IIiv
Standards

11 THE CIMI STANDARDS FRAMEWORK

And the Interchange of Multimedia Information

John Perkins
CIMI
Canada

Increasingly, museums are looking towards "multimedia" to use in documenting collections and interpreting them for visitors. Like any of the other common types of monomedia data such as text, still images, sound etc., there are requirements for the interchange of "multimedia information objects"[1] that can be achieved through the application of standards to their structure, description, and representation.

While internationally accepted multimedia standards are in a fledgling stage and far from being universally accepted, there are aspects of them and principles underlying their adoption and use that can profitably be applied for current as well as for future benefit.

This paper presents the CIMI Standards Framework[2] as a context for a landscape view of the issues and standards relevant to the interchange of multimedia and as a focus of community effort for further development. In doing this it assumes that museums are at the beginning of the consensus building process on the adoption of standards which includes, identifying the range of applicable standards; determining those with relevance, and establishing methodologies for their successful implementation.

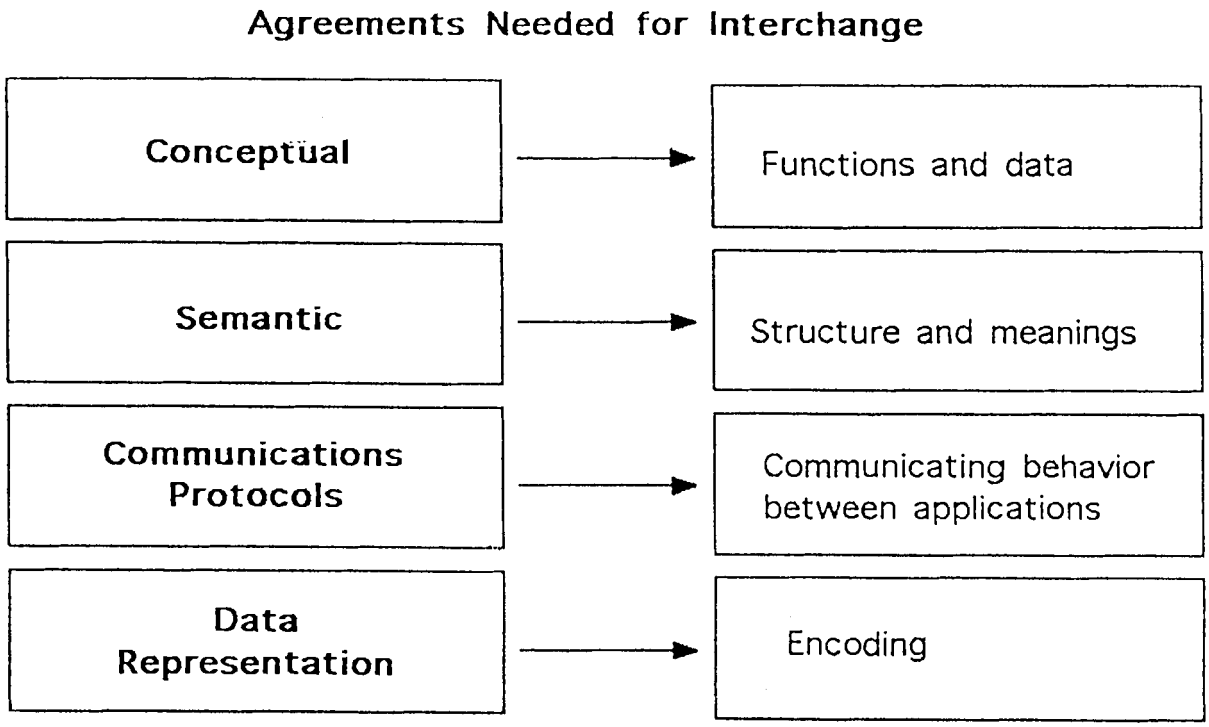
Rational for interchange standards

The basis of the standards framework is the need museums have to interchange information. This may be for: 1) application portability for example when a museum wishes to move from one computing system to another or wants to use the same application on more than one type of platform; 2) data re-use as in sharing information between applications; 3) the protection of investments in creating digital data; or 4) linking to external (to the application) sources of information such as reference databases and authority files.

1 The object-oriented approach is suggested by the ISO/IEC JTC1 SC29 in its draft standard for coded representation of multimedia and hypermedia information objects (MHEG) because it fits requirements for describing data that is active, autonomous, and reusable. It does not mean that applications or any structures must be object-oriented.

2 The Museum Computer Network (MCN) launched its initiative for Computer Interchange of Museum Information (CIMI) to develop standards that could support museum requirements.

Fig.1 Areas where standards are necessary to enable interchange.



Without a standard for such interchange, each instance of interchange requires preparation and programming. This may result in lost data, or the cost of one-off interchange outweighing the benefits. With interchange standards, basic management responsibility for preservation of museum information and integration of museum functions will be met while enhancing the potential for scholarly information exchange.

The type, quantity and degree of formalisation of the agreements necessary for interchange vary depending on the complexity of the processing functions being supported and the technical dependencies of the data. For example, for a simple exchange of ASCII text files on disk, the Conceptual Agreement is based on the inference that there is a document to exchange. The Semantic Agreement that it is a document, in English, with paragraphs etc. is also likely to have been inferred. The Representation and Communication Protocol agreements were made when it was proposed to exchange ASCII files for the DOS or Macintosh operating system on disk by mail. It is relatively simple to have interchanges of ASCII files on diskette because the necessary agreements are fully supported by widespread standards and because our processing requirements are of a very low level.

For interchange of structured data more has to be done. There is a need to indicate the relationships in the structure such as fields in records, links between records, and links between contents of fields. For multimedia even more structure is required to describe the links between information and its orientation in time and space to other objects. For applications to be ported seamlessly between platforms extensive additional agreements about the application operating environment are necessary.

The Open Systems Environment (OSE)

The OSE reference model is described in detail elsewhere[3, 4] but in summary can be described as presenting a framework for understanding all aspects of a variety of technologies that are needed to have true openness. These are presented in OSE as five issues:

- Human/Computer user interface
- Platform internal operating system
- Information interchange
- Communication and networking
- Data and systems management

Requirements and standards have to be developed in all five areas.

Fig. 2 CIMI in the context of OSE

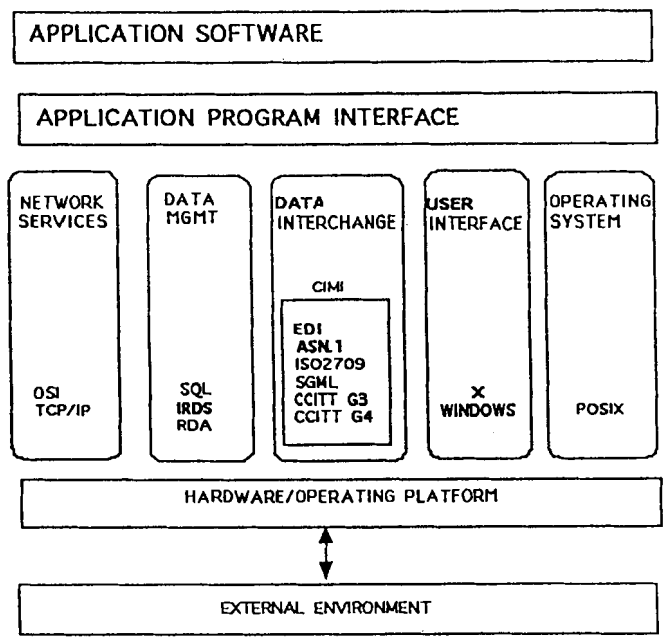


Fig.2 shows the OSE context as adopted by CIMI standards framework. CIMI addresses the interchange aspects of the OSE model that includes; how data is represented; how various data types are identified; and how data content objects are presented. CIMI has a special interest in network services but otherwise only monitors developments in the other areas.

3 The OSE originated with the IEEE and nurtured by the US national institute for standards and technology, NIST. It is well described in "Multimedia Courseware in an Open Systems Environment: A Federal Strategy", Judi Moline, Allen L. Hankinson, Lawrence A. Welsch, NISTIR 4484, NIST, December 1990.

4 "Working Draft of the Technical Report on Multi and Hypermedia: Model and Framework", ISO/IEC JTC1/ SC 18/WG 1 N1444, October 1992. Secretariat: Charles L. Doty, IBM MS 06-03-70, 5 West Kirkwood Blvd., Roanoke, TX 76299-0001, USA. Tel: 1-817-962-5150 Fax: 1-817-962- 3480.

The CIMI Standards Framework

The Standards Framework for CIMI[5] is a model that suggests what standards should be used in what applications to best assure the interchangeability of data between museum applications, migration of data across generations of museum software, and exchange of information among museums and other institutions so it can be used directly by the recipient independent of kind of software, hardware system or network service vendor.

As shown in Fig.3, it is a suite of communications protocols, interchange formats and data representation schemes to support museum application interchange needs.

It employs pre-existing and standard methods for data representation (ASCII and the ISO 7 and 8 bit sets along with extension mechanisms, MPEG, JPEG, CGM), ISO 10162/10163 for information retrieval, EDI/EDIFACT for business transactions, FTAM for file transfer, X.400/500 for messaging, ISO 9049/41 for terminal access. Transport services can be provided by OSI or an appropriate alternate such as TCP/IP. Database building, reference file construction, and the interchange of collections management data can be handled by ISO 2709, ISO 8879 SGML, ISO 8824 ASN.1, and ISO 9735 EDIFACT.

The standards framework recognises that museum applications will require agreements on data content standards still to be developed. Nevertheless it recommends adoption of appropriate carriers for application data and the beginning of data interchange service definitions.

Fig. 3 Museum applications and their interchange requirements

Museum Application	Activity	Examples
Database Building	creating records, editing, deleting	Union catalogues, cooperative cataloguing, scope of collections, authority files, directories
Information Retrieval	accessing stored information	public access catalogues, bibliographic databases, full text, directories, reference databases
Business Transactions	electronic data interchange	purchasing, order fulfilment, loans processing, financial transactions, exhibition management
Messaging	messaging services	email, conferencing, document exchange
File Transfer	binary transfer	exchange of scientific data, data migration, media exchange
Document Handling	document processing	correspondence, cooperative publishing, research
Real-time Links	machine-to-machine	interactive training, interactive exhibits, monitoring

5 "Standards Framework for the Computer Interchange of Museum Information", David Bearman and John Perkins, Spectra 20:2/3, July 1993.

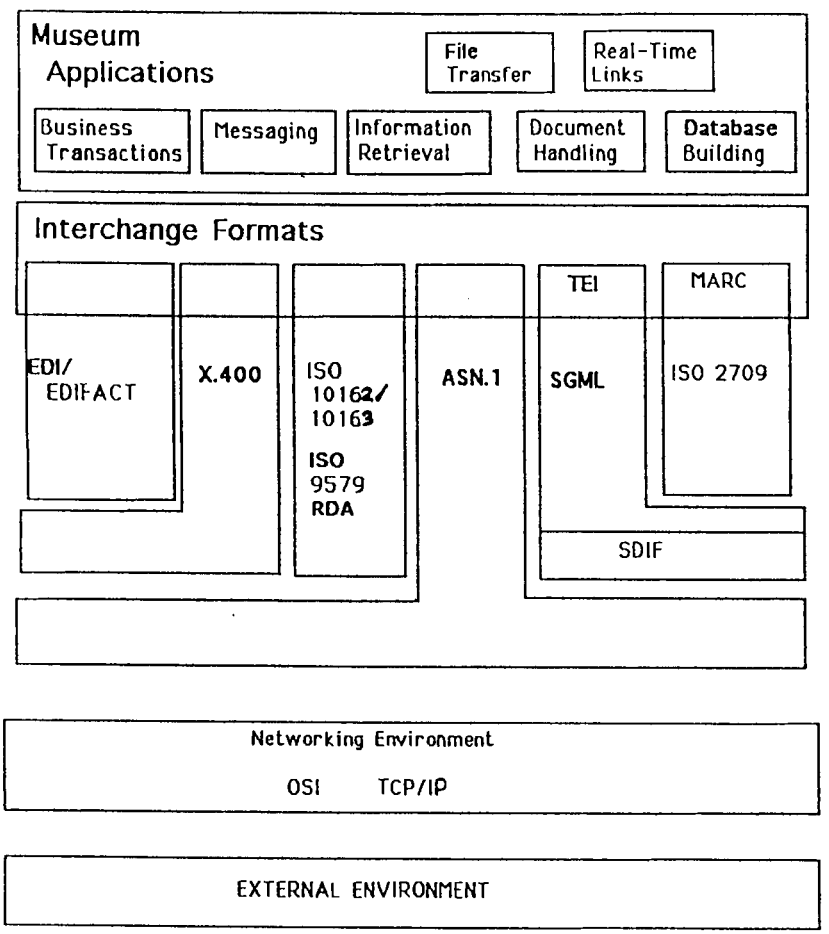
Multimedia in the CIMI Standards Framework

Fig.4 shows the classes of museum applications identified in the CIMI standards framework. Multimedia interchange seems to best be encompassed by the Document Handling and Realtime Links application areas.

This is because multimedia - a somewhat imprecisely defined term - can be thought of simplistically as a collection of different data objects (eg. text, graphics, sound) that have been integrated into a single higher level object (eg. a document) for use[3]. (Judi Moline, Allen L. Hankinson, Lawerence A. Welsch, 1990).

While multimedia was not identified specifically in the framework as a class of application[6] it was identified as a service that should be supported by adopting the external standard MHEG. This endorsement anticipates that MHEG (or an equivalent) will be advanced to an international standard and will have extensive acceptance.

Fig. 4 CIMI Standards Framework



6 Multimedia was not considered or examined as a separate class or entity by the CIMI committee, or by CIMI staff.

MHEG (Coded Representation of Multimedia and Hypermedia Information Objects ISO CD 13522).

MHEG, originally an acronym for Multimedia, Hypermedia information coding Expert Group, has come to mean multimedia standards. MHEG is described in first draft form as CD 13522 (Document N354) from JTC1 SC29 February 1993[7]. A higher-level model and framework is presented as MHMF (Multimedia Hypermedia Model and Framework)[4] (1992).

Fig. 5 shows multimedia and hypermedia services from an application's perspective taken from the MHMF[4] (P.19, 1992)

Of these, CIMI assumes some will be completely developed externally to museums (Operating system, User Interface, Modelling, Data Management, Security, Systems Management, Application Management) and will be market-driven. Museums, like the rest of the world will use what's sold to work with. Others, like network and communication services will require strategic decisions to be made about existing, competing technologies such as OSI or TCP/IP for network transmission. A third group, Intellectual Property Management and Interchange Services will require direct, coordinated participation by museums if effective standards are to be negotiated[8].

Interchange issues in MHEG

MHEG addresses the need to go beyond standards at the monomedia level to ensure multimedia application portability. Fig.6 shows the interchange requirements model from MHEG Draft Version 1 and the standards needed to support the features. It depicts interchange between Application A and B and identifies the levels at which interchange takes place.

MHEG does not address standards at the Application level other than to say it may make use of a script interchange standard at the next lower level. Similarly, it does not address specific standards at the Other Protocol level encompassing housekeeping messages required by the exchanging applications, although it might be assumed the use of ASN.1 would be preferred.

At the non-MHEG Content Level MHEG assumes the use of externally defined monomedia standards such as: ASCII, EBCDIC, telex for text; NAPLPS, CEPT, CGM etc. for graphics; CCITT G711, MIDI, MPEG for audio; FAX Gr3, JPEG for still picture, etc.

The Script level is where sequencing and relationship information between MHEG objects and calls to external processes are articulated. MHEG says script-level interchange should be accomplished by standards for hyperdocument exchange HyperODA (extensions to Office Document Architecture ISO 8613-1-1989) and HyTime (ISO 10744 Hypermedia/ Timebased Structuring Language) and may make use of standards at the next lower level - the MHEG Object level[9].

---

7 "Coded Representation of Audio, Picture, Multimedia, and Hypermedia Information Objects", ISO/IEC JTC1 SC29 WG12 N354. Secretariat: Information Technology Standards Commission of Japan, Kikai Shinko Kaikan, 3-5-8, Shiba-Koen, Minato-ku, Tokyo 105, Japan. Tel: 81-3 3431 2808 Fax: 81-3 3431 6493.

8 This is one of the roles CIMI hopes to undertake.

9 It is important to note that neither ODA nor HyTime are very "real" at the time of writing. The both provide excellent theoretical standards but it remains to be seen if they are accepted and if implementations are made.

Fig. 5 Multimedia services from and application's perspective (after MHMF)

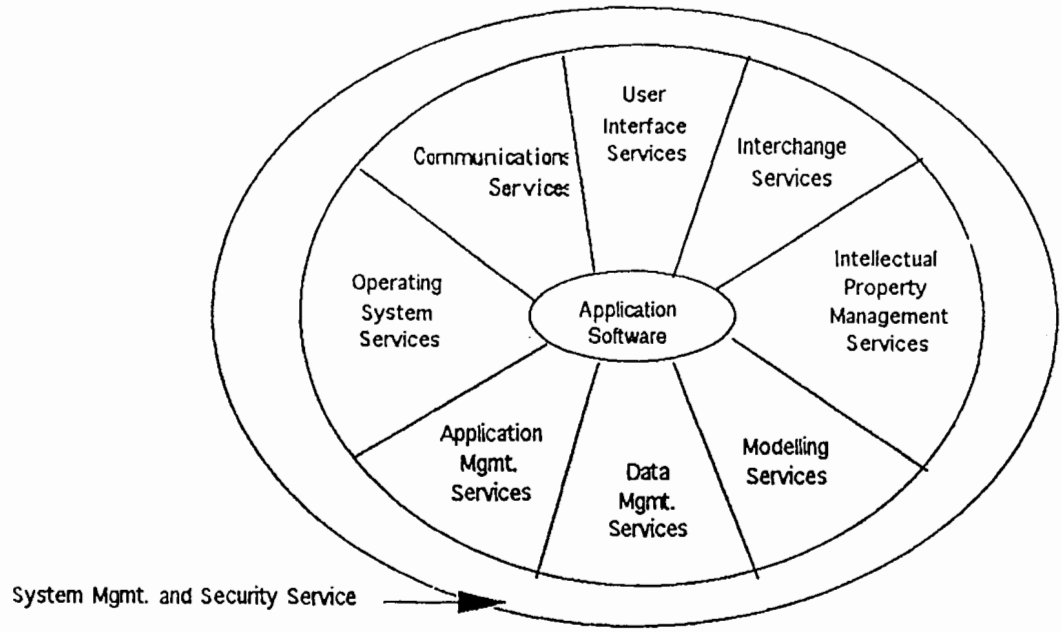
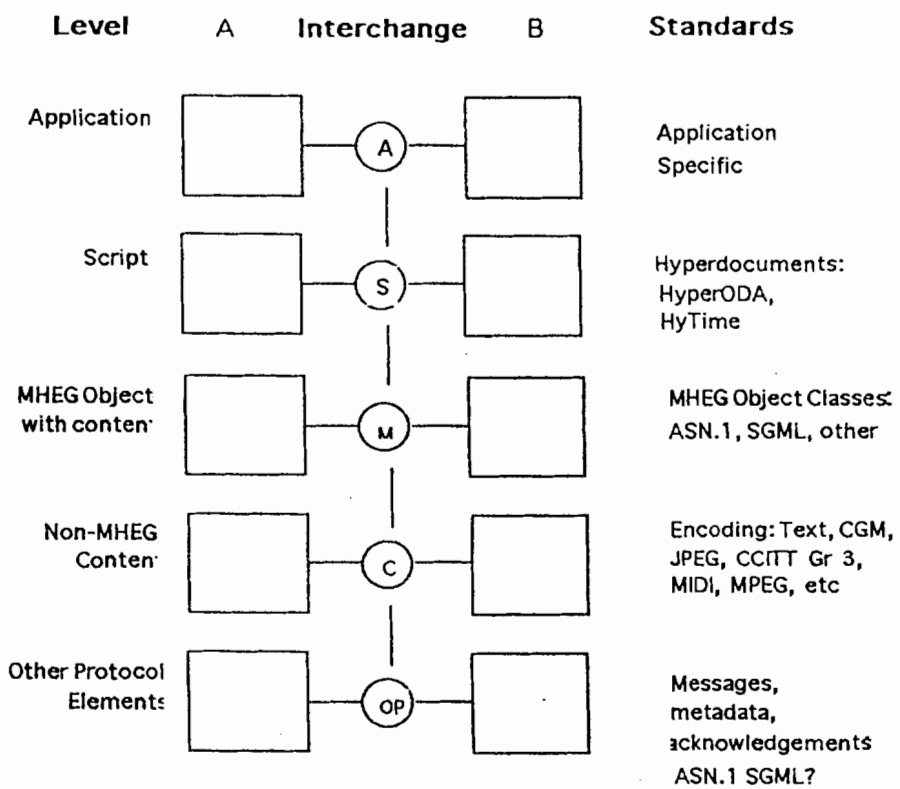


Fig. 6 MHG interchange requirements model



MHEG components: object classes and viewer

MHEG objects are organised into classes. One set of classes contain inter-object relationship information (action, link, script). A second deals with presentable objects that contain mono, hyper and multimedia information that can be presented by an MHEG engine (content, composite, selection, modification). A third group has classes that contain useful information for realtime and optimised handling of objects.

MHEG also defines the concept of viewer which allows various virtual views of object related in different ways. Through these MHEG provides generic structuring of multimedia that will allow a wide variety of multimedia applications to use the information objects.

MHEG object representation by ASN.1 and SGML

To do this MHEG provides two alternate methods of representing MHEG objects. ASN.1 and SGML[10].

ASN.1 (Abstract Syntax Notation One ISO 8824 and 8825) is a flexible, technical language for specifying data structures. Together with its companion Basic Encoding Rules (ISO 8825) it prescribes how information units can be taken from the local system's encoding scheme (called an Abstract Syntax), transformed to a mutually agreed upon (by the systems in question) encoding scheme (called the Transfer Syntax) for transmission to the destination system. There it is retransformed into that local Abstract Syntax.

SGML (Standard Generalised Markup Language ISO 8879) is primarily used in the publishing and online information industry to describe how text should be formatted, printed, processed or laid out in a system independent way. SGML, and its associated standard SGML Data Interchange Format (ISO 9069 SDIF), are methods to denote the structure of free, fielded, or formatted text and to link these with associated images and sounds. The strength of the SGML-based standards is that they permit logical content designation of any data, including data represented in other standards, and can specify specific physical presentations of that data as well.

MHEG prescribes ASN.1 as the base notation for MHEG objects and ASN.1's basic encoding rules (BER) as the base encoding representation. It allows the use of SGML and SDIF as an alternate notation and encoding representation and indicates that other notations will be acceptable as well as long as they are isomorphic (equivalent) to the base notation and representation.

Positioning museums

Although multimedia standards are still in development and great uncertainty exists museums can position themselves for future benefit by:

- specifying museum applications support the CIMI Standards Framework and subscribe to the philosophy of the OSE
- negotiating community-wide monomedia data representation standards
- negotiating guidelines for structuring data element (and object) dictionaries

10 A more detailed description is in the Glossary

- reaching consensus on data element and information object naming conventions
- adopting SGML as a notation language for text, image and multimedia
- developing guidelines for structuring interchange services and interchange formats

In the coming years CIMI hopes to continue its work in each of these areas.

Future work for CIMI and Multimedia

The CIMI Management Committee and MCN has recommended that a Consortium for CIMI be formed to promote further research and development of the Standards Framework for CIMI. If funding can be secured, a research and development agenda will be pursued over the coming years to promote the implementation of standards in museum software and network services, including standardisation needs for multi and hypermedia interchange.

At a general level CIMI will continue to promote the idea of standards, and especially of a standards framework built on existing internationally accepted information systems standards and architectures, as a method for achieving museum data interchange objectives.

Within this CIMI intends to look at the body of monomedia standards and work to define preferred and alternate standards. CIMI will continue its collaborative work with CIDOC (the Documentation Committee of the International Council of Museums) on standards for data element naming conventions. Consistency in naming conventions and element/object dictionaries will also be an objective of work with CIMI's Task Groups.

CIMI will continue to encourage groups, including those whose interest is multimedia, within the museum community with data interchange needs to define the data content and service requirements of their applications and to provide the necessary technical support for such efforts.

Glossary

ASN.1: Abstract Syntax Notation One (ASN.1 ISO 8824 and 8825) ASN.1 is a flexible, technical language for specifying data structures. Together with its companion Basic Encoding Rules (ISO 8825) it provides a means of specifying, encoding & decoding messages (Protocol Data Units or PDUs) to be transmitted.

Its strength lies in the ability to carry any kind of data, in any quantities and to maintain data relationships. A drawback is that implementing ASN.1 interchange, even using available utilities, is a highly complex technical task.

CIMI: An acronym and logotype for all the efforts on behalf of the Computer Interchange of Museum Information undertaken by the Museum Computer Network to support the development and implementation of standards for automated recording and retrieval of museum information.

CIMI Task Group: purpose-directed subsets of the museum profession that work directly with CIMI to develop museum information interchange in a system-independent way.

EBCDIC: a proprietary IBM data encoding scheme for text characters.

ELECTRONIC DATA INTERCHANGE (EDI): EDI is the exchange of routine business transactions in machine readable format. There are two competing standards: EDIFACT

and ASC X12. ASC X12 and EDIFACT consider their format differences to be minor and are pursuing reconciliation.

EDIFACT: The United Nations Economic Commission for Europe EDI for Administration, Commerce, and Transport. The standard is specified as ISO 9735 Electronic Data Interchange for Administration, Commerce, and Transport 1988 - Application Level Syntaxes Rules. This standard is in competition with ASC X.12 but there is an international push to harmonise the two approaches.

FTAM: ISO 8571 File Transfer, Access, and Management (FTAM) provides a full service environment for the transfer of whole or partial files, different file types (flat or hierarchical structured), and file manipulations such as file access (eg locate, erase,) and file management (eg create, delete, read, write).

Hypermedia/Time-based Document Structuring Language: see HyTime

HyTime: Hypermedia/Time-based Document Structuring Language ISO 10744: Based on SGML, HyTime extends the concept of markup of single documents to those of multiple data objects or documents with multiple parts.

Interchange Service: a detailed specification of the structure and meaning of the information being interchanged and the purposes of the interchange. If this includes agreement on the consecutive processes of an interchange transaction, particularly in the formal OSI usage, it is called a Service Definition. Interchange Service is used in the CIMI context to distinguish it from the formal OSI concept of Service Definition.

Interchange format: a formal, ordered expression of the nature and structure of data elements, relationships between data elements, and metadata information needed as part of an Interchange Service.

ISO 2709: ISO 2709-1981 and its US equivalent ANSI Z39.2-1985 are representative of bibliographic and information (textual, descriptive data) interchange formats. ISO 2709 is specifically intended for communications between data processing systems, not a processing format within systems.

Joint Technical Committee 1 of ISO and IEC (JTC1), was formed to address areas of overlapping interest between IEC and ISO. It is responsible for CD-ROM (SC23), OSI (SC16), Office Document Interchange Formats and Standard Generalised Markup Language, SGML (SC18), telecommunications and ISDN or Integrated Services Digital Network (SC6), as well as multimedia and hypermedia (SC29). JTC1/SC18 (Text and Office Systems) has developed the standard ISO 8879 for Standard Generalised Markup Language (SGML). SGML allows the definition of the logical structure of printed documents independent of their published appearance and is widely used by the publishing industry for document formatting and interchange. Those interested in exchanging hypertext have also been examining SGML for its potential usefulness. Standards for the coded representation of still and moving images are being developed by JTC1/SC29. JTC1/SC29/WG10, commonly known as JPEG, is concerned with the digital coding of still images. This working group has proposed a draft standard for the coding of images in a compressed form along with the necessary instructions for decompression that is gaining wide acceptance. The draft JPEG standard solves both problems of coding and creating manageable file sizes without information loss which in turn makes it possible for images to be handled and interchanged in similar ways to textual information in digital form.

SC29/WG11 is dealing with moving picture encoding (MPEG) and WG12 with multi and hypermedia issues.

METADATA: Information about data. In interchange it is the additional information about the content that is necessary for a receiving system to understand and intelligently deal with the exchange.

OSI: Open Systems Interconnection. An ISO set of standards specified in the base standard ISO 7498 1-4 which is the basic description of the seven layered model for effecting open communications between two computer systems. Standard Generalised Markup Language (SGML) ISO 8879-1986

SGML began (and is still primarily used) in the publishing industry to describe how text should be formatted, printed, processed or laid out in a system independent way. ISO 8879 defines two parts, a Prologue and a Document Instance. The prologue contains the SGML declaration and a document type declaration (DTD); the document instance is the content. The SGML Declaration specifies facts about the characters set, the delimiter codes, and the length of identifiers. The DTD contains the description of elements the description of attributes and entities which are named parts of a document. As is the case for other international standards, SGML provides a generalised structure that is given meaning by implementation standards.

SGML, and its associated standard SGML Data Interchange Format (ISO 9069 SDIF), are methods to denote the structure of free, fielded, or formatted text and to link these with associated images and sounds. Museums should use SGML for most object description, especially if it involves use of extended text in fields in which the contents within text should be identified or if it involves text associated with graphics, images, sounds or multimedia, or if the ultimate purposes of the interchange will involve paper of electronic publishing.