

# **Report of the Ad Hoc Committee for Development of a Standardized Tool for Encoding Archival Finding Aids**

## **1. Introduction**

### **Mandate**

The International Council on Archives has asked the Ad Hoc Committee to make recommendations for the development of a standardized tool for encoding archival finding aids, based on the application of ISAD(G) and ISAAR(CPF). This is consistent with one of the stated objectives of ISAD(G), specifically, to "make possible the integration of descriptions from different locations into a unified information system." (1.5 d.)

To this end, the committee was expected to include in its report:

- a) a preliminary design model for a prototype system for encoding archival finding aids based on the application of ISAD(G) and ISAAR(CPF), and,
- b) cost estimates associated with personnel necessary to finalize the functional requirements of the preliminary design model and the development and delivery of a working prototype.

### **Committee Response**

The Ad Hoc Committee agrees that it would be highly desirable to develop a modular, open source software tool that could be used by archives worldwide to manage the intellectual control of their holdings through the recording of standardized descriptive data. Individual archives could combine their data with that of other institutions in regional, national or international networks. Researchers could access this data either via a stand-alone computerized system or over the Internet. The model for this software would be the successful UNESCO-sponsored free library program, ISIS, which has been in widespread use around the developing world for many years. The software, with appropriate supporting documentation, would be freely available via an ICA or UNESCO web site or on CD-ROM. Unlike ISIS, however, the source code and not just the software should be freely available.

### **Scope of the Report**

This report focuses on the development of tools for the description and intellectual control of archives and the discovery of relevant resources by users. Other archival functions, such as appraisal, acquisition, preservation, and physical control, are beyond the scope for this project. The system developed as a result of this report should be useable on stand-alone computers in small institutions, by multiple users in larger organisations, and by local, regional, national, and international networks. The development of such a system should take into account the strategies, experiences, and results of other initiatives such as the European Union Archival Network (EUAN), the Linking and Exploring Authority Files (LEAF) initiative, the European Visual Archives (EVA) project, and the Canadian Archival Information Network (CAIN).

## **Organization of the Report**

This report is divided into five sections. A description of the conceptual structure of an archival information system, described as six layers of services and protocols, follows this introduction. Section three details the functional requirements for the software tool and is followed by a discussion of the relationship of these requirements to existing archival software application. The report concludes with a series of recommendations that provide a strategy for the successful development, deployment, and maintenance of an Open Source Archival Resource Information System (OSARIS). There are two appendices: a data model and a comparison of the functional requirements statements to several existing archival systems.

## **2. Archives Information System Profile**

This profile identifies and organizes the data models, services, standards, and functionality for a system that supports archivists in the work of recording information about archival materials in automated systems. The system is defined to meet the specific needs of users who wish to discover the location all archival resources that meet their research or administrative needs, to identify the resources their search has uncovered, to understand the content and navigate the structure of the information presented, and to retrieve directly any archival documentary materials that exist in electronic form.

This profile represents the components of this system as hierarchical layers of standards and services that begin with the most fundamental concepts and proceed up to working applications that implement the functional requirements defined in section three of this report. There are six layers to the profile.

### **1. Information Model Layer**

- Defines the high-level functional components of the system. There are four data stores whose interrelationships define the fundamental characteristics of archival information systems. These relationships are depicted graphically in the data model in Appendix A. Systems that manage other aspects of the archival enterprise, such as acquisition, storage and conservation, would be represented by additional data stores.
- Descriptive records provide information about particular units of archival description such as fond, series, files, and items.
- Context authorities are data stores that identify and describe persons, families, and organizations responsible for the creation of archival records as well as the functions and activities that produced those documents. These include both standardized data in traditional authority files and non-standardized sources such as biographical directories.
- Content authorities consist of thesauri and sources of data such as geographic place names, topical subjects, documentary forms, and genres.
- External entities include images, data files, and other data stores such as databases and library catalogues.

## **2.Data Structure and Content Layer**

Defines the data elements and their content for each of the components of a system. Different standards-setting organizations are responsible for each of the four components described by the profile.

- ISAD(G) identifies twenty-six data elements for describing archival entities in the context of multi-level archives. For example, it defines title as one of the essential element for understanding the nature of particular archival materials. Existing national standards, such as the Canadian Rules for Archival Description, define the formulation of the content of these elements. As a subset of elements common to most national descriptive standards, ISAD(G) will be appropriate as the base element set for the record descriptions defined in this profile.
- ISAAR(CPF) defines a generally acceptable set of data elements for identifying persons, families, organizations responsible for the creation of archival records, information essential for establishing their context. The contents of such authority records are based on national or international content standards found in specialized files. The Anglo-American Cataloguing Rules and the U.K.National Council on Archives Rules for the Construction of Personal, Place, and Corporate Names are examples of such authority rules.
- Descriptions of archival materials also incorporate terms such as topical subject headings, geographical place names, form and genre terms, and other concepts that are used to explain the content of the records and are essential for indexing and retrieval. The thesauri from which these terms are taken are created by archivists or others.
- The structure of external entities such as other databases, image file formats like JPEG and MPEG, and text files are specified outside the archival domain. Archival descriptive systems must be capable of linking these related components.

## **3.Data Storage Layer**

Defines the electronic syntax in which data are logically and physically organized and stored. These include Extensible Markup Language (XML), relational or object-relational tables, comma separated value (CSV) files, the ISO 2709 (MARC) structure, and word processing formats.

## **4.Data Communication Layer**

Describes the standards used for transporting data between computers and their applications. ISO 2709 (MARC) and the hypertext transfer protocol (http) are examples of such communication protocols.

### **5. Local Application Layer**

Defines the functionality of a software application or applications used by local archivists and end-users to create or import data; store, index and query data; and display and/or export query results. The functionality for applications that create and access data as local, stand-alone systems are described in section 3 of this report. Applications at this layer should be modular to accommodate additional data stores and functions required by other activities of the archival enterprise such as the acquisition, storage, and conservation of collections.

### **6. Union Application Layer**

Defines the functionality of software applications used to provide access to the descriptive records of multiple archival institutions. In addition to the basic functionality found in the local application layer, the union layer includes the services necessary to support either a centralized or distributed method for the storage of and user access to descriptive information. A directory service that identifies archival institutions is a necessary requirement at this layer.

## **3. Functional Requirements**

This section identifies the general functional specifications for a unified archival information system (OSARIS), as well as describing specific requirements in three areas:

1. Data input (data capture in a way that makes exchange possible);
2. Data output (presentation environment);
3. Information exchange (communication environment).

### **General Requirements**

**G 1:** The system must be compliant with both ISAD(G) and ISAAR(CPF).

The system must accommodate both multilevel description and many-to-many relationships between entities, for example, part to whole, record to record, provenance to record relationships. The system will therefore not have a flat data structure, but rather one characterized by relationships.

**G 2:** The system must be extensible to accommodate the additional requirements of national standards.

The system will be based upon the core essential elements of the two ICA standards, but will also include all the optional data elements from ISAD(G) and ISAAR(CPF) and have the capacity for additional data elements to be incorporated from national, specialized or in-house standards.

**G 3:** The data format must be software and operating system independent but should be in compliance with the high-level data model found in Appendix A. The software must be able to be run on a wide range of systems and platforms.

**G 4:** Software and explanatory documentation for the system must be freely available, that is, downloadable over the Web or distributable on CD-ROM(s) or diskette(s).

**G 5:** The system must be able to accommodate multiple languages and multiple scripts

**G 6:** The system must be as scalable as possible.

It should be useable on standalone computers in small institutions, by multiple users in larger organisations and by local, regional, national and international networks. This might include the development of a simple ground-level version first that can be upgraded by tool-kits to incorporate more complex features that require greater technical expertise. In terms of data capacity, the system should be able to cope with medium to large-scale, item-level description projects.

**G 7:** The system must be designed to allow transfer of data structures and functionality to other systems, either to permit replacement versions of the same system or facilitate migration to other, more sophisticated systems yet to be developed.

**G 8:** The system must be web-enabled.

The system should be able to function either independently of the web or in a web-enabled environment. This means that the system should be able to support the input, output and exchange of data via the Internet.

**G 9:** The system must provide mechanisms for two-way links with external resources.

Resources in the system should be addressable by external systems and, in turn, have the capacity themselves to link to external sources.

**G 10:** The system must be able to incorporate or link to controlled vocabularies.

During data input and searching, it should be possible to select and/or validate data values from a pre-existing list of approved terms. The system should be capable of generating and maintaining thesauri.

### **Data-Input-Requirements**

**1.1:** The system should be as user-friendly as possible. Data entry screens should be simple and unambiguous; help screens should be available.

**1.2:** The system should provide a basic set of input-templates and the option to generate others as required. It should be possible to create templates that have built-in quality assurance mechanisms.

**1.3:** The system must accommodate simplicity as well as complexity of descriptions, providing at least an essential set of descriptive elements.

**1.4:** The system must support input of descriptions of archives of any provenance, both private and public, and descriptions of all forms of records and records in all types of media, regardless of their physical and logical structure.

**1.5:** The system must accommodate tools for import-generation of data from a variety of external sources in a standard format.

External sources may include legacy and records-management-systems; authority, bibliographical, existing descriptive, biographical, and Geographic Information System (GIS) data.

### **Data-Output-Requirements**

**2.1:** The system must be user-friendly for users of archives.

Users should be able to navigate intuitively through and around the system. Screens and system messages should be unambiguous. Help screens should be available.

**2.2:** The system must have the capacity to create a variety of end-user-interfaces.

It should provide a basic set of display-screens with an option for local customization or creation of new screens. Data must be displayed in a readable and attractive form.

**2.3:** Users must be able to save, print out or download the results of their searches.

**2.4:** The system must support different navigation and search strategies and retrieval mechanisms, e.g. free-text, controlled language, Boolean logic, and index searching.

### **Requirements for Information Interchange**

**3.1:** The system needs to support the exchange of data between local systems and networks, regardless whether the network is based on a distributed or centralized architecture.

**3.2:** The system must support the current archival standards for machine-readable data communication, Encoded Archival Description (EAD) and Encoded Archival Context (EAC). A subset of elements found in EAD may be used to exchange descriptions based on ISAD(G) while elements in EAC may be used to exchange ISAAR(CPF)-based authority data.

**3.3:** The system should also be flexible in its ability to use different data syntaxes, e.g. XML and HTML.

**3.4:** The system must accommodate wider information management data interchange standards, e.g. Dublin Core and MARC.

The system should have the capacity to convert selected data into different metadata formats for wider data exchange purposes. It is important for data in archival systems to be made available for use by the wider information management community.

**3.5:** The system must accommodate directory information including contact and access to each individual repository. This information can be managed in the form of authority records for each repository.

#### **4. Mapping**

There are a number of systems and software that can be identified as potentially satisfying the functional requirements in whole or in part. However, they are all proprietary in some respect and cannot now be freely distributed. They do provide a basis for a systems analysis and their creators could be approached to discover whether the applications might become freeware and what costs would be involved.

The chart in Appendix B lists several systems that should be examined, their country of origin, and the extent to which they satisfy the key functional requirements specified in the previous section.

#### **5. Recommendations, Priorities and Strategies**

The following statements take into account the data model, the functional requirements, and the preliminary results of the mapping. The Ad Hoc Committee recommends:

##### **Application Development**

1. That the ICA endorses the functional requirements presented in this document as the basis for moving the initiative forward.
2. That the functional desiderata and technical specifications for the software applications, such as user requirements, business rules, and detailed data models, should be developed further by a team of experts from both ICA/CDS and ICA/ITC as the next stage of this project.
3. That following the finalization of the technical specifications for OSARIS, the requirements should be compared to existing systems and a decision made to adopt or adapt existing software or to build new applications. At that point in time, it will then be possible to estimate project costs.
4. That a solution that incorporates the functional requirements result in the development of several modular software applications.

5. That the implementation of the system should follow a modular strategy.
6. That the development of software applications must include a thorough investigation and assessment of existing solutions beginning with those identified in section four and Appendix B of this document.
7. That the ICA develop a strategy for communicating the progress of this project to members of the international archival community on a regular basis. This would include the distribution of progress reports in multiple languages. The communication strategy must include a two-way exchange of ideas. The project will benefit strongly from the ongoing comments, suggestions, and input of the members of the international archival community.
8. That a test-bed be developed to allow the testing of software solutions in a realistic archival environment.

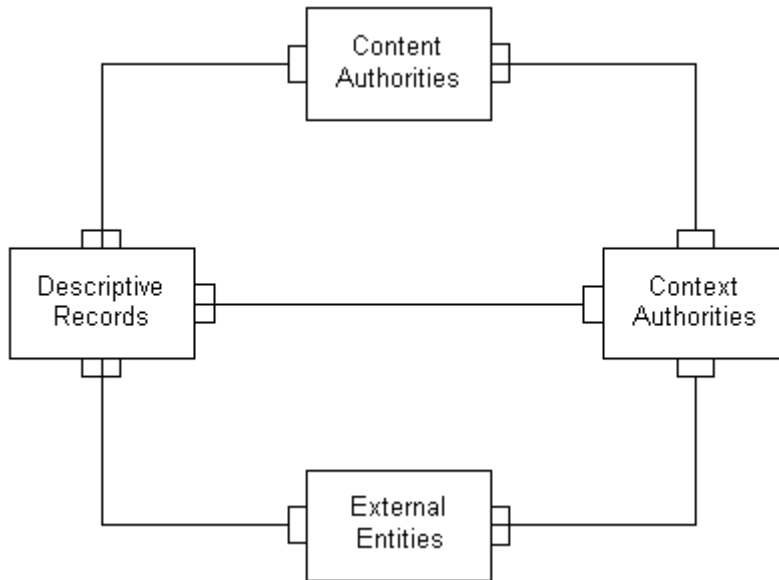
### **Deployment**

9. That the system specifications, its documentation, and the source codes for the applications be freely available.
10. That training courses for new users, ongoing education, and web-based support groups be established.
11. That promotion of the software be carried out through the existing regional infrastructure of ICA and through UNESCO.

### **Maintenance and Enhancement**

12. That an infrastructure for ongoing maintenance, distribution, and technical support be developed. This should include a web site to download software and supporting documentation. The ICA should also establish and maintain a mechanism for end-users to recommend changes and enhancements to the software.
13. That the ICA establishes and maintains an official mechanism for regular review of the software by an advisory committee that includes technical and archival experts.

**Appendix A**  
**Data Model for OSARIS**



**Appendix B**  
**Existing Archival Information Systems**

	G1 ISAAR ISAD	G6 Scalable	G8 Web	G9 External	G10 Controlled vocabulary	1.2 Input template	1.3 Description	1.4 range descriptions	1.5 input generators	2.2 user interface	2.3 down load save	2.4 search strategies	3.2 EAD EAC	3.3 XML HTML
n	Y	Y	Y	Y	Y	Y/N	y	Y	N	N	N	Y	Y	Y
	Y	Y	Y	Y	N	Y	y	Y	N	N	N	Y/N	Y	Y
ny	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
ia	Y	Y	Y	Y	In devel- opment	Y	Y	Y	N	Y/N	Y/N	Y	Y	Y
	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N	Y
ia	Y	Y	Y	Y	N	Y	Y	Y	MS Access	Y	Y	N	N	Y

This table provides a preliminary assessment of the compliance of certain existing systems with the functional requirements outlined in section three.

ARKIS2: Database of the National Archives of Sweden. See <http://www.ra.se/nad/arkis2/>

NRA: National Register of Archives. United Kingdom. See <http://www.hmc.gov.uk/nra/nra2.htm>

MIDOSA- A license-free, open software system for the production of archival finding aids for navigating and structure oriented or full-text based research. The program is freely available in English and German. See <http://www.midosa.de>

OHRM/HDMS: Online Heritage Resource Manager and the Heritage Documentation Management System. Australia. Developed and distributed by the Science and Technology Heritage Centre at the University of Melbourne. HDMS manages archival documentation while OHRM is a resource discovery system. Both utilize Microsoft Access and available under license.

\*The Sistema Unificato per la Soprintendenze archivistiche italiane (SIUSA) is a system developed by Italian Archival Administration in collaboration with Centro di Ricerche informatiche per i Beni Culturali (CRIBECU: Center for application of Computer Science to Cultural Assets) in Pisa for accommodating description for private and non-governmental public archives (local administration, semi public organizations and so on) in a local or network environment.

*Tabularium* is a freeware archives control system designed for small archives operations based on the Australian 'series system' of archival control and description. It is a control system, rather than a finding-aid system, although the production of finding aids is an important part of its functionality. It is not a standalone software product but requires a licensed copy of Microsoft Access to run. It was originally developed as a model structure for a relational database for archival control and description and be used as a the model for building archival control systems for database platforms other than Microsoft Access.

In addition to these systems, the Committee has identified other systems that might profitably be studied. They include-

United Kingdom: A2A- Access to Archives. Public Record Office, United Kingdom. See <http://www.a2a.pro.gov.uk>

United Kingdom: Higher Education Hub, Manchester University, United Kingdom. See <http://www.archiveshub.ac.uk>

United Kingdom: AIM25 Archives. University of London Computing Centre. AIM25 is a major project to provide electronic access to collection level descriptions of the archives of over fifty higher education institutions and learned societies within the greater London area. See <http://www.aim25.ac.uk/>

SCAN: Scottish Archive Network, National Archives of Scotland. See <http://www.scan.org.uk/index.html>

ASTA (Archives System To All): Norway

SCOPE: Switzerland See <http://www.scope.ch>

ISIS: UNESCO