Comparative Study of Metadata for Scientific Information:

The place of CERIF in CRISs and Scientific Repositories

Keith G Jeffery CLRC, Andrei Lopatenko MU, Anne Asserson, UiB

Summary

Metadata provides the human- and machine-accessible gateway to data, improves data to information, and provides the semantic context within which knowledge can be induced from information. Metadata is the means for using together scientific data from heterogeneous sources. A CRIS commonly holds data which, while useful in itself, commonly is also metadata describing more detailed data and information on projects, persons, organizations, products of R&D (patents, products, publications) equipment used for R&D and R&D funding. It is important, therefore, to classify the metadata formats used in various scientific repositories in order to understand their scope and interoperability, and their relationship to CERIF representing CRISs. Metadata formats are reviewed according to intention, abstraction level and technology criteria. The place of CERIF in CRISs in this wider sense (including scientific repositories) is considered and compared with other metadata models and formats. The superiority of CERIF (in formalism and flexibility) is demonstrated.

1 Introduction

1.1 Data, Information and Knowledge

Europe is encouraged to evolve to 'the knowledge society'. In order to reach this state it is necessary to take data and convert it to information by structuring within context. For example '06-03-2002' would mean to a USA user 3 June 2002 but to a European 6 March 2002. Indicating which digits represent day, month and year by use of metadata (e.g. mm-dd-yyyy or dd-mm-yyyy) structures the data and makes it comprehensible. Of course, the unambiguous 20020306 is preferred, and is reliable information. Then knowledge (defined as commonly accepted belief) can be produced by induction over the information – sometimes described as generating the intension from the extension. This process is analogous to the classical scientific mind-process of building hypotheses based on observed patterns in information. As an example, if one observed the timetable for planes from London to Frankfurt to be :

Start-Time	Departureairport	Flight	Arrivalairport	End-Time
0800	LHR	BA123	FRA	1000
0900	LHR	BA125	FRA	1100
1000	LHR	BA127	FRA	1200
1100	LHR	BA129	FRA	1300
etc				
etc				
1800	LHR	BA137	FRA	2000

one might induce that, during normal working hours, a BA flight leaves LHR every hour on the hour for Frankfurt and takes 2 hours (including any time-change).

Knowledge applied to a situation provides insight, and this fuels the process of technology transfer, of inspiration to do further research and of education.

1.2 The Data Deluge or Information Explosion

The exponential growth of scientists, and especially of the power of data collection devices they use, has led to problems in dealing with data and information, and finding the needed knowledge within information. Earth observation satellites provide megabytes per minute. The LHC (Large Hadron Collider) at CERN will generate approximately 5 petabytes per year. Scientists publish vast numbers of peer-reviewed papers and even more grey literature material. One of the possible approaches to solve these problems is classification and structuring of information. Metadata provides interpretation or meaning of data and facilitates information and knowledge management.

The question facing us concerns the relationship of all this scientific repository data to conventional CRISs and how to manage the data together. In general the end user requires lateral browsing over a CRIS and in-depth searching for the detailed scientific repository data associated with a record in the CRIS. For example, the user might search for all projects investigating the link between smoking and lung cancer, and then wish to dive deeply into the analysis software used, the original raw scientific data and the publications from one particular project. The answer lies in the use of metadata to provide an integrated information space and to provide access: in this way scientific repositories and bibliographic repositories (among others) - while continuing to be used for their normal purposes - come within the sphere of access of CRISs.

1.3 Data and Metadata

Metadata is the key to achieving interoperation of CRISs and of CRISs with scientific and bibliographic repositories. Metadata has been divided into three main types (Jeffery 1998):

1. schema metadata is an intensional description of extensional instances. Typically a schema consists of: database {name, size, security authorisations}, attributes {name, type, constraints}. Some of the constraints concern the attribute domain, some are inter-attribute and as such may express relationships. The schema (intension) has a formal logic relationship to the data instances (extension). This is important in ensuring data quality. It also provides a formal basis for systems. In short, schema metadata constrains the data it describes to ensure its integrity.
2. navigational metadata provides information on how to get to an information source. Mechanisms include: filename, DB name + navigational algorithm, DB name + predicate (query), URL (Uniform Resource Locator), URL + predicate (query) or various combinations of them. Navigational metadata has no formal logic relationship to the data instances, however attributes of the navigational metadata may describe the name of the collection of data instances (e.g. filename).
3. associative metadata provides additional information for application assistance. The assistance may improve performance, accuracy or precision of the system and / or provide

assistance to the end-user through a domain aware supportive user interface. The main kinds of associative metadata are:

a. descriptive: catalogue record (eg (Dublin Core))
b. restrictive: content rating (eg PICS) or security, privacy (cryptography, digital signatures) (W3C) or rights usage
c. supportive: dictionaries, thesauri, hyperglossaries (VHG), domain ontologies eg (PROTÉGÉ)

Associative metadata usually does not have a formal logic relationship to data instances although there may be systematic association relationships.

2 Usage of metadata in CRISs

2.1 Introduction

All types of metadata are used in CRISs and also in the extended sense of CRISs including scientific repositories. To assess the different metadata formats, we evaluate them on three main dimensions – (1) kind (schema, navigational, associative), (2) intended use (3) CERIF compatibility. The compatibility with CERIF is subdivided into two parts 1) if a metadata format is compatible and an interoperable solution can be developed; 2) if a metadata format describes entities different from CERIF and those can be used (extending CERIF) to provide additional services for CRIS.

2.2 Schema metadata and CRIS

The schema metadata in CRIS can be used in implementing distributed information access solutions, integration of data, building up directories of CRISs (systems assisting user to find CRIS relevant to his data requirements) and describing web resources generated by CRIS for data collecting by agents.

CERIF RDBMS is defined in SQL DDL. The EuroCRIS CERIF Task Group (ECTG) developed an (XML) Schema for basic CERIF entities (projects, persons, orgunits and relations between them). Due to a weakness of XML Schema, integrity rules like uniqueness, referential integrity are not describable. Universal Modeling Language (UML) is powerful enough to describe CERIF. In terms of schema metadata, CERIF is as fomally defined as it can be. This means it can be co-utilised with schemas from scientific repositories and bibliographic repositories; the quality of the co-utilisation depending on how formally their schemas are defined.

2.3 Navigational metadata and CRIS

The possible uses of navigation metadata for CRIS include description of data location for distributed systems, providing persistent data access for information sources, and resolving location of data by its characteristics. Examples include the University of Bergen index of CRISs (graphical interface to URLs) (BergenCRIS), DOI (Document Object Identifier) and its use in the publishing industry (a number of URLs for the same resource for different aims). CERIF uses URIs to provide navigational metadata referencing (pointing to) more detailed sources of information. In terms of navigational metadata, CERIF is as complete as it can be. This mechanism provides the primary extension of conventional CRISs to include scientific and bibliographic repository data.

2.4 Associative-descriptive metadata and CRIS

CERIF provides – when used as metadata – rather complete associative-descriptive metadata for CRISs. This CERIF metadata provides a formal information context for consideration of data from scientific or bibliographic repositories, by relating the detailed repository data to a canonical model describing research information.

2.4.1 Metadata for scientific repositories

These metadata describe scientific data or scientific data collections and represent scientific objects or results of scientific experiments, allowing them to be exchanged between applications, stored in databases or published using WWW. In CRISs they can: 1)assist in the seamless extension of the user search space from the conventional CRIS information to detailed scientific and technical information; 2)serve for precise and formal description of subjects of research information to improve search effectiveness. Some examples are:

(1) Spatial Data Standards for Facilities, Infrastructure and Environment (SDSFIE): can be used in CRIS which needs geospatial description of research information entities or scientific results

CERIF compatibility: the standard maybe used to extend the location entity in CERIF. No technologically compatible version now.

(2) Content Standard for Digital Geospatial Metadata (CSDGM): CERIF compatibility: CERIF is able describe location of objects. CERIF location consists of country (2-letters code), region code, city and mail address. CSDGM maybe used as extension to CERIF notion of location.

(3) OpenMath is a standard for representing mathematical objects with their semantics. CERIF compatibility: OpenMath can be used to describe formulae or used in mathematics research information and then used to search information.

(4) CSCM (Content Standard for Computational Model (Hill et. al. 2001) CERIF compatibility: CSCM can be used to describe computational models and then used to search information. Also it may be used to describe in details product results in CRIS oriented on computational models

(5) GILS (Guideline Interchange Standard) CERIF compatibility: GILS can be used to extend result product description in CERIF for special purpose CRISs.

2.4.2 Metadata for research information

These metadata describe sources (such as bibliographic repositories) with similar entities to those described by CERIF e.g.Digital Libraries. Translating CERIF metadata into other metadata schemas allows re-use of CERIF data by other systems: 1) to attract new users of research information; 2) to provide additional services to work with research information; 3) to organize multi-system information processing. Translating research information expressed in other metadata formats to CERIF allows: 1) one to load new data into CERIF CRIS; 2) to interoperate by exchange between CRISs through the canonical CERIF definition; 3) to interoperate by access to heterogeneous distributed CRISs using CERIF as the canonical model to drive the portal. Such metadata formats are compared:

(1) CERIF ontology is a formalized CERIF-based metadata format. CERIF2000 specification defines metadata format for CERIF in full-text and diagrams and formal RDBMS schema. CERIF ontology is a formal specification in (DAML) of what CERIF metadata means and how they maybe encoded into (RDF). CERIF ontology was developed by the euroCRIS CERIF Task Group. (RDF) maybe used to encode metadata. CERIF TG provides a toolset to import/export data from CERIF RDBMS, use them

for integration of distributed sources, provide intelligent information search. CERIF compatibility: CERIF ontology is formalized CERIF metadata definition. The final aim is full CERIF implementation..

(2) Dublin Core (DC) was adopted and applied for a national distributed CRIS in the Safari project in Sweden (SAFARI). DC is machine readable but not machine understandable (Jeffery 1999) and a formalized DC was developed. The current thinking is that DC provides a high-level associative-descriptive metadata and that more detailed, formal and domain specific associative-descriptive metadata sets are required for real interoperability or advanced information processing including query improvement and results explanation (Lagoze 2000).

CERIF compatibility: The mapping from CERIF entities and attributes to Dublin core is provided by the ECTG.

(3) Math-Net application profile : CERIF compatibility: in description of CERIF entities Math-Net mostly is a subset of CERIF. Mapping between CERIF and Math-Net is provided by CERIF TG.

(4) Grey Literature metadata (GL) includes adaptation of CERIF and DC metadata formats to describe grey literature. The importance of Grey Literature for science (Wentraub 2000) and CRIS (Jeffery 1999) has been identified. The proposed Grey Literature metadata format is developed as extended formalized DC based on Extended E-R-A modeling and may be coded as RDBMS DDL. It may also be encoded in (RDF) and (XML). CERIF compatibility: the format is developed with CERIF compatibility as a key requirement. It provides for extended description of publication entities. A formalized model of attribute encoding to preserve the referential integrity of CERIF data is proposed.

(5) SWRC (Semantic Web Research Community) SWRC is a formal (DAML) ontology. CERIF compatibility: highly intersected with CERIF.

(6) Common European format for curricula vitae (CV) CERIF compatibility: describe person expertise mostly like CERIF.

IN addition there are exist a large number of other formats to describe scientific publications: ETD-ms: an Interoperability Metadata Standard for Electronic Theses and Dissertations, Euler's metadata for electronic journals etc.

2.5 Associative-restrictive metadata and CRIS

These metadata restrict the access of users (or software systems) to sources, modes of use or alternatively provide security services like preserving intellectual rights, or encryption. CERIF handles associative-restrictive metadata by placing constraints (mapped as attribute values together with temporal constraints) in the linking relations which represent the relationship between, for example, an author and a publication or a user and a publication. This provides much more flexibility than storing associative-restrictive attributes with the publication or with the person (in role author or user) since it is easily changeable and extensible.

2.5.1 Intellectual rights and secure policies for Research Information

The intellectual rights issues for CRIS were investigated in (Seipel 2000); (Losano 1995); (Beren & Rubert 1995). Such restrictions require a metadata to describe information access/copy/use policies, for users, and for information. A general description of the application of Digital Rights technologies in CRIS-like applications can be found in (Ianella 2001); (Erickson 2001).

(1) XrML (eXtensible rights Markup Language) is a Markup Language to describe intellectual rights and conditions associated with digital content, resources, and services. XrML was developed at PARC

(Xerox Palo Alto Research Center) and governed by ContentGuard, Inc. CERIF compatibility: CERIF entities can be described as XrML resources.

(2) ODRL (Open Digital Rights Language) CERIF compatibility: ODRL can describe intellectual right for research information. CERIF entities can be modeled as ODRL asset elements.

2.6 Associative-supportive metadata and CRIS

These types of metadata are being used to create domain values for descriptive metadata and other support services and include dictionaries, thesauri, hyperglossaries, and domain ontologies. CERIF provides a framework whereby almost any associative-supportive metadata can be used. The use of separate relations (lookup tables) to hold permitted values of an attribute together with a term description provides maximum flexibility in both classifying and reclassifying an entity by its attributes. CERIF has built-in multilinguality features. Furthermore, terms in CERIF can be referenced to and from externally established thesauri and / or domain ontologies.

2.6.1 High-level Thesauri

The objective is to increase search effectiveness (Buckland 1999); the use of controlled terms improves both precision and recall, and if the attribute of the term is also a foreign key it can improve greatly referential integrity (and therefore relevance and recall of deep searches):

(1) Library of Congress Subject Headings (LCSH) is a high level thesaurus used by many libraries.

(2) Cyc is a formal ontology, covering a large set of terms with formal description of their meaning. The best use of Cyc for CRIS is likely to be development of CRIS for knowledge management, and as a terminology server for CERIF to extend such existing CERIF vocabularies as the role of persons in projects and organizations.

(3) WordNet is a terminological ontology and could be useful for CRIS as a metadata for assisting in query building or information search. Example: "deontic logic" is a hyponym of "modal logic" - and for a particular kind of search the hyponym term may give improved precision and recall.

2.6.2 Thesauri for Science

Thesauri for science are thesauri of scientific terms and research areas which can be used for indexing research information about projects or expertise to improve search effectiveness

(1) CERIF2000 / ORTELIUS, has wide multilinguality, but only very basic terms The CERIF Task Group provides a distributed version of ORTELIUS as SQL DDL for immediate installation into any CERIF-compatible database and also the (XML) encoding of ORTELIUS according to VocML.

(2) (COS Keywords) is a controlled vocabulary of terms for science which is used to describe research expertise and funding opportunities.

(3) AMS MSC – The Mathematical Subject Classification of American Mathematical Society categorises items in mathematics; there are other ontologies and thesauri in this field

(4) Research Methods Glossary – an index of terms to describe research methods of a wide set of sciences.

(5) Physics and Astronomy Classification Scheme® (PACS) is a thesaurus developed by the American Institute of Physics, and has been used in Physical Review since 1975.

2.6.3 User-Oriented Presentation

The following metadata types are used to support additional value-added services which could make CRIS more usable by providing higher-level entry-points into CRIS systems.

(1) Metadata for collection descriptions serves for description of collections. There are several developing standards in this area :

(a) (RSLP) (Research Support Libraries Programme) Collection Description is based on modeling of collections and their catalogues. Has a metadata schema and associated syntax using (RDF)

(b) (SCD) Simple Collection Description. A format of eLib Collection Description working Group.

(c) (ISAD(G)) (General International Standard Archival Description) – provided a general framework for description of archival collections

(d) (DCMICD) A standard for collection description as a set of elements as extension to Dublin Core to describe and share information about collections

CERIF, because of its naturally recursive nature, can describe collections and reference the individual items within the collection. The proposed extension to CERIF2000 to handle publications in more detail within a formal Dublin Core framework, would ensure that CERIF provides a compatible superset capability over the majority of these proposed standards.

(2) Metadata for quality rating serve to describe the quality of information resources to help users in filtering qualitative resources and evaluation of resources. Description of the quality of research information resources is created by the information provider or an expert. Value-added services can help information consumers in getting only high-relevant and qualitative information according their needs. A review of quality control procedures for research information in the medical field is provided in (Eysenbach & Diepgen 1998).

(a) PICS – Platform for Internet Content Selection, is a recommendation developed by W3C. Although developed initially to provide parental (or other) control over access by minors to unsuitable material on the WWW, an example of PICS use in the CRIS field is "Critical Appraisal of Medical Information on the Internet" (CCAMI) – a collaborative project of physicians dedicated to medical quality on the Internet.

(b) There are attempts (led by ACM) to improve quality of scientific publications in the IT field by having online digital review, where known and trusted reviewers are encouraged to provide commentary attached to any digital publication. The aim is to assist readers to judge the quality. It is not unlike the peer review process but is not anonymous. The commentary can be regarded as 'annotation metadata'.

2.6.4 Workflow and data processing

The following metadata types described workflow data: description of routines for data and document processing which maybe required for CRISs of policy-maker organizations or CRISs with a high demand for data quality, in which research information is processed by experts according to some corporate policies. Some examples of using workflow processes for research information are described in (Lindgren 2000),(Shyu-2000). Very often workflow routines are used in digital libraries providing access to theses, publications, and preprints with research information. The Workflow Management Coalition metadata [WFMC] - a metadata set allowing representation of many facets of workflow, is one of the most popular among workflow standards. It has specified a reference model, vocabulary and XML Schema to describe workflows

A particular Schema for SGML ETD Workflow Record[SGMLETD] exists and is a SGML Schema to describe the workflow processes of submitting, approving, publishing electronic thesis and dissertations. CERIF2000 provides classification schemes related to results (products, patents, publications) by linking relations such that different classification schemes may be used to apply to the same result (e.g. a publication). However, a more general treatment using annotation metadata may be more applicable.

3 Metadata and CERIF

The current (CERIF2000) definition includes metadata of the three types:

- Schema metadata to describe information sources, CRIS data models, metadata formats;
- Navigational Metadata with URLs to more detailed sources
- Associative Metadata of the kinds:
  - Associate-descriptive metadata – CERIF metadata model, and CERIF database to describe entities of research information;
  - Associative-restrictive metadata - CERIF metadata model, and CERIF database to describe entities of research information;
  - Associate-supportive metadata – CERIF vocabularies to classify entities or values.

Looking forward, potential solutions to future CRIS requirements (including scientific and bibliographic repositories) are considered to lie in different architectures, depending on the then user requirements, but utilising (an extended) CERIF.

(a) Distributed information access solutions aim to provide transparent access to data in a distributed heterogeneous environment where there is a strong need to have a semantic and structural description of the data model in repositories for CRIS. CERIF2000 is an example.

(b) Directories of CRISs are systems which assist the user or machine mediator in choosing the information source from which to obtain data. To reason about whether an information source satisfies user requirements, the data of the information source should be described by their content, actuality and other characteristics – i.e. metadata. Examples include NIWI (DRIS), California Digital Library – Directory of Collections, OBSERVER (UGA) (and scientific publications data), TSIMMIS.

(c) Describing web resources requires that research information can be collected from metadata scattered on the web. As different organizations or communities use different formats to describe research, the metadata describing the meaning of attributes, objects, properties and constraints for valid data should be defined. ECTG defined a mapping between ontologies.

One of the key requirements for CERIF was the compatibility with other metadata formats on the basis of the layered information model (Jeffery et. al. 1994) with distinct semantic, conceptual, intensional, logical and physical levels. This is to simplify the technologies to implement solutions and allow better technologies to be utilized independently of each other by using clean interfaces between the layers. To satisfy the requirement that CERIF is based on an open, widely-used standard or formats (RDBMS, (XML), (DAML + OIL), (RDF)), the CERIF2002 model is developed from the CERIF2000 model by agreement of experts from different countries after a long analysis stage. The

CERIF2002 model thus extends CERIF2000 and contains formal machine-understandable definitions of entities, properties and constraints to make the compatibility study of CERIF with other metadata formats more automatic and formally provable. This will allow also automatically to integrate CERIF based CRISs with other, non-CERIF-compatible, CRISs. Furthermore it provides the foundation for integrating scientific repositories and bibliographic respositories seamlessly into the information space of a CRIS. The vision of CERIF – in addition to its canonical data content and structure for exchange and access - is for it to provide the common vocabulary for research which can be used to create global schemas, to be a kernel (or reference) ontology for building CRIS ontologies, and to be a shared vocabulary for federation systems or systems based on harvesting of metadata.

4 Conclusions and recommendations

CERIF2000 provides an excellent basis and is demonstrably mappable to formal metadata classifications of other CRISs. CERIF2002 as proposed by the ECTG goes further and provides a basis for open interoperability providing seamless acces for users across CRISs and associated scientific and bibliographic repositories.

Contact Information

keith.g.jeffery@rl.ac.uk
alopatenko@cs.man.ac.uk
anne.asserson@ub.uib.no;

References

BergenCRIS http://www.nsd.uib.no/english/research/eucris/
Beren, M; Rubert, C; (1995) The new Spanish data protection act, EuroCRIS-1995, Milan, Italy.
Buckland, M (1999) Vocabulary as a Central Concept in Library and Information Science in: Digital Libraries: Interdisciplinary Concepts, Challenges, and Opportunities. Proceedings of the Third International Conference on Conceptions of Library and Information Science (CoLIS3, Dubrovnik, Croatia, 23-26 May 1999)
CCAMI Collaboration for Critical Appraisal of Medical Information on the Internet, http://www.dermis.net/medpics/
CORBA http://www.corba.org/
COS Keywords http://keywords.cos.com/
DAML http://www.daml.org/
DAML+OIL see [W3C]
DC http://purl.oclc.org/metadata/dublin_core/
DCMIC] http://dublincore.org/groups/collections/
DRIS http://www.niwi.knaw.nl/cgi-bin/nph-dris_search.pl?language=en
Dublin Core http://purl.oclc.org/metadata/dublin_core/
Erickson, J; (2001) A Digital Object Approach to Interoperable Rights Management, D-Lib Magazine, Vol. 7 Num. 6 http://www.dlib.org/dlib/june01/erickson/06erickson.html
Eysenbach, G; Diepgen, T; (1998) Towards quality management of medical information on the internet: evaluation, labelling, and filtering of information. BMJ 1998;317:1496-1502 (28 November 1998)

Hill, L; Crosier, S; Smith, T; Goodchild, M (2001) A Content Standard for Computational Models, D-Lib Magazine, Vol 7 Num. 6., Jun. 2001. http://www.dlib.org/dlib/june01/hill/06hill.html
Ianella R. (2001) Digital Rights Management (DRM) Architectures, D-Lib Magazine, Vol. 7 Num. 6 http://www.dlib.org/dlib/june01/iannella/06iannella.html
ISAD(G) http://www.ica.org/biblio/com/cds/isadg_2e.pdf
Jeffery, K; Hutchinson, E; Kalmus, J; Wilson, M; Behrendt,W; Macnee, C (1994) A Model for Heterogeneous Distributed Databases Proceedings BNCOD12 July 1994; LNCS 826 pp 221-234 Springer-Verlag 1994
Jeffery, K (1998) The Future of CRIS Invited Paper CRIS98 Conference, March 1998, Luxembourg. http://www.cordis.lu/cybercafe/src/jeffery.htm
Jeffery, K (1999) An Architecture for Grey Literature in a R&D Context Proceedings of Grey Literature Conference, 1999, Washington, US
Jeffery K.; Asserson A.; Revheim J; (2000) CRIS, Grey Literature and the Knowledge Society, Proceedings CRIS-2000, Helsinki ftp://ftp.cordis.lu/pub/cris2000/docs/jeffery_fulltext.pdf
Lagoze, C (2000) pers.comm.
Lindgren N.; Rautamдki A.; Managing Strategic Aspects of Research. Proceedings CRIS-2000, Helsinki, ftp://ftp.cordis.lu/pub/cris2000/docs/rautamdki_fulltext.pdf
Losano, M (1995) The Schengen Treaty and Italy - Some problems in transnational data flow, Proceedings CRIS-95, Milan, Italy
MARC http://minos.bl.uk/services/bsds/nbs/marc/commarcm.html
PROTÉGÉ http://smi-web.stanford.edu/projects/protege/Hpkb-web/MusenWestKickoff/tsld001.htm
(RDF) see (W3C)
RSLP http://www.ukoln.ac.uk/metadata/rslp/
SCD http://www.ukoln.ac.uk/metadata/cld/simple/
Seipel P., (2000) Copyright, Information Technology and the Edifice of Knowledge, Proceedings CRIS2000, Helsinki ftp://ftp.cordis.lu/pub/cris2000/docs/seipel_fulltext.pdf
Shyu Y.-M.; Kao C.-F.; The Integrated Research Information System: Government Research Bulletin (GRB), Proceedings CRIS2000, Helsinki, ftp://ftp.cordis.lu/pub/cris2000/docs/shyu_fulltext.pdf

SAFARI http://www.lub.lu.se/~colm/safari/safari.html

SGMLETD http://www.ndltd.org/workflow/workflow.htm

UKOLN http://www.ukoln.ac.uk/metadata/interoperability/

UML http://www.omg.org/technology/uml/

van Woensel, L: (1988) CERIF Manual October 1988

VHG http://www.venus.co.uk/vhg/

W3C www.w3.org

W3Cmetadata http://www.w3.org/Metadata/

Wentraub I., (2000) The Role of Grey Literature in the Sciences, http://academic.brooklyn.cuny.edu/library/access/greyliter.htm

WFMC http://www.wfmc.org

(XML) see (W3C)