InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems

International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records

APPENDIX 18

Case Study Data Relating to Metadata

Status: Final (public)

Version: Electronic

Publication Date: 2008

URL: http://www.interpares.org/display_file.cfm?doc=ip2_book_appendix_18.pdf

How to Cite: Description Cross-domain Task Force, "Appendix 18: Case Study Data Relating to Metadata," [electronic version] in *International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records*, Luciana Duranti and Randy Preston, eds. (Padova, Italy: Associazione Nazionale Archivistica Italiana, 2008). <http://www.interpares.org/display_file.cfm?doc=ip2_book_appendix_18.pdf>

Case Study Data Relating to Metadata

Focus 1. Artistic Activities

General information regarding metadata

CS01, Arbo Cyber, théâtre (?)

The report states that no descriptive schemas and metadata are employed. However, records are classified by date of the performance (not by their date/time of digitization) to which they are linked. Individual practices are used to relate to the functional and technological needs of the *Ludosynthese*. However, the report reveals that if Arbo Cyber, theatre (?) decides to enter digital information, these properties will be limited to the programs' capabilities.

CS02, Performance Artist Stelarc

The report states that Stelarc has no recognized system of organization from an archival point of view with regard to his digital materials. Instead, the materials are arranged according to Stelarc's performance and publicity needs. No documented processes or procedures are used to identify, retrieve or access his digital materials. Although some records access and modification restrictions are in place, these do not appear to be formally documented. In effect, there are few, if any, formal recordkeeping practices and no metadata are consciously or intentionally recorded.

CS03, *HorizonZero*/ZeroHorizon Online Magazine & Media Database

The report states that the organization of the files pertaining to each issue of *HorizonZero* is ad hoc and is generally organized by the issue for which they were created. These files are accessible through a shared space that can be navigated using tracking software that organizes the posting into threads.

These tracker entries are saved using an archival function implemented in the tracker software (Mantis 0.18.0A4).

CS09(01), Altair4 di Roma

The report states that there is neither a recordkeeping system nor metadata schemas; however, Altair4 uses the "Where is it" program to reorganize and retrieve digital entities. To use them, it is necessary to know the filename, path and approximate date of production.

CS09(02), National Film Board

The report states that all work done using the computer as an intermediary [...] is kept on the server system and is related to a given project by the project number (assigned before a production is given the go-ahead; once a production is approved, it is given a unique production number).

CS09(03), Commercial Film Studio

The report states that only those digital entities that are archived have metadata. The standards used are Dublin Core, the Thesaurus for Graphic Materials I & II and AACR2.

CS09(04), WGBH Boston

It is important to note that the current production entity investigated during the case study consisted of a mixed analogue/digital production system. At the time of the case study, the creator was in the process of converting to a digital asset management (DAM) system, while at the same time maintaining its collection of analogue film, tapes and audio content that dates back to the 1950s. Catalogue records for these materials are kept in a FileMaker Pro 7.0

database designed and developed in-house. The DAM system is an Artesia TEAMS product that has been customized by WGBH.

CS10, Danube Exodus

File naming is largely ad hoc and some individuals develop their own system. Therefore, there is no formal recordkeeping system; furthermore, there is no system to track the changes, actions or transactions to the digital files.

CS13, *Obsessed Again...*

The report implicitly states that no metadata schemas or standards are employed. There is no formal recordkeeping system. All digital entities are stored on computer disks, which remain in the possession of the composer. These entities are only identified through the assignment of a semi-descriptive filename.

CS15, *Waking Dream*

The reports states that metadata is not consciously captured. The digital entities are kept in simple directories and are not entered in any sophisticated recordkeeping system. Professor Fels wrote the code used in *Waking Dream* and maintains it on his computer. Thus, retrieval and access of these digital entities is dependent on whether or not the computer in question contains the necessary application.

Metadata information in the 23 questions:

4d. How are the digital entities identified (e.g., is there a [persistent] unique identifier)?

CS01, Arbo Cyber, théâtre (?)

Arbo Cyber, theatre (?) does not make use of a persistent or unique identifier for electronic records, but they do use a naming convention. This was referred to during the interviews as the "nomenclature": it makes use of a strict set of punctuation and spelling rules and relies on signifying and representative values[24]. This abbreviation code is very important in the *Ludosynthese*, as it indicates location within the site.

CS02, Performance Artist Stelarc

The digital entities are identified under project titles, event series and biographical content on the Web site.

CS03, *HorizonZero/*ZeroHorizon Online Magazine & Media Database

The digital entities are identified by naming conventions that are ad hoc, though some staff members have evolved consistent naming conventions for their own work.

CS09(01), Altair4 di Roma

These conventions comprise the folder with project name/file object name/number of version and the last version file object name/final version.

CS09(02), National Film Board

No information provided.

CS09(03), Commercial Film Studio

Strict naming conventions are used to identify the digital entities, and all those having a role in manipulating the file are required to adhere to these conventions. Among other elements, the name of the file contains information on the sequence, the scene, the name of the object as well as numerical information to identify the version. The sequence of information in the file name is:/studio/title/sequence/scene/object/version.

Interpretation of this information is as follows: "Studio" refers to the name of the studio that owns the artwork, since occasionally artwork is outsourced to another studio or a subsidiary. "Title" refers to the working title of the film being produced. "Sequence" and "Scene" refer respectively to these parts of the film (in the parlance of the studio studied, "scene" is the equivalent of "shot"). "Object" refers to the particular piece of artwork in hand. Finally, a version number is added to identify the precise iteration of the file. Sometimes in PODS (a proprietary system) or at the story stage, there is also an abbreviation for information such as the sequence date and the name of the artist. There has been some attempt to develop a consistent taxonomy. Specific terms to describe each object in development are selected in the brainstorming stage by the production team. Thus there is agreement by committee on the naming conventions to be used for each production. These, however, do not extend from one production to another.

CS09(4), WGBH Boston

Current: Yes, and the unique identifier links the catalogue red in the log with the original footage. The original footage and logs follows naming conventions that link them together and to the final production. Please see question 4(f).

DAM: Same as above.

CS10, *Danube Exodus*

No alternative attempt to apply persistent unique identifiers was noted. Most files were organized in folders whose directory structure seemed to follow the intellectual conceptualization of the project.

CS13, *Obsessed Again...*

The report states that the format of each digital file is dictated by the specifications of the individual software programs with which they were created. The NoteWriter, Max/MSP and Editor/Librarian files are proprietary, binary formats and, as such, their specifications are unreleased. The MIDI files used by the Max/MSP patches are standard text files following the MIDI specification.

CS15, *Waking Dream*

The report states that the digital entities are uniquely identified with file names and, when changes have been made, with version numbers.

18b. From what applications do the recordkeeping system(s) inherit or capture all digital entities and the related metadata (e.g., e-mail, tracking systems, workflow system, office system, databases, etc.)?

CS01, Arbo Cyber, théâtre (?)

This question does not really apply to Arbo Cyber, theatre (?), but it can be said that the documents are influenced by the programs used by the artists, such as Photoshop, Illustrator or Flash. However, the properties gained through these programs have no real significance and therefore cannot be seen to have any real value for the recordkeeping system.

CS02, Performance Artist Stelarc

The applications that Stelarc captures their digital entities and related metadata are from the following, the mail system, Web-driven database operated by Web host, Internet networks, public databases functioning as sources for data mining and conversion into performance images.

CS03, *HorizonZero*/ZeroHorizon Online Magazine & Media Database

The report states that the recordkeeping system is not an RMA; the documents are "captured" by transferring them from individual hard drives to the shared server space. Metadata are attached to those documents (once again, not automatically) that are subsequently transferred to the ZeroHorizon database.

CS09(01), Altair4 di Roma

Not applicable.

CS09(02), National Film Board

The records in the recordkeeping system come primarily from office systems such as Microsoft Office, as well as from various graphics systems (for photography and posters, for example).

CS09(03), Commercial Film Studio

Another database, built on FileMaker Pro and called ArchiveWorks, is used for tracking physical pieces of artwork that are not digital.

CS09(04), WGBH Boston

Current: Productions stand alone FileMaker databases feed into the Archives database.

DAM: Same as above and through direct user input.

CS10, *Danube Exodus*

None of the subjects have a formal or automated recordkeeping system, though all have some process by which records are kept. There is therefore no system in place to track changes, actions or transactions to digital files, beyond renaming by individuals and such strategies, and, as far as can be ascertained, none of the subjects employ any kind of digital or media asset management system that could perform similar functions. (It has not been possible to confirm this with C_3.) All the subjects stated that they attempted to keep all relevant files, despite only really being concerned about the fate of work files and any secondary files that would allow them to remain functional. What constituted relevant or important files was largely left to the discretion of whatever individual was regarded as responsible for the project; for instance, the Project Manager at the Labyrinth Project.

CS13, *Obsessed Again...*

None.

CS15, *Waking Dream*

Not applicable.

18d. Does the recordkeeping system provide ready access to all relevant digital entities and related metadata?

CS01, Arbo Cyber, théâtre (?)

The report states that access is not direct, because the preservation strategy involves transferring records and placing them on external storage devices. Furthermore, Arbo controls their own entities without any need for particular measures of control.

CS02, Performance Artist Stelarc

Yes. Links are also present to make collaborators' Web sites and other relevant internet locations accessible. If general links become obsolete, the webmaster will keep them on the Web site as dead links. If important links become obsolete, new links will be set up to make that information accessible.

CS03, *HorizonZero*/ZeroHorizon Online Magazine & Media Database

Yes.

CS09(01), Altair4 di Roma

Not applicable.

CS09(03), Commercial Film Studio

Yes, access is maintained for all relevant digital entities and their metadata. Everything in the system that can be opened can be downloaded.

CS09(04), WGBH Boston

Current: No, the analogue/digital hybrid nature makes access cumbersome, though possible.
DAM: The fully digital nature of the recordkeeping system allows for greatly improved access, as well as the implementation of automatic standard language applications and thesaurus capability.

CS10, *Danube Exodus*

The report does not explicitly state how it provides access to the digital entities.

CS13, *Obsessed Again...*

Again, no system exists, but Dr. Hamel currently has ready access to all relevant digital entities.

CS15, *Waking Dream*

Not applicable.

18e. Does the recordkeeping system document all actions/transactions that take place in the system re: the digital entities? If so, what are the metadata captured?

CS01, Arbo Cyber, théâtre (?)

The lack of a true recordkeeping system makes it difficult to apply this question. The entities are saved on external storage devices; thus, it is impossible to modify them or for the system to document these modifications.

CS02, Performance Artist Stelarc

No, the Web master does not keep a record of specific updates to the Web site. The report states that the metadata are unknown.

CS03, *HorizonZero*/ZeroHorizon Online Magazine & Media Database

The report states there are no recordkeeping system.

CS09(01), Altair4 di Roma

Not applicable.

CS09(02), National Film Board

If different versions of digital entities are created by the animator, these must have a separate identification in order that they be retrievable. However, it is not known what metadata are captured as the NFB's Synchrone system (an intranet comprised of an integration of multiple databases created through in-house software developments) used is unique to the National Film Board and the subject was not queried during the interview process.

CS09(03), Commercial Film Studio

No, for the moment only the check-in and check-out transactions are documented. Some transactions modify a record's metadata, but these are not documented at present.

CS09(04), WGBH Boston

Current: Partially. Use of tapes is tracked in a FileMaker database, but re-use of shots is not tracked. DAM: Yes, each use will be noted along with versioning.

CS10, *Danube Exodus*

No.

CS13, *Obsessed Again...*

No such documentation exists.

CS15, *Waking Dream*

No metadata is consciously captured.

22. What descriptive or other metadata schema or standard are currently being used in the creation, maintenance, use and preservation of the recordkeeping system or environment being studied?

CS01, Arbo Cyber, théâtre (?)

The report states that FLA files in Flash allow for notes in a "grey-zone" that are inaccessible to users. They are used as memory aids, and no specific data is required. Furthermore, the notes only deal with content. These "grey-zones" also fail to capture information concerning the records themselves. The informant also did not see the use in identifying metadata. The informant had no knowledge of the information that can be captured in digital images. The only data attached to these images was that created automatically by the computer at the moment of creating and saving files.

CS02, Performance Artist Stelarc

This is unknown.

CS03, *HorizonZero*/ZeroHorizon Online Magazine & Media Database

No descriptive or metadata schema are consistently used for the records of *HorizonZero* pertaining to the production of each issue. There are naming conventions that describe the content of some records, but most records can be identified only by their context in the filing system.

CS09(01), Altair4 di Roma

There are no standards for activity of a creative nature. Since Altair4 uses no recordkeeping system, no reference is made to standards of description and/or indexing.

CS09(02), National Film Board

The NFB is introducing the use of MPEG-7 and MPEG-21 as standards for encoding content and rights about films. These are being introduced to simplify commercialization.

CS09(03), Commercial Film Studio

There are no standards for creation of the assets in the workflow pipeline. However, the archivist has introduced standards for description and indexing which cover those assets that make it to the archive. These include the Categories for the Description of Works of Art (CDWA), the Dublin Core (DC), the Thesaurus for Graphic Materials I: Subject Terms (TGMI), the Thesaurus for Graphic Materials II: Genre and Physical Characteristics Terms (TGM II). The Anglo-American Cataloguing Rules are used to describe scripts, manuscripts, partial notes and such. Some tracking information about other documentation is recorded using the Turabian Style Guide and The Chicago Manual of Style.

CS09(04), WGBH Boston

Current: In-house descriptive standards combined with modified Library of Congress Subject Headings. DAM: The above plus Dublin Core and PBCore (i.e., Public Broadcasting Core) compliant.

CS10, *Danube Exodus*

The interim report states that neither standards nor schemas are being used consistently in the environments studied. Forgács does capture metadata in the course of his work, but it is a system largely based on individual need, as informed by standard professional filmmaking practice. However, to date it is uncertain to the extent to which any metadata schema is currently used within the institution.

CS13, *Obsessed Again...*

There are no descriptive or other metadata schemas or standards currently being used.

CS15, *Waking Dream*

No descriptive or metadata standards are currently being used. There is no recordkeeping system being used.

23. What is the source of these descriptive or other metadata schema or standards (institutional conventions, professional body, international standard, individual practice, etc.?)

CS01, Arbo Cyber, théâtre (?)

Arbo does not use any descriptive or metadata standards. The report states that the "grey-zones" list information; thus, are not standardized.

CS02, Performance Artist Stelarc

The final report states that it is likely individual practice by Stelarc and his Web master that are the sources for any descriptive standards.

CS03, *HorizonZero*/ZeroHorizon Online Magazine & Media Database

The CanCore standard is derived from the Dublin Core metadata set, and is based on and fully compatible with the IEEE Learning Object Metadata standard and the IMS Learning Resource Meta-data specification. Other metadata sets are the result of individual practice.

CS09(01), Altair4 di Roma

The final report states that only material that is archived are then governed by international standards.

CS09(02), National Film Board

The NFB participates in international standards making bodies and is in some instances responsible for either assisting in developing these or in adapting them to the Canadian scene. These standards are, however, technical rather than descriptive.

CS09(03), Commercial Film Studio

Institutional convention governs practice during the workflow stage for any particular production. A snapshot of the entire directory structure for each production is kept, but users trying to access materials from even recent productions have been unsuccessful because of hardware and software changes that occurred in the meantime. Material that is archived is done so using the tools listed in the answer to Question 22, so professional bodies and international standards govern these activities.

CS09(04), WGBH Boston

Current: In-house data entry personnel with professional archives and library training, Library of Congress published and on-line sources. DAM: The above plus Dublin Core and PBCore (i.e., Public Broadcasting Core) reference resources.

CS10, *Danube Exodus*

The interim report states that this is not applicable.

CS13, Obsessed Again...

No such schema or standards are employed.

CS15, *Waking Dream*

Not applicable.

Focus 2. Scientific Activities

General information regarding metadata

CS06, Cybercartographic Atlas of Antarctica

The final report states that metadata in the field of geomatics are critical to business processes. The Cybercartographic Atlas of Antarctica acquires data from a number of organization and these data sets are accompanied by metadata (see Appendix K of the final report for details). The Atlas itself adheres to the ISO19115 geographic metadata standard for each module that has been entered into the MADRAS Registry developed at UCLA. Digital multimedia information objects (e.g., video clips, photos, audio, webcams, etc.) are also fully referenced and include metadata embedded into the object and/or accompanying the object and/or referenced as a caption and acknowledged in the bibliography of each content module. CS06 includes metadata-specific documents as follows:

- Excerpt - *Elements of geospatial data quality, March 8, 2002*
- *Multimedia Metadata Discussion Document, December 2003*
- Appendix P, List of Standards Adhered to on the Project

CS08, Mars Global Surveyor Data Records (NASA)

The PDS (Planetary Data System) uses self-describing data files as a preservation strategy. The labels of self-describing files describe the file format of attached data as well as the context in which the data were created. The PDS is referred to as an "active archive," whereas the National Space Science Data Center's (NSSDC's) repository is referred to as a "deep archive." The PDS is the entrance for Planetary Science data into the NSSDC archives for long-term preservation.

CS14, Archaeological Records in a GIS

The final report states that process for creating and maintaining the digital entities is ad hoc, even though GIS dynamically links geospatial data and descriptive attribute data from a wide variety of sources, and thus is a spatially referenced data set with specific metadata.

CS19, Preservation and Authentication of E-Engineering and Manufacturing Records

The main purpose of the engineering experiment examined by CS19 was to develop an open-source preservation format for digital computer-aided design (CAD) records of solid models used in high-tolerance manufacturing of complex assemblies. The experiment used Web Ontology Language (OWL), a W3C specification that extends XML to allow representation of semantics within metadata schemas, to persist the geometry, topology and functional characteristics of CAD model objects. The semantic format enabled automated querying of the digital entity's meaning, expressed in its metadata in order to assess its authenticity. The CAD model objects were developed using proprietary reasoning programs and instantiated in accordance with ISO 10303, Standard for the Exchange of Product Model Data (STEP), AP 203 and part 21 EXPRESS. STEP is ISO's metadata standard for the representation and electronic exchange of industrial product data between computer-based product life-cycle

systems. AP 203 specifies the complete boundary representation of a solid model and EXPRESS defines its elements and attributes using an object-oriented approach (see 4a, below). Metadata elements were stored in the metadata catalogue management system (MCAT) of the ISO 14721, Open Archival Information System-compliant pilot preservation system managed by CS19 partners the Electronic Records Archives (ERA) Program of the U.S. National Archives and Records Administration (NARA), the University of Maryland and the San Diego Supercomputer Center (SDSC). This preservation system also incorporated SDSC's Storage Resource Broker technology, a middleware application that uses grid and metadata technologies to transparently manage data. The intent of the experiment was to preserve not only the geometric specifications of the model but also its semantically encoded metadata, joined to make a "new logical preservation format" for archival purposes. By logical preservation format, the experiment partners in CS19 meant a format encompassing not only the fixed form and content of information representing the model, but also instructions encoded within its metadata in a way that reasoning engines of the future can conduct "proofs" against the object to authenticate it as fit to support the procedural action for which it was designed to be used.

CS26, MOST Satellite Mission

The final report states that the MOST researchers chose file formats based upon best practice; thus, resulting in metadata based upon the file format chosen.

Metadata information in the 23 questions:

4a. What are the key formal elements, attributes, and behaviour (if any) of the digital entities?

CS06, Cybercartographic Atlas of Antarctica

The information expressed is primarily cartographic, according to the functionality of each of the file types below.

Text

- HTML
- XML with XSL style sheets
- Feedback / comment forum or blog
- Databases
- PostgreSQL—open source
- PostGIS (e.g., polygons, etc.)—open source
- Excel spreadsheet (scientific numeric data—e.g., local databases)
- ESRI EOO (e.g., Antarctic Digital database)
- Flat binary (e.g., National Snow and Ice Data Center (NSIDC) at NASA)
- Graphics (e.g., remote sensing data, terrain models, Digital Elevation Models
- (DEM), Radar data, pictures, etc.)
- 2-dimensional—BMP
- 2-dimensional—GIF
- 2-dimensional—JPEG
- 2-dimensional—TIF
- 2-dimensional—PNG
- 2-dimensional—GEOTiff

- 3-dimensional—VRML and the viewer(s) required to access it (e.g., Cortona or other that works with Firefox and Mozilla browsers)

Sound

- OGGVorbis—open source
- WAV
- AIF
- AU
- Moving images
- Quicktime
- MPEG4
- Animation
- SVG (Scalable Vector Graphics)—open standard
- Flash
- Virtual reality fly-through
- VRML or video
- Games
- Online quizz
- Programming languages and technical specifications
- Javascript
- Java
- SVG
- DHTML
- XML (schema files)
- GML (Geographic Markup Language)
- VRML (Virtual Reality Modeling Language)
- Haptics (e.g., a vibrating mouse, shaking chair, force feedback devices)
- Feasible if creator wishes to do so
- Operating System, Middleware
- Linux Redhat Enterprise V4
- Apache Server—open source
- TomCat - Java—open source
- PROJ—open source
- GEOS—open source
- Geoserver—open source
- Deegree—open source
- Java SDK—open source
- XML Libraries—open source
- WFS
- WMS/WCS

For additional details about the digital entities in use, please consult the following report appendices:

- Appendices H & J: Hardware and software lists
- Appendix J: Mime Encoding of Project Software
- Appendix K: List of Data Sources for the CAA

- Appendix M: Atlas Framework, Model and File Types - Freiburg Paper and Presentation

Given the complexity of the CAA, it is not possible to list all the digital components or their individual specifications. Please refer to Figure 2 in the report for a diagram of the overall technical architecture of the CAA.

CS08, Mars Global Surveyor Data Records (NASA)

The PDS data nomenclature standards define the rules for constructing Data Element and Data Object names. A keyword (standard data element name) is an element of the Planetary Science Data Dictionary (PSDD) that defines a named property of an object. The keyword plus its value is an attribute. A label (product label) is a resource description stored in a file. If the label is in the same file as the resource, it is called an attached label. If it is in a separate file, it is called a detached label. Labels also describe the structure or format of the data. Object Description Language (ODL) is used to create labels (data descriptions) for data files and other objects such as software and documents. The PDS labels contain the key attributes of the digital objects. The behaviours of a digital object consist of the various operations that can be performed on the object. For instance, an image object is a regular array of sample values. Image objects are normally processed with special display tools to produce a visual representation of the sample values. This operation on the digital object to produce a visual representation is a behaviour of an image object. Other behaviours of these digital objects consist of the processing and analytic tools that can be used to create other objects, e.g., a tool to produce an image histogram from an image.

CS14, Archaeological Records in a GIS

The final report states that the core data set is represented in both text and numeric characters, while the outputs are textual and graphic in nature (map(s) alongside tabulated data). Furthermore, the process for creating and maintaining these entities is ad hoc.

CS19, Preservation and Authentication of E-Engineering and Manufacturing Records

There were five (5) digital entities in the CS19 engineering experiment. The first two entities listed below are produced during actual computer-aided design (CAD) and computer-aided manufacturing (CAM) processes of the original experiment partner. In the actual business process, these entities are stored with a TIFF rendition of designs as an archival aggregate in a product data management system. They were extended in CS19's engineering experiment by three additional digital entities. Each iteration of format in the experiment was chosen to either strengthen semantic expressiveness or to capture knowledge representation in a persistent, open source format:

1. The original entities (1) are created by product designers using proprietary Pro-Engineer CAD systems and are provided to colleagues charged with computer-aided manufacturing of high-tolerance, high-assurance objects used in complex assemblies. There is no formal definition of this format in the public domain as the file has a proprietary format.
2. Corporate business rules of the original experiment partner ensure that the proprietary CAD design record (1) is translated into (2) Standard for the Exchange of Product Model Data (STEP) AP203 format (ISO 10303). The formal element, attribute and behaviour definition of the objects in the STEP file are contained in ISO 10303 AP 203. The standard describes the formal representation of the Euler complete boundary representation definition of a solid model. The definition of the elements and attributes are described in an object-oriented representation language called

EXPRESS that is ISO 10303 Part 21. EXPRESS schemas are computer-processable and can be verified automatically for syntactical correctness and for the existence of appropriate links to other schemas. Instances of the defined entities form the actual exchanged data. Entity definitions include rules that can be checked at translation time to verify certain aspects of semantic validity of the transferred instances.

3. From there, the experiment took the logical form of this STEP record (2) and enhanced it into another logical form (3) that supported the delineation of additional geometric relationships and reasoning about part shape and action or process semantics using C++ based knowledge representation tools. The derived features and action semantics able to be represented by this format allow for their automated interrogation by reasoning programs, establishing semantic metadata to enable automated archival authentication of the digital solid model.
4. These entities (3) were then taken through a proprietary reasoning engine (Logistica) to complete rendition of a format (4) with all required attributes and metadata, including the formulation of logical predicates. Although the Logistica format is proprietary, it can be said that it contains a knowledge component and a procedural reasoning component.
5. The Logistica entity (4) was converted to Web Ontology Language (OWL) format (5), an open source, public domain XML specification of the World Wide Web Consortium (W3C) for persistent archiving purposes. The OWL form is in ASCII. The logical components of this form are defined mathematically by concepts of descriptive logic, and the syntax of this form is defined by the W3C in the specifications. OWL is a semantic XML format to represent machine interpretable content when the content needs to be processed by applications rather than just structured for presentation to humans. This requirement applies not only to the World Wide Web but to the digital holdings of any given domain within it, including records repositories. OWL can be used to explicitly represent the meaning of terms in vocabularies and the relationships between those terms; in other words, to operationalize an ontology. OWL has more facilities for expressing meaning and semantics than XML, RDF and RDF-S, and thus OWL goes beyond these languages in its ability to represent machine interpretable content.

CS26, MOST Satellite Mission

The key elements are mainly textual, but there are graphic elements as well.

4d. How are the digital entities identified (e.g., is there a [persistent] unique identifier)?

CS06, Cybercartographic Atlas of Antarctica

There are no unique and/or persistent identifiers, and there is no formal ID lookup system.

- The digital objects are identified by a unique combination of a file name and a location in the system
- Some objects are identified in databases, with location information included with other metadata (see Question 22 below).
- There are also some metadata embedded within some digital objects. The modules are associated with metadata. Within a module, metadata are available to reference any entity via a citation.

- Some of the maps will have embedded Geographic Markup Language (GML) to link to and describe related geo-referenced objects, such as images or sounds.
- A multimedia metadata schema is being developed. Some of the elements will be embedded within the information objects themselves and some will be linked to the object. This will become a part of the Authors' Toolkit, which includes a template of the XML schema that is completed by the content creators.

CS08, Mars Global Surveyor Data Records (NASA)

A product ID data element represents a permanent, unique identifier assigned to a data product by its producer. In the PDS, the value assigned to a product ID must be unique within its dataset. The PDS Standards Reference also specifies the rules for dataset and volume names and IDs. Each PDS dataset must have a unique data set name and unique data set ID, both formed from up to seven components. Within datasets, there are unique volume IDs. Within volumes, the file names are unique.

CS14, Archaeological Records in a GIS

Digital entities are identified through file naming conventions. Aggregations of files within certain folders can also create an associative identity of their own.

CS19, Preservation and Authentication of E-Engineering and Manufacturing Records

In the business activities of the originating experiment partner, digital entities (1) and (2), along with a TIFF version of a solid model design, are stored according to documented company policies in a proprietary product data management system (PDM). The PDM in use is a commercial records management application. This aggregate, termed a "bill of materials," is filed in the PDM according to a numbered schema corresponding with design/manufacturing procedures and there under by project number. Within digital entities (3), (4) and (5), the underlying format allows the assignment of unique identifiers at the file level depending upon business needs. This is especially true of files formatted according to the ISO 10303 STEP AP203 and part 21 EXPRESS metadata schemas, which among their functions support specification of the bond between components in complex mechanical assemblies. It also should be noted that within individual CAD files and the semantic extension formats the representation of each individual attribute or element also has persistent unique identifiers. However, the protocol of the engineering experiment did not require the unique identification of each digital entity, since there was only one instance of each of the five entities. Furthermore, CS19 is founded on the proposition (already operational in the Semantic Web) that simple enumeration of discrete identity and integrity metadata is inadequate to the demands for discovery and authentication facing the future of archives. The conception of intrinsic documentary form needs to go much further into recognizing the characteristic patterns (classes, relationships, constraints) that cohere among and between otherwise static identity attributes.

CS26, MOST Satellite Mission

Digital entities are uniquely identified by file names [managed by 1) primary target (star) and 2) date]. In addition to this, the metadata provide another set of unique identifiers. The report does not explain what these identifiers are.

18b. From what applications do the recordkeeping system(s) inherit or capture all digital entities and the related metadata (e.g., e-mail, tracking systems, workflow system, office system, databases, etc.)?

CS06, Cybercartographic Atlas of Antarctica

The final report states that CAA relies on the XML-tagged content modules for the creation of metadata. CAA content modules are developed by content creators in such a way that the linkages between the information objects, their functionality and associated metadata are described in an XML document (created within the specified XML project schema), where the markup language indicates what to display. Subversion maintains all code, and all versions of that code are tracked. Subversion is from Tigris.org—an open-source content versioning system (CVS) for use with the most popular operating systems. The Subversion database is backed up regularly.

CS08, Mars Global Surveyor Data Records (NASA)

Instrument measurements are sent as data packets from spacecraft through the Deep Space Network to computers at the Mission Ground Station at JPL (Jet Propulsion Laboratory). Computer workstations of the various project institutions are connected via NASCOM and Ethernet links to a project database (PDB) at JPL. The workstations are used to create standard data products, documentation and index tables. These are packaged into archive volumes and sent to the Science Data Validation Team (SDVT) for validation. The SDVT transfers the archive volumes to the PDS where there is additional validation.

CS14, Archaeological Records in a GIS

The final report states that there is no recordkeeping system external from the applications; therefore, no formal capture activity. There are numerous capture activities within the GIS. Other than other elements of the Microsoft Office Suite, there are no collective capture tools for the information within the GIS. Groups of data are captured temporarily within the GIS application, ArcView while analysis is being conducted, but then is exported to its appropriate areas outside of the GIS application, either from Microsoft Excel or Access files.

CS19, Preservation and Authentication of E-Engineering and Manufacturing Records

In the business activities of the originating experiment partner the digital entities created in the CAD system are captured in the corporate PDM, which is a commercial records management application system (cf. 4d, above). The expression of the experiment digital entities into the final logical preservation format was a process of derivation and extension from both proprietary and open source systems, as detailed in 4a, above. Within the protocol of the CS19 engineering experiment, the digital entities and related metadata were captured by SDSC's Storage Resource Broker and NARA-ERA's Metadata Catalog Management System.

CS26, MOST Satellite Mission

The final report states that there is no formal capture system in place, beyond the tools within Microsoft Windows.

18d. Does the recordkeeping system provide ready access to all relevant digital entities and related metadata?

CS06, Cybercartographic Atlas of Antarctica

The final report states that a "multimedia metadata schema is being developed where some of the elements will be embedded within the information objects themselves and some will be

linked to the object and these will become part of the Authors' Toolkit, which includes a template of the XML schema which is completed by the content creators." The ISO19115 metadata standard will be adhered to at the module level.

CS08, Mars Global Surveyor Data Records (NASA)

Presumably, but this is not clarified in the final report.

CS14, Archaeological Records in a GIS

No. As mentioned earlier, the recordkeeping environment is a dispersed and does not provide organized access. The creator is the intermediary between the files when access is needed, especially because the majority of the files are in the file directory or on the hard drive of the creator.

CS19, Preservation and Authentication of E-Engineering and Manufacturing Records

In the business context of the originating experiment partner the PDM system allows ready access to all digital entities and related metadata. Access is accomplished through standard queries invoked by menu picks by such attributes as procedure number, job, creator, design-change number, design release version number, etc. For the CS19 engineering experiment the SRB and MCAT systems provide a variety of means to access digital entities and any combination of metadata. In addition, the experiment protocol called for the logical querying of the semantic metadata of formats (3), (4) and (5), to authenticate the digital entity's identity, integrity and suitability for the manufacturing process for which it was designed.

CS26, MOST Satellite Mission

The final report reveals that it is possible to access all digital entities via Windows Explorer but does not actually mention how access is provided to the metadata prescribed by the MOST researchers.

18e. Does the recordkeeping system document all actions/transactions that take place in the system re: the digital entities? If so, what are the metadata captured?

CS06, Cybercartographic Atlas of Antarctica

Although this question is not directly applicable to CS06, answers to the following of the 23 research questions provided in the report do touch on this issue.

Question 8: Any digital object that forms part of the CAA must be described by the creator, using metadata standards adopted or developed by the project. See Question 20 in the report and Appendix P, which includes the project's metadata standards. Retrievability of, and access to, the digital objects are based on a number of adopted OGC interoperability specifications (see Appendices P and N in the report).

Question 10: Data are acquired from authoritative sources and are peer-reviewed (e.g., British Antarctic Survey, Scientific Committee on Antarctic Research, scientific and academic journals and books). Each is assessed against the Elements of Spatial Data Quality, which include:

- Lineage
- Positional accuracy
- Attribute/thematic accuracy
- Completeness
- Logical consistency

- Semantic accuracy
- Temporal information

See Appendices T and K in the report for the list of data sources.

Authenticity in geography is captured in standard metadata as data lineage. Quality measures are dependent on the type of data and their function (e.g., the acceptable margin of error for the precise location and size of a particular ice flow to inform tourist ships is smaller than fish counts to inform fisheries and ecological modeling). In addition, each scientific domain is governed by its particular data quality standards, measures and assurances and these are included in the metadata. Appendix P in the report includes a list of such standards.

Question 13: Changes to the code are captured in Subversion, a source repository system used by the project. Subversion maintains all code, and all versions of that code are tracked. Subversion is from Tigris.org—an open source content versioning system (CVS) for use with the most popular operating system. The Subversion database is backed up regularly. Other digital objects that form part of the CAA are not captured by Subversion.

The Authors' Toolkit will eventually allow changes to associated metadata to be tracked as well. Also see:

- Excerpt—*Elements of geospatial data quality, March 8, 2002*
- *Multimedia Metadata Discussion Document*, December 2003

CS08, Mars Global Surveyor Data Records (NASA)

The PDS logs accesses to restricted areas of the system. User ID, date, time and operation are logged.

CS14, Archaeological Records in a GIS

The report explicitly states that there is no audit trail. The GIS Specialist is in the process of creating metadata relating to the source of the data, including the original author, date or recording, etc.

CS19, Preservation and Authentication of E-Engineering and Manufacturing Records

The PDM system used by the originating research partner in actual business processes captures actions, events and changes to the digital entities (1), (2) and the bill of materials aggregate. Metadata is typically name of creator, release version numbers, date of release, etc. The SRB and MCAT systems captured all changes to the representation of the CAD solid model as it migrated through the semantic format extensions (3), (4) and (5), including the formulation of metadata that support querying by automated reasoning programs.

CS26, MOST Satellite Mission

The final report states that there is no audit trail.

22. What descriptive or other metadata schema or standard are currently being used in the creation, maintenance, use and preservation of the recordkeeping system or environment being studied?

CS06, Cybercartographic Atlas of Antarctica

The final report states that the CAA has solid metadata practices in place; these metadata practices include the following: FGDC and/or British Antarctic Survey DIF (Directory Interchange Format), OGC interoperability specifications and the International Standards

Organization 19115 Geomatics Standards. The report also indicates that the ISO 19115 metadata standard for digital mapping data has been explored (see *Multimedia Metadata Discussion Document, December 2003*).

CS08, Mars Global Surveyor Data Records (NASA)

The final report states that the *Planetary Science Data Dictionary* (PSDD) is used in the creation, maintenance, use and preservation of the PDS. The PSDD contains definitions of the standard data element names and objects.

CS14, Archaeological Records in a GIS

The final report states that they are interested in using ArcCatalogue, a metadata tool that is in the new version of ArcView. Their main goal relating to metadata capture surrounds source information relating to CC Database data. The metadata would indicate from what source (publication, repository, Web site, database, etc.) the data was retrieved. In addition, time tagging of georeferenced information is part of the documentation of the processes of creating online digital maps, models and georeferenced visualizations.

CS19, Preservation and Authentication of E-Engineering and Manufacturing Records

The final report states that the first digital entity (1), produced during actual computer-aided design (CAD) and computer-aided manufacturing (CAM) processes of the original experiment partner, originates in a proprietary software tool, thus the precise metadata schema is unavailable. However, the tool produces models in conformance with the ANSI Y-14.5 tolerance standard and provide export files (2) compliant with ISO 10303 Standard for the Exchange of Product Model Data (STEP), AP 203 and part 21 EXPRESS. Corporate metadata standards and procedures govern the filing of these two digital entities with a third TIFF export of the model view into a commercial Product Data Management System. The formats of CS19's digital entities (3) and (4) included the formulation of additional semantic metadata by in-house computer scientists expert in knowledge representation systems that supported the delineation of additional geometric relationships of the CAD solid model and reasoning about part shape and action or process semantics. Although some of the metadata supporting action semantics was lost in the translation to digital entity (5), OWL XML, it was able to persist and authenticate precise specifications about part shapes and relationships, including the classes, predicates and constraint rules that govern the identity and behavior of the CAD solid models.

CS26, MOST Satellite Mission

The metadata schema that is used was created by the MOST researchers and is specific for the data/files that are created in the MOST project. The metadata refer to information such as orbital parameters, observational parameters, telemetry information and target image information. The report notes that some of the metadata/descriptive fields in the FITS files are mandatory, due to the file format. In general, no metadata standards are used; the MOST researchers have created their own scheme of important descriptive fields.

23. What is the source of these descriptive or other metadata schema or standards (institutional conventions, professional body, international standard, individual practice, etc.?)

CS06, Cybercartographic Atlas of Antarctica

The source of metadata comes directly from professional bodies, institutional conventions, as well as international standards. The Atlas adheres to ISO 19115 at the modular level and additional research is ongoing regarding metadata at the granular level.

See response to question 22, above.

- International Standards Organization (ISO)
- Open Geospatial Consortium (OGC)
- Scientific Committee on Antarctic Research (SCAR)
- Geomatics and Cartographic Research Centre (GCRC)
- DIF Format (see http://gcmd.gsfc.nasa.gov/User/difguide/difman.html for details)

The CAA project itself: Y. Zhou, MA thesis on this topic entitled "Profiling and Visualizing Metadata for Multimedia Information in a Geospatial Portal."

CS08, Mars Global Surveyor Data Records (NASA)

The *Planetary Science Data Dictionary* is a NASA institutional standard for Planetary Science Metadata. The PDS procedures for assigning standardized names to data elements follow closely the NBS Guide on Data Entity Naming Conventions.

CS14, Archaeological Records in a GIS

Within ArcCatalogue, the user could create, manage and edit metadata based on the Federal Geographic Data Committee (FGDC) Content Standards for Digital Geospatial Metadata or the ISO 19115 Metadata Standard. These metadata would be stored in XML.

CS19, Preservation and Authentication of E-Engineering and Manufacturing Records

ANSI, ISO, W3C, corporate business rules

CS26, MOST Satellite Mission

The metadata that are used for the various files are based on experience and best practice in the astronomical community and on the foreseeable use of the records in the future. There is an internal MOST document that describes the descriptive fields of the FITS files.

Focus 3. Governmental Activities

General information regarding metadata

CS05, Archives of Ontario Web Exhibits

The final report states that the Ontario government has developed a standard look and feel to which all government Web content must adhere. These are standards created or adopted within the Ontario Public Service. One such standard is the metadata, which refers to title, keyword and description and classification metatags.

CS12, Antarctic Treaty Searchable Database

The final report states this "is an automated technology that objectively integrates digital-record entities without markup, metadata or databases." However, the report further states that "unlike subjective content descriptions in metadata or controlled vocabularies, DIGIN® comprehensively searches both the contents of the granules and their categorical tags to objectively identify those granules that match the search queries. DIGIN® is interoperable with metadata, mark-up and databases."

CS17, New York State DMV On-line Services System

The final report states that the DMV provides a highly interactive online system featuring a complex set of interwoven electronic activities, which collects information about the user via cookies, Web protocols and transactional metadata. A third party digital signature company, VeriSign, is used to make transactions legally binding.

CS18, Alsace-Moselle's Land Registry

The final report states that there are no descriptive or other metadata schemas or standards; however, within the relational database the data are linked together through queries. It has been explained that there is a metadata schema that will be completed for the second phase of the project.

CS20, Revenue On-Line Service (ROS)

The final report states that data mining of ROS-created data is used to audit tax details, improve efficiencies, increase customer service and enable fraud detection.

CS21, Electronic Filing System, Supreme Court of Singapore

The law firm is expected to enter information on their cases through a prescribed documentary template in EFS. Some of the metadata elements are fixed as there is a pull-down menu for law firms to select. Some of the metadata elements the file has to enter include the firm's file reference number and party details, which include the party type, name and address of parties and name of solicitor. EFS captures both the metadata of the record and the actual record itself; the court must check both the metadata and the record.

CS24, City of Vancouver GIS (VanMap)

The interim report states that metadata is not a means of tracking how the information is used, but it does reveal what information is being used and when; this is conducted through the generation of statistical reports.

CS25, Legacoop of Bologna Web Site

The final report explains that all paper mail (what is sent to the organization and what is sent directly to the single functionaries) is registered. And, that the electronic mail sent to the organization official email-system is registered when it is determined that the message is of a certain importance. The registry system uses an automated application to register the records. This application provides a profile of the registered incoming and outgoing documents, including the following: classification code, recipients, object, date and type of document.

Metadata information in the 23 questions:

4a. What are the key formal elements, attributes, and behaviour (if any) of the digital entities?

CS05, Archives of Ontario Web Exhibits

Elements and attributes that are considered integral to the validity and completeness to the Web exhibits Elements were determined based on how the exhibits are normally accessed.

Key intrinsic elements include:

- navigation links from the institutional home page to a listing (with or without a précis of the exhibit);
- exhibit content, normally comprised of Web pages containing text, images, and occasionally with sound or video files;
- government visual identity signs, especially the provincial and city logos and the institutional;
    - Provided by a central body for all Ontario Web sites are:
        - Standard disclaimers
        - Instructions for accessing and installing plug-ins
        - Copyright statements
        - Privacy statements
        - Graphics (.gif format) provided for every ministry name

- Graphics for mandatory toolbars are provided
- Ontario logo, mandatory for every government Web page, and footer graphics are provided

The last three are compliant with the W3C's WAI (Web site Accessibility Initiative) requirements, and all text is provided in English and French.

Key extrinsic elements include:

- A corporate standard Web page template
- The cascading style sheet created for the Web site as a whole
- The institutional Web site (contains other exhibits, links to databases, external links, etc.)
- The corporate Web environment (contains links to all government Web sites, news releases, etc.)
- HyperText Markup Language, specification version 4.01
- Navigation bars required at the top/bottom/side of each Web page
- A "feedback form" that utilizes Common Gateway Interface (CGI) script to interface with an email application

Behaviour of the rendering platform takes place on two levels:

1. the feedback form is a CGI program executed in real-time; and
2. the way the user's browser interacts with the HTML coding of the exhibits.

CS12, Antarctic Treaty Searchable Database

Indeterminate from answer provided.

CS17, New York State DMV On-line Services System

The records are live records and have the ability to change over time. They can be placed into a status where they are no longer alterable, as when a driver dies or a vehicle is junked. The official records contain an official crest or logo.

CS18, Alsace-Moselle's Land Registry

The ordonnances take the form of XML files, containing tagged information relative to landowners, land parcels and rights/obligations relative to the property. Associated with the ordonnance is a digital signature of the judge authoring the ordonnance. The structure of the ordonnance is defined through a DTD. The scanned images of the register take the form of TIFF files.

CS20, Revenue On-Line Service (ROS)

There are some elements and attributes common to the three "classes" of records [digital certificates/signatures, tax forms and debit instructional forms] for presentation, Revenue logo, font and style, certification practice statement, privacy policy, terms and conditions, copyright statements, contact details and standard Web page templates.

CS21, Electronic Filing System, Supreme Court of Singapore

The EFS is composed of standardized HTML style sheets, XML files, Visual Basic, PDF and graphic files for the EFS logo.

CS24, City of Vancouver GIS (VanMap)

The data sheets describe VanMap's data layers, features and functionalities; each layer typically contains, layer name, group name, scale, data currency status, responsible department and definition.

CS25, Legacoop of Bologna Web Site

All the entities have at least a title and a body text and a date. Each element is numbered sequentially according to the chronological order.

4d. How are the digital entities identified (e.g., is there a [persistent] unique identifier)?

CS05, Archives of Ontario Web Exhibits

Each Web exhibit is identified by its title and a URL, which has been assigned within the institution's Web domain. When viewing the source coding for each Web page within each exhibit, each page is also titled.

CS12, Antarctic Treaty Searchable Database

Each of the information granules or digital-record entities in the current *Database* contains unique provenance information in a categorical header tag as well as in the title. Unlike metadata, which are stored in repositories separately from the digital entities, the unique identifiers are part of each granule in the *Database*. Thus, with the categorical header tags, there is never a risk for decoupling the unique identifiers and the digital entities.

CS17, New York State DMV On-line Services System

There is a unique identifier connected to each transaction. The transaction and its identifier are stored with the core record, as a result of the transaction. Different sets of identifiers exist for each of the three file types: license, registration, and title.

CS18, Alsace-Moselle's Land Registry

Every inscription in the database is numbered with a persistent, unique identifier and dated. Ordonnances are also numbered and dated. Each scanned image of the registers is numbered according to the system already in place for numbering individual pages of the registers.

CS20, Revenue On-Line Service (ROS)

There is no need for specialized codes and keys beyond those normally used by the Revenue.

CS21, Electronic Filing System, Supreme Court of Singapore

The case number is a unique identifier, which is automatic generated number assigned by the courts.

CS24, City of Vancouver GIS (VanMap)

The HTML and CML pages and embedded GIF images are identified by unique URLs. The data fields, layers and groups are also identified by field names, layer names and group names.

CS25, Legacoop of Bologna Web Site

A primary key in the form of a progressive number (managed as a key field in the database) is the main identification attribute.

4e. In the organization of the digital entities, what kind of aggregation levels exist, if any?

CS05, Archives of Ontario Web Exhibits

The Web exhibits and the Web pages reflect aggregations of text, images and other components of the exhibit which are conceptually linked. The institutional Web sites and the Web exhibits are grouped together for the navigational convenience of the user.

CS12, Antarctic Treaty Searchable Database

The aggregation levels among digital entities are based on the inherent parent-child relationships within the policy documents. In general, the aggregation levels or hierarchy levels reflect the granularity of a digital collection. This collection granularity is represented specifically by:

- Antarctic Treaty Searchable Database > Year > Major Document or Antarctic Treaty Consultative Meeting > "measures"

Dynamic aggregation of digital-record entities with DIGIN® facilitates the discovery of relationships within and between the digital-record series.

CS17, New York State DMV On-line Services System

The DMV does not use directories or subdirectories, but keeps everything in tables and databases. The individual transactions in the audit trail are organized by date and time, category and current status.

CS18, Alsace-Moselle's Land Registry

The database aggregates the data according to the main categories: parcels, persons, rights and obligations. The presentation of data is organized in the same way as the paper register; that is, a single *feuillet* contains information relative to all the properties of a person within a given administrative territory (usually, a commune, or part thereof).

CS20, Revenue On-Line Service (ROS)

All tax records and debit instructions are saved chronologically and are viewable within the Revenue Customer Information Service. They can be sorted and viewed depending upon the field type selected. Regarding digital certificates and signatures: Metadata surrounding the older Digital Certificates, in addition to the security wrapper, are maintained with ROS. Revenue has a separate Archiving Policy for Certificates, but this is considered beyond the remit of ROS.

CS21, Electronic Filing System, Supreme Court of Singapore

The main case files are divided onto various sub-folders based on the type and nature of records filed, such as affidavit, draft order, minute sheet and summon in chambers.

CS24, City of Vancouver GIS (VanMap)

HTML and related pages are grouped into folders for storage, identification and retrieval purposes. The data are organized into layers, with each layer including a single data source or multiple data sources.

CS25, Legacoop of Bologna Web Site

The entities are aggregated according to the main logical categories of the Web site (documents of the association, news from the cooperation world, CVs and announcements, other services related to the Bologna business area).

18b. From what applications do the recordkeeping system(s) inherit or capture all digital entities and the related metadata (e.g., e-mail, tracking systems, workflow system, office system, databases, etc.)?

CS05, Archives of Ontario Web Exhibits

The exhibits are created using Dreamweaver and PageMaker software applications.

Metadata captured would normally be what are automatically captured by the default settings of those applications. None of the interviewees commented that they used the document properties function to add any specific metadata. Metadata captured would normally be what are automatically captured by the default settings of the applications used to create supporting documentation, such as Microsoft Word.

CS12, Antarctic Treaty Searchable Database

According to the final report, this question is irrelevant since, after the initial implementation of the *Database* in 1999, the only captured files are the entire Antarctic Treaty Consultative Meeting (ATCM) Final Reports without metadata that have been published on the ATCM Web sites of the host nations. The new *"measures"* that have been adopted by the Antarctic

Treaty Consultative Parties are then extracted and added to the *Database* with header tags that define their unique location in the overall collection.

CS17, New York State DMV On-line Services System

The system that the DMV uses captures IP addresses, system dates and session cookies. The cookies are used only to maintain the session state; they are not stored on the hard drive of the patron.

CS18, Alsace-Moselle's Land Registry

Both ordinances and inscriptions are captured through custom applications. The scanned images of the register were captured once at the onset of the computerization process.

CS20, Revenue On-Line Service (ROS)

Databases and other systems – ITP is held on an Ingress II mainframe back-end system.
ITP – Integrated Taxation Processing [System]

CS21, Electronic Filing System, Supreme Court of Singapore

The EFS captures digital entities from an oracle database, Filenet (document management system), jukbox (for CDs) and visual basic software. See answer to question 5a in final report for a detailed list of the application systems.

CS24, City of Vancouver GIS (VanMap)

DOMINO, PRISM, License and other systems produce the data.

CS25, Legacoop of Bologna Web Site

The report states that the recordkeeping system does not have any relationships with the Web site system.

18d. Does the recordkeeping system provide ready access to all relevant digital entities and related metadata?

CS05, Archives of Ontario Web Exhibits

The report states that related metadata is not readily accessible, even if it has been captured. This is due to the absence of a recordkeeping system and lack of consistent recordkeeping processes around the provision of access to Web exhibits within the two institutions.

CS12, Antarctic Treaty Searchable Database

The final report says yes, through providing comprehensive integrated access to the digital-record entities. And, that the Antarctic Treaty Searchable Databases does not require metadata.

CS17, New York State DMV On-line Services System

If the mainframe system is equivalent to the recordkeeping system, then the answer to this question is yes. Although customers and third party users have access to only a small portion of the digital entities, the system provides DMV personnel with access to all aspects of the digital entities.

CS18, Alsace-Moselle's Land Registry

Yes, as relevant for each category of user.

CS20, Revenue On-Line Service (ROS)

Yes, to both Revenue employees and ROS users. Not all users will view metadata.

CS21, Electronic Filing System, Supreme Court of Singapore

Yes, as relevant for each category of user. Authorized court users can view both the record profile and the PDF record. Specified group of users can, based on their job competency, view certain categories of audit logs.

CS24, City of Vancouver GIS (VanMap)

Metadata in the form of data sheets is also readily available. This study has not yet investigated the types of metadata that may be generated automatically by the various technological processes used to create VanMap.

CS25, Legacoop of Bologna Web Site

Not applicable.

18e. Does the recordkeeping system document all actions/transactions that take place in the system re: the digital entities? If so, what are the metadata captured?

CS05, Archives of Ontario Web Exhibits

The Web logging software documents aspects of all interactions with the institution's Web site. The report presents 21 reports generated by *Analog* based on the data it gathers. Please refer to page 46 of the report for this list.

CS12, Antarctic Treaty Searchable Database

The report states metadata are not captured. However, the report also states that all queries of the Web site version of the *Database* are automatically logged.

CS17, New York State DMV On-line Services System

The recordkeeping system at the DMV tracks all changes to records in the mainframe system through audit trails and user logs.

CS18, Alsace-Moselle's Land Registry

The report reveals that the system includes extensive login capabilities for recording all actions and transactions taking place in the system. Logs may be used for two distinct purposes.

CS20, Revenue On-Line Service (ROS)

The report states that all changes are noted and logged with time/date stamp and name of Revenue employee making change.

CS21, Electronic Filing System, Supreme Court of Singapore

Yes, all actions and transactions are documented in various audit logs, including:

- Transaction Log: Records user ID of user who activates the change, function name, date./time of the change, data items before and after the change
- Financial Audit Log: Records user ID, function name, date/time of the action, case number/document, control number/unique reference number, amount of fees before the change, amount of fees after the change, remarks, approval for exemption/waiver of court filing fees, approval for request of waiver of hearing fees, and approval for refund of hearing fees.
- Violation Log: Records user ID of user who attempts to access functions he or she is not granted access to, unsuccessful log in attempts, function name, date/time of the action, and brief description of the nature of the violation.

CS24, City of Vancouver GIS (VanMap)

The interim report states that the use of the data can be tracked by unique client IDs randomly generated when users download the MapGuide ActiveX Viewer to their workstations. For example, whenever a user accesses VanMap and issues a request for data the transaction results in a log file record containing his or her ID, the date and time of access and strings of numbers representing specific data layers used.

CS25, Legacoop of Bologna Web Site

Not applicable.

22. What descriptive or other metadata schema or standard are currently being used in the creation, maintenance, use and preservation of the recordkeeping system or environment being studied?

CS05, Archives of Ontario Web Exhibits

The Web site coordinator was unfamiliar with metatags, and initially ignored metadata standards. Metatags are only applied to "key pages"; therefore, is nothing that distinguishes an exhibit page from any other page on the Web site. The report states that only comprehensive source of metadata governing the entirety of an exhibit appears to be the "definition document" created for *The War of 1812* exhibit. This document includes the title, reference code, image number (where applicable), location information and a summary of the document/image.

CS12, Antarctic Treaty Searchable Database

Descriptive metadata, as conventionally applied with templates and attributes that reside in repositories, are not used to implement the Antarctic Treaty Searchable Database. However, there is the use of header tags that describe the parent-child provenance. Also, conventional metadata regarding the portal for the Antarctic Treaty Searchable Database are being added to the National Science Digital Library and Digital Library for Earth System Education. The metadata format for these submissions is a modified Dublin-Core metadata with additional fields for the education audiences that are being addressed by these digital libraries.

CS17, New York State DMV On-line Services System

Although data layouts and schema are used, the DMV respondents indicated that they did not feel comfortable revealing specifics about such information to the InterPARES research team.

CS18, Alsace-Moselle's Land Registry

No descriptive nor metadata schema is currently being used.

CS20, Revenue On-Line Service (ROS)

Twenty-two schemas for the tax forms have been made available in XML DTDs for inclusion in the third party compatible software; view www.ros.ie/downloads.html and Appendix IV. All schemas include a DTD and element definitions and explanations. Although an Irish Public Service Metadata standard exists, it is not used with ROS.

CS21, Electronic Filing System, Supreme Court of Singapore

The schemas for the documentary templates of the EFS are based on the workflow and juridical requirements of the court.

CS24, City of Vancouver GIS (VanMap)

The metadata applied by the VanMap developers include layer name; group name; scale at which the data become available; data currency status; responsible department, branch or division; and definition. Not all of these metadata are applied to all of the data layers. Metadata generated automatically upon creation of the data have not yet been investigated.

CS25, Legacoop of Bologna Web Site

Basic metadata related to the registry system are required in the recordkeeping system, but not exported to the Web site management, which handles only a numbering system for each digital entity and a date.

23. What is the source of these descriptive or other metadata schema or standards (institutional conventions, professional body, international standard, individual practice, etc.?)

CS05, Archives of Ontario Web Exhibits

There is no identified source for the Government of Ontario Category Metadata rules. The City's Web coordinator stated that the metadata tags he uses do not conform to any standards.

CS12, Antarctic Treaty Searchable Database

The final report states that conventional metadata are unnecessary with the DIGIN® technologies, which can interoperate with or without metadata to integrate "structured" as well as "unstructured" information. The sources of the descriptive schema are the persistent digital-record entities themselves.

CS17, New York State DMV On-line Services System

The source of these standards was not mentioned or discussed during the case study interview.

CS18, Alsace-Moselle's Land Registry

No descriptive nor metadata schema is currently being used.

CS20, Revenue On-Line Service (ROS)

Institutional practice. Form design and structure is based on existing paper-based forms. Field selection and management is based on requirements and format of ITP applications and data flow to this and other back-end systems. The XML schemas may include other descriptive standards such as ISO Year Standard.

CS21, Electronic Filing System, Supreme Court of Singapore

Institutional practice. The metadata used in the documentary template are based on common data elements associated with the court records that have to be converted into PDF.

CS24, City of Vancouver GIS (VanMap)

Metadata are based on what the VanMap Team thinks will be useful information for the end user.

CS25, Legacoop of Bologna Web Site

The metadata included are strictly related to the professional standard followed for building the Web site (SQL for the database and HTML for the Web site pages).