	24	15	2	0	4	0	3	17%
Drifters	9	4	3	3	3	3	1	31%
Waterchemistry	8	7	2	1	3	2	2	35%
Tsprofiles	11	10	6	6	6	6	2	55%
UnderwayCurrents	5	2	0	0	2	0	0	13%
Tsseries	9	9	5	5	5	4	4	59%
Tsunderway	5	3	4	4	3	0	2	53%
MooredCurrents	6	5	4	3	5	3	3	64%
RemoteSensing	4	2	1	2	3	0	3	46%
Optics	2	2	0	0	0	0	0	17%
Acoustics	2	2	0	0	0	0	0	17%
Waves	1	1	1	1	1	1	0	83%
Waterlevel	1	1	1	1	1	1	0	83%
Overall	86	63	29	26	36	20	20	
		73%	34%	30%	42%	23%	23%	

Physical Oceanographic Archives: Annex IV – Priorities

Parameter Group	#1 Designated Data Center	#2 Formal Backups	#3 Standard Processing	#4 Data Versioning	#5 Metadata std.	#6 Designated Access	#7 Nat- zonal coordination	#8 Web / ftp portal	Total
Biology	9	11	13	12.5	12	12.5	13	11	94
Waterchemistry	6	4	13	9.5	7	9.5	11	10	70
T/S profiles	4	4	10	8.5	7	2.5	4	10	50
T/S series	4	3	7	8.5	5	4.5	8	5	45
MooredCurrents	3	2	6	6	3	3	6	3	32
Drifters	4	3	5	3	2	4	5	4	30
UnderwayTS	4	3	6	3	1	3	4	3	27
UnderwayCurrent	3	2	5	3	3	3	4	3	26
Optics	3	3	3	2	2	3	2	2	20
RemoteSensing	1	1	2	1	1	2	0	1	9
WaterLevels	1	0	1	2	0	0	1	2	7
Acoustics	1	1	1	0	0	1	0	1	5
Waves	0	0	0	0	0	0	0	0	0
OVERALL	43	37	72	59	43	48	58	55	

Objectives

1. Data are maintained by a designated data center (not an individual)
2. Data are maintained in a managed environment with formal backup and archival procedures
3. All data subjected to standard processing procedures
4. Data versions are controlled
5. Metadata conform to international standards
6. Data access is provided by a designated data center (not an individual)
7. Data access is co-ordinated nationally, or at a minimum, zonally
8. Data are accessible through a web/ftp portal to the public, or at minimum, to all DFO.

Physical Oceanographic Archives: Annex V – Regional Priority Response

- O1. Data are maintained by a designated data center (not an individual)
- O2. Data are maintained in a managed environment with formal backup and archival procedures
- O3. All data subjected to standard processing procedures
- O4. Data versions are controlled
- O5. Metadata conform to international standards
- O6. Data access is provided by a designated data center (not an individual)
- O7. Data access is co-coordinated nationally, or at a minimum, zonally
- O8. Data are accessible through a web/ftp portal to the public, or at a minimum, to all DFO.

NFLD Region

Parameter Groups	1	2	3	4	5	6	7	8	Comments
Biology	H	H	M	M	H	H	H	H	
T/S profiles	L	L	M	M	M	L	L	H	Standards, access
TSseries	M	M	M	H	M	H	H	H	Zonal coordination desirable??
Drifters	M	M	M	M	L	M	M	M	Data is(??) given to MEDS
Waterchemistry	M	M	M	M	M	M	M	M	
MooredCurrents	M	M	M	M	M	M	M	M	Zonal coordination desirable??
UnderwayCurrents	M	M	H	M	M	M	M	M	ADCP about to be reviewed
UnderwayTS									Not collected
RemoteSensing	L	L	L	L	L	L	L	L	No longer collected, archived could do with work
Optics									PAR no sure how to use it yet

Maritimes Region

Parameter Groups	1	2	3	4	5	6	7	8	Comments
Biology	M	H	H	H	H	H	H	M	our main problem area
T/S profiles	L	L	M	M	M	L	L	M	standards and versioning
TSseries	L	L	M	H	M	L	H	L	need to share within the zone
Drifters	M	M	M	L	L	M	M	M	we don't have a process in place
Waterchemistry	L	L	H	M	H	M	M	M	lot of progress made, require standards
MooredCurrents	L	L	M	H	M	L	H	L	need to share within the zone
UnderwayCurrents	M	M	M	L	L	M	M	L	we don't have a process in place
UnderwayTS	M	M	M	L	L	M	M	L	we don't have a process in place
RemoteSensing	L	L	M	L	L	L	L	L	process in place functions well
Optics	M	M	M	L	L	M	L	l	we don't have a process in place
Acoustics	M	M	M	L	L	M	L	M	we don't have a process in place
WaterLevels	L	L	L	L	L	L	L	L	limited data

Gulf Region

Parameter Groups	1	2	3	4	5	6	7	8	Comments
Biology:	M	L	M	M	H	H	M	M	Marc Ouellette (mollusks)
	M	L	L	M	L	H	M	L	Marc Ouellette (mollusks video)
	L	L	L	L	L	L	L	L	Susan Bates (MES)
T/S profiles:	M	L	M	M	H	H	M	M	Marc Ouellette (mollusks)
	L	?	?	?	?	M	M	M	Doug Swain (MFD) nice to have a one-stop shop instead of both BIO & IML
T/S Series:	M	L	M	M	H	H	M	M	Marc Ouellette (mollusks)
	L	L	M	L	M	L	M	M	Denis Gagnon (Lobs)
	L	L	M	L	M	L	M	H	Elmer Wade (crab)
	L	?	?	?	?	M	M	M	Doug Swain (MFD) nice to have a one-stop shop instead of BIO & IML
Waterchemistry	L	?	?	?	?	M	M	M	Doug Swain (MFD) nice to have a one-stop shop instead of BIO & IML

Québec Region

Parameter Groups	1	2	3	4	5	6	7	8	Comments
Biology	L	H	H	H	M	L	H	M	1 and 6 are linked. 8 is M to all DFO and L to the public. Also linked to 1 and 6.
T/S profiles	L	L	M	M	L	L	M	L	Well managed. Multiple access point.
T/S series	L	L	M	M	L	L	M	L	Although they look like T/S, they need more works. Multiple entry point.
Drifters									We don't have this type of data.
Waterchemistry	L	L	M	M	L	L	M	L	Zonal coordination is required like Biology.
MooredCurrents	L	L	M	M	L	L	M	L	Idem T/S profiles and T/S series
UnderwayCurrents	L	L	L	L	L	L	L	L	Marginal data set. Don't know how to prioritize
UnderwayTS	L	L	M	L	L	L	L	L	We routinely collect thermosalinograph data. Well managed and accessible.
RemoteSensing	L	M	L	L	L	L	L	L	Formal backup procedure but always faced with the problem of high volume

MEDS

Parameter Groups	1	2	3	4	5	6	7	8	Comments
Biology									We only have some JGOFS data and some chlorophyll, etc.
T/S profiles	L	L	M	M	L	L	L	H	Standards and versions not perfectly controlled
TSseries									We have some thermograph data and some from lighthouses. We are not big players in this type.
Drifters	M	L	H	M	M	M	H	M	The concern centres around data residing in the regions that do not come to MEDS
Water chemistry	M	L	H	H		H	H	H	
Underway TS	H	M	H	H		M	H	M	MEDS handles data arriving from the GTS and some delayed data.
Optics									MEDS has some PAR, turbidity data now
Acoustics									MEDS has some sound velocity profile data
Waves	L	L	L	L	L	L	L	L	MEDS acquires the data mostly now collected by MSC
Waterlevels	M	L	L	M	L	L	M	H	

Central and Arctic (Sault Ste. Marie and Burlington)

Parameter Groups	1	2	3	4	5	6	7	8	Comments
Biology	H	M	H	H/M	H	H/M	H	H	Meta data has to be assessed against international standards. Overall, Burlington and Sault Ste. Marie are in it's infancy. We are just starting to work on achieving the above mentioned objectives for data sets here in this region.
T/S profiles	H	M	H	H/M	H	H/M	H	H	Same as above
T/S series	H	M	H	H/M	H	H/M	H	H	Same as above
Drifters									
Waterchemistry	H	M	H	H/M	H	H/M	H	H	Same as above

Central and Arctic, Winnipeg

Parameter Groups	1	2	3	4	5	6	7	8	Comments
Biology	M	H	H	H	H	H	H	M	Data in the High Canadian Arctic (2001-2002) and in Beaufort sea (Cases 2002).
T/S profiles	M	H	H	M	M	L	L	H	Two years of data in the High Canadian Arctic (2001-2002)
TSseries									No priorities define. Limited data. Two sets of data, two months mooring under a first year sea-ice cover.
Waterchemistry			H	M	L	H	H	H	Data in the High Canadian Arctic (2001-2002) and in Beaufort sea (Cases 2002).
Optics									No priorities define. Limited data. Under ice PAR profiles, Solar Radiation

Pacific

Parameter Groups	1	2	3	4	5	6	7	8	Comments
Biology (Phyto)	H	H	H	H	H	H	H	H	No processes in place
Biology(Zoop)	L	L	M	M	L	M	L	M	Standards and versions not perfectly controlled
T/S profiles	L	L	M	M	L	L	L	L	Standards and versions not perfectly controlled
TSseries	L	L	M	M	L	L	L	L	Standards and versions not perfectly controlled
Drifters	M	M	M	M	M	M	M	M	Drifting buoys only as Palace and Profiling buoys are dealt with at MEDS
Waterchemistry	H	M	H	M	H	M	M	M	Our highest priority, especially in value of impending retirement of key personnel
MooredCurrents (conventional)	L	L	M	M	L	L	L	L	Standards and versions not perfectly controlled
MooredCurrents (ADCP)	H	M	H	M	M	H	H	H	No processes in place. Our second priority
UnderwayCurrents	M	L	H	H	H	M	H	H	Question about scientific return on investment required.
UnderwayTS	M	M	H	M	M	M	M	H	Demand dependent as to whether it gets collected or processed.
RemoteSensing	M	L	M	M	M	H	L	M	Requires work on determining useful products for range of clients
Optics	H	H	H	H	H	H	H	H	Small volume now but building in importance. Our third priority
WaterLevels	L	L	M	M	L	L	L	L	Standards and versions not perfectly controlled. CHS is primary agency for this data in Pacific

Annex III. The State of Fisheries Data Archives in 2002.

Data Management Policy Implementation – Fisheries

I. Introduction

During the National Science Data Management Workshop - September 2002, a number of sub-groups were formed to report on Oceanography, Fisheries, Aquaculture, Hydrography and Environmental Science. These reports were to be used to develop a data policy implementation strategy. This report is for Fisheries.

The terms of reference (Annex I) were established during the workshop. The emphasis for this report has been to focus on a set of problems and priorities that are common to a number of regions, and propose specific initiatives to address these problems. Many of the objectives in the initial TOR will be addressed in these initiatives.

In order to keep the main body of the report as brief as possible, much of the detail is contained in a series of annexes at the end of the report.

II. Membership

The names of the committee members are listed below.

Sylvain Hurtubise Quebec - Lead	Darlene Fiander - Newfoundland	Shelley Bond - Maritimes
Bob Branton - Maritimes	Gloria Poirier - Gulf	
Bruce Patten - Pacific	Eugene Murphy - Newfoundland	

III. Process

The initial task was to first survey the regions to prepare an assessment of the current situation. A parallel exercise established objectives to allow us to assess the current situation against rated criteria. Regions then prepared a set of priorities based on the ratings. Ratings were collated to establish national priorities. Information describing the state of fisheries data management was provided by all the regions. The priorities were used to generate “proposals of intent” which are targeted at specific problem areas. These proposals, designed to more closely integrate to data management activities across regions, form the basis of the implementation plan for fisheries data.

IV. Principles and Objectives

In carrying out the assessment, we developed criteria on which we can measure our performance in adhering to the principles of the data policy.

Principles

The data policy establishes the twin pillars of **Data Archival** and **Availability of Access.**

The objectives define the goals to be achieved to satisfy the principles of the policy. An over-riding aim is to provide some convergence in the way we handle and distribute our data across the regions. Objectives 1-5 are primarily concerned with Archival. Objectives 5-8 focus on Access. Note that conforming to international standards (#5), is common to both activities.

Objectives

1. Data are physically and logically secured by a designated data center
2. Data are maintained in a managed environment with formal backup and archival procedures
3. All data subjected to standard processing procedures
4. Data versions are controlled
5. Metadata conform to international standards
6. Data access is provided by a designated data center
7. Data access is co-coordinated nationally, or at a minimum, zonally
8. Data are accessible through a web/ftp portal to the public, or minimum, DFO.

V. Assessment

The first phase of the project was to establish baseline information on the current situation within each region. Because of the wide range of data encompassing the fisheries data sub-group, the focus was data management on a parameter basis. The information compiled is described in Annex II Assessment Process. The detailed survey results are available from Hurtubise as an Excel spreadsheet.

A total of 356 entries were returned from the Atlantic regions. To help in compiling, these entries were combined into "parameter groups" which were based on the different sources of fisheries data within the Department. A brief description of each category is given in Section V.2.

The individual entries were all rated against the objectives in section IV. A simple satisfies / does not satisfy rating was used. The results are tabulated in Annex III Assessment Results. It should also be noted that, due to the large number of datasets (near 1200), the Pacific information has not been included in the Annex III table. Otherwise, the summary table would have been much like the Pacific table. The Pacific table is however presented in the Annex IV.

V.1 By Objective

It is obvious from looking at the results in Annex III that we have a long way to go to meet the overall objectives. The concept of a designated data centre or data manager was new in the policy, and with a 16% rating it is clear that some steps has still to be taken to ensure this happens.

The best results are obtained with the "Formal backups" and "Standard processing" objectives, although there is still a lot of work to do.

V.2 By Category/Parameter Group

A. Assessment - This category is meant to include all the data that are collected through DFO surveys.

- **Catch & effort** – Well managed as far as the formal backups are concerned. Results related to the other objectives are relatively poor.
- **Biological data** – The largest parameter group in terms of number of entries. The best results are obtained with the "Formal backups" (45%) and "Standard processing" (47%) objectives.
- **Acoustics** – Even if the number of entries was very low, this parameter group could not be included into the two previous ones. It follows the overall pattern.

B. Assessment/Commercial - This category is dealing with datasets from the Maritimes Region that include data from assessment and commercial categories that are linked together.

- **Biological data** – The number of entries is low. The results are a little bit above the average.

C. Recreational - This category is meant to include all the data that comes from recreational fisheries.

- **Catch & effort** – The number of entries is low. It follows the overall pattern.
- **Biological data** –There is still no entry for that parameter group. We would likely get some results if Pacific datasets were to be considered. It was left in the table because it was assessed within the overall priorities.
- **Economic impact data** – There is still no entry for that parameter group. We would likely get some results if Pacific datasets were to be considered. It was left in the table because it was assessed within the overall priorities.

D. First nation - This category is meant to include all the data that comes from First nation fisheries.

- **Catch & effort** – There is still no entry for that parameter group. We would likely get some results if Pacific datasets were to be considered. It was left in the table because it was assessed within the overall priorities.
- **Biological data** –There is still no entry for that parameter group. We would likely get some results if Pacific datasets were to be considered. It was left in the table because it was assessed within the overall priorities.

E. Commercial - This category is meant to include all the data that comes from the traditional commercial fisheries.

- **Catch & effort** – The number of entries is low. The best results are obtained with the "Formal backups" (71%) and "Standard processing" (43%) objectives.
- **Biological data** – The second largest parameter group in terms of number of entries. Again, the best results are obtained with the "Formal backups" (56%) and "Standard processing" (59%) objectives.

VI. Priorities

Each region was asked to assign a **H**igh, **M**edium, or **L**ow priority to each objective within a parameter group. To assist in the collating, a High response was re-assigned a rating of 2; Medium was given a value of 1. Low or no response was assigned a zero. In some cases, certain parameter groups within a region were assigned a H/M/L priority. Those specific cases were re-assigned a value of 1.

The collated priorities are reported in Annex V Priorities. The individual responses on a regional basis are reported in Annex VI.

Within the objectives, the requirement for a designated data center and formal backups are the main priorities. Designated access and common access through a Web/FTP Portal come in a second group. Standard processing and national/zonal co-ordination are forming a third group and finally data versioning and metadata standardization would be the last objectives to set priorities on.

Some of the overall results are surprising as, for instance, one could expect that a minimum of metadata standardization is performed before having a common access to the data.

Commercial and assessment categories were the #1 and #2 priorities. For each parameter group of these

categories, at least a medium priority was associated to all objectives, with a stronger signal for a designated data center.

Fisheries Data Archives: Annex I – Terms of Reference

- Describe the data processing/management and dissemination system, and the roles & responsibilities of each region, for each subset of the data in that group.
- Document the QC procedures for each, and assess consistency among regions.
- Develop the strategy for respecting the designated authority/ownership of the region collecting data while managing databases as a national asset with access / delivery through national or zonal portal.
- Address the life-cycle management issues for the system(s) that is (are) selected for each data type.
- Objectively assess the best system for management of each data type, designated archival centre, designated data manager and the associated responsibilities.

Fisheries Data Archives: Annex II – Assessment Process

Survey Information – For each of the parameter group described in part V.2 of the document, regional representatives were asked to report on a number of topics consisting of**:**

- Parameter name
- Data Manager – name of primary contact
- Primary Archival location – i.e. national, zonal, regional, individual
- Secondary Archival Location - i.e. national, zonal, regional, individual
- Archival Format System – brief description of software / formats
- Archival Protocol – description of archival process
- Primary Access – description of who can access and how
- Access Format System – brief description of software / formats
- Access Protocol – description of how to access the data

Fisheries Data Archives: Annex III – Assessment Results

Objectives 4 and 5 were not rated due to insufficient information in the initial survey. The number in each column refers to the number of times an individual entry met an objective.

Category	Parameter group	No of entries	Designated Data Center	Formal Backups	Standard Processing	Designated Access	Nat/zonal cordination	Web/FTP Portal	Compliance
Assessment	Catch & Effort	20	3	16	5	3	0	1	27%
	Biological data	257	32	115	122	14	0	9	19%
	Acoustics	4	0	2	2	0	0	0	17%
Assessment/ Commercial	Biological data	5	2	2	3	1	0	0	27%
Recreational	Catch & Effort	4	1	2	1	1	0	0	21%
	Biological data								
	Economic impact data								
First nation	Catch & Effort								
	Biological data								
Commercial	Catch & Effort	7	1	5	3	1	1	0	26%
	Biological data	59	19	33	35	4	0	0	26%
Overall		**356**	**58**	**175**	**171**	**24**	**1**	**10**	
			16%	**49%**	**48%**	**7%**	**0%**	**3%**	

Fisheries Data Archives: Annex IV. – Pacific Fishery Dataset Assessment - Best Guess Results

Objectives 4 and 5 were not rated due to insufficient information in the initial survey. The number in each column refers to the number of times an individual entry met an objective.

Category	Parameter group	No of entries	Designated Data Center	Formal Backups	Standard Processing	Designated Access	Nat/zonal Coordination	Web/FTP Portal	Compliance
Assessment	Catch & Effort	50+	5	3	2	3	3	2	6%
	Biological data	500+	3	?	2	3	2	0	0.5%
	Acoustics	10+	1	?	?	0	0	0	2.5%
Recreational	Catch & Effort	20+	1	?	0	1	1	1	4%
	Biological data	100+	0	?	0	0	0	0	0%
First nation	Catch & Effort	10+	0	1	0	0	0	0	2%
	Biological data	10+	0	?	0	0	0	0	0%
Commercial	Catch & Effort	20+	5	3	5	5	3	3	20%
	Biological data	500+	3	2	3	3	?	1	0.5%
Overall		**2020**	**18**	**8**	**12**	**15**	**9**	**7**	

Fisheries Data Archives: Annex V – Priorities

Category	Parameter group	#1	#2	#3	#4	#5	#6	#7	#8	**Total**
Assessment	Catch & Effort	7	7	4	3	2	6	4	5	**38**
	Biological data	6	6	4	3	2	5	4	5	**35**
	Acoustics	6	6	4	3	2	5	4	5	**35**
Assessment/Commercial	Biological data	2	2	1	0	0	2	1	2	**10**
Recreational	Catch & Effort	4	4	3	2	1	3	3	4	**24**
	Biological data	4	4	3	2	1	3	3	4	**24**
	Economic impact data	2	2	1	0	1	2	1	2	**11**
First nation	Catch & Effort	4	4	2	1	2	4	2	3	**22**
	Biological data	4	4	2	1	2	4	2	3	**22**
Commercial	Catch & Effort	10	10	7	5	4	5	5	4	**50**
	Biological data	9	9	6	4	4	5	5	4	**46**
Overall		**58**	**58**	**37**	**24**	**21**	**44**	**34**	**41**	

Objectives

1. Data are physically and logically secured by a designated data center
2. Data are maintained in a managed environment with formal backup and archival procedures
3. All data subjected to standard processing procedures
4. Data versions are controlled
5. Metadata conform to international standards
6. Data access is provided by a designated data center
7. Data access is co-ordinated nationally, or at a minimum, zonally
8. Data are accessible through a web/ftp portal to the public, or at minimum, to all DFO.

Fisheries Data Archives: Annex VI – Regional Priority Response

O1. Data are physically and logically secured by a designated data center
O2. Data are maintained in a managed environment with formal backup and archival procedures
O3. All data subjected to standard processing procedures
O4. Data versions are controlled
O5. Metadata conform to international standards
O6. Data access is provided by a designated data center
O7. Data access is co-coordinated nationally, or at a minimum, zonally
O8. Data are accessible through a web/ftp portal to the public, or at a minimum, to all DFO.

NFLD Region

Category	Parameter group	#1	#2	#3	#4	#5	#6	#7	#8
Assessment	**Catch & Effort**	L	L	L	L	L	L	M	M
	Biological data	L	L	L	L	L	L	M	M
	Acoustics	L	L	L	L	L	L	M	M
Recreational	**Catch & Effort**	L	L	L	L	L	L	M	M
	Biological data	L	L	L	L	L	L	M	M
	Economic impact data								
First nation	**Catch & Effort**	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	Biological data	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Commercial	**Catch & Effort**	H	L	M	H	H	L	H	L
	Biological data	H	L	M	H	H	L	H	L

Maritimes region

Category	Parameter group	#1	#2	#3	#4	#5	#6	#7	#8
Assessment	**Catch & Effort**	H	H	L	M	M	H	M	H
	Biological data	H	H	L	M	M	H	M	H
	Acoustics	H	H	L	M	M	H	M	H
Assessment/Commercial	**Biological data**	H	H	L	M	M	H	M	H
Recreational	**Catch & Effort**	H	H	L	M	M	H	M	H
	Biological data	H	H	L	M	M	H	M	H
	Economic impact data	H	H	L	M	M	H	M	H
First nation	**Catch & Effort**	H	H	L	M	M	H	M	H
	Biological data	H	H	L	M	M	H	M	H
Commercial	**Catch & Effort**	H	H	L	M	M	H	M	H
	Biological data	H	H	L	M	M	H	M	H

Gulf region

Category	Parameter group	#1	#2	#3	#4	#5	#6	#7	#8
Assessment	**Catch & Effort**	H	H	H	H	L	L/M/H	M	L/M/H
	Biological data	H	H	H	H	L	L/M/H	M	L/M/H
	Acoustics	H	H	H	H	L	L/M/H	M	L/M/H
Recreational	**Catch & Effort**	H	H	H	H	L	M	M	L/M/H
	Biological data	H	H	H	H	L	M	M	L/M/H
First nation	**Catch & Effort**								
	Biological data								
Commercial	**Catch & Effort**	H	H	H	H	L	L/M/H	M	L/M/H
	Biological data	H	H	H	H	L	L/M/H	M	L/M/H

The **designated data center**, is not necessarily understood to be anything other than a regional centre in these responses. The **standard processing procedures** are standard to at least a sectional or regional level. People seem to understand the commercial data to be data collected and managed by other branches within the department; but in as much as they are very important to us, it is important that they are well-managed and accessible when necessary. We're not sure we have any recreational data - we understand a research/commercial data split

H: high priority;
L: L priority;
M: priority somewhere between L and high;
L/M/H: priority pretty much split along section lines

Quebec region

Category	Parameter group	#1	#2	#3	#4	#5	#6	#7	#8
Assessment	**Catch & Effort**	H	H	M	M	M	H	M	M
	Biological data	H	H	M	M	M	H	M	M
	Acoustics	H	H	M	M	M	H	M	M
Recreational	**Catch & Effort**								
	Biological data								
	Economic impact data								
First nation	**Catch & Effort**	H	H	M	M	M	H	M	M
	Biological data	H	H	M	M	M	H	M	M
Commercial	**Catch & Effort**	H	H	M	M	M	H	M	M
	Biological data	H	H	M	M	M	H	M	M

Pacific Region

Category	Parameter group	#1	#2	#3	#4	#5	#6	#7	#8
Assessment	**Catch & Effort**	M	M	L	L	L	M	L	L
	Biological data	L	L	L	L	L	L	L	L
	Acoustics	L	L	L	L	L	L	L	L
Recreational	**Catch & Effort**	L	L	L	L	L	L	L	L
	Biological data	L	L	L	L	L	L	L	L
	Economic impact data	L	L	L	L	L	L	L	L
First Nation	**Catch & Effort**	L	L	L	L	L	L	L	L
	Biological data	L	L	L	L	L	L	L	L
Commercial	**Catch & Effort**	H	H	M	M	L	L	L	L
	Biological data	M	M	L	L	L	L	L	L

Annex IV. A Project Data Management Plan

Last updated: 22 Feb, 2006

A Project Data Management Plan

Projects are of such wide variety that it is not possible to provide detailed specifications that will cover all cases. What can be said is that there needs to be collaboration between data managers and scientific staff to manage the resulting data from a project.

We use the term "data system" to describe the formal procedure of managing data. The procedures include the routine capability to accept, process, archive and provide data to other parties.

The plan should address the following points in sufficient detail that it can form the framework for developing the detailed implementation specifications for managing the data.

- What are the variables to be measured, at sea or in the lab, or produced by the project and for how long must they be kept?
- Are there value added products that must be archived and for how long?
- Will there be physical samples collected and archived?
- Is the instrumentation to be used similar to instrumentation used to collect data already managed by the data system?
- What volumes of data are expected, and when, how and how often will they be presented to the data system?
- How many different data streams will provide the data to the data system?
- Is there a native format for the data coming from the instruments or is there some processing that will be done before the data are presented to the data system?
- Are there procedures and archives already in place within the data system to receive the data from the project? Which ones do you intend to use?
- Are there required associations between the data collected by this project and data collected by another or an existing archive?
- What information is needed to describe the data collection so that another user will be fully informed of the characteristics of the data?
- What, if any, are the restrictions in distribution of the data?
- Are there established and documented procedures to assess the quality of the returned data and will this procedure be managed by the data system, by the project, or by some combination?

Generally, managing data can cost somewhere between 5-10% of the project activity. These costs will be at the lower end if the data are of a type already managed and the processing streams are already in place and running. When the data have new characteristics that require changing or building new components in a data system, the costs will be at the higher end of this range.

Annex V. Terms of Reference of the National Science Data Management Committee

Reporting to NSDC:

- Implement and Maintain to date the National Sciences Data Management policy
- Develop national goals for science data management
- Develop annual work plans and allocate funding to implement them.
- Develop and implement accountability mechanisms and metrics to demonstrate improvements in data management.
- Identify the range of data (numeric and non-numeric) now held, develop and implement plans for appropriate data management procedures for each type.
- Develop as needed and implement national standards for data processing, archiving, and access. Where possible, international standards should be adopted.
- Promotes the broad diffusion of data, both internally and externally through the use of appropriate technologies.
- Work with other sectors, other government departments, universities, the private sector and international colleagues to improve timeliness of data processing, data quality and reliability, and access to data.
- Develop and implement a plan for protecting data at risk of loss and assist as appropriate to move project data into the national data system.
- Provide a point of contact between Science and IM&TS so that Science and IM&TS better communicate their respective needs.
- Develop and implement appropriate national solutions (including infrastructure, software purchase or licensing) for managing data so that resource costs can be shared.
- Ensure appropriate linkages and representation on national and international committees that connect DFO activities to other programs with similar goals and interests.
- Devote an appropriate level of effort to exploring new solutions to data management problems
- Work with appropriate HR representatives to incorporate valuation of data management for SE-RES staff.

Annex VI. Data Management Proposals Instructions and Template

Last updated: 1 Dec, 2005

Proposal approval process

There are two aspects to this process. The first is the process of preparing, submitting and approval of proposals. The second is defining the criteria against which proposals will be judged.

Process

Proposals should be succinct and limited to a maximum of 3 pages. Longer proposals will be turned back by the regional NSDMC member.

Proponents should prepare proposals that address one or more of the categories indicated. In the proposal, they should indicate the relative proportion (as a percentage) of the categories to which the proposal applies. The template for proposals provides opportunity for proponents to describe how the project meets the approval criteria. Though the template seeks information for funding beyond one fiscal year, there is no mechanism at present to approve funding for more than one year at a time.

At any time during the course of a fiscal year, one or more regions may prepare a proposal for submission to the NSDMC. The proposals should be passed through the regional member to the chair of NSDMC. The proposals will be considered during the next round of project approvals for the coming fiscal year. If a project receives no funding in the approval process, the proponents will be notified through their NSDMC member and they can choose to resubmit at another time.

Depending on when the project is received and if funds are available, a project may be approved in the same fiscal year as it is submitted. The Chair of NSDMC will determine if the funds requested by a project are available and if not the proposal will be held for consideration for the next fiscal year. If funds are available, the proposal will go through the same screening process used for all proposals.

NSDMC will set criteria no later than 31 January each year. The criteria are stated under different categories with the desired relative effort indicated by the figures for each category. Each member of NSDMC will score each proposal and totals will be accumulated. Proposals scoring below a 50% threshold will be removed from further consideration. NSDMC will take the remaining proposals and weigh them to ensure a balanced effort in the categories within the funding limits of each category.

Proposal approvals will take place in March.

Criteria

Categories and weightings:
Archives - 20
Access - 25
Standards - 15
Governance – 5
Data rescue – 20
Inventories – 10
Other - 5

Approval factors:

1. The project promotes multi-regional or national solutions
2. The project improves efficiencies in managing data
3. The project expands the availability or accessibility of data
4. The data managed by the project are of high importance to Science
5. The project promotes the use of standards

DFO National Science Data Management
PROJECT PROPOSAL

PROJECT TITLE:

PROJECT MANAGER AND PARTICIPANTS (*by regions***):**

OBJECTIVES:

DESCRIPTION:

DELIVERABLES *(itemize and indicate participants if appropriate*)**:**

PROJECT STATUS *(if an ongoing project, give a brief description of present status)***:**

RELATIONSHIP TO DATA MANAGEMENT OBJECTIVES *(reference deliverables*)

1. Governance (%):

2. Data Archival %):

3. Data Rescue (%):

4. Access to Data and Information (%):

5. National Inventory (%):

6. Standards (%):

7. Other (explain) (%):

LINKAGES TO OTHER INITIATIVES

FUNDING:

1. Requirements *(break down the requested funding by deliverable)*

2. Identified Resources *(list any other resources, in kind etc. to apply to project)*

3. Requested Funding

	2006-2007	2007-2008	2008-2009
O&M			
Capital			
Total			

Annex VII. Regional Reporting Template

DFO National Science Data Management
ANNUAL REGIONAL REPORT – *region name i.e. Pacific*

Summary

Governance

National
Regional
Issues

Data Archives

National Programs
Regional Programs
Issues

Data Rescue

National Programs
Regional Programs
Issues

Access to Data and Information

National Programs
Regional Programs
Issues

National Inventory

National Programs
Regional Programs
Issues

Standards

National Programs
Regional Programs
Issues

Non-Numeric Assets

National Programs
Regional Programs
Issues