The Services:
Data discovery, harmonisation, analysis and dissemination

Uwe Jensen
GESIS Data Archive (Germany)

30.05.2008 H1 – CESSDA ESFRI Project
Setting up a one-stop-shop portal for European Data

linking
european
data
resources

Vision of the CESSDA RI upgrade
System with integrated services to allow

- open, extensible, evolvable **access** to a wide range of data collections

- **dissemination** of data and metadata for scientific analysis supported by a flexible technical architecture using standard, open protocols and interfaces,

- a new generation of tools for advanced **data discovery**, **integration** and **harmonisation** based on metadata standards for data documentation

linking
european
data
resources

The upgrade of services

1. **Infrastructure services for data discovery**
   - CESSDA portal & multilingual metadata
2. **Infrastructure for content harmonisation & conversion**
3. **Data and metadata infrastructure**
4. **DDI 3.0 evaluation**

1. **Data discovery & CESSDA One-stop portal**

- **Conceptual work** to upgrade a large-scale documentation- and portal software to plan & implement a One-stop portal

- **Investigations on** data storage, discovery and retrieval questions to allow extensive access to several types of complex data

- **Consider** complex data collection processes & modular metadata organisation of DDI 3.0

linking
european
data
resources

1. Data discovery I: Upgrades to extend CESSDA portal services

- To integrate and extent the access to complex data types functionality of the portal needs several enhancements

Data documentation and data publishing software
- functional specifications based on experience from the developments in the MetaDater project and the recent NESSTAR publisher

Data location services to discover data for comparative research:
- in deep evaluation on metadata harvesting, indexing and search technology in the context of multi-lingual metadata

Data access and data exploration:
- Particular investigations will focus on problems of data comparability and relationships over time and space,
- linking of data over the micro-meso-macro axis & inclusion of macro-oriented data.

1. Data discovery II : Multi-lingual thesaurus & controlled vocabs

Support for an increased number of European researchers in resource discovery using native languages

Extension of the current CESSDA thesaurus

- identify further European languages to integrate
- exchange expertise; integrate approaches; project monitoring

Provision of controlled vocabularies

- Support harmonisation & comparability of datasets (Classifications)
- English translation of DDI key fields from recent portal records; options
- Create source controlled vocabularies (English & translations)

Implications: Financial, technical, intellectual property rights

- future use, maintenance & extension of the thesaurus

linking
european
data
resources

2. Infrastructure - Content harmonisation & conversion

Identification of the demands of the research community for harmonisation to common standards and options to meet these demands

Concepts on infrastructure elements & functional requirements

- Define structure to document & develop harmonisation rules and conversion keys
- Evaluate strategies & practical procedures to collect & create conversion keys
- technical & organisational conditions for a long-term infrastructure

Database & middleware to facilitate harmonisation techniques at the portal

conceptual work & functional specifications to prepare a technical prototype:

- DB for compiling & re-distributing concept, classification & conversion inform. (CCCDB)
- DB for collecting, managing & re-distributing information on survey questions (QDB)
- Specifications for integrated service of both databases & linking with CESSDA RI

Organisational & financial sustainability of the harmonisation services

- Evaluate options for cooperation with other infrastructures
- funding models for continuous input to the information system,
- development, installation and maintenance of central facilities & staff requirements

linking
european
data
resources

3. Enhancement of data & metadata infrastructures at the CESSDA RI

Setup strategic plans for general developments on metadata, data models and software upgrades.

Conceptual analysis and functional specifications

- requirements to upgrade backend software for data and metadata capture, processing and management which will complement the guiding frontend user services for complex data types (online publishing, discovery, access, analysis)
- determine the metadata and data model requirements of the CESSDA RI in handling complex dataset types throughout the entire data life-cycle.
- beyond cross-national surveys and time-series support for user services providing complex dataset types not currently present in the CESSDA RI portal

Concepts on future organisation & funding

- future resourcing & sustainable CESSDA infrastructure for the development of metadata standards and software tools (Funding models – National, European funding agencies)
- future strategic organisation of technical developments and training for the metadata technicians within CESSDA RI (Software development scenarios)

linking
european
data
resources

4. Evaluation of DDI version 3.0

Central issue upgrading the technical resources and to provide advanced data services in the CESSDA RI is to evaluate DDI 3.0 regarding its capacity to support the study life-cycle processes of complex datasets

Specials issues to evaluate DDI 3.0 and to plan potential solutions

- problems of a common identifier system for data resources, with version identification for version control purposes
- Concept & elements required to construct controlled vocabularies in multiple languages
- Concepts & elements for technical representation of the metadata from harmonisation and conversion DB and the Question database
- Facilities to document complex data types and data collections supporting manifold relationships linking research concepts, study level and dataset information with the single question - variable relations covering involved space and time dimensions as well as multi-lingual metadata of the study life-cycle

Cooperation

- DDI experts and the involved work packages of the CESSDA PPP project:
  WP4 (controlled vocabularies; multi-lingual thesaurus) - WP5 (CESSDA portal software) - WP8 (Data and metadata enhancements) - WP9 (harmonisation) - (WP12 technical prototyping)

linking
european
data
resources

Thank you!

Contacts & information:
http://www.nsd.uib.no/cessda/project/

linking
european
data
resources