

Scalable Models of Data Sharing in Earth Sciences

John Helly
San Diego Supercomputer Center

Hubert Staudigel, Anthony A.P. Koppers
Institute of Geophysics and Planetary Physics

University of California, San Diego

Abstract

The ability of investigators to share data is essential to the progress of integrative scientific research both within and across disciplines. This paper describes the main issues in achieving effective data sharing based on previous efforts in building scientific data networks and, particularly, recent efforts within the Earth sciences. This is presented in the context of a range of information architectures for effecting differing levels of standardization and centralization both from a technology perspective as well as a publishing protocol perspective. We propose a new Metadata Interchange Format (.mif) that can be used for more effective sharing of data and metadata across digital libraries, data archives and research projects.

Introduction

Most data and metadata in Earth sciences are published in the context of traditional, peer-reviewed publications in paper journals, or electronic journals that are functionally quite similar to paper journals. The practice of publishing data electronically is poorly developed, in particular, since electronic journals are still functionally similar to the paper journals with few important exceptions in large scientific projects that developed their own data sharing mechanisms. Peer reviewed publications in printed media are not well suited for data publications, because the layout is optimized for viewing by human eyes, and not for optimum information transfer. The confinement of most data publication to paper journals has the result that authors rarely publish data and even less frequently their metadata (Helly 1998). In geochemistry, this has resulted in a crisis in data publications (GERM Steering Committee, 2001) where a large amount of legacy data are in danger of being permanently “buried” in private research files (see Figure 1a). Earth sciences are likely to gain substantially from introducing new means of publishing data. There is a wide array of electronic data publications that range from strict peer-reviewed, central publication to entirely non-reviewed, decentralized publication on personal websites (Table 1).

In this paper, we discuss two important information technology aspects of the electronic publication of data in the Earth sciences, metadata, and a variety of different concepts of electronic data publication. Metadata are the foundation of electronic data publications

and they are determined by needs of archiving, the scientific analysis and reproducibility of a data set, and the interoperability of diverse data publication methods. We use metadata examples drawn from the companion paper by Staudigel et al. (this issue) to illustrate the issues involved in scaling-up the publication of data and metadata by individual scientists, disciplinary groups, the Earth science community-at-large and to libraries in general. We begin by reviewing current practices and considering a generalized alternative.

The current practice of scholarly publication is illustrated in Figure 1a. Research produces effectively three types of products: (1) data, (2) metadata and (3) the interpretations of these data in the form of various writings and illustrations organized as manuscripts. Our current paper publishing protocol (Figure 1a) is mostly biased towards the visual products, the writing and the illustrations. Digital data and metadata tend to be rarely published on paper copy or in its electronic equivalents (Fig 1a, dashed lines). Page limitations in high profile journals actively work to eliminate or greatly reduce the actual publication of data and metadata. As a result, interpretations and figures based on data are widely published and archived in libraries, while most of the primary data are confined to research files of investigators. These private archives, however, do not provide sufficient access for future research that might result in a re-interpretation of the data. There is no return flow of data into community-wide research activities and, therefore, these data are effectively lost. In the worst possible case, the data will have to be re-analyzed, forcing another cycle of data generation (and destruction). This places the Earth sciences in a situation where the transient interpretations of data are kept in reliable archives while the actual permanent measurements and records have an uncertain fate, at best. This effective loss of data and metadata due to non-publication cannot be reconciled with our principles of scholarly science. Without them, scientific work cannot be duplicated or falsified.

In Figure 1b, we depict an alternative publishing protocol in which the generation of scientific products and resultant publication remains the same, but data and metadata are published in a scientific data network. This network provides a library-like function for the archiving and serving of data to the scientific community in order to support future research. This resolves the key problem of Figure 1a, where the data and metadata are inadequately archived and published. In Figure 1b, this new protocol augments the traditional publication of scientific results by publication of the data and metadata; making the “complete” scientific product available in a consistent and coherent manner (Helly, Elvins et al. 2002). Such a data network may include a range of data publication options, including the attachment of electronic data supplements to publications (GERM Steering Committee, 2001).

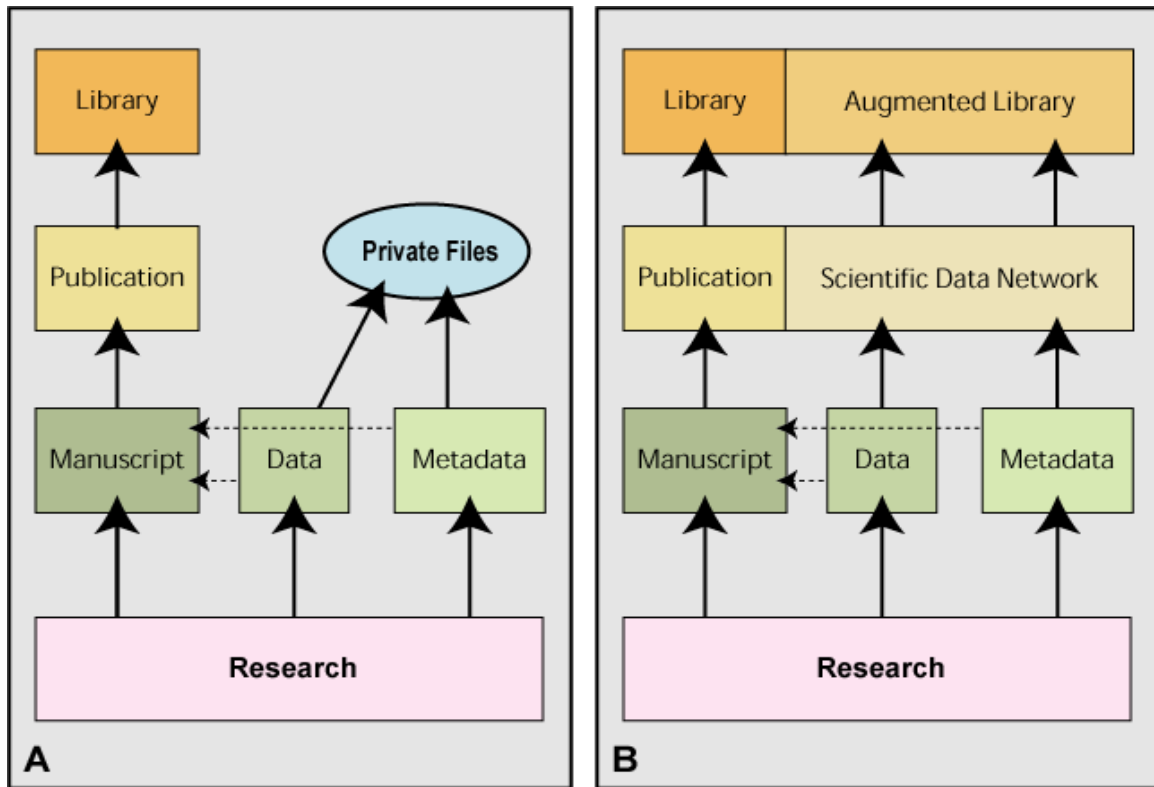


Figure 1a. Schematic depiction of the flow of scientific information from research to published library resources as currently practiced. **Figure 1b.** Potential approach based on interoperable resources contained within a scientific data network but leading to library-grade products analogous to the formally published journals we are familiar with today. This augmented library will protect the future of science against any loss of valuable research data that normally resides in private files. Note that the dashed lines from data and metadata to the manuscript reflect the limited publication of these information sources in our conventional manuscripts.

Electronic data publication in Earth sciences is still in an emerging phase, where many options may be considered that have particular advantages to particular Earth science communities. Metadata play an important role in such data networks, but as different Earth science disciplines may choose different data publishing protocols, it becomes extremely important to have data and metadata in formats that are portable between protocols. For this reason, we will first discuss different methods of data publishing via a scientific data network followed by an inventory of desirable characteristics of such a network. Then, we will introduce a method for generating a highly portable metadata interchange format we call *.mif* (pronounced dot-mif) and conclude with a discussion of how this metadata format can be scaled to support the diversity of interests within the Earth science community and other scientific communities. The most recent contributions to this discussion may be found at <http://Earthref.org/metadata/GERM/>.

Methods of Data Publishing

Information Technology (IT) offers a range of publication options that may be classified by their respective publishing protocols and their system architecture. In Table 1 we have illustrated a few combinations using these two parameters. The choices between these protocols are mainly driven by the need for centralization and peer-review, the complexity of data and its volatility, the cost of publication and longevity. However, some of the most difficult choices often involve socio-political issues, such as which group or organization gains authoritative control over a central data resource or the functioning of the network. These choices are difficult because they have to be made competitively where the winner is expected to deliver the best product for the community, in exchange for funding and scientific rewards or influence. A careful evaluation is needed to balance the benefits of the freedom of managing one's own data versus the costs of maintaining public data servers at individual laboratories that are secure, reliable and available to all. While these costs are continually decreasing as more and more scientists become trained in maintaining computer systems and managing data as a routine part of doing their science, they are still significant and likely to remain so until the methodological issues we are discussing in this paper become standard practice.

The classical model of scholarly publishing is the scientific journal with a protocol based on peer-review and editorial boards. The information architecture supporting this is centralized; typically located at the publisher's facilities where the copyright and copy-of-record are retained. In the past, libraries were also keepers of the copy-of-record and the redundancy of libraries provided insurance that there would "always" be a copy available. Now, with the shift to e-publishing, it is the publishers who increasingly hold the copy-of-record and the libraries themselves never physically hold anything.

In contrast, a de-centralized system architecture united only by the basic *http* protocol is exemplified by individual web-sites without any publishing protocol per se; what you see is what you get and it may not be there tomorrow. As an intermediate case, federated networks may be thought of as locally-structured but de-centralized. A particularly interesting and recent example is the music distribution network approach of Kazaa at <http://www.kazaa.com>. Kazaa uses a peer-to-peer approach but selects a subset of peers to also act as search-server nodes according to their exceptional computational and communication resources. This has the particular advantage of federating the server function that supports searching (based on the metadata of song, artist and peer-node) without centralizing it. This approach is also capable of dynamically reconstituting the subset of servers if one or more is removed from the network for any reason. Consequently, it is referred to as a self-organizing network. This is a very attractive and, apparently, durable network approach.

The Kazaa music distribution network also depends on the relatively homogeneous universe of data formats (e.g. .mp3 and .wav) and the limited metadata needed to conduct a successful search and maintain synchronized metadata catalogues on the variable population of query servers. Consequently, there is a design-space created by the cross-correlation of publication protocol versus system architecture (i.e. the centralization of data, metadata and the resources that house and manage them) as illustrated in Table 1.

		PUBLISHING PROTOCOL		
		Closed	Mediated	Open
SYSTEM ARCHITECTURE	Centralized	<ul style="list-style-type: none"> ○ Editorial Board ○ Peer Review Journal <i>e.g. G-cubed</i> 	<ul style="list-style-type: none"> ○ Napster 	<ul style="list-style-type: none"> ○ CEED ○ EarthRef.org
	Federated	<ul style="list-style-type: none"> ○ Airlines <i>e.g. Reservation System</i> 	<ul style="list-style-type: none"> ○ GNU ○ Kazaa 	<ul style="list-style-type: none"> ○ FTP ○ Open Source <i>e.g. LINUX</i>
	Peer-to-Peer	<ul style="list-style-type: none"> ○ Private Networks 	<ul style="list-style-type: none"> ○ GNUTella ○ FreeNet 	<ul style="list-style-type: none"> ○ Personal Web

Table 1. Trade-off between system architecture and publishing protocol providing possible models of data publishing and sharing.

In the Earth sciences, disciplinary boundaries naturally suggest important data collections for a scientific data network with disciplinary-specific publishing protocols and conventions suited to the discipline but recognizing and supporting inter-disciplinary data sharing and interoperability. Such a network could conceivably be operated as a self-organizing network, like Kazaa, given sufficient standardization of both the data and metadata formats (see also Staudigel et al., this volume) and it is for this reason that we later describe a new approach to metadata standards as a step in that direction. These protocols and conventions must provide local structure for the operation of the scientific data network while supporting the diversity that is the very heart of specific Earth science research. Such protocols and conventions would include metadata standards, conventions for naming arbitrary digital objects (ADO) so they can be uniquely cited and versioned, to enable verification or falsification, and conventions for protecting intellectual property rights. However, to clarify this notion we consider next the features that a scientific data network should have in order to effectively support scientific research the future.

Desirable Characteristics of A Scientific Data Network

Fundamental to science is the verification and falsification of results. Consequently, an analysis and interpretation of data has to be supported by the exhibition of the data as well as the metadata required for proper understanding of the data by others. In the conventional publishing model, the metadata are often found in figure and table captions or in narrative descriptions of methods. In our increasingly digital, data-intensive scientific culture, the scientific method demands reliable access to the data supporting published results and the ability to unambiguously identify the data and analytical techniques used in any given analysis. To support these requirements, we have reproduced a table of essential functions needed in a scientific data publication system (Table 2). We have found that these functions must be explicitly addressed in any system architecture, regardless of the degree of its centralization, in order to preserve scientific data and metadata.

FUNCTION	PURPOSE
User Registration	Assignment of an user ID and password to a given user while acquiring their email address and related contact information. This is used for auditing data access and communication with users.
Data Acquisition	Data contribution or submissions through uploading while acquiring a minimal set of metadata . This initiates the automatic creation of an unique and persistent name for the ADO and a transportable metadata file which is directly bundled with the uploaded ADO.
Search	A search system providing for spatial , temporal and thematic queries based on the content of the uploaded metadata.
Deletion Control	The ability to delete an ADO must be tightly controlled to prevent the arbitrary deletion of data that users already have downloaded. In a manner analogous to journal articles, one should not un-publish data. Errata can be accomodated by publishing a revision of the data as a new ADO version. A special case to consider is the editorial peer-review process. This requires confidentiality and the ability to remove an ADO if not accepted for peer-reviewed publication. A looser deletion policy would allow deletion of data if it had never been downloaded by an user.
Assignment of Persistent Names	ADO's within a data repository should have persistent names. This allows for monitoring when updates of ADO's come online. It helps identifying which ADO's are most frequently downloaded or may be used to notify users of anomalies/issues related to a particular ADO. Finally it establishes precedence by publication date and enables citability in publications.
QA/QC	Quality assurance and control can exist to varying degrees. It is exemplified by peer-review or non-peer-review and by anomaly detection and reporting. It must be explicitly stated. Some investigation is beginning on approaches to semi-automate this function for specific types of data.
Access Control	Access control enables the data contributor to specify a password known only to him/herself for an uploaded ADO. This password may subsequently be provided to other users to assure limited access. This enables data submitters to independently control access to their published data. Any user attempting to retrieve a password-protected ADO from the system must first obtain that password from the contributor of the data.
Versioning Traceability	Versioning data is required in order to prevent in situ changes to data. This enables the establishment of data heritage which is required to inform the user of the emperical, derived or computed nature of the data. This is necessary to ensure the reproducibility of results and it reserves intellectual property rights and facilitates proper attribution.

Table 2. Basic Functions for controlled publication of scientific data after Helly et al. (2002). ADO refers to an Arbitrary Digital Object which is a general term to describe any digital object; general a file of some kind, that can be stored in a computer file system. As the name implies, it may contain anything that can be digitally encoded.

In addition to these functions, we must consider the reliability and availability of such a data network to ensure ready and repeatable access to its contents. These system engineering concerns are frequently overlooked in discussions about scientific publishing whether for data or scientific manuscripts. The reliability and availability of the electronic resources are vital to the conduct of science and to ensuring equitable access to

data throughout the research community. The software community has achieved this, albeit imperfectly, by a high degree of redundancy through the use of mirror sites for software, and a pains-taking emphasis on clearly articulating the software version and hardware dependencies directly in the name of a software distribution file. This has resulted in a thriving open-source software community for a wide range of software systems and tools. We should be able to do the same thing for data.

Designing Metadata

Metadata is many things to many people (Baru; Michener, Brunt et al. 1997; 1998; Daniel and Lagoze 1998; Federal Geographic Data Committee 1998; PURL 1998; Weibel 1998). Our focus here is on key functions of metadata that are relevant to research in the Earth sciences. First, metadata is used for discovering the existence of data by searching a metadata catalogue or its equivalent. Second, metadata is documentary information describing the content, context, quality, structure, accessibility and so on of a specific data set. In this latter respect metadata plays a role in the integration of data, analysis and modeling. It is also essential for the proper interpretation by others who were not directly involved in the original research efforts as well as by the original investigator(s). In this paper we, therefore, refer to metadata either as *cataloguing* metadata (i.e. archiving) or as *application* metadata (i.e. scientific).

Cataloguing metadata is used for the search and retrieval functions in a scientific data network; analogous to the role played by the card catalogue of a library. Application metadata provides information about how to use the data once it is found and retrieved. These are not necessarily distinct subsets in the metadata space. For example, a metadata parameter might contain the word *basalt* to facilitate searching for *basalts* while the same metadata parameter may later be used to sort the data files or papers found by the search into classes according to *rock type*. In the first case, a keyword parameter would be used as cataloguing metadata while in the second as application metadata. So we have the same metadata available to both functions: cataloguing and application. We distinguish them in order to draw attention to the differences in how they are used in a digital library or scientific data network.

Nevertheless, discussions of metadata can be confusing, subjective and discouraging. To make any progress on this front one seeks a systematic way to approach the definition of a reasonable set of metadata parameters. We begin this by recognizing that metadata is fundamentally arbitrary in nature. What's important to one individual may be of little consequence to another and who's to say what is more important? However, there usually exist subsets of metadata that are clearly required for any scholarly publication and more or less effortlessly agreed upon and that can be augmented to accommodate more specialized metadata requirements as they emerge. So we have designed an approach that accommodates this essential arbitrariness but enables compliance with whatever metadata standards or conventions emerge within the scientific, information technology, library

and publishing communities. Some of the diversity of efforts that are relevant to the Earth science community, and that must be accommodated, is reflected in the list presented below:

- Federal Geographic Data Committee <http://www.fgdc.gov/metadata/constan.html>
- United States Geological Survey <http://www.usgs.gov/tools/metadata>
- World-Wide Web Consortium (W3C) <http://w3.org>
- Earth Science Markup Language <http://esml.itsc.uah.edu/products.html>
- Canadian Geochemistry Online <http://geochem.gsc.nrcan.gc.ca/>
- Ecological Metadata Language <http://knb.ecoinformatics.org/>
- Object Management Group (OMG) <http://www.omg.org>
- Corporation for National Research Initiatives <http://www.handle.net>
- Dublin Core: <http://www.dublincore.org>

Each of these approaches has strengths and weaknesses and, since they coexist, we need a means to comply with those that are important at any given time and for any given purpose without interference between them. For example, the Federal Geographic Data Committee (FGDC) standard was not designed for survey-type data and there are no provisions in the FGDC standard for laboratory data although some extensions have been added to the original specification to make it more useful to researchers in ecological disciplines. Nonetheless, it has been required by law for government agencies.

In order to take advantage of the strengths and minimize the effects of the weaknesses in these and other, approaches it is helpful to think in terms of categories of metadata and to organize them into a hierarchy (i.e., an acyclic, directed graph) (Figure 2) and then to cast this tree into a form that can be conveniently represented in a computer file. Such a hierarchy provides a structure that enables the addition or deletion of sub-trees that can be of arbitrary depth and breadth without affecting other parts of the tree. Each leaf on the tree is a metadata parameter and inherits all of the description that results from following the tree from the root node to the leaf.

If we consider the tree shown in Figure 2 and root it to a unique identifier for the metadata interchange format (*.mif*) and augment the tree with metadata developed for other applications and catalogues (e.g., water analyses and the Dublin Core library standard) we arrive at the extended tree shown in Figure 3. As we will show below this type of tree structure can be encoded into a form that can be conveniently represented in a computer file.

Metadata Interchange Format

As discussed above, sharing data can be accomplished in various ways across the spectrum from centralized to de-centralized system architectures. To the extent that one chooses to decentralize the sharing of data, it becomes increasingly important to be able to disseminate metadata about the various de-centralized holdings just as there are many instances of library catalogues throughout the world with common metadata. It is the

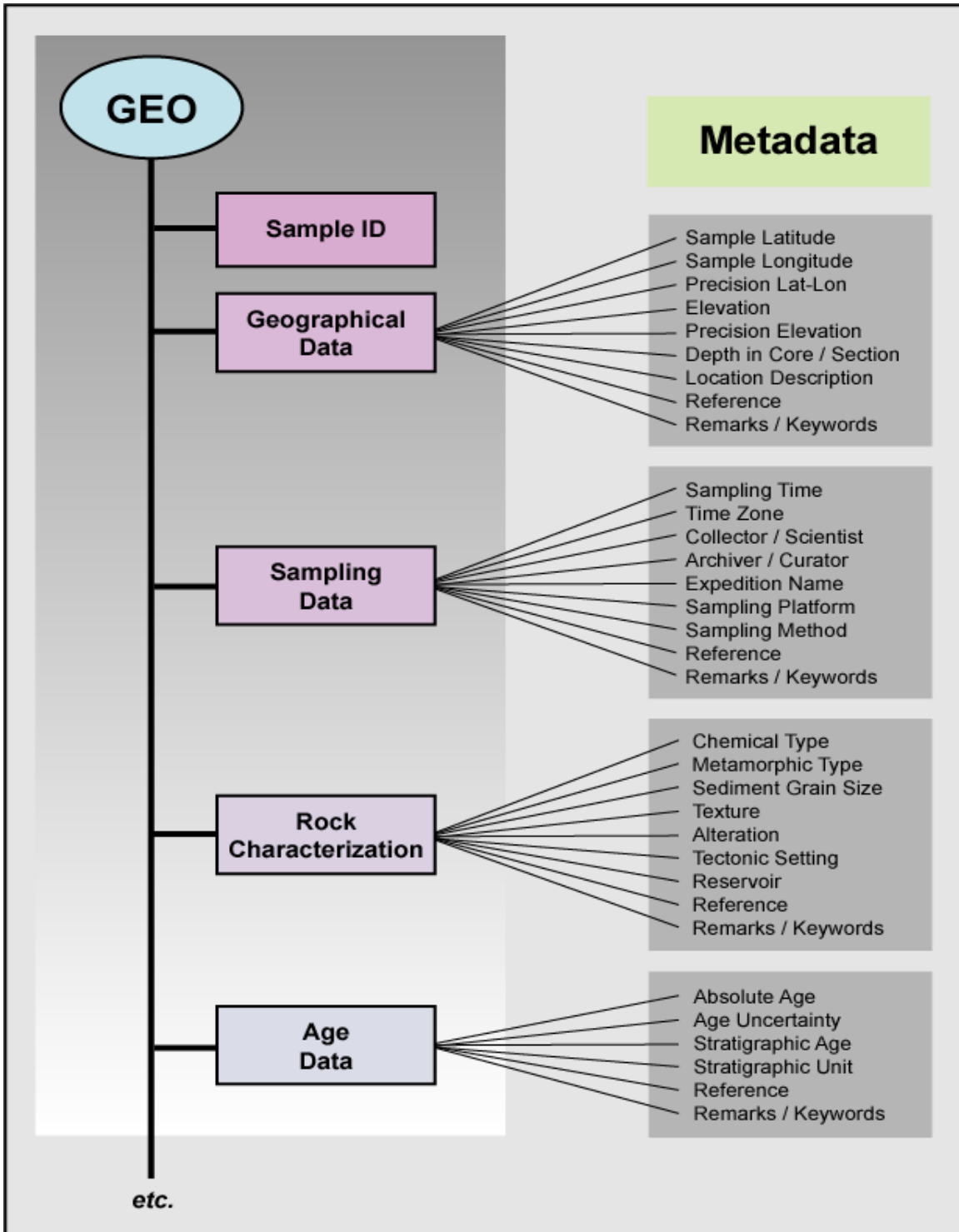


Figure 2. Geochemistry metadata tree. Each metadata field (e.g. Absolute Age) is part of a node (e.g. Age Data) on the main metadata tree (GEO). Representing this path as a string of characters provides a unique name for the metadata parameter as discussed in the text (see also Figure 4). Example after Staudigel et al.

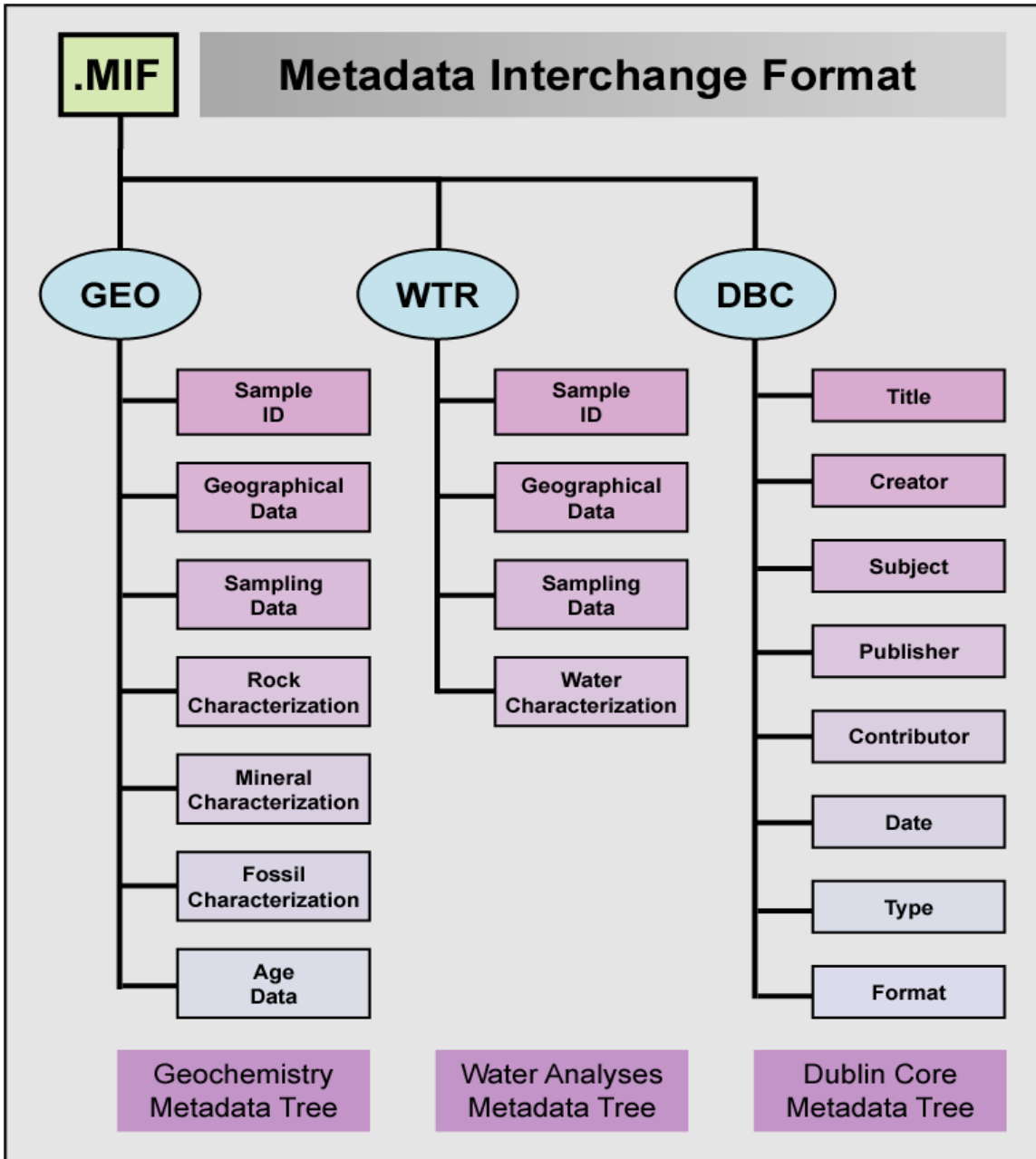


Figure 3. Geochemistry (GEO) metadata tree integrated into the *.mif* or metadata interchange format. Note that the second column is for water analyses (WTR) metadata and the last column for Dublin Core (DBC) library archiving metadata.

standardization of catalogue entries that enable libraries to interoperate and commonly refer to their holdings despite the fact that the holdings are non-identical and geographically separate. This is the basis of the inter-library loan system. The problem of sharing scientific data is both the same and different. The capability to search for and locate a particular data object is based on information that is similar to that needed to locate a book. However, the need to describe the nature and meaning of the data object's

content is quite different and this is why we have *catalogue* metadata and *application* metadata.

We can borrow from the library community the methods by which to search for the existence and location of data (e.g., Dublin Core <http://www.dublincore.org>) but we must invent new ways to document the metadata needed within the Earth sciences and to comply with other metadata standards such as the Federal Geographic Data Committee (FGDC). To accomplish this, we propose a metadata interchange format that we call *.mif* that enables interoperability and an open architecture that is maximally independent of computer systems, data management approaches, proprietary software and file formats, while encouraging local autonomy and community cooperation. Any individual or group can store metadata in any form at a home institution but share it using an interchange format. This further encourages innovation in both information technology and metadata and provides a fail-safe exchange mechanism in case currently popular methods of information exchange prosper or fail according to commercial or other interests that may not always be consistent with scientific interests. Such well-structured and openly-available metadata allow for the development of metadata export/import tools (for examples see Appendix 1 and 2) written to be consistent with the *.mif* syntax, and can be jointly developed and shared by the community. This type of development would be similar to NetCDF (network Common Data Form) a freely available interface for array-oriented data access at UCAR (<http://www.unidata.ucar.edu/packages/netcdf/>). The community can achieve the benefits of a scientific data network to the extent desired as well while retaining the benefits of autonomy and independent action so close to the heart of most scientists. In this way, the data network can be scaled.

In Figure 4 we describe the *.mif* syntax. It encodes the hierarchical structure of the metadata tree in Figure 3 into a set of triples (name, value, label) with each triple corresponding to a text row in a flat comma-delimited ASCII file.

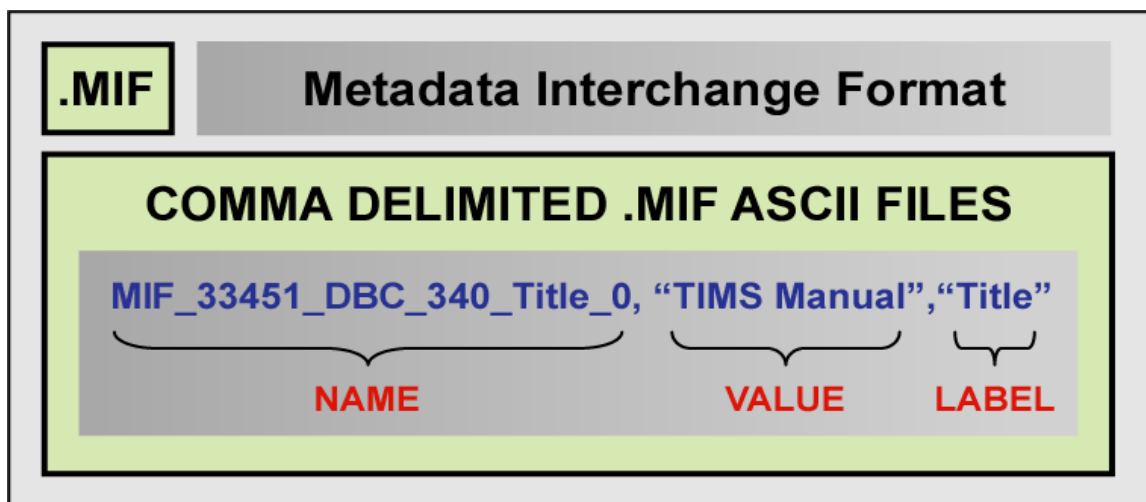


Figure 4. Proposed *.mif* metadata interchange format. See Appendix 1 and 2 for examples.

This structured convention enables procedures to be written to recover the hierarchical information in the metadata design along with the metadata itself. The name field uniquely encodes any metadata parameter. It is an ordered set of rows with each row corresponding to a (name, value, label) triple and each name constructed as an ordered sequence of node acronyms (e.g., MIF, DBC) followed by their sequence numbers (e.g., 33451, 340) within the corresponding block of metadata. This enables the nodes of the hierarchy to be re-sorted into their original hierarchical order if they are sorted for other purposes. This concatenation of NODE_NUMBER (e.g., MIF_33451) can be arbitrarily long to accommodate the depth of the tree and, equivalently, the extent of nesting within the file. In this way the *.mif* format complies with the convention referred to as the Resource Description Framework (1998). In all cases the “name” field ends with the metadata field name and its sequence number (e.g., Title_0). The “value” and “label” fields describe the actual value and name of the metadata field that has been stored in the scientific data network. Text strings have to be delimited by double-quotes. Nodes within the hierarchy are delimited by the underscore character (`_`) while blanks that would normally appear in parameter names are substituted with the hyphen (`-`).

The *.mif* convention also includes a master template file describing a particular version of the interchange metadata format standard. This template is itself a *.mif* file that can be used to control the current metadata definition within a browser or editor. It allows for the addition or deletion of metadata nodes or parameters, as interoperability requirements in the scientific data network evolve. In this master template each row is comprised of another triple (label, name, value) containing (1) a humanly meaningful label for the metadata parameter that can be used for presentation purposes, (2) a parameter name and (3) a value field that contains information about the metadata parameter itself. All in all, this master *.mif* file guides the database engineers in how particular metadata fields are to be displayed or processed. These structuring conventions enable procedures to be written to recover the hierarchical information in the metadata design along with the metadata itself (Figure 5).

Extensibility and Related Applications of the *.mif* Approach

The *.mif* approach can be extended to accommodate an arbitrarily complex hierarchical metadata structure. In Figure 5 we illustrate these hypothetical extensions showing how one might go from version n of the *.mif* structure to version $n+1$. However, there are other interesting possibilities for using this type of master control file and we have illustrated these in Appendix 1 and 2. There we show how the master control file can be produced from a spreadsheet as input to a Java-based metadata editor and browser application for the purpose of configuring the editor or browser to dynamically reflect the structure of the master template file currently in use. This approach can be used to provide backward compatibility of software tools with earlier versions of the metadata convention.

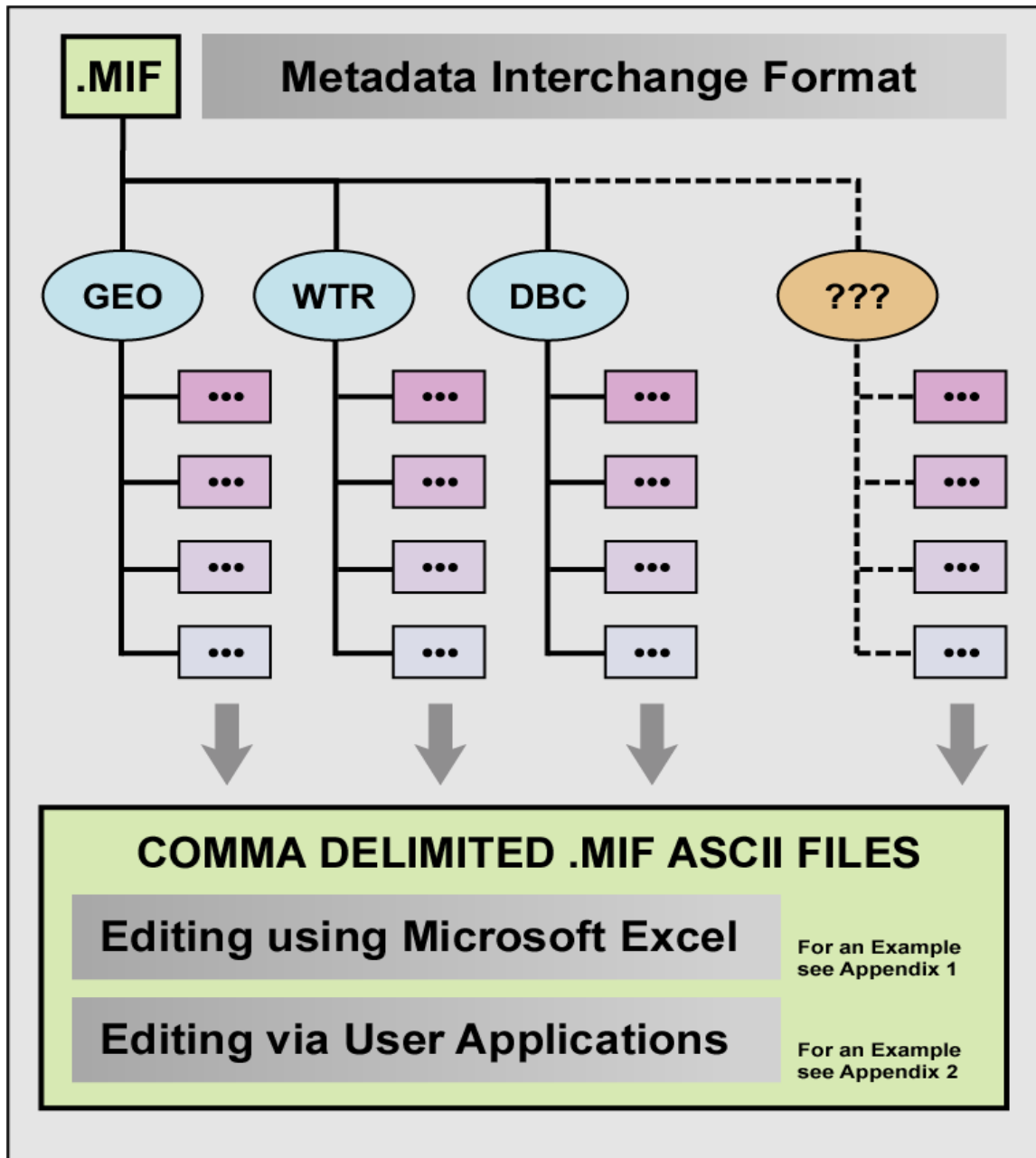


Figure 5. Illustration of the relationship between the graph and the file representations. Color is used to aid understanding of the block structure of the *.mif* file. The bottom half of the illustration shows the Excel spreadsheet used to generate the *.mif* master template file. This is a convenient mechanism for maintaining the various versions of the master template that will emerge over time as the metadata conventions come and go. A more readable version of this file can be found in Appendix 1. The file can be enlarged by zooming to read its contents but the purpose here is to illustrate the relationship to the tree above by color-coding.

Summary and Conclusions

The ability of investigators to share data is essential to the progress of interdisciplinary and integrative scientific research. This even is true within individual disciplines. Here we have described a range of information architectures for effecting differing levels of

standardization and centralization. We have proposed a generic *.mif* metadata interchange format for use in the Earth sciences and potentially other disciplines. While using a highly-structured but openly-available metadata format we are able to achieve an effective data sharing across digital libraries, data archives, libraries and research projects. These scalable techniques are being used in the development of a project we call SIOExplorer that can found at <http://sioexplorer.ucsd.edu> although we have not discussed that project in any detail. The most recent contributions to this discussion and *.mif* applications and examples may be found at <http://Earthref.org/metadata/GERM/>.

Literature Cited

(1998). Resource Description Framework efforts. <http://www.w3.org/RDF>. O. L. <ora.lassila@research.nokia.com> and R. R. S. <swick@w3.org>.

Baru, C., R. Frost, R. Marciano, R. Moore, A. Rajasekar, M. Wan "Metadata to Support Information-Based Computing Environments." IEEE International Conference on MetaData 97 September 1997.

Daniel, R. and C. Lagoze (1998). Extending the Warwick Framework: from Metadata Containers to Active Digital Objects. <http://www.dlib.org/dlib/november97/daniel/11daniel.htm>: 1-13.

Federal Geographic Data Committee, M. A. H. W. G. (1998). Content Standard for Digital Geospatial Metadata. Reston, Virginia, U. S. Geological Survey.

Helly, J. (1998). "New concepts of publication." Nature **393**: 107.

Helly, J., T. T. Elvins, et al. (2002). "Controlled Publication of Digital Scientific Data." Communications of the Association of Computing Machinery.

<http://www.fgdc.gov/metadata/constan.html> (1998). "Content Standard for Digital Geospatial Metadata (CSDGM)." Federal Geographic Data Committee.

Michener, W. K., J. W. Brunt, et al. (1997). "Nongeospatial Metadata for the Ecological Sciences." Ecological Applications **7**(1): 330-242.

PURL (1998). Dublin Core Metadata. http://purl.org/metadata/bublin_core.

Weibel, S. (1998). Dublin Core Metadata. http://purl.org/metadata/dublin_core.

Appendices

Appendix 1: Spreadsheet template to generate *.mif* ASCII files.

Appendix 2: Illustration of the use of the *.mif* master file to configure an user application for browsing and editing metadata (developed by Andrew Chase at the San Diego Supercomputer Center using an approach developed by J. Helly, D. Sutton, T. T. Elvins and D. Martinez for the CEED data archive <http://ceed.sdsc.edu>).