Multi-Institution Testbed for Scalable Digital Archiving

<MMI /> Marine Metadata Interoperability

Earth Reference Data and Models

National Science Foundation
WHERE DISCOVERIES BEGIN

Report from the 3rd Interoperability Workshop

November 14-15, 2006
La Jolla, California

The University of Arizona

University of California, Santa Barbara

UCSD

Scripps Institution of Oceanography

SDSC San Diego Supercomputer Center

SOEST School of Ocean and Earth Science and Technology

Lamont-Doherty Earth Observatory, The Earth Institute at Columbia University

insight4

Monterey Bay Aquarium Research Institute

Marine Program, University of New Hampshire

GSO Graduate School of Oceanography at the University of Rhode Island

Unidata

TPAC

Woods Hole Oceanographic Institution

Prepared by

John Helly (hellyj@ucsd.edu)
Bob Arko (arko@ldeo.columbia.edu)
Luis Bermudez (bermudez@mbari.org)
Peter Cornillon (pcornillon@gso.uri.edu)
Ian Cumming (ian@insight4.com)
Ben Domenico (Ben@unidata.ucar.edu)
Stephen Miller (spmiller@ucsd.edu)
Caryn Neiswender (cneiswender@ucsd.edu)

TABLE OF CONTENTS

Executive Summary 1
1.1 Observations 1
1.2 Recommendations for the Scientific Community 2
1.3 Recommendations to the National Science Foundation 2
2 Workshop Results 3
2.1 Significance to Science 4
2.2 Conceptual Model 5
2.3 Search Space Equivalence and Standardization 8
2.4 Automation Building Blocks 8
2.4.1 Controlled Vocabulary 8
2.4.2 Schemas and not-Schemas 10
2.4.3 Dictionaries 11
2.4.4 Controlled Nomenclature 12
3 Literature Cited 16

LIST OF TABLES

Table 1. Workhsop Agenda with Hyperlinks to Presentations 3
Table 2. Tabulation of resources of interest identified by workshop participants. 13
Table 3. List of participants with links to related resources. 15

LIST OF FIGURES

Figure 1 Version 1 data and metadata harvesting model presented at the Workshop. Considered insufficiently general by workshop participants, this was revised to the Version 2 presented in Figure 2. 6
Figure 2 Version 2 conceptual model of the steps in producing interoperable data and metadata products. 7
Figure 3 Relating the automation building blocks to each other in an example of how the information can be used to combine metadata content to establish new semantica relationships across data resources. 12

Executive Summary

It is generally accepted that we now are reaping benefits from the efforts of a broad range of researchers across institutions contributing to the development and acceptance of the use of metadata for documenting scientific research; especially for data resulting from the research. As these benefits accrue, it is clear that a major limitation to realizing the full potential of these advances is the labor-intensive nature of generating high-quality metadata at the scales needed to match the production of new data and to retroactively document priceless legacy data.

To address this limitation, we convened the 3rd Interoperability Workshop funded by the National Science Foundation to address Automated Methods for Harvesting Data and Metadata. The workshop was held at the University of California, San Diego and co-hosted by the San Diego Supercomputer Center and the Scripps Institution of Oceanography. The workshop ran for one and one-half days and was attended by 20 participants from ten institutions.

The objectives of the workshop were to 1) inventory and describe the problems that exist and methods that are employed in automatic harvesting of data and metadata and to 2) develop a community approach to the development of automated methods for harvesting data and metadata from diverse sources ranging from federal agencies to individual investigators, laboratories and institutions. The scope included everything from in-situ, real-time data acquisition systems, laboratory systems and Internet-connected resources.

This report is intended to provide a review of the subject area as a guide to practitioners and to encourage a community of interest related to advancing this important process for the future of integrative science. Although the workshop participants herald predominantly from marine-related research areas and organizations, the issues addressed span the scientific domains and build on the results of the prior two workshops (cf. http://bbe.sdsc.edu/InteropWorkshop).

The workshop began by postulating a general model for the harvesting process to provide a common nomenclature and frame of reference for discussion among the participants and resulted in the model that is presented in this report. This provided a focus for discussion of the processing steps required as well as the identification of interfaces between processing steps and the information communicated between the steps and approaches for its generation. The workshop recommendations have been distilled from that discussion and are presented below.

1.1 Observations

There is now a critical mass of investigators with projects that would benefit from the adoption of community standards and conventions for automated metadata harvesting. Some relatively straightforward steps can produce significant benefits to the community but they require advocacy and encouragement and need continued funding and community support in order to achieve community awareness and acceptance.

- The adoption of standards and common practice seems to require generational-scale changes in order to reach into the standard practices of collaborating laboratories and individual investigators. This relatively slow adoption can be fostered and accelerated by the examples set by developers and early adopters but these must be similarly encouraged and supported by appropriate funding opportunities and long-term funding commitments. Funding for maintenance activities is needed.

- It is well-recognized that different research groups have developed a number of useful and widely-applicable methods for representing metadata. Independent sharing of schemas and controlled vocabularies is good but doing it through established community projects, such as the *Marine Metadata Interoperability Project* (cf. http://marinemetadata.org/) and the related Ocean Science Interoperability Experiment (cf. http://opengeospatial.org/projects/initiatives/oceansie), is better both for continuity and consistency. It is recommended that community-led activities be fostered and advocated to minimize duplication of effort and divergence of interpretations.
- There is a need for a registry of published data sources that can be harvested by evolving automated methods. The registry does not have to do anything more than declare the existence of a data resource and provide a standardized set of access methods. However, it requires a home and a stable funding level in order to become a persistent and reliable data resource. It is recommended that funding for disciplinary registries be offered to provide community-led foci to emerge for more effective harvesting of data and metadata from recognized sources.

1.2 Recommendations for the Scientific Community

The recommendations include the utilization of the generalized model of metadata harvesting presented here to encourage the adoption of a common nomenclature for the sequential processing stages in the model. This should facilitate a community discussion of the pros and cons of different aspects of the model and invite suggestions for improvements, alternatives and modification by as broad a constituency as possible.

In conjunction with the recommendation that the conceptual model be used as a frame of reference, the workshop participants identified the need for openness and collaboration in the definition of terms and development of common methods. In this spirit, the participants defined recommendations to the National Science Foundation in order to draw attention to those things that are respectively ready areas for progress as well as those that will require new approaches. Both sets of items will require additional funding but the participants recommend that funding opportunities address these items appropriately in their announcements.

1.3 Recommendations to the National Science Foundation

The recommendations presented here require actions that are beyond the capability of the community or of individual investigators to achieve on the time-scales that are possible with NSF attention.

- It is recommended that NSF continue to participate in inter-agency efforts to develop metadata and data conventions and emphasize the need for approaches that facilitate the automation of harvesting from national and international data sources.

- Since major data sources for the research community are provided by non-NSF agencies, it is recommended that NSF encourage steps by other federal agencies to facilitate the exploitation, through automated methods, of data and metadata extraction to enable scaling and the consistency of information content and value to the community.
- Because greater exploitation of important research data can be accelerated by the encouragement of the use of community-based methods, it is recommended that NSF advocate the use of the community-based resources in proposal opportunities and encourage their recognition in publications resulting from NSF funding.

2 Workshop Results

Over the past ten years the disciplinary science communities, beginning with the Ecological Society of America (Gross, Allen et al. 1995), Federal Geographic Data Committee (cf. http://www.fgdc.gov/), NASA Global Change Master Directory (GCMD) (Olsen 1996), Marine Geology & Geophysics data management workshop held in La Jolla, 14-16 May, 2001, the SIOExplorer National Science Digital Library project (Helly 1998; Helly, Elvins et al. 2002), the Earthref project in geophysics and geochemistry (Helly 2003; Staudigel, Helly et al. 2003) and more recently the Marine Metadata Initiative (Watson 2004), have awakened to the importance of the relationship of metadata to the preservation and utilization of data and interoperation of data systems. We have documented much of this prior work in the reports of the other Interoperability Workshops (cf. http://bbe.sdsc.edu/InteropWorkshop/) and we encourage the reader to consult these for that detail.

Table 1. Workshop agenda with hyperlinks to presentations

Nov 14		
When	**What**	**Who**
0800-0830	Morning Refreshments	All
0830-0900	Welcome and Introductions	Helly
0900-0930	Draft Model(s) of Harvesting Process: Framework for Discussion http://magma.sdsc.edu/files/Presentations/3rdInteropWorkshop_Helly.key.pdf	Helly
0930-1000	Discussion	All
1000-1030	Break	All
1030-1045	Automatic harvesting of metadata using OGC Sensor Observation Services in OOSTethys http://magma.sdsc.edu/files/Presentations/Bermudez_oostethys.pdf	Bermudez
1045-1100	Oceans and Climate Digital Library Portal http://magma.sdsc.edu:8080/files/Presentations/Cummins_TPAC Workshop Talk.ppt	Cummins
1100-1115	THREDDS Update http://magma.sdsc.edu:8080/files/Presentations/Domenico_THREDDS.ppt	Domenico
1115-1130	Staging and Controlled Vocabulary Strategies for Auto Harvesting http://magma.sdsc.edu:8080/files/Presentations/MIller_strategies_20061113b.ppt	Miller

1130-1145	Automatic Non-intrusive Recovery of Provenance Metadata http://magma.sdsc.edu:8080/files/Presentations/hardy_talk07.pdf	Hardy
	Automatic Non-intrusive Recovery of Lineage Metadata http://magma.sdsc.edu:8080/files/Presentations/Frew.ppt	Frew
1145-1200	Automated Methods for Harvesting Data Provenance http://magma.sdsc.edu:8080/files/Presentations/Frew.ppt	Ram
1200-1330	Lunch	All
1330-1400	Revisit Model(s) Before Workgroups	Helly
1400-1600	Breakout: 2 Workgroups, same topics	All
1600-1730	Plenary Reports	Workgroup Leaders
Nov 15		
When	**What**	**Who**
0800-0830	Morning Refreshments	All
0830-1000	Review and Discussion	Helly
1000-1030	Break	All
1030-1200	Inventory of Models for Report	Helly
1200-1300	Lunch	All
1300-1430	Writing Assignments and Discussion	All

2.1 Significance to Science

Data and metadata interoperability is the ability to transport, understand and utilize data in differing computing environments and for different purposes. It is essential to the broader and more effective exploitation of data fusion and data mining across scientific disciplines and research activities. To emphasize the potential benefits, we want to draw attention to the following key aspects of the current state-of-practice. Metadata is currently very difficult to generate. There are many reasons for this but generally it is because data are produced by software and instruments that have undergone a great deal of effort to produce for a primary or dedicated purpose without sufficient regard for the documentation, that is metadata, that is necessary for the use of those data by others not involved in its production. This is slowly changing but must change more rapidly; and it can.

One of the key reasons that automated processes for the production of metadata are not as widespread as those for the production of data itself is that the requirements for metadata are poorly defined. Metadata requirements are determined by the different uses to which data they describe will be put. For example, to fuse a set of temperature values from one source with those of another into a more complete time-series requires one set of information. To enable a researcher to search for and find all of the temperature data that might be of interest, other metadata is needed. *Most importantly, these metadata are needed in common across the data sources. That is what interoperability is about.*

The principal reason for investing in the development of automated methods of data and metadata harvesting is that we cannot possibly expect do keep up with the growth in digital data production manually nor to exploit existing collections of legacy data as effectively as possible. The key hindrances to generating metadata are the difficulties in the decision processes relating to the establishment of acceptable terminology, methods of quality control and peer review and attribution and a merit-based system of reward for producing high quality data collections. This is a temporary problem but its solution needs to be accelerated and guided. At the heart of this problem is the shifting character of science from individual investigator to intensely collaborative research programs necessitating the sharing of data for scientific progress. Improvements in data and metadata interoperability and automated methods to produce them will go a long way in accelerating scientific progress.

2.2 Conceptual Model

A model of the steps required to produce interoperable data products was provided as a starting point as well as to help establish a vocabulary for the ensuing discussion. This Version 1 model is shown in Figure 1. This initial version was felt to be insufficiently general. However, it served adequately in stimulating vigorous discussion and led, ultimately, to the Version 2 model presented in Figure 2 which was generally agreed upon and satisfied workshop participants. Version 2 represents a consensus of the participants and serves to identify a sequence of named processing steps considered essential to the interoperability process. The name for each step was discussed in considerable detail because of the overloaded meanings that some of the terms have and this was the final agreement on terminology as described below.

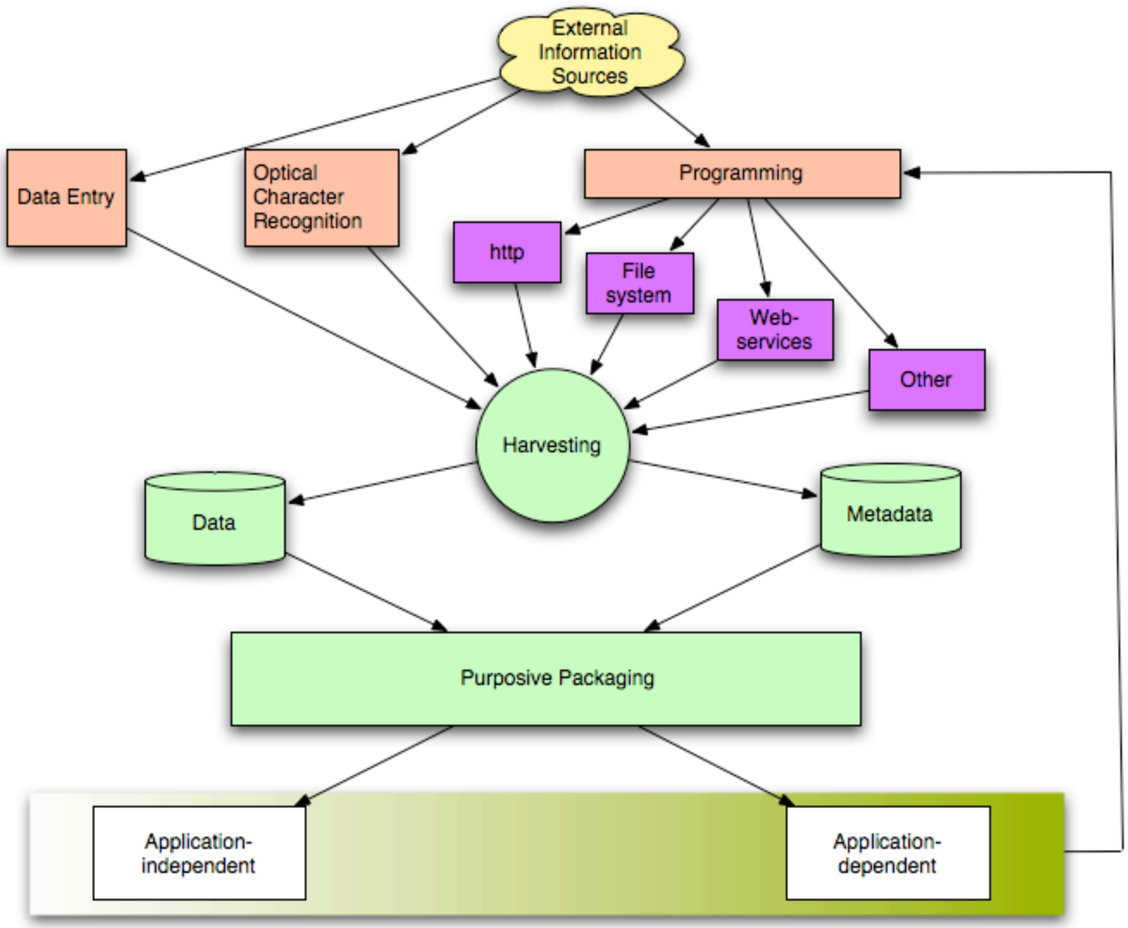

Figure 1 Version 1 data and metadata harvesting model presented at the Workshop. Considered insufficiently general by workshop participants, this was revised to the Version 2 presented in Figure 2.

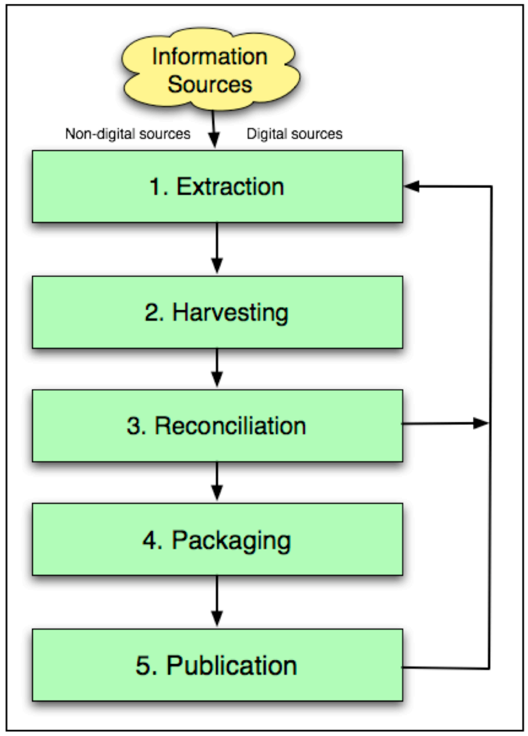

Figure 2 Version 2 conceptual model of the steps in producing interoperable data and metadata products.

The key aspects of this model of automated harvesting of data and metadata can be viewed in sequential order as extraction, harvesting, reconciliation, packaging, and publication. These sequential functions are defined as follows.

1. *Extraction* is the process of obtaining data and metadata from one or more sources.

2. *Harvesting* is specifically the gathering together of extracted metadata and data into a staging location.

3. *Reconciliation* is the comparison and analysis of the source metadata schema with a target schema and controlled vocabulary. This process may require reorganization of the data by grouping based on metadata or data content. This might proceed, for example by sorting, clustering or classifying methods, into subsets in order to understand characteristics of the data and to organize it properly for description by metadata and for subsequent processing. This may reveal the need for additional metadata and therefore may be an iterative process. Ultimately, the resultant metadata considered validated against the target schema and controlled vocabulary and is ready to be packaged with its related data for publication.

4. *Packaging* is the process of grouping the metadata with associated data objects together for distribution. The methods of grouping may include clustering or classifying as well as the physical bundling of data and software

into conventional distribution forms such as tarballs and RPMs. Many of the workshop participants felt strongly that this is not just the distribution of files but extends to include a wide range of digital products that services as well as individual database records.

5. *Publication* is the process of declaring a package complete for distribution to the extent that the published item is uniquely identifiable, attributable and transportable.

This simple model of the processing steps was considered to be necessary and sufficient as a community reference by the workshop participants. We recommend that others consider it and begin to describe their methods with reference to these steps as an aid to correlating and describing our individual efforts towards improved automated harvesting of data and metadata. The subsequent discussion provides more detail on the motivation for the development of a model in the hope that others will recognize a common interest and personal interest in contributing to the advancement of this effort.

2.3 Search Space Equivalence and Standardization

One of the key goals of the automation of data and metadata harvesting is to provide the equivalence and standardization of the search space that is implicitly defined by the universe of metadata surrounding a data collection. The sought-after equivalence is such that the same search, operating on the same data in similar or different environments will produce the same results. We have faced this problem many times and an example of the problem is illustrated by the cross-walk approach. This problem is to resolve the similarities and differences between two or more different metadata schemas describing similar, or even identical, data. For example, can the term *'research ship'* and *'submersible'* both be used to describe the same thing? How can we be sure that if we search for all the data from cruises by research ships that it will include the data collected from submarines as well? Of course, one can do both searches and compare them but this is potentially extremely computationally costly and prone to errors of its own.

The mapping of an attribute-value pair to one or more other attribute-value pairs, sometimes called cross-walking and often associated with ontologies, attempts to ensure that metadata in one representation means the same thing in another application. This is the problem of semantics and one of the goals of semantic analysis is to reduce ambiguity; in this case, the ambiguity between two uses of the same information. This is one of the problems automation of data and metadata harvesting can help to resolve by providing the tools to minimize ambiguities.

2.4 Automation Building Blocks

Attempts to reduce ambiguity are aided by the use of controlled vocabulary and common or standard schemas combined with dictionaries to aid in the interpretation of the meaning and use of the attributes and their values.

2.4.1 Controlled Vocabulary

In its simplest form, a controlled vocabulary is a list of valid values for an attribute (i.e., field, parameter, variable). In a more advanced form, it is a list of attributes paired with another list of valid values. The same value may appear in any list but it is the association of the value with a specific attribute that is the valuable step. This form helps reduce the ambiguity about when to use elements of the simple list and how to extend the controlled vocabulary to sets of attributes.

In order to be used effectively to reduce ambiguity, the controlled vocabulary must be used during the creation of the metadata as well as during the searches. So the effort required to define the controlled vocabulary should be invested at the beginning of a data project to gain the maximum benefit. The list of valid values for each attribute can be used to populate pull-down menus or other widgets in a user or application interface to enable the effective searching of any given search space. These lists have a wide range of uses. One of the most interesting is to invert the list such that each unique value is associated with a list of attributes that use it. Using an inversion of this type, one can make a graph, semantic nets or networks, of attributes that are potentially related to each other in some way because they have one or more of the same potential values.

For example, the attributes ***sun*** and ***star*** can both be hot and so have the similar property that they are both sources of heat energy. This can lead to advanced methods of semantic resolution across metadata domains leading to a systematic method for automating the cross-walking problem by investing sufficient effort in the development of the necessary controlled vocabulary as illustrated in Figure 2. It also can be applied in the harvesting process to provide standardization of the search space. The trick that requires innovative approaches in harvesting is to increasingly automate the provisioning of the appropriate value, or set of values if the cardinality is greater than one, for any given attribute.

Figure 3 The metadata schema shown is common to both SIOExplorer, DIGARCH and IODP projects and is an example of an interoperable and scalable schema. Note that the Canonical branches are the minimal subset required. Each of the other branches are blocks of metadata that are arbitrary in number and composition. For example, Instrumentation, is included or excluded based on the analysis of the data being described. This schema is used to automatically harvest data and metadata for tens of thousands of heterogeneous data files. (courtesy C. Neiswender)

2.4.2 Schemas and not-Schemas

Schema has a number of meanings but here we use it to mean the members of a set of metadata attributes that are organized into a set because they jointly describe the same digital resource. When we say not-schemas, we want to specifically account for the objection that the W3C Resource Description Framework (http://www.w3.org/RDF/) is a schema-less method of providing machine-translatable metadata. RDF uses a convention *'…based on the idea of expressing simple statements about resources, where each statement consists of a subject, a predicate, and an object. In RDF…'.* There is no requirement that there be a schema in the sense of a relational model with tightly controlled primary and secondary keys, for example. Our set-oriented approach to defining schema includes an unstructured set of attributes whose only required relationship is that they describe a common resource.

For example, there may be a schema for instrumentation and a schema for a research cruise or for water samples or for trawls (cf. Figure 3). The list of possibilities is often long but not endless. For the Integrated Ocean Drilling Project we have one schema and for the SIOExplorer and DIGARCH projects, we have others. However, we have been able to distill out from each, since they all involve ships and instruments, common schemas that we can combine into a project-level schema for each and an interoperability schema for both.

Schemas and associated controlled vocabularies are central to automated harvesting because some of the information they provide can be used to define a search space for discovery of the digital resources. That is not all they could or should contain because metadata is also used in an important way inform the post-processing of the data they describe and to establish relationships between digital objects, intellectual property rights and so forth. However, they are most powerful in concert with controlled vocabularies since metadata attributes are the set of thing that controlled vocabulary terms provide the values for. Controlling the set of values that metadata attributes can have is essential to ensuring the ability to have a complete and consistent search of the universe of digital resources.

Metadata are also used to provide information that is important in machine-processing of data they describe such as datatype, ownership and whether they are required or optional fields. Cardinality is, for example, also a key ancillary piece of information which helps the developer know when to stop searching for additional values for a given attribute; or how many may be created. For example, it occurs that there may be many values for a given metadata attribute associated with a given digital resource. The cardinality of a metadata attribute can provide *a priori* information about the way a search should be conducted before concluding or assuming it is complete.

2.4.3 Dictionaries

For human languages, a dictionary is a set of definitions of the values that the parts of a valid sentence can assume. The dictionary is created for that particular purpose. Metadata dictionaries are created for a similar purpose; to provide a set of definitions for the values that metadata attributes can assume within a domain. Although we generally do not have the notion of sentences in the same way, it might be interesting to begin to think of how narrative English, as one example language, could be automatically parsed to extract metadata using a controlled vocabulary, and a dictionary. Useful dictionaries can be constructed in much the same way as controlled vocabularies. Start with a list of attribute values, sensibly taken from the union of all the values in the controlled vocabulary, and create a new list of lists such that each attribute has at least one definition. Then the schema, the controlled vocabulary, and the dictionary can be connected as shown in Figure 3 by common keys.

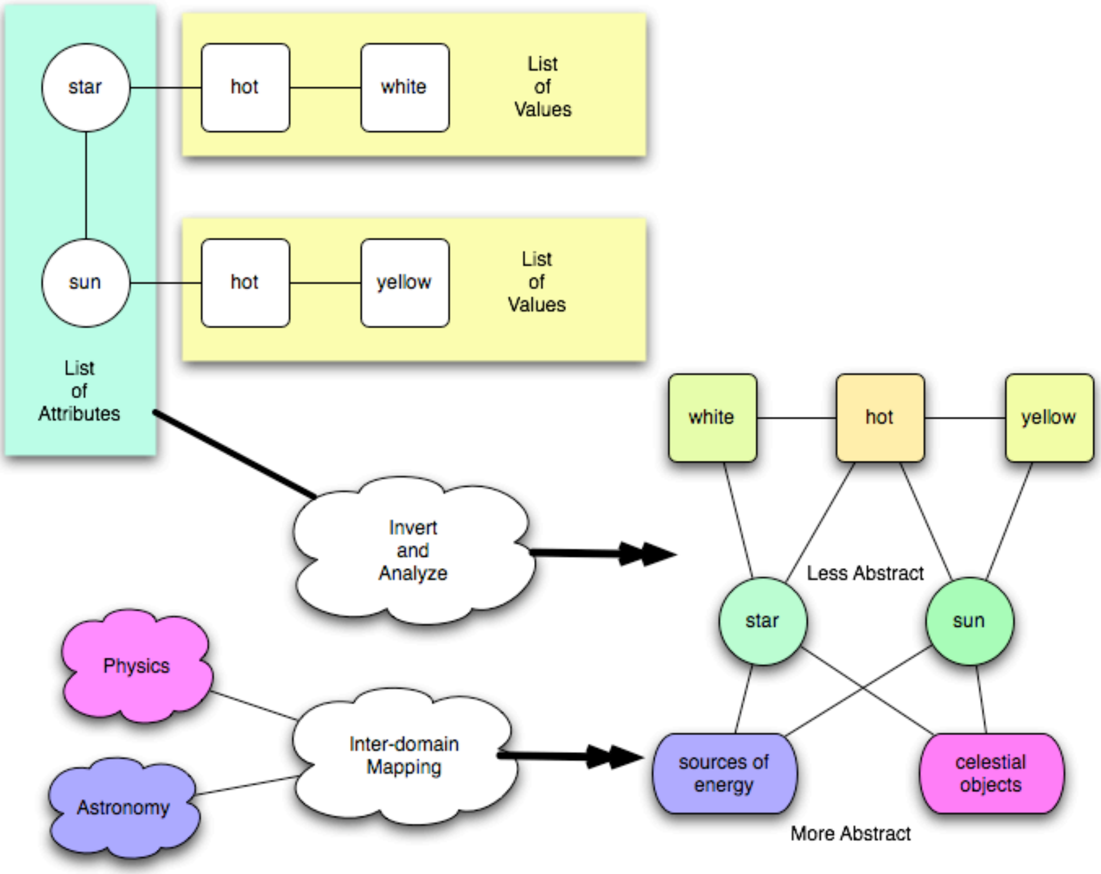

Figure 4 Relating the automation building blocks to each other in an example of how the information can be used to combine metadata content to establish new semantic relationships across data resources.

2.4.4 Controlled Nomenclature

Controlled nomenclature is inherent in all of the steps that we have discussed so far but it is especially important for the naming of digital objects. This is a topic of endless opinion and methods but its importance cannot be overemphasized as it limits or enables complete and coherent searches and is essential to the unique identification of digital objects and their updating or versioning. There is no general approach that suits all purposes and none has yet surfaced to become dominant. In our workshop we recognized its significance but did not further address it. Table 2 is a list of activities and resources that workshop participants identified as important to consider for future reference in establishing standards and conventions.

Table 2. Tabulation of resources of interest identified by workshop participants.

Category	Common Name	Resource Description or Reference
Standards and Conventions	Climate and Forecast (CF) Standard Names	http://www.cgd.ucar.edu/cms/eaton/cf-metadata/CF-current.html#sname
	COARDs	ftp://ftp.unidata.ucar.edu/pub/netcdf/Conventions/COARDS
	Digital Object Identifier System	http://www.doi.org/
	Dublincore	http://dublincore.org/index.shtml
	EML	http://knb.ecoinformatics.org/software/eml/eml-2.0.1/index.html
	Fedora	http://www.fedora.info
	FGDC	http://www.fgdc.gov/
	Grib	http://www.wmo.ch/web/www/WDM/Guides/Guide-binary-2.html
	Geographic Markup Language	http://www.opengis.net/gml/
	Geoscience Markup Language	http://www.opengis.net/GeoSciML
	HDF	http://hdf.ncsa.uiuc.edu/
	ISO 19115	Defines the schema required for describing geographic information and services.
	ISO 19139	Geographic information -- Metadata -- XML schema implementation
	MMI	http://marinemetadata.org
	NetCDF	http://www.unidata.ucar.edu/software/netcdf/
	OAI	http://www.openarchives.org/
	OBIS	http://www.iobis.org/
	OpenGIS	http://www.opengeospatial.org/
	SensorML	http://www.opengeospatial.org/projects/groups/sensorweb
	STD-DOI	http://www.std-doi.de
Transport Protocol	GridFTP	http://www.globus.org/grid_software/data/gridftp.php
	HTTP	http://www.w3.org/Protocols/
	OPeNDAP	http://www.opendap.org/
	REST	http://en.wikipedia.org/wiki/Representational_State_Transfer#References
	SOAP	http://www.w3.org/TR/soap/
Catalog Services	NOESIS Ontology Search	http://noesis.itsc.uah.edu
	OGC CSW/ebRim	The Open Geospatial Consortium, Inc.® (OGC) selected the OASIS standard ebRIM (electronic business Registry Information Model) as the preferred cataloguing metamodel for profiles of the OpenGIS® Catalogue Service Web (CS-W) specification. The catalogue specification defines the information required to support discovery and searchingand ebRIM defines requirements for registration of services.
	TDS (THREDDS	http://www.unidata.ucar.edu/projects/THREDDS/tech/TDS.html

	Data Server)	
	TPAC Digital Library	http://digitallibrary.tpac.org.au
Access Protocol	Sensor Observation Service	http://www.opengeospatial.org/projects/groups/sensorweb
	W3C XML Schema	http://www.w3.org/XML/Schema
	WCS	http://www.opengeospatial.org/standards/wcs
	WFS	http://www.opengeospatial.org/standards/wfs
	WMS	http://www.opengeospatial.org/standards/wms

Table 3. List of participants with links to related resources.

Name	email	Organization	URLs
Arko, Robert	arko@ldeo.columbia.edu	LDEO	http://www.marine-geo.org
Bermudez, Luis	bermudez@mbari.org	MBARI	http://marinemetadata.org/; http://www.oostethys.org
Case, Jim	casej@ccom.unh.edu	UNH	N/A
Clark, Dru	dru@sdsc.edu	UCSD/SIO Geological Data Center	http://gdccoll.ucsd.edu
Cornillon, Peter	pcornillon@gso.uri.edu	URI	N/A
Cumming, Ian	ian@insight4.com	Insight4 Pty Ltd & TPAC	http://www.insight4.com http://www.tpac.org.au
Domenico, Ben	bendomenico@gmail.com	UCAR	N/A
Ferrini, Vicki	ferrini@ldeo.columbia.edu	LDEO/WHOI	http://www.marine-geo.org
Frew, Jim	frew@bren.ucsb.edu	UCSB	N/A
Hardy, Darren	dhardy@bren.ucsb.edu	UCSB	N/A
Helly, John	hellyj@ucsd.edu	UCSD/SDSC	http://ssdb.iodp.org http://sioexplorer.ucsd.edu http://gdcoll.ucsd.edu:8080/digarch
Janee, Greg	gjanee@alexandria.ucsb.edu	UCSB	N/A
Johnson, Paul	paul@hawaii.edu	UH	N/A
Koppers, Anthony	akoppers@ucsd.edu	UCSD/IGPP	http://earthref.org
Miller, Steve	spmiller@ucsd.edu	UCSD /SIO Geological Data Center	http://ssdb.iodp.org http://sioexplorer.ucsd.edu http://gdcoll.ucsd.edu:8080/digarch http://gdccoll.ucsd.edu
Neiswender, Caryn	cneiswender@ucsd.edu	UCSD/SIO Geological Data Center	http://gdccoll.ucsd.edu:8080/digarch
Ram, Sudha	ram@eller.arizona.edu	U of Arizona	http://kartik.eller.arizona.edu/index.html;http://adrg.eller.arizona.edu
Stocks, Karen	kstocks@sdsc.edu	UCSD	N/A
Sutton, Don	suttond@sdsc.edu	UCSD	N/A
Weatherford, John	jweather@ucsd.edu	UCSD	N/A