DASISH

Digital Services Infrastructure for Social Sciences and Humanities

Workshop
Trust and Certification

Vigdis Kvalheim

Norwegian Social Science Data Services (NSD)

16th -17th October – NWO, Den Haag

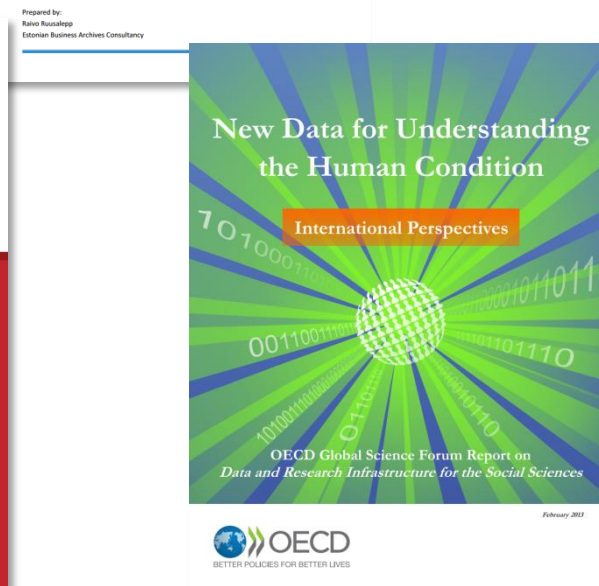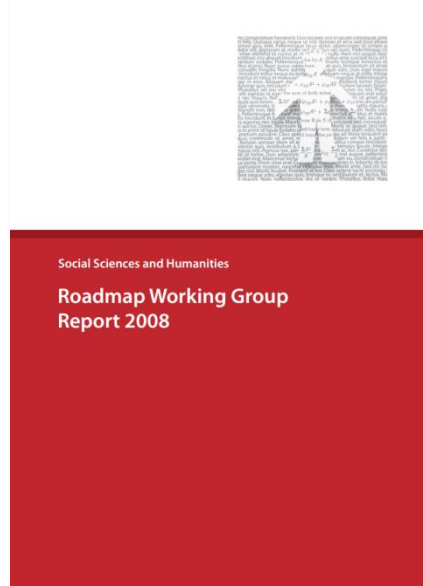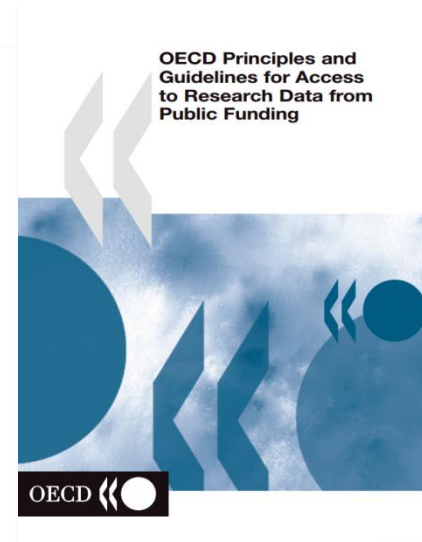# The context is changing- new policies, new instruments

Table 1: Summary of challenges and recommendations from the OECD Global Science Forum Report

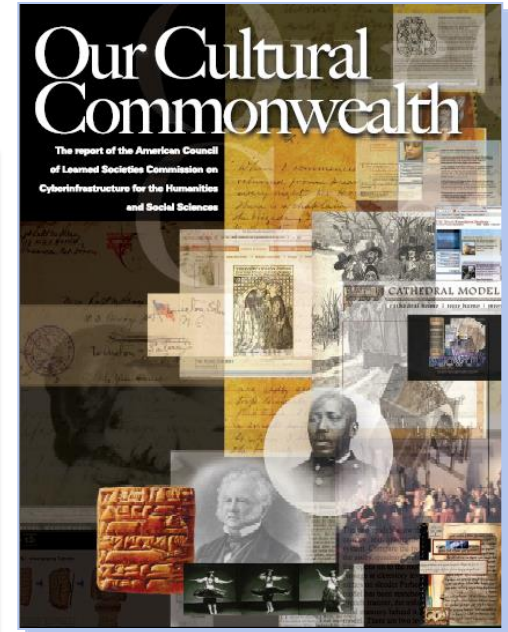| Challenge	Recommendation	Main actors
Reliability, statistical validity and generalizability of new forms of digital data	**Collaborate internationally to provide resources (methods and tools) for researchers**	National research funding agencies
Information about the existence of micro-data and their availability	Data should be processed and **documented to agreed and common standards**,	National and international statistical organisations
Privacy of individuals	**Collaboration** to develop an internationally recognised **framework code of conduct for research** use of new forms of personal data	Research funding agencies and data protection authorities
Barriers to access (legal, cultural, language, proprietary rights)	Cooperation on mechanisms to improve access for comparative research	National statistical agencies and international agencies (e.g. World Bank, WHO, OECD)
Increasingly interdisciplinary research agenda	**Share expertise**, knowledge and resources, particularly in the areas of data access, linkage, integration and analysis	Data producers and data users
Increased focus on cross-country comparative research	**Harmonise social and economic** data; foster an integrated approach to data design and harmonisation, access and sharing	National and international statistical agencies, researchers and research funders
Re-usability of data	Requirements for researchers to **make data management plans**	National funding agencies
Resources and infrastructures connected to professional data preservation and curation	Assessment of national needs and assets that will contribute **to national plans of action, Roadmaps**	Social science research communities
Ensure effective data sharing	Incentives to ensure **data sharing**, through common systems for referencing and recognizing data resources in research publications	Research funding agencies, publishers of research and employers of researchers

Common Needs and Requirements

- Open access policy rely on good data management, including long-term preservation and providing access to data resources

- The international push towards open access increases the need for research data infrastructures and deposit services, such as the CESSDA archives

WP4 Data Archiving

"*Preservation was once an obscure background operation of interest chiefly to conservators and archivists. It is now widely recognised as one of the most important elements of a functional cyberinfrastructure.*"

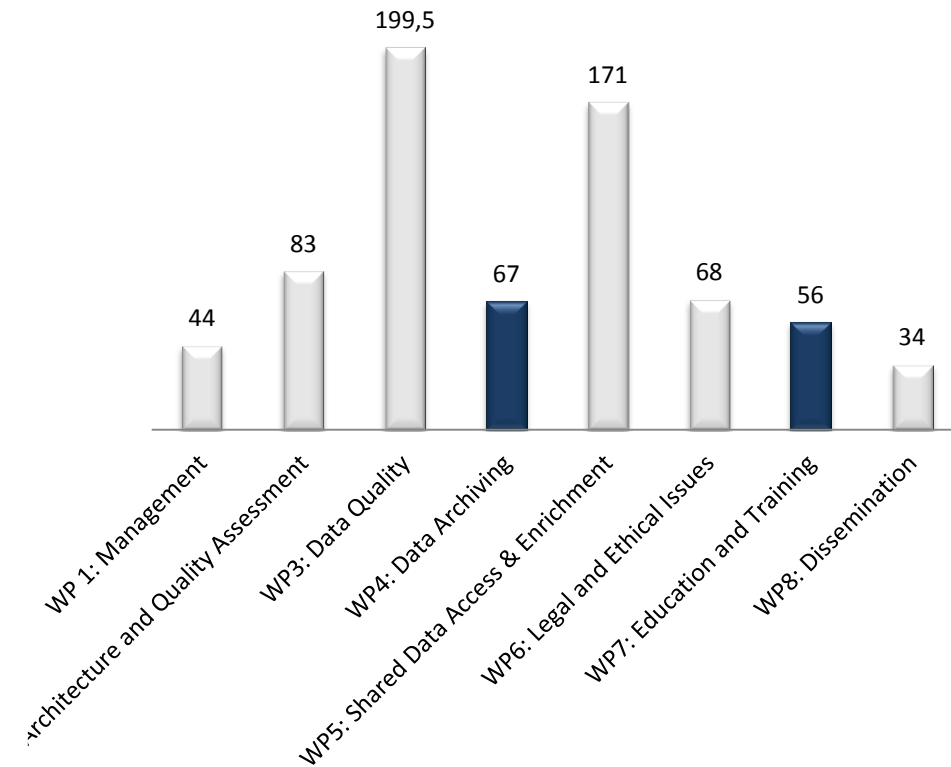(Our Cultural Commonwealth, ACLS, 2006)

DASISH – Data Service Infrastructure for the Social Sciences and Humanities

- EU initiative based on ESFRI's Roadmap for Research Infrastructures

- Brings together all 5 ESFRI research infrastructure initiatives in the social sciences and humanities

- Develop common solutions on common problems

- The rationale behind this idea is that
  - double work and financing should be prevented,
  - initiatives should mutually benefit from the advanced work of the others and
  - to establish joint integrated domains where this makes sense for the SSH users.

DASISH PM Distribution and Partners

PM

CESSDA	NSD , Norwegian Social Science Data Services (15 PM)
	FSD, Finish Social Science Data Archive (2 PM)
	SND, Swedish National Data Services (5 PM)
	GESIS - Leibniz Institute for the Social Sciences, (6 PM)
CLARIN	MPG , Max Planck Institute for Psycholinguistics (6 PM)
	UiB, University of Bergen (7 PM)
DARIAH	OEAW, Austrian Academy of Sciences (5 PM)
	DANS, Data Archiving and networked services (5 PM)
	UGOE, Goettingen University (6 PM)
ESS	CITY, City University, London (2 PM)
SHARE	CentERdata, The Netherlands (7 PM)

NSD © 2014

WP4 Data Archiving to Secure Access

"DASISH will rely on common data services offered by a network of *strong data centres* with *national backing*"

- Assess and discuss the state of data and deposit services in the SSH domain and identify bottlenecks and requirements (variations and similarities)

- Develop/suggest a model for common deposit service which handle various types of data, and could be provided to all SSH researchers

- Work out and suggest policy rules and guidelines for proper data management, that can be taken up by a common data service layer providing long term preservation and curation services

- Information material, training and dissemination in cooperation with WP7 and WP8.

NSD © 2014

WP4.1 and WP4.2: Reports

Data Service Infrastructure for the Social Sciences and Humanities

EC FP7

Grant Agreement Number: 283646

Deliverable Report

Deliverable: D4.1
Deliverable Name: 4.1 Roadmap for Preservation and Curation in the SSH
Deadline: 31 December 2012
Nature: R

Responsible: NSD, CentERData
Work Package Leader: NSD

Contributing Partners and Editors: NSD, CentERData, GESIS, MPG (TLA), CITY, OEAW

www.dasish.eu   GA no. 283646   1

WP4.1 Roadmap for Preservation and Curation in the SSH

Data Service Infrastructure for the Social Sciences and Humanities

EC FP7

Grant Agreement Number: 283646

Deliverable Report

Deliverable: D4.2
Deliverable Name: 4.2 "Report about Preservation Service Offers"
Deadline: 31 December 2012
Nature: R

Responsible: NSD, UiB
Work Package Leader: NSD

Contributing Partners and Editors: NSD, UiB, DANS, GESIS, MPI (TLA), CITY, OEAW

www.dasish.eu   GA no. 283646   1

WP4.2 "Report about Preservation Service Offers"

NSD © 2014

D4.1 and D4.2: Fact Sheets – First Year

Roadmap for Preservation and Curation in the SSH*

Data Service Infrastructure for the Social Sciences and Humanities (DASISH) brings together all 5 ESFRI research infrastructure initiatives in SSH area. The goal of DASISH is to determine areas of cross-fertilization and synergy in the infrastructure development and to work on concrete joint activities related to data, such as data access, data sharing, data quality, data archiving and legal and ethical aspects.

Background and description

In the Social Sciences and Humanities research is increasingly driven by the availability of a variety of digital resources, which exhibit an escalating internal complexity as well as diverse external relationships. The data production, management and dissemination processes are organized in a distributed manner, both within and between data producing institutions and data repositories. This fragmentation should be taken into account when designing or developing research infrastructures and data repositories for the corresponding scientific disciplines. The data produced within different data communities should be made available to their respective designated communities via environments that implement discipline specific workflows in a *trustworthy* manner.

A well-defined set of guidelines for digital curation is crucial to the continued viability and trust of digital materials. This report assesses existing data repository models, or frameworks that provide checklists or guidelines to data preservation entities. Based on the assessment of the various models it can be considered as a "guideline of guidelines" that can be used for a broad spectrum of digital long-term repositories that aim to retain their validity over a longer period. By identifying and discussing selected benchmark guidelines and standards for trusted digital repositories, the report can work as an implementable resource for existing and emerging repositories that seek to provide trustworthy long-term data repository services.

Findings

The guideline and standards are organized in a five-step 'trust maturity' model. This model is based on the five organizational stages of digital preservation (Kenney & McGovern, 2003), the CMMI five organizational maturity levels (SEI/Carnegie Mellon, 2010), and the Trusted Digital Repository framework that consist of three levels of trustworthiness.

The assumption behind the model is that all data repositories that seek to become trusted digital repositories should assess their organization and aim for a certain level of trust maturity.

Level 3, peer-reviewed self-assessment through the Data Seal of Approval (DSA), is considered the benchmark level as it contains trust requirements that are both necessary and sufficient for most data repositories within the SSH area. CLARIN and CESSDA require DSA level and have some additional community specific requirements.

External review and formal certification at level 5 is fairly expensive, time consuming and requires a high level of expertise. ISO 16363 certification should therefore only be considered by repositories that have special data holdings and special needs.

*Deliverable 4.1 of WP 4: Data Archiving in the Social Sciences and Humanities. Available at: http://dasish.eu/publications/projectreports/D4.1_-_Roadmap_for_Preservation_and_Curation_in_the_SSH.pdf

http://dasish.eu/publications/projectreports/D4.1_-_Roadmap_for_Preservation_and_Curation_in_the_SSH.pdf

Report about Preservation Service Offers*

Data Service Infrastructure for the Social Sciences and Humanities (DASISH) brings together all 5 ESFRI research infrastructure initiatives in SSH area. The goal of DASISH is to determine areas of cross-fertilization and synergy in the infrastructure development and to work on concrete joint activities related to data, such as data access, data sharing, data quality, data archiving and legal and ethical aspects.

Background and Description

Europe has a number of facilities for storage of data. Some have a national mandate while others are based in local and independent institutions and have a narrower mandate. Some deposit services are designed to meet the specific needs of the research community, while others are general public services. The report describes and analyses a selection of existing institutional and academic deposit services within the Social Sciences and Humanities (SSH) based on a common analytic framework and a set of information categories. By analysing practices, policies and available documentation, it provides a set of best practices and guidelines for the key activities within SSH data deposit services. The services are selected within the five ESS ESFRI initiatives and the commercial sector.

Services analysed:

- UK Data Archive (UKDA)
- Norwegian Social Science Data Services (NSD)
- GESIS Data Archive (GESIS)
- Data Archiving and Networked Services (DANS)
- The Language Archive (TLA)
- European Social Survey (ESS)
- Survey of Health, Ageing and Retirement in Europe (SHARE)
- Figshare
- Flickr
- Dropbox
- Youtube

Findings

The report aims to answer a basic question: can services offered guarantee that data deposited today keeps its value in the future? The answer is that even though we can never be absolutely certain, it is easier to rely on the deposit services if they have an explicit set of basic and trustworthy guarantees in place.

One of the main findings is that the institutional and academic deposit services offered to scholars across Europe are more developed within the Social Sciences than within the Humanities, much thanks to the long history of the CESSDA (Council of European Social Sciences Data Archives). The CESSDA members (e.g. UKDA, GESIS and NSD) share much of the same organizational characteristics. For example, they all have a trusted and acknowledged long-term funding basis through national commissioning. The services within the Humanities are more fragmented and the funding is based more on projects and independent institutions. High-quality services are, however, emerging within the Humanities, both through the ESFRI-processes and through the fulfilment of international well recognized standards of trust.

None of the commercial services analysed are regarded or recommended as trusted long-term repositories and are thus not designated for research infrastructural purposes. The main concern is that there are no guarantees of longevity and continuity in terms of access, data sharing and use. It is also difficult to be confident that data is stored in a way that makes it possible to retrieve and read after long term storage.

*Deliverable 4.2 of WP 4: Data Archiving in the Social Sciences and Humanities. Available at: http://dasish.eu/publications/projectreports/D4.2_-_Report_about_Preservation_Service_Offers.pdf

http://dasish.eu/publications/projectreports/D4.2_-_Report_about_Preservation_Service_Offers.pdf

WP4.3 and WP4.4: Reports

WP4.3 Scope and Characteristics of Data Archive Services within the DASISH Communities

WP4.4 Report about Preservation Policy-Rules (Preservation Challenges)

Scope and Characteristics of Data Archive Services within the DASISH Communities

- **The questionnaire**; based on the results and recommendations of D4.1, D4.2 and the DADS

- **The purpose**; to gain broader and more detailed insights about the organization, the state of and the degree to which data archive solutions exists across Europe and across scientific fields.

- Point of departure for the next steps: having in-depth interviews with **selected** data archive services

WP4 Data Archiving

DASISH survey findings

SSH are well placed to develop common solutions for data management, access and sharing

All fields face challenges in data preservation and data curation, particularly in a distributed environment.

Variations across research fields and countries regarding the existence of research infrastructures and deposit solutions

Important recommendation

"Specialised support services should be considered as means to ensure the cost-effective production, use, management and archiving of research data."
(OECD Principles and Guidelines)

Support deposit services with national backing, long-term funding and transparency in procedures and policies.

Thank you for listening!