



CESSDA

Council of European Social Science Data Archives



CESSDA-PPP Update

John Shepherdson

April 2010



What is CESSDA?

CESSDA - Council of European Social Science Data Archives

“The CESSDA network provides efficient data services to support European Social Research by facilitating network access to more than 25,000 data collections for thousands of users world-wide.”

Source: <http://www.cessda.org/>

What is the CESSDA-PPP?

- PPP – Preparatory Phase Project;
- Funded by EC 7th Research Framework Programme;
- Runs from January 2008 to June 2010 (includes a 6 month extension);
- Facilitates transition of CESSDA from informal grouping of data archives to formally constituted, fully integrated European Research Infrastructure (ERI).

PPP Members

- Austria
- Denmark
- Czech Republic
- Finland
- France
- Germany
- Greece
- Hungary
- Italy
- The Netherlands
- Norway
- Romania
- Slovenia
- Spain
- Sweden
- Switzerland
- UK

Plus associates from:

Poland; Russia; Luxembourg; Estonia; Portugal; USA; Canada; Croatia; Ireland; Ukraine; Latvia; Bulgaria; Slovakia; Serbia; Macedonia; Belarus; Lithuania.

Rationale

"The present major task is ... to create pan-European infrastructural systems that are needed by the social sciences ... to utilise the vast amount of data and information that already exist or should be generated in Europe. Today the social sciences ... are hampered by the fragmentation of the scientific information space. Data, information and knowledge are scattered in space and divided by language, cultural, economic, legal and institutional barriers"

Source: ESFRI European Roadmap for Research Infrastructures, Report 2006

Vision

- Integrated resource discovery tools
 - Multilingual searching
- Integrated common Authentication & Access
 - Single sign-on
 - Single access protocols
- Extensible system
 - nationally & internationally

Achievements 1

- Web-based management tools for multi-lingual thesaurus maintenance and development
 - Researchers can suggest new terms
 - Translators can do local translations
 - Managers control structure and content

<http://elsst.esds.ac.uk/login.aspx>

Achievements 2

- Prototype Single Sign On (SSO)
 - Shibboleth-based
 - Protecting data resources in UK and Norway

<https://shib-portal-cessda.data-archive.ac.uk/Cessda-Test/>

Id: test

Password: 098asd

Achievements 3

- Proposal for new architecture to support enhancements to services
 - Registry for service provider data collections
 - Standard metadata (DDI and SDMX)
 - ELSST* for multilingual discovery
 - Question data bank
- (*European Language Social Science Thesaurus)

Next Steps

PPP work plan includes technical and non-technical activity including:

- Final refinements to thesaurus
- Extend and test SSO prototype to real resources for real researchers, (Germany/UK), collaborating with eduGAIN (*in collaboration with other SSH Infrastructures*)
- Secure remote access to sensitive data between Germany/UK (*begin Autumn – possible legal constraints will be addressed*)
- Enhance existing portal to support harvesting and discovery using a wider range of standards (e.g. SDMX)
- Longer term development of more sophisticated architecture for wider range of services

Common Needs

CESSDA needs High Throughput Computing more than High Performance Computing

- systems that handle lots of simultaneous users, instead of systems that offer a lot of computing power for complex problems

Plus following 8 areas:

1. Authentication – is the user who they say they are?

- Requires co-ordinated involvement with national Federations. Some countries do not have a Federation in place. Several cross-national initiatives are developing (e.g EduGain)

2. Authorisation – is the user authorised to use this resource/file?

- Resources (datasets) often have legal access and IPR conditions attached
- Researchers/users have to agree to formally acknowledge acceptance of these conditions before being authorised to use the resource. This is managed through registration services

3. Access – system by which the user can access the resource on their desktop from a remote site
- Having identified a resource they want to use, the researcher needs to download it to their desk, regardless of its location
 - Requires technical system to identify the resource and permit downloading from diverse sites

4. *Audit* – information about users and use of resource. Required for:

Usage reports to resource (dataset) producers

Reports for funding bodies (e.g. research councils)

Information to support EC applications

Information should cases of misuse of resources need to be investigated (e.g. data protection breaches, commercial use of data available for research only)

1 - 4 above need to be interoperable to enable researchers to discover, locate, register and download resources

A national SSO system exists for social science data in the UK which the CESSDA PPP has demonstrated can be extended for use in other countries.

Extending SSO would also require:

5. PIDs - System of persistent identification of resources

Required to ensure that the resource that is received is the one that was requested

Extending SSO would also require:

6. *Common, cross-border licence agreement* would enable ESS and SHARE to be available within a single system rather than, as now, from completely independent systems

CESSDA PPP has drawn up suggested licences for its members. When finalised, this could certainly be applied to ESS and SHARE and if needed might be appropriate for use by CLARIN & DARIAH

Possible additions:

7. Adoption of common metadata standards for
a) cataloguing and
b) describing resources

8. Common tools
for cataloguing and resource discovery, thesaurus ...

- Any Questions?

Please contact

Hilary Beedham - Project Manager

(beedh@essex.ac.uk)