



Title	Funding models for the future development of metadata standards and software tools (D8.2a)
Work Package	8
Authors	Uwe Jensen (GESIS)
Source	WP8
Dissemination Level	PU (Public)

Summary/abstract

The core focus of this report concerns the particular considerations and requirements of the cessed-ERIC in terms of future resourcing and sustainability of the development of metadata standards and software tools.

The respective needs of the cessed-ERIC relate to four central aspects:

- Data model and standards for developing the technical infrastructure at large (OAIS, PREMIS, METS; DDI, SDMX – Service Oriented Architecture on the Web; PID, SSO);
- Particular support for specific technical metadata standards for operational archive processes, data documentation, the publications and retrieval process and respective mappings among involved standards;
- Interoperable standards and harmonised operational archival processes are required to implement and maintain seamless ingest, documentation, publication, retrieval and access to data and metadata of the distributed sources of the cessed-ERIC infrastructure;
- Maintenance of ongoing requirements and changes of standards and tools over time.

The conclusion and recommendations consider particular core objectives and options in cooperating with international initiatives and bodies working on metadata standards and related tools. Secondly options for resourcing and sustainable development of metadata standards within the cessed-ERIC are discussed focussing on needs for a Metadata manger, experts groups and proactive initiatives in the concerned fields.

Table of Contents

1. Introduction	1
2. cessda-ERIC requirements to develop metadata standards and software tools	2
2.1. Data model and standards for developing the technical infrastructure	2
2.2. Four central needs on metadata and services in the cessda-ERIC	2
2.3. Extend interoperability of standards and harmonised operational processes.....	2
2.4. Maintaining ongoing requirements and changes of standards and tools over time ...	3
3. Conclusions and recommendations	4
3.1. Core recommendations on developments for standards and software	4
3.2. Particular cooperation in implementing and developing technical standards	4
3.3. Particular cooperation in developing research related metadata standards.....	5
3.4. Optional models for resourcing and sustainable development of metadata standards.	5
3.4.1. Investments and organisation models within the cessda-ERIC	6
3.4.2. Particular cooperation need with the DDI to implement DDI version 3.....	6
3.4.3. Cooperation with further technical standardisation bodies and initiatives	6
4. Online references and glossary of abbreviations.....	8

1. Introduction

This report is dedicated to investigate and make recommendations on options for the future resourcing and sustainability of the development of metadata standards and software tools. Therefore a small number of recommendations were defined based in particular on a statement about the necessity of metadata development.

These requirements are discussed in relation to software tools to develop for the cessda-ERIC data infrastructure. Insofar there is a strong relation to recommendation and models on how to organise such technical developments as provided with D8.4. Report "Models for future organisation of technical R&D developments and training for the metadata technicians". In additions to this perspective no individual funding model will be proposed in this report as it appears more feasible to handle such issues within the general funding model of the cessda-ERIC. Furthermore the preparation of proposals for funding of technical developments will necessarily include coordinated actions related to particular technical standards and to the needs for specific software developments.

Quality driven performance of the CESSDA total data infrastructure must be guided by technical standards and appropriate procedures to be implemented according to the lifecycle and in particular to achieve compliance with the OAIS reference model.

The present situation represents still great variety and the further involvement of technical and substantial standards definitely will become one core issue for several dimensions managing the whole cessda-ERIC. The use of standards implies as well standardisation and thus changing of present archive procedures step by step.

In saying this political, judicial, economic and further organisational questions are to consider with the development of required software and the related implementation of technical metadata standards. However the core focus of this report concerns the particular considerations and requirements of the cessda-ERIC in terms of future resourcing and sustainability of the development of metadata standards and software tools.

2. cessda-ERIC requirements to develop metadata standards and software tools

The development of the total technical infrastructure and respective software tools will be driven by state of the art technologies and in applying most recent standards in managing studies, data and metadata within a common but distributed archival framework and in sharing knowledge products from social science research and re-use of research data in the SSH community.

2.1. Data model and standards for developing the technical infrastructure

The development of the technical infrastructure at large and the software requires the application of appropriate technical standards to acquire, document, store, preserve, disseminate data as well as to promote re-use of research results across many contexts. The future data infrastructure has thereby and beyond survey data to consider further or new complex data and metadata needs from different strands, formats (rectangular, cubes, text, pictures, sound, data visualisation etc) or data sources.

Requirements:

- To bring together these facets requires a common information and data model for the whole technical infrastructure is necessary considering relationships between and layers of particular standards and their respective place in the organisational environment;
- To facilitate interoperability of metadata standards require specifications on its usages as well as standardization of metadata at the semantic, structural and syntactic level;
- To support the practical implementation of agreed standards and as a consequence the standardisation of processes within a distributed framework requires a phased roadmap to harmonise step-by-step current differences of the local practise;
- To support operational work from testing to practical use of applied standards, tools and functions as well as changing organisational workflows and principles requires a robust information and communication platform for developers and metadata technicians.

2.2. Four central needs on metadata and services in the cessda-ERIC

Requirements:

- Ingest, Preservation and operational archive processes (OAIS; PREMIS, METS);
- Technical data documentation standard DDI version 3 ;
- Mapping between DDI 3 and other standards (e.g. SDMX);
- Publication, retrieval and access to data (web based SOA; SSO, PID).

2.3. Extend interoperability of standards and harmonised operational processes

Beyond the core task of data documentation from several study types the exchange of metadata, internal archive processes and long-term preservation need clear operational principles and advice within a common but distributed framework. It must allow for seamless publication, retrieval and access to such data and metadata finally.

Requirement:

- The general development around data management (in a broad understanding) requires the archives to have a dynamic relationship to data treatment, following technology

and analytic development. It will be a continuous phased process to implement interoperable standards and harmonised operational processes among the distributed sources.

2.4. Maintaining ongoing requirements and changes of standards and tools over time

To manage a virtual system with distributed resources provided by different vendors or operated by different organizations requires overall a sustainable environment to upgrade, develop and implement the required technical tools in accordance to standards used in particular areas. But standards are always fixed for a certain frame of time and need to be flexible and extensible to meet the dynamics in the respective field.

Requirements:

- It is an ongoing requirement to maintain and develop implemented standards over time to keep compliance with new metadata needs in extending the scope or range of social science research data. As such it necessitates sustainable resources both for daily support and proactive actions to improve or extend scope and functionality of standards and tools;
- Involvement of technical standards implies standardisation of operational processes and requires adequate procedures in detail. Thus provision of guidance and advice to the cessda-ERIC members and in particular needs of the daily operators an important accompanying mean to the technical developments;
- Vice versa: The social sciences need to influence as well as make its needs known to the standards development community. It might be considerable to support this needs within the cessda-ERIC. In initialising workshop or meeting of suitable format with researcher and research project on extended support for complex data and metadata overall could also focus needs for extended items like at the DDI standards and beyond.

3. Conclusions and recommendations

Development of standards is often undertaken in cooperation within a worldwide network of interested partners to solve the need for a particular substantial or technical standard like the DDI. Sometimes standards or more precise frameworks, developed in other research areas like natural sciences, are adoptable for social science data infrastructures like the OAIS framework (Open Archival Information System). In saying this optional models for resourcing and sustainable development has to include and reference to related consortia, projects, or initiatives working in particular standardisation areas.

3.1. Core recommendations on developments for standards and software

- To avoid duplicated efforts and to reduce costs for large scale developments on standards, technical infrastructures and software for processes internal or external to the data service provider;
- To increase developments based on open source and community efforts to share solutions for system architectures and software to allow large scale service for the cessed-ERIC based data infrastructure and beyond;
- To foster best practice solutions by use of advanced technologies according to the needs of main customer groups acting along the whole lifecycle;
- To achieve interoperability of standards and software developments on international level and to harmonise operations and processes among the cessed-ERIC members and partner not becoming member of the cessed-ERIC.

3.2. Particular cooperation in implementing and developing technical standards

The following initiatives related to the development of software and standards are of importance.

- **IDATA** (International Data Technology Alliance) supports worldwide cooperation on broad areas as discussed in Canberra for particular areas like OAIS and specified data modelling, core architecture with tools and service for social science data archives (FEDORA etc.).
- **DDI 3** and in particular the DDI TIC (Technical Implementation Committee) and expert groups strongly related to tasks from the current CESSDA-PPP agenda to establish a DDI3 compliant environment and to migrate existing content and workflows respectively. This work requires stable support for expert groups and training facilities to enhance expertise and knowledge to apply the standard within the cessed-ERIC.
- The **OAIS** reference model (Open Archival Information System) is considered the blue-print for archival processes where long-term preservation of digital objects plays a central role. One particular task of WP8 regards the definition and use of Preservation metadata. On this background the DDI 3 and two further standards are of core interest for the future:
- The **PREMIS** (PREservation Metadata: Implementation Strategies) Working Group provided the PREMIS Data Dictionary version 2 as of March 2008.
- **METS** (Metadata Encoding and Transmission Standard) provides XML based schema to encode descriptive, administrative, and structural metadata for objects within a digital repositories. It is developed as an initiative of the Digital Library Federation

and maintained in the Network Development and MARC Standards Office of the Library of Congress.

- **SDMX** (standards for **Statistical Data and Metadata eXchange**) was launched in 2002 as an initiative from seven international organisations (BIS, ECB, EUROSTAT, IMF, OECD, UN, World Bank). Its mission is the development of technical standards for the formatting and exchange of aggregate statistics and to provide guidelines with mainstream technology to support statistical processes. The present specifications are contributed with SDMX Standards Version 2.0. Based on a common information model the specification defines respective formats to exchange aggregated statistical data and related metadata to understand how the data are structured as well as reference metadata to provide information on methodology, quality etc. of the data and formal objects from the model (ranging from organisations to code lists etc.). There is a strong relation to the DDI standard with overlapping functions and direct mappings for complementary use in a single system.

This technical standard will have high future relevance like in relation to the WP10 recommendation providing secure access to data from official statics via the CESSDA portal.

3.3. Particular cooperation in developing research related metadata standards

Particular cooperation needs regard the research on and developing of substantial content of documentation standards.

- The harmonisation platform performs with the step-wise technical implementation related (community based) development of substantial standards in topical fields;
- Further documentations needs as well as development of related standards and specialised software tools are to consider in cooperation with cross-national projects and survey programs hosted in or supported by the cessda-ERIC data infrastructure;
- Further implications will be given considering extensions of data and metadata to support interdisciplinary use of research results (compare INSPIRE - Infrastructure for Spatial Information in Europe) or multi-level analyses research approaches (see also Report D8.1 "Data and Metadata Extensions of the CESSDA RI").

3.4. Optional models for resourcing and sustainable development of metadata standards

The proposal of optional models has to consider some major criteria and respective needs funding human resources and to develop, implement and maintain related facilities. Of particular importance appear the extension of cooperation with project at the ESFRI roadmap and international organisation, research projects and initiatives.

The models require review with the proposed models on organisation of technical developments provided with the Report D8.4 "Models for future organisation of technical R&D developments and training for the metadata technicians": Basic options concern the cessda-ERIC as:

- Developer;
- Customer;
- Co-ordinator.

Organisational principles and regimes will care for consistent and efficient regulations in the cessda-ERIC overall.

3.4.1. Investments and organisation models within the cessda-ERIC

It is recommended to employ one central position in the future cessda-ERIC responsible for all issues related to technical standards (Manager for metadata standards). The presentation of cessda-ERIC interests in respective standard bodies is to consider within the frame of the overall organisation of duties and division of labour.

This work is supported by expert groups to discuss and prepare proposals on maintenance, development and implementation of technical standards for dedicated projects and tasks. Where appropriate common work on requirements from other operative fields like best practice guidance, training issues etc. may be integrated. Vice versa subgroups will work on dedicated (sub-) tasks where necessary.

Expert groups to operate particular technical and conceptual tasks are to consider the following areas:

- Information Technology and Network Infrastructure (portal functions, infrastructure, at large, secure remote access system and the potentials of grid);
- Implementation and maintenance of technical standards (DDI, SDMX; PREMIS; METS) and PID system;
- OAIS compliant implementation for operational processes, preservation metadata and data format issues (Reference projects like TRAC, DRAMBORA, NESTOR, and Data Seal of Approval);
- Single-sign-on (AAA rules; monitoring) for both data protection and open access;
- Software development: Harmonisation platform & QDB; DDI editor; VCC functions for the Virtual Centre of Competence; ELSST management tool;
- Controlled vocabularies: Thesaurus management team with DDI expert group;
- CESSDA VCC – Virtual Centre of Competence content and maintenance.

Further details are specified in final reports and recommendations from WP5 to WP12.

A particular proposal on dedicated group settings has to be decided in the broader context in formulating the technical development plan for the cessda-ERIC.

3.4.2. Particular cooperation need with the DDI to implement DDI version 3

In preparing and guiding the implementation of DDI ver.3 it is recommended to establish a formal cooperation between cessda-ERIC and the DDI Technical Implementation Committee (TIC) to manage and solve effectively needs and questions in this process.

3.4.3. Cooperation with further technical standardisation bodies and initiatives

General resources are necessary to support the future development of the DDI standard as well as further requirements for cooperation with standard bodies or workgroups as mentioned for SDMX, the OAIS model and the particular preservation related standards (METS, PREMIS).

This regards also the further cooperation like for common implementation issues for archival processes, operations and tools with organisations outside CESSDA like IDATA (International Data Technology Alliance) and ICPSR (University of Michigan Institute for Social Research Inter-University Consortium for Political and Social Research).

Furthermore the liaison and/or alignment with global developments in grid and social science arena is recommended in promoting the design of web oriented SOA infrastructure to allow maximum flexibility towards other infrastructures and technologies (compare EGEE (Enabling Grids for E-science); PARSE.Insight (Permanent Access to the Records of Science in Europe)).

- An official delegation from the cessa-ERIC is proposed for expert group members to participate in operative tasks at standardisation bodies or software development initiatives;
- Sensible financial support should consider sustaining the work of Standardisation bodies with high professional relevance to realise standard related objectives of the cessa-ERIC;
- The cessa-ERIC proactively initialises project and / or participates in appropriate activities to gain funding to realise innovative technologies for long-term preservation services, to set high professional standards in data management and to open new data sources for shared access by the social science research community.

4. Online references and glossary of abbreviations

Technical Standards

DDI - Data Documentation Initiative <http://www.ddialliance.org/>

METS - Metadata Encoding and Transmission Standard www.loc.gov/standards/mets/

PREMIS - PReservation Metadata: Implementation Strategies;

Data Dictionary version 2 (March 2008) <http://www.loc.gov/standards/premis/>

SDMX - Statistical Data and Metadata Exchange www.sdmx.org/

Reference models & long term preservation

OAIS - Reference Model for an Open Archival Information System

<http://public.ccsds.org/publications/archive/650x0b1.pdf>

Provider: Consultative Committee for Space Data Systems <http://www.ccsds.org/>

Data Seal of Approval www.datasealofapproval.org/

DRAMBORA - Digital Repository Audit Method Based on Risk Assessment toolkit

<http://www.repositoryaudit.eu/>

NESTOR - the german Network of expertise in Digital long-term preservation

<http://www.langzeitarchivierung.de/index.php?newlang=eng>

TRAC - Trustworthy Repositories Audit & Certification – see

<http://www.dcc.ac.uk/tools/trustworthy-repositories/>

Tools for managing and providing access to digital content

DuraSpace – Joint organization of Fedora Commons and the Dspace

<http://duraspace.org/index.php>

DataVerse network Project <http://thedata.org/>

NESSTAR <http://www.nesstar.com/>

E-infrastructure and the Social science and Humanities

EGEE - Enabling Grids for E-sciencE <http://www.eu-egee.org/>

PARSE.Insight – Project to develop a roadmap and recommendations for developing the e-infrastructure in order to maintain the long-term accessibility and usability of scientific digital information in Europe <http://www.parse-insight.eu/>

AVROSS - Accelerating Transition to Virtual Research Organization in Social Science – reported 2007 “on the state in applying e-Infrastructure to social science and humanities (SSH).” (M4 Final Report page V) [Final Report](#) at AVROS website

<http://web.fhnw.ch/plattformen/avross>