February 13 (Friday), 2015
14:00~17:15

YHTEISKUNTATIETEELLINEN TIETOARKISTO
FINLANDS SAMHÄLLSVETENSKAPLIGA DATAARKIV
FINNISH SOCIAL SCIENCE DATA ARCHIVE

International Conference on Data Preservation and Dissemination, 2015

Opening Remarks

Hiroshi Ishida, Director, Institute of Social Science, University of Tokyo

The New CESSDA: milestones and challenges of the European Social Science Data Archives

Sami Borg, Director, Finnish Social Science Data Archive

Providing service for Finland, CESSDA and others from the perspective of IT practitioner

Matti Heinonen, Senior Systems Analyst, Finnish Social Science Data Archive

Moderator: Yukio Maeda, Associate Professor, Interfaculty Initiatives in Information Studies/Institute of Social Science, Univerisy of Tokyo

Akamon General Resarch Building, Room 549

Center for Social Research and Data Archives
Institute of Social Science
The University of Tokyo

The New CESSDA: milestones and challenges of the European Social Science Data Archives

Sami Borg
director, Finnish Social Science Data Archive
International Conference on Data Preservation and Dissemination, Tokyo, February 13, 2015

YHTEISKUNTATIETEELLINEN TIETOARKISTO
FINLANDS SAMHÄLLSVETENSKAPLIGA DATAARKIV
FINNISH SOCIAL SCIENCE DATA ARCHIVE

Why data archives?

Without long-term data management and preservation plans data will sooner or later deteriorate → no re-use options

For too many this comes still as a surprise…

Sometimes consequences can be just harmful…

But unfortunately, every now and then they are irrecoverable…

Luckily assistance is available! Something just need to enter in….

Providing evidence based knowledge on society and on human behaviour and attitudes

To begin with some quotes from a Finnish sociologist Erik Allardt…

- Allardt's words are from an article where he analyses Stein Rokkan's role and influence on the development of political sociology from 1950s to 1970s ("Stein Rokkan and the Twentieth Century Social Science")
- "Social improvements can not be accomplished by socialization and learning alone, but that they also required substantial reforms both in economy, polity and social policy."
- "…study of political institutions and the creation of welfare policies …led to a strong interest in comparative research, systematic comparisons of societies"
- "…→ development of sophisticated quantitative methods and…technical and administrative tools for gathering reliable data. →… the great methodological value of data archives."

Emergence of Social Science Data Archives

- First social science data archives emerged in early 1960s (ICPSR in the U.S., ZA in West Germany, UKDA in Britain, then soon also NSD in Norway)
- All this was after the launch of the first international large scale comparative research project in sociology and political science
- Increasing popularity of mass surveys and computer-assisted methodologies
- Social sciences were early forerunners in providing open access to research…
- …and their data archives are still pretty well developed in providing well standardised metadata to research data for open scientific use

Council of European Social Science Data Archives ("old" CESSDA)

- 1976 - CESSDA was founded (sufficient number of archives had emerged)
- 1980 – Standard Study Description (early agreement on harmonising contents of documentation)
- 1987 – First expert seminar (professional training)
- 1998 – Starting to build Nesstar (online tool for the Internet age)
- 2000 – DDI 1.0 (major step towards more detailed and professional data documentation)

The old CESSDA?

- An informal group of about 20 European data organisations
- Members were quite different, with varying funding structures
- More than 35 years of successful voluntary cooperation
- In 2010 CESSDA hold jointly over 25000 datasets

Objectives of the old CESSDA

- 1. To promote the acquisition, archiving and distribution of data throughout Europe.
- 2. To promote projects and procedures for enhancing exchange of data and technologies among data organizations.
- 3. To stimulate the development and the use of these procedures throughout Europe.
- 4. To encourage new data organizations to further these objectives.
- 5. To promote the integration of the European database.
- 6. To associate and cooperate with other international organizations sharing similar objectives.

Then, in early 2000s something happened outside the CESSDA networks…

European Strategy Forum of Research Infrastructures (ESFRI) was established

ESFRI

- ESFRI Is about supporting Policy-Making on Research Infrastructures in the European Research Area
- Started in 2002…
- …to support a coherent and strategy-led approach to policy-making on research infrastructures in Europe and…
- …to facilitate multilateral initiatives leading to a better use and development of research infrastructures
- Have produced and updated regularly European level ESFRI roadmaps on research infrastructures
- Have encouraged ESFRI member countries to plan and publish their national RI roadmaps

Example of a national roadmap of RIs

Roadmap 2014–2020: National infrastructures 31, potential ones 2

Social sciences and humanities
- ESS Finland
- FIN-CLARIN
- Finna
- FinElib
- FMAS
- FSD and CESSDA
- TTA-KDK-PAS

Environmental sciences
- EISCAT_3D Finland
- FIN-EPOS
- FINMARI
- ICOS Finland HQ in FI
- INAR RI
- oGIIR (potential)

Biological and medical sciences
- BBMRI.fi
- Biocenter Finland
- EATRIS Finland
- ELIXIR Finland
- EuBI Finland
- EU-OPENSCREEN Finland
- INFRAFRONTIER Finland
- Instruct Finland
- NaPPI
- NVVL

Material sciences and analytics
- MAX IV
- OMN
- XFEL and XBI

Natural science and engineering
- BIOECONOMY
- CTA
- Euclid
- JYFL-ACCLAB

E-infrastructures and mathematics
- CSC RI
- PRACE Finland
- FGCI (potential)

ESFRI roadmaps

- The ESFRI Roadmap identifies new Research Infrastructures (RI) of pan-European interest corresponding to the long term needs of the European research communities, covering all scientific areas, regardless of possible location.
- First ESFRI roadmap in 2006
- Updates in 2008, 2010 and 2016 (in process)
- http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri-roadmap

CESSDA and ESFRI

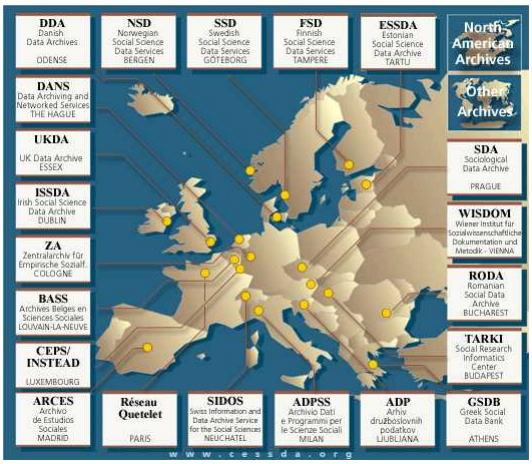
- CESSDA was listed on the first ESFRI roadmap
- This helped to get EU funding for planning more close cooperation
- CESSDA PPP project 2008-2010

A Preparatory Phase Project for a Major Upgrade of the Council of European Social Science Data Archives (CESSDA) Research Infrastructure

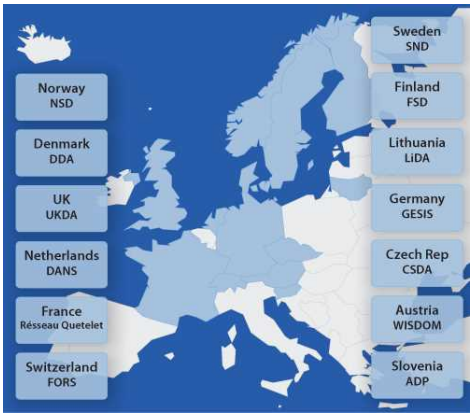
Towards the new CESSDA

- To join the new CESSDA countries (their governments) must now sign for sustainable cooperation and guarantee funding for sufficient national services
- 2010 – Signing the MoU for the new CESSDA
- Special legislation at the EU level for the forthcoming new RIs (the so called ERIC legislation)
- 2010 – 2013 CESSDA Steering Committee led planning of the new CESSDA

Council of European Social Science Data Archives (1976-2013) = the old CESSDA



The new CESSDA (5/2014)



Funding and administration of the new CESSDA

- Norway is the host country for new central services, which are situated in Bergen
- Germany plays a significant role in providing services for training
- Together the governments of Norway and Germany pay about 75 % of the annual 2 M euro costs of running the international RI
- Other countries pay the rest as membership fees according to model based mainly on their GDP
- All member countries are expected to choose and fund a national service provider to take care to national obligations in the CESDDA cooperation

CESSDA administration and organisation

- General Assembly
- Board of Directors
- Managing director
- Scientific Advisory Board
- Service Providers' Forum

The new CESSDA: towards an European Research Infrastructure Consortium ERIC

- CESSDA is now a non-profit limited company registered in Norway (CESSDA AS)
- As soon as the Norwegian parliament will approve some legislative amendments, CESSDA will change from ltd company to an ERIC (similar to EU legislation)
- Central office in Bergen has been operative from autumn 2014
- Managing director mr. Paul Jackson (Paul comes from the Office of National Statistics (ONS) in the UK where he was the Head of ONS Knowledge and Information Management. Paul has a background from statistics, research, legal issues as well as from research infrastructure projects.)
- Currently three other staff members are working in the hub
- In 2015 they are likely to have a couple of staff members more

How CESSDA runs its services?

- http://www.cessda.net/
- National Service Providers provide many of their services as before
- The idea is not to expand rapidly centralised CESSDA services
- Rather the hub and training services in Germany will coordinate and harmonise European wide activities with several instruments
- Implementation of the CESSDA work plan (with CESSDA's "own" money) will be based or targeted calls → many groups formed from service providers will work for specific duties
- The hub applies EU funding and funding from other sources as a legal entity (such funding can be redirected to national service providers participating in projects)

CESSDA work plan and obligations to national service providers

- Implementation of the new CESSDA work plan starts in spring 2015
- The work plan is closely linked to obligations of Service Providers, which are part of the CESSDA statutes

Annex 2: Obligations of Service Providers

All Members and Observers must appoint Service Provider(s) able to meet these obligations:

1. be fully compliant with the elements of the DDI metadata standard that are required to enable the member/observer to contribute fully to CESSDA activities and which will be identified by CESSDA;
2. adopt and apply the CESSDA common single sign-on user authentication system when it exists;
3. enable the harvesting of all their resource discovery metadata and relevant preservation metadata for inclusion in the CESSDA data portal;
4. make their data holdings downloadable through common data gateways;
5. ensure that their local language(s) within the multi-lingual thesaurus are maintained;
6. share their data archiving tools (under the IP conditions described in the CESSDA Statutes, Article 11);

CESSDA work plan and obligations to national service providers

Obligations of Service Providers continue…

7. adhere to the principles of the OAIS reference model and any agreed CESSDA requirements for operating trusted repositories;
8. contribute to CESSDA's cross national data harmonisation activities;
9. contribute material and/or expertise to the cross-national question bank;
10. provide mentor support for CESSDA Observers and their representative Service Providers to achieve full Membership;
11. provide member support for countries with immature and fragile national infrastructures to help them build up needed competence later to be able to fulfil tasks as Members;
12. facilitate access to national government and research funded relevant data, dependent on national legal systems;
13. adhere to CESSDA's Data Access and Dissemination Policy;
14. adhere to the provisions of the entire Organisation's policies as required.

Current members and plans for widening and strengthening CESSDA

- The old CESSDA had more countries than the current new CESSDA
- Many activities to encourage new countries to join in
- Having more countries to join CESSDA needs much lobbying at the political level as ministerial level decision are needed to become a full member
- There are, however, other membership options than just the full member status

International cooperation of CESSDA / who can become a member

- CESSDA Statutes: "The following entities, which adhere to the MoU establishing CESSDA, may become Members or they can join as Observers without voting rights of CESSDA":
- EU Member States
- Associated countries
- Third countries other than associated countries
- Intergovernmental Organizations

For instance, ICPSR from the U.S. has shown such interest

Milestones and challenges?

- Emergence of data archives
- Long history of international cooperation → tools, services & mutual trust
- Services for the social sciences now viewed as large-scale research infrastructures
- Sustainable funding schemes for international data services for social sciences
- Favourable to national funding schemes
- Careers of data professionals

The values of CESSDA (according to the CERSSDA Statutes)

To provide effective leadership and be a catalyst for change

To provide support for its Member and partner organisations in order to enable them to derive maximise benefit from the Consortium

To engage constructively and openly with the varied user communities served by its Member and partner organisations, and to work with other stakeholders to mutual benefit

To demonstrate professionalism of the highest standards in all its activities and to promote and enhance professionalism amongst its Members and related Service Providers

To promote and facilitate research, teaching and learning of the highest quality

To provide a positive, productive and rewarding working environment for all staff in which the rights of staff are respected and upheld

Thank you!

1

Providing service for Finland, CESSDA and others

IT Practitioner's viewpoint
Matti Heinonen / Head of IT at FSD

International Conference on Data Preservation and Dissemination, Tokyo, February 13, 2015

YHTEISKUNTATIETEELLINEN TIETOARKISTO
FINLANDS SAMHÄLLSVETENSKAPLIGA DATAARKIV
FINNISH SOCIAL SCIENCE DATA ARCHIVE

YHTEISKUNTATIETEELLINEN TIETOARKISTO
FINLANDS SAMHÄLLSVETENSKAPLIGA DATAARKIV
FINNISH SOCIAL SCIENCE DATA ARCHIVE

2

SIX YEARS AGO …

Europe

- CESSDA: loose ”club” of entities with common interests
- Bigger archives leading the way

Finland

- Infrastructure
  - Entities with common interest but few common services
  - Loose co-operation
- Funding
  - We’re not a research organisation so not many external funding sources available to us

FSD

- Solid foundation for archiving
  - IT: basic archival services in good working order
- Too small for big projects
  - IT: 4 FTEs in the team
  - IT: focusing on the fundamentals – keeping things running

…BUT THINGS CHANGE

- Project funding has become available
  - More people → bigger projects
  - Consistent service development
- Increasing interest in data and data management in Finland
  - Partnerships possible
  - Services to use and services to offer
- From CESSDA to CESSDA-ERIC
  - Common European goals

LANDSCAPE NOW

Europe

- CESSDA becoming full blown ERIC
- CESSDA is a major player in its own right
  - Interoperability: common practices & tools will emerge
  - Leadership by the central hub
    - Clearer goals & guidance
  - FSD as official service provider for Finland
  - Service levels that must be met
  - Definitions and precise requirements not available yet

Finland

- Rising interest in good data management
  - Partners! Common goals! Services!
- Funding structure clearer
  - Must compete (good!) and thrive with merit (super!), but can plan ahead (excellent!) and consistent goals possible (outstanding!)
- On-going service development
  - Not everybody must ”invent the wheel”
  - Using resources wisely

FSD

- OAIS compliant
- DSA certified
  - First in Finland!
- Strong teams for service development
  - Service teams and IT teams
- Solid foundation not only for archiving but to build new services

PROJECTS BUILDING TOWARDS CESSDA

CESSDA PPP

- Active 2008-2010
  - http://cessda.org/project
- Preparatory work towards CESSDA ERIC
  - Plans, draft legal documents, specifications
  - Software prototypes
    - CESSDA portal, PIDs and versioning, Authentication
  - Thesaurus management software
    - http://elsst.esds.ac.uk/

Data without Boundaries

- Active 2011 – 2015
  - http://www.dwbproject.org/
- Lowering barriers for research done on official microdata
  - Easier access
  - Structure for cooperation
- ONSs and data archives

Outcomes of DwB

- Reports, training, standards development
- DwB Discovery Portal
  - http://dwb-dev.nsd.uib.no/portal/#/home

DASISH

- Active 2012-2014
  - http://dasish.eu/
- Building common tools for SSH ESFRIs
  - CESSDA, CLARIN, DARIAH, ESS, SHARE
  - Reports, tools, training

Tools by DASISH

- Many available from Github
  - https://github.com/dasish
  - Open Source
- TERESAH – tools registry tool
- Annotation framework backend
- DASISH Metadata Portal
  - http://ckan.dasish.eu/

Tools by DASISH

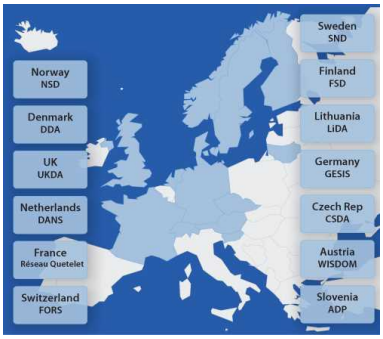
- QDDT - Questionnaire Design and Documentation Tool
- QVDB - Question and Variable Database
- Joint Metadata Domain tools

CESSDA Tasks

- Active 2015 →
- Commissioned by CESSDA Central Hub
- Implemented by CESSDA Service Providers & partners
- Will implement CESSDA work plan
- Building & operating services for CESSDA
- Work-in-progress, details available later

Speculation: Probable topics for CESSDA tasks in the near future

- CESSDA catalogue & tools
  - Access Gateway
- Common PID-tools
- Single Sign On for CESSDA
  - AAI infrastructure
- Data Harmonisation tools



Techie's Viewpoint to

CESSDA REQUIREMENTS

Membership Obligations

- No specific requirements available yet
- Generally we are as prepared as we can
  - Amount of work to be done?
    - We know once final requirements are available
    - Expect small effort here and there
    - Expect medium effort on some topics
  - Tech work vs. content work vs. administrative work
    - Fitting it all together is not just tech work
    - E.g. 13 member countries, 13 legislations

Obl. 1: Metadata DDI-compliant?

"be fully compliant with the elements of the DDI metadata standard that are required to enable the member/observer to contribute fully to CESSDA activities and which will be identified by CESSDA"

- Our metadata is based on DDI-C

Obl. 2: CESSDA Common SSO?

"adopt and apply the CESSDA common single sign-on user authentication system when it exists"

- We are
  - Service Provider (SP) for Haka Identity federation
    - Finnish HE institutes, SAML/Shibboleth based
  - Identity Provider (IdP) for clients outside of Haka
    - SAML/Shibboleth based

Obl. 3: Enable Harvesting Metadata?

"enable the harvesting of all their resource discovery metadata and relevant preservation metadata for inclusion in the CESSDA data portal"

- It is already possible to harvest FSD metadata using
  - OAI-PMH @ http://services.fsd.uta.fi/oai
  - NESSTAR
  - DDI-C downloadable as files @
- Our OAI-PMH not yet supporting DDI
  - should be simple enhancement

Obl. 4: Data holdings downloadable?

"make their data holdings downloadable through common data gateways"

- Aila service portal already does this for us
  - https://services.fsd.uta.fi

Obl. 6: Share your tools?

"share their data archiving tools (under the IP conditions described in the CESSDA Statutes, Article 11)"

- Some tools already released as open source
  - https://github.com/tietoarkisto
- Have code, will share
  - But we have to check any IP restrictions before we do

Obl. 7: OAIS & CESSDA Compliance?

"adhere to the principles of the OAIS reference model and any agreed CESSDA requirements for operating trusted repositories"

- FSD is OAIS compliant
- FSD is DSA certified
- FSD's metadata conforms to existing CESSDA recommendations

Obl. 9: Question Bank?

"contribute material and/or expertise to the cross-national question bank"

- Question bank and tools for FSD under development
  - Our major effort for 2nd half 2015

Reaching CESSDA Service Levels

- Again, no exact requirements given yet
- Our internal goal: robust & agile IT services platform
  - Doing quite nice already but we want to reach higher
  - Should help to meet any Service Level requirements from CESSDA

YHTEISKUNTATIETEELLINEN TIETOARKISTO
FINLANDS SAMHÄLLSVETENSKAPLIGA DATAARKIV
FINNISH SOCIAL SCIENCE DATA ARCHIVE

Image by Ssolberg CC BY-SA 3.0

PROJECTS BUILDING TOWARDS FINNISH NATIONAL INFRASTRUCTURE

YHTEISKUNTATIETEELLINEN TIETOARKISTO
FINLANDS SAMHÄLLSVETENSKAPLIGA DATAARKIV
FINNISH SOCIAL SCIENCE DATA ARCHIVE

Haka Identity federation

- Finnish universities, polytechnics and research institutions
  - Since early 00's
  - Operated by CSC - IT Center for Science
  - Shibboleth/SAML based
  - https://confluence.csc.fi/display/HAKA/In+English
- Reliable identities for ~90% our current customers
  - The remaining 10 % we need to manage ourselves

Yhteiskuntatieteellinen tietoarkisto | www.fsd.uta.fi

KDK (NDL) – National Digital Library

- Improving accessibility and preservation of digital cultural heritage
  - Including research data
- http://www.kdk.fi/en

KDK (NDL)– Finna portal

- Managed by National Library of Finland
- Search and find all material in libraries, museums, archives
  - https://www.finna.fi/
  - OAI-PMH harvesting from FSD
  - FSD metadata included (at the moment only in finnish)
- Available from github
  - https://github.com/NatLibFi/NDL-VuFind
- Relay to Europeana in the future
  - http://www.europeana.eu/

KDK (NDL) – Long-term preservation

- Bit-level long term preservation
  - For ”gatekeeper” organisations, not individuals
- Managed by CSC – IT Center for Science
- FSD will use the service once ready for us
  - pilot planned for 2015

TTA and ATT Projects

- Open Science and Research Initiative
  - http://openscience.fi/
- Two successive projects commissioned by Ministry of Education
  - TTA 2013 – 2014
  - ATT 2014 – 2017

Tools by TTA and ATT

- Etsin
  - Search and find research data
  - FSD metadata available via Etsin
  - https://etsin.avointiede.fi/
- IDA
  - Storage service for research data and metadata
  - https://www.tdata.fi/ida

REMS

- Resource Entitlement Management System
  - Developed & operated by CSC - IT Center for Science
  - https://confluence.csc.fi/display/REMS/Home
- FSD will start using REMS in 2015

Data Policies at Universities

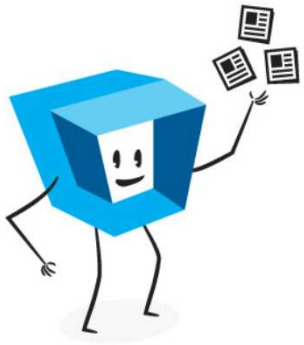
- Finnish universities are formulating Data Policies
  - Need services and tools to support them
  - What will emerge? How the existing ones like FSD will be utilised?
- We got a good service concept to utilise

ADDRESSING FINNISH REQUIREMENTS

- Meeting CESSDA reqs addresses fundamental issues
- Need to address requirements specific to Finland separately
  - Danger! What if we need to implement same req's for CESSDA and Finland using different technologies → wasted resources
  - Try to reuse as much as possible

National requirements?

- What are they?
  - There's no definite list of what-to-do's
  - See and wait to see what emergences
- Adhere to requirements by Finnish policy makers when arise
- Serve the Finnish Scientific community
  - If you see a need in your own field, try to address it



THANK YOU!

ANY QUESTIONS?