The International Journal of Digital Curation
Issue 2, Volume 2 | 2007

Report from the PV 2007 Conference, DLR Oberpfaffenhofen, October 9–11, 2007

Alexander Ball,
UKOLN, University of Bath

November 2007

Summary

This is a report from the conference PV 2007, held at the German Remote Sensing Data Centre, DLR, Oberpfaffenhofen, on October 9–11, 2007. The PV conference is held every two years, and is concerned with the preservation of scientific and technical data, and the activity of adding value to the data to ensure it can be widely used and re-used. The theme for 2007 was 'the Challenge of Heterogeneity', not only in the data being supported but in the tools used to access and process the data, and in the uses to which the data may be put.

The International Journal of Digital Curation is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. ISSN: 1746-8256 The IJDC is published by UKOLN at the University of Bath and is a publication of the Digital Curation Centre.

Introduction

The PV conference series — otherwise known as *Perennisation et Valorisation* or, in full and in English, *Ensuring the Long-Term Preservation of and Adding Value to Scientific and Technical Data* — began in 2002, with the aim of promoting good practice with respect to the preservation of scientific and technical data along with the means of understanding and interpreting it. The fourth conference in this series, with the theme 'the Challenge of Heterogeneity', was held on October 9–11, 2007 at the German Remote Sensing Data Centre, German Aerospace Centre (DLR), Oberpfaffenhofen near Munich.

Over the course of the three days, thirty-eight papers and fourteen posters were presented on the topics of preservation, adding value to data, lessons learned from current practice and prospects for future research. There were no parallel sessions, which meant rather long days to fit it all in, but on the other hand there were no difficult choices to be made on which tracks to follow. The proceedings of the conference have been made available online, supplemented with the presentations used by the speakers (Mikusch & Reck, 2007).

Opening Keynote Presentations

David Giaretta of the Science and Technology Facilities Council (STFC) structured his talk around three themes: data, preservation and adding value.

Data. Giaretta noted that the distinction between data and documents ('rendered objects') does not exist in the Open Archival Information System (OAIS) Reference Model (Consultative Committee for Space Data Systems [CCSDS], 2002): they are both 'reinterpretable representations of information'. Nevertheless, there appears to be a divide in practice between those who habitually deal with data and those who habitually deal with documents. While 'data people' are all too aware of the fragility of the semantics of the data, 'document people' tend to assume that the semantics of documents is robust enough without further effort.

Preservation. In the OAIS Reference Model, preservation is the act of keeping an object independently understandable by the designated community (the main consumers of the information) in the face of changing technology and communities: as such it is a hard term to pin down. Knowledge of format ('structural representation information') is sufficient to render an object, but not necessarily understand it, especially if the object is a data set. It is not unknown for formats to be used in odd ways; Giaretta gave the example of audio files that are used to hold configuration parameters for scientific instruments. XML does not guarantee the semantics will be preserved, especially if the schema is undocumented and the XML documents permit the inclusion of, say, Base64 encoded streams. Thus the metadata needed for preservation goes beyond description and format-related information.

Adding value. While preservation activities are important, they are not always easy to justify financially, since the beneficiaries exist in the future and have no influence now. Giaretta argued that since much of the information needed for preservation can

also aid contemporary value-adding activities and improve interoperability, an alternative source of funding or information for preservation activities could be through the cyberinfrastructure supporting e-Science, for which financial justification is rather easier.

In response to questions following the talk, Giaretta noted that preservation aspects need to be taken into account at the very earliest stages — for scientific data sets, this means from the design of the instruments — and that while software is often overlooked, it too can be preserved.

Carlos Oliviera of the European Commission talked on two main themes. One was the policies under which the organization is addressing preservation issues. 'i2010 Digital Libraries' is looking at the digitization of library collections and associated issues of access and preservation. 'Scientific Information in the Digital Age', meanwhile, is attempting to identify issues of access to, dissemination of and preservation of scientific information, although one of problems with this is that many researchers and organizations are unwilling to admit to problems and failures they have experienced.

Oliviera's second theme was the programmes and initiatives under which the European Commission is funding research in the area: most notably FP6-IST, FP7-ICT and FP7 'capacities'. Among the notable projects funded in this way are PLANETS,[1] CASPAR[2] and Digital Preservation Europe.[3] Oliviera concluded his talk with the news that the third call for proposals under FP7 is due, and that proposals dealing with databases and outreach would be particularly welcome.

Ensuring Long-Term Data Preservation

Brought together under the umbrella of ensuring long-term data preservation, the papers of the first block tended to have one of three different themes. The most popular theme was opening up data repositories using web services and loosely-coupled client-server architectures.

Lyndon R. Oleson talked about increasing the visibility and re-usability of the metadata associated with Earth Observing Satellite (EROS) datasets. **Tobias Schluach** presented DLR's in-house Data Finder data management system, which can handle many different transport protocols, including GridFTP, and distributed storage resources and archives. **Christophe Arviset** of the European Space Astronomy Centre talked about contribution made by the European Space Agency (ESA) to the worldwide Virtual Observatory (VO) initiative, noting in particular the architecture that allows legacy ESA data to be interpreted by and integrated into the VO. **Freddy Fierens** and **Jan Dries** of the Flemish Institute for Technological Research (VITO) presented Flexsys, a modular data archiving system designed to cope with heterogen-

[1] Preservation and Long-term Access through Networked Services (PLANETS), http://www.planets-project.eu/

[2] Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval (CASPAR), http: //www.casparpreserves.eu/

[3] http://www.digitalpreservationeurope.eu/

eous Earth observing satellite data. **Stephan Schneider** of DLR related how two data archives were implemented using the commercial HyperTest platform; each archive is adapted to the peculiarities of the data (and project documentation) it stores, and dependencies between data files and documentation are recorded as metadata.

The second most popular theme concerned the organizational procedures and processes behind the preservation process. **Thomas C. Stein** of Washington University talked about the processes used by the Geosciences Node of the Planetary Data System (PDS) at the National Aeronautics and Space Administration (NASA), including co-operation with mission-specific data archiving working groups, setting stringent minimum standards for ingested data, and using pre-launch dry runs to ensure data is generated and channelled correctly. At the other end of chain, data is disseminated using the same standards-based, mission-independent web services approach promoted in some of the other talks.

Robert H. Rank of the National Oceanic and Atmospheric Administration (NOAA) reported on a pilot project for transferring data between NASA's Earth Observing System (EOS) and NOAA's Comprehensive Large Array-data Stewardship System (CLASS). The result was the EOS Data Submission Handbook which contained a submission agreement (declaring what data was to be transferred and what form the submission and dissemination information packages should take), an interface control document (defining the transfer protocols, data volumes and so on to be used), an operations agreement (hosts, personnel involved), along with contextual information and lessons learned from the pilot.

Kathrin Höppner talked about the work of the World Data Centre for Remote Sensing of the Atmosphere (WDC-RSAT). The Centre holds many different data sets, most of which are freely accessible but some of which are partly classified (e.g. data relating to renewable energies, climate change and air quality), and provides a number of derived information services, such as delivering safe UV exposure durations (based on location) via SMS.

Michael Lautenschlager of the World Data Centre for Climate (WDCC) presented the strategy being adopted by WDCC and the German Climate Computing Centre (DKRZ) to cope with the vast increase in data to be stored. From now on, project data will have an expiry date and will only be moved into permanent long-term storage (with bit preservation and quality assurance) if they are fully documented to WDCC standards, which should ensure the data can be browsed, discovered and used without difficulty.

The third theme in this section was specialist formats for data preservation. **Gian Maria Pinna** of ESA introduced the Standard Archive Format for Europe (SAFE). SAFE is a profile of the specialist data content packaging standard XFDU (CCSDS, 2007), and must itself be specialized one or more times before it can be implemented practically. The SAFE specializations already defined are specific to an individual mission and processing level (e.g. ENVISAT Level 0, ENVISAT Level 1, ENVISAT Level 2). In terms of support, two Java APIs with C++ wrappers are available for

working with the format, along with a set of format converters; a Toolbox for creating, editing and visualizing product instances is under development.

Matthew Dunckley of STFC discussed four different data file format description languages — EAST, DRB, PADS/ML and DFDL — their advantages over other file format identification techniques, such as file extensions or signatures, and their areas of relative strength and weakness. Briefly: EAST (CCSDS, 2000) is most suited to rigidly structured formats, and can be coupled with DEDSL (CCSDS, 2001) to provide semantic information; DRB (Gael Consultant, 2005) expands on EAST using XML Schema (World Wide Web Consortium [W3C], 2004b) and XQuery (W3C, 2007), and is easier to process automatically but harder to use manually; PADS/ML (Mandelbaum, Fisher, Walker, Fernandez, & Gleyzer, 2006) is designed for automatic processing, but is light on semantic description; while DFDL (Beckerle, 2007), also based on XML Schema, has no native support for semantic description and is not yet mature.

Nicolas Lormant of Silogic introduced a custom plasma physics data format devised for the Plasma Physics Data Centre (CDPP). The format comes in binary and ASCII flavours, and is optimized for time-series data. Tools available include a Java API with C wrapper and converters for interoperability with CDF (National Aeronautics and Space Administration [NASA], 2005), NetCDF (Rew, Davis, Emmerson, Davies, & Hartnett, 2007) and CEF (Allen et al., 2004) formats.

The remaining paper in this section was the odd one out: **Elizabeth Griffin** of the Dominion Astrophysical Observatory related her experiences of digitizing astronomical log books and photographic plates from the period 1900–1980. The data held in analogue form is vital for analysing slow moving effects such as long-term ozone trends and stellar evolution, but is very hard to digitize: logbooks are often filled out messily, while commercial scanners are not accurate enough to get reliable scans from the photographic plates.

Adding Value to Data

The second block of papers, concentrating on processes that enhance the usefulness and accessibility of data, was also largely dominated by presentations on architectures, systems and software.

Pierre-Henri Cazes of CS-SI Toulouse gave an overview and comparison of two CNES information systems, SITools and SIPAD-NG. The primary difference is that SIPAD-NG is off-the-shelf software for use in data centres, and only works with an Oracle database backend, whereas SITools is aimed at laboratories wanting to build their own system, and works with a range of database backends.

Patrick Harms of Werum Software and Systems introduced Earth Observation on the Web (EOWEB), a data catalogue that supports multiple different views on the same underlying data. New interfaces can be generated automatically from the native data model, the access protocol data model and the mapping between them.

Bernard Pruin of Werum Software and Systems gave an update on ESA's Multi-Mission Facility Infrastructure (MMFI) which went live in late 2006. The MMFI's

Processing Facility Management (PFM) now has a generic processor interface so that only a few standard processes need be supported natively; mission-specific processing is now accomplished by plug-in Instrument Processing Facilities (IPFs).

Mihai Datcu of DLR presented the thinking behind the proposed Earth Observation Image Librarian, software that would learn how to analyze images and deduce the meanings of the various components from the context.

Sergio d'Elia talked about ESA's Service Support Environment (SSE), which can coordinate the processing of a set of data by a chain of difference service providers, and its Knowledge-centred Earth Observation (KEO) distributed component-based processing environment, which supports the feature extraction algorithms and probabilistic data mining needed for image information mining.

The other papers in this block considered ways in which ontologies could be used to support interoperability.

Santa Martinez of ESA described a new approach for processing telemetry data that allows all the instruments in a mission (and across missions) to use the same generic data pipe. The new technique also allows the archive data product to be generated as part of the telemetry data processing, as opposed to being a separate process conducted later. Key to the whole approach are a telemetry description language, a data description language and Planetary Data System (PDS) product configuration files, which allow the specific data to be handled by the generic processors.

Nicolas H. Younan of Mississippi State University presented the case for using application specific ontologies that are linked by a common (domain) ontology. While the specific ontologies allow flexibility and efficiency, the shared ontology allows the specific ontologies to be mapped together, allowing semantic interoperability between the different applications.

Stephen Hughes of the Jet Propulsion Laboratory described the recent cleaning and redefinition of the PDS data model, the most widely used standard for archived planetary data sets. The original standard sprawled over more than eight hundred pages and contained over four thousand product types. The new standard uses less than two hundred pages and one hundred product types, and is entirely expressed in Universal Modelling Language (UML) class diagrams and related engineering notation, making it independent of the quirks of any given implementation method. This is now the definitive expression of PDS3, and forms the basis of the new International Planetary Data Alliance Information Model. It also promises to make the task of producing future editions of the PDS model rather more tractable than previously.

Poster Session

There were fourteen posters due to be displayed at the Conference. Sadly, not all of them made it to the venue, but the papers associated with them were available in the proceedings (Mikusch & Reck, 2007). The posters covered a mix of topics: particular systems and software, some of which tied in with presentations already given, digital archive management and data standards. There were also posters promoting the work of the German cluster of World Data Centres and the International Planetary Data Alliance.

Lessons Learnt

The papers in this block reported on the lessons that had been learned from research projects, development projects and system administration activities.

The first of the papers from the research and development side of things was presented by **Catherine Jones** of STFC, who reported on the progress of the CLADDIER project.[4] CLADDIER is working on a way of maintaining automated links between scientific data sets and the papers based on them. Jones noted that while the tools and mechanisms produced by the project so far are promising, there are still some issues that need to be resolved: preventing abuse of the blog-style trackbacks used to monitor which papers cite a dataset, balancing comprehensiveness and comprehensibility in the citation information, and mismatches between the metadata required in citations and the metadata that can be advertised through OAI-PMH (Lagoze, Sompel, Nelson, & Warner, 2004).

Focussing in on a single tool, **Christoph Reck** of DLR introduced The Operating Tool (OT), a front-end GUI for the DLR's Data and Information Management System (DIMS). The OT can be used to monitor and control the operation and configuration of the various services with the DIMS, as well as perform data management functions. It works similarly to an all-in-one remote control, inasmuch as selection of a particular service brings up customized displays and controls for working with that service.

Turning to the more administrative side of preservation, **Lyndon Oleson** introduced the appraisal system used by the US Geological Survey (USGS) to decide whether and how to preserve a given data set. The heart of the process involves collecting a comprehensive set of background documentation for the data set, and then subjecting it to peer review to determine whether other scientific organizations would be interested in the records, what the original scientific use of the records were and what possible future uses the records might have. The results of both of these stages are recorded electronically. Since 2005, twenty-three appraisals have been conducted, resulting in USGS refusing two offered datasets, disposing of ten legacy datasets and retaining eleven legacy datasets.

Kathy Fontaine of NASA reported on a similarly administrative tool, this time a Cost Estimation Toolkit (CET) for predicting the likely life-cycle cost of new long-

[4] Citation, Location, And Deposition in Discipline and Institutional Repositories (CLADDIER), http://claddier.badc.ac.uk/

term archival facilities. The toolkit is an Excel spreadsheet with VBA scripting, and works by comparing the proposed facility with twenty-nine existing data centres, then drawing analogies and making adjustments. There are ninety-four points of comparison, derived from about five hundred input values. The toolkit also has tools for checking the effect of modifying operational parameters and for displaying results graphically.

The remaining papers in this section provided an insight into the practical business of migrating data and upgrading systems within data centres.

Claude Huc of CNES reflected on recent medium and format migrations undertaken on legacy space mission data. The first major medium migration (starting 1992) was too late to save some data on obsolete or deteriorated tapes, which taught a valuable lesson about planning migrations. Format transformations were conducted initially to make the formats neutral, and later the data was migrated to a new standard format. Huc suggested that a standard data format within a discipline is key to accelerating the development of both open source and commercial tools for working with the data.

Wilhelm Wildegger of DLR related experiences of migrating the Data and Information Management System (DIMS) to new hardware and media. The process involved eight members of staff and three years of planning, while the migration itself took two weeks. Bernd Ritschel of GFZ Potsdam talked about the transformation of metadata files from DIF ASCII format to DIF XML (NASA, 2007) in the context of the Centre's Information System and Data Centre storage management architecture. Stephan Kiemle of DLR talked about the implementation of a distributed archive system, working between the two sites at Oberpfaffenhofen and Neustrelitz. Replication of data between the two sites is automatic and takes account of load balances and the respective roles of the two sites.

Future Prospects

The final block of papers enabled researchers to promote ongoing research and new projects in the field of digital curation. I only managed to catch one of these papers as I had to depart for to the airport, but as the presentations were available from the Conference website a few days later, I did not have long to wait to catch up with the remaining papers.

Three of the papers dealt with developments in specific systems. **Didier Giacobbo** of Spot Image presented the work begun by the Heterogeneous Mission Accessibility (HMA) project to improve online access to ESA's Earth Observation (EO) data. Currently, data retrieval works according to an Online Data Access – Transfer (ODA-T) concept, where an entire dataset has to be downloaded before it can be processed; HMA is looking at an alternative, Online Data Access – Consumption (ODA-C), where processing can begin before the complete dataset has been transferred. This has several advantages, such as the ability to patch together the most detailed information for a given geographical area from different data providers, each with different levels of detail and coverage. HMA is also considering the use of OGC WCS (Open Geospa-

tial Consortium, 2007) for describing services accessing the data, and JPEG2000 (ISO 15444, 2003–2007) for encoding raster image data.

Toshihiro Ashino of Toyo University presented early work on an information platform for exchanging data between heterogeneous material data resources. The platform will use RSS 1.0 (Beged-Dov et al., 2001), suitably extended with a materials module, to describe material databases and the data they hold. The feeds will use a controlled vocabulary, initially drawn from two existing online databases (Material Data Management Consortium, n.d.; National Institute for Materials Science, 2007), and extended by community participation through a web interface. Mappings between the vocabularies and data models used by different databases will be accomplished using OWL (W3C, 2004a).

Jessie Hey of the University of Southampton gave a tour of the various projects and initiatives the University is involved in — including OAI-PMH (Lagoze et al., 2004), OAI-ORE,[5] PRESERV[6] (now PRESERV2), CLADDIER, KULTUR,[7] DataShare[8] — and the different repositories it manages, including eCrystals,[9] SERPENT[10] and EdSpace.[11] Hey's talk also took in various tools, services and projects related to institutional repositories, such as OpenDOAR,[12] ROAR,[13] OAIster,[14] Repository66.org Repository Maps,[15] SciVee[16] and the Depot.[17]

John Rumble of Information International Associates took a step back from specific projects and reflected on the role of standards in modern science. While the challenges are obvious — getting wide enough participation in drawing up standards to ensure wide acceptance; overcoming conceptual, linguistic and disciplinary boundaries; etc. — the benefits for interoperability are invaluable. Rumble suggested that the keys to success are: phased development to allow for evolution in understanding; participation from a wide number of interested parties; use of standards frameworks, allowing one to concentrate on matters of immediate importance while permitting later expansion; and recognition that not all fields require extensive standards, especially those that deal with transitory data.

[5] Open Archives Initiative – Object Reuse and Exchange (OAI-ORE), http://www.openarchives.org/ore/
[6] http://preserv.eprints.org/
[7] KULTUR Consortium, http://www.jisc.ac.uk/whatwedo/programmes/programme_rep_pres/repositories_sue/kultur.aspx
[8] http://www.disc-uk.org/datashare.html
[9] http://ecrystals.chem.soton.ac.uk/
[10] Scientific and Environmental ROV Partnership using Existing iNdustrial Technology (SERPENT), http://archive.serpentproject.com/
[11] http://www.ecs.soton.ac.uk/research/projects/EdSpace/
[12] Directory of Open Access Repositories (OpenDOAR), http://www.opendoar.org/
[13] Registry of Open Access Repositories (ROAR), http://roar.eprints.org/
[14] http://www.oaister.org/
[15] http://maps.repository66.org/
[16] http://www.scivee.tv/
[17] http://depot.edina.ac.uk/

David Giaretta of STFC gave some very practical examples of standards-related work. The main focus of Giaretta's talk was the CASPAR testbeds, which aim to discover which preservation approaches are effective for which types of data. The techniques investigated will include emulation, hardware archiving, source code escrow, and rewriting software using data models and other documentation. Semantic representation information will be tested as well, by non-experts trying to understand the data. Giaretta also mentioned a new effort to draw up an ISO standard for Repository Audit and Certification (RAC).

The remaining two papers in this block presented preliminary explorations of the theoretical underpinnings of digital curation.

Giorgos Flouris of CNR/ISTI presented some work on constructing a formal theory of preservation. Flouris sees the act of preservation as a three-level process, a bit like laughing at a joke: one has to hear the joke (bit preservation), understand the words of the joke (data preservation) and also understand the context of the joke (information preservation). He argued that digital objects can be thought of as a set of answers to questions, written in the language (logic, logical theory) of an Underlying Community Knowledge base (UCK), with no meaning apart from that UCK. The act of preservation may be reduced to mapping the meanings within the original UCK to the same meanings within a new UCK; in simple cases this may be modelled using belief revision (Gärdenfors, 1992) or ontology evolution (Haase & Sure, 2004) formalisms, but in more complex cases the concept of an intermediate Expanded UCK must be employed.

Yannis Tzitzikas of FORTH-ICS presented a model of preservation which regards representation information networks as dependency trees. The model also introduces converters as a different sort of edge on the same dependency graph. The task of keeping an object intelligible for a user is modelled as ensuring that all the dependencies of the object are satisfied, either by the user's own knowledge profile (i.e. the set of dependency 'modules' known by the user) or by modules supplied by the repository.

Conclusions

As in previous years, the Conference managed to take in a wide range of different topics, although where papers did overlap it was interesting to see a broad consensus emerging in the responses to certain common pressures and concerns. The speakers showed that heterogeneity — the theme for the conference — could be found in the various forms that a single type of data could take, the types of data that a single system could be called upon to handle, the systems that a single interface would need to handle, and the knowledge and experience of the users of the interfaces. In all the tone was positive, and suggested that with enough careful thought, planning and funding, many of the problems faced by the digital curation community can be dealt with sufficiently well, at least for now.

Acknowledgements

This work is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) and the Economic and Social Research Council (ESRC) under Grant Numbers EP/C534220/1 and RES-331-27-0006.

References

Allen, A. J., Schwartz, S. J., Harvey, C., Perry, C., Huc, C., & Robert, P. (2004). *Cluster Exchange Format — data file syntax. Issue 2, revision 0.3* (Technical Note No. DSQMW-TN-0010). London: Queen Mary, University of London. Retrieved October 23, 2007, from http://www.space-plasma.qmw.ac.uk/csds/ ExchangeFormat/

Beckerle, M. J. (2007). *Data Format Description Language (DFDL) v1.0 core specification* (Internal Committee Working Document Version 025). Lemont, IL: Open Grid Forum. Retrieved October 23, 2007, from http://forge.ogf.org/sf/ docman/do/downloadDocument/projects.dfdl-wg/docman.root.current_0/ doc14764

Beged-Dov, G., Brickley, D., Dornfest, R., Davis, I., Dodds, L., Eisenzopf, J., et al. (2001). *RDF Site Summary (RSS) 1.0. Version 1.3.4.* RSS-DEV Working Group. Retrieved October 23, 2007, from http://purl.org/rss/1.0/spec

Consultative Committee for Space Data Systems. (2000). *The data description language EAST specification (CCSD0010)* (Blue Book No. CCSDS 644.0-B-2). Reston, VA: Author. (Also published as ISO 15889:2003.) Retrieved October 23, 2007, from http://public.ccsds.org/ publications/archive/644x0b2.pdf

Consultative Committee for Space Data Systems. (2001). *Data Entity Dictionary Specification Language (DEDSL) — abstract syntax (CCSD0011)* (Blue Book No. CCSDS 647.1-B-1). Reston, VA: Author. (Also published as ISO 21961:2003.) Retrieved October 23, 2007, from http://public.ccsds.org/publications/archive/647x1b1.pdf

Consultative Committee for Space Data Systems. (2002). *Reference model for an Open Archival Information System (OAIS)* (Blue Book No. CCSDS 650.0-B-1). Reston, VA: Author. (Also published as ISO 14721:2003.) Retrieved October 23, 2007, from http://public.ccsds.org/ publications/archive/650x0b1.pdf

Consultative Committee for Space Data Systems. (2007). *XML Formatted Data Unit (XFDU) structure and construction rules* (Red Book No. CCSDS 661.0-R-1). Reston, VA: Author. Retrieved October 23, 2007, from http://public.ccsds.org/ sites/cwe/rids/Lists/CCSDS%206610R1/Attachments/661x0r1.pdf

Gael Consultant. (2005). *DRB 2.2 RC-12 users' manual.* Champs-sur-Marne, France: Author. Retrieved October 23, 2007, from http://www.gael.fr/drb/site/drb-sum.pdf

Gärdenfors, P. (1992). Belief revision: An introduction. In P. Gärdenfors (Ed.), *Belief revision* (p. 1-28). Cambridge, UK: Cambridge University Press.

Haase, P., & Sure, Y. (2004). *State-of-the-art on ontology evolution* (No. SEKT/2004/D3.1.1.b/v0.5). Karlsruhe, Germany: SEKT Consortium. Retrieved October 23, 2007, from http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/SEKT-D3.1.1.b.pdf

ISO/IEC 15444. (2003–2007). *Information technology — JPEG 2000 image coding system.* Geneva, Switzerland: International Organization for Standardization. (Multipart standard)

Lagoze, C., Sompel, H. Van de, Nelson, M., & Warner, S. (2004). *The Open Archives Initiative Protocol for Metadata Harvesting. Protocol version 2.0 of 2002-06-14.* Open Archives Initiative. Retrieved October 23, 2007, from http://www .openarchives.org/OAI/openarchivesprotocol.html

Mandelbaum, Y., Fisher, K., Walker, D., Fernandez, M., & Gleyzer, A. (2006). *PADS/ML: A functional data description language* (Tech. Rep. No. TR-761-06). Princeton, NJ: Department of Computer Science, Princeton University. Retrieved October 23, 2007, from ftp://ftp.cs.princeton.edu/techreports/2006/761.pdf

Material Data Management Consortium. (n.d.). *Material Data Network.* Author. Retrieved October 23, 2007, from http://www.matdata.net/

Mikusch, E., & Reck, C. (Eds.). (2007). *PV 2007: Ensuring the long-term preservation and value adding to scientific and technical data: Conference proceedings.* Oberpfaffenhofen, Germany: German Remote Sensing Data Centre, DLR. Retrieved October 23, 2007, from http://www.pv2007.dlr.de/main/results_ en.htm

National Aeronautics and Space Administration. (2005). *CDF internal format description. Version 3.0.* Greenbelt, MD: Space Physics Data Facility, National Aeronautics and Space Administration. Retrieved October 23, 2007, from ftp://cdaweb.gsfc.nasa.gov/pub/cdf/doc/cdf31/cdf31ifd.pdf

National Aeronautics and Space Administration. (2007). *Directory Interchange Format (DIF) writer's guide.* Greenbelt, MD: Author. Retrieved October 23, 2007, from http://gcmd.nasa.gov/User/difguide/

National Institute for Materials Science. (2007). *NIMS Materials Database (MatNavi).* Tsukuba, Japan: Author. Retrieved October 23, 2007, from http://mits.nims .go.jp/matnavi/

Open Geospatial Consortium. (2007). *Web Coverage Service (WCS) implementation specification. Version: 1.1.1c1, corrigendum 1* (Document No. 07-067r2). Wayland, MA: Author. Retrieved October 23, 2007, from http://portal.opengeospatial.org/files/?artifact_id=22560

Rew, R., Davis, G., Emmerson, S., Davies, H., & Hartnett, E. (2007). *The NetCDF users' guide: Data model, programming interfaces, and format for self-describing, portable data. NetCDF version 3.6.2.* Boulder, CO: Unidata Program Center.

Retrieved October 23, 2007, from http://www.unidata.ucar.edu/software/netcdf/ docs/netcdf

World Wide Web Consortium. (2004a). *OWL Web Ontology Language overview* (W3C Recommendation). Author. Retrieved October 23, 2007, from http://www.w3. org/TR/owl-features/

World Wide Web Consortium. (2004b). *XML Schema part 0: Primer* (2nd ed.; W3C Recommendation). Author. Retrieved October 23, 2007, from http://www.w3.org/TR/ xmlschema-0/

World Wide Web Consortium. (2007). *XQuery 1.0: An XML query language* (W3C Recommendation). Author. Retrieved October 23, 2007, from http://www.w3.org/TR/ xquery/