DASISH

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.

Report about Preservation Service Offers*

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

Recommendations

The report publishes its own agenda through a set of ten high-level recommendations for ensuring long-term preservation of digital information within the SSH area.

- Specialised deposit services should replace "private" or project based deposit solutions in order to support and enhance long-term preservation and open access to research data.
- Deposit services should have a clear mandate, operational status and responsibilities as a data-archive and communicate these internally and externally, to all repository stakeholders.
- The repository should have a mission statement that reflects its goals and its commitment to the preservation of, long term retention of, management of and access to digital information.
- The repository should have written policies or other documents that specify the type of information it will accept, preserve, retain, manage, and provide access to.
- The repository should identify and comply with roles, rights and obligations concerning use of various types of data.
- The repository should create and maintain business and financial plans and a have a clearly defined funding model.
- The repository should facilitate solutions that gather necessary information, establish contact with data producers, and provides the data producers with all necessary information.
- The repository should process and validate the received information from the data producer and initiate appropriate follow-up action with the data producer if necessary.
- The repository should perform data cleaning, validation, assigning preservation metadata, assigning representation information and ensuring acceptable data structures or file formats to ensure trustworthy long-term preservation and retention of data.
- The repository should provide tools and interfaces that makes the archived data easy accessible, by using unique persistent identifiers for each available data package.