cessda
Consortium of European Social Science Data Archives

CESSDA Persistent Identification Task Force established

Being able to unambiguously locate and access digital data resources and then associate them with the related metadata is becoming more and more essential in the current digital era. This will allow the citation, retrieval and preservation of these data in a reliable manner. Past CESSDA projects have shown that there is a need to introduce the use of persistent identification (PID) for data since maintaining and resolving PID is seen as a must for repository systems with a long-term strategy.

With regard to this, CESSDA held a workshop on PID within CESSDA. The workshop took place on 18-19 February 2016 at GESIS in Cologne, Germany.

The workshop aimed to set the stage for developing a joint CESSDA PID Policy, which not only contributes to the cooperation and exchange of data between the different research institutes active in the social sciences, but also facilitates the access to data for data users all over Europe. The workshop also served as a platform to discuss the results of work carried out in recent months on the use of PID by GESIS, as well as to provide an overview of the general status quo concerning the use of PID within CESSDA.

Representatives of CESSDA Service Providers and other social science data archives from ten different European countries participated in the workshop and discussed the implementation of a common PID policy for data within CESSDA. The archives and countries represented at the workshop were: ČSDA (Czech Republic), DDA (Denmark), FSD (Finland), GESIS da|ra (Germany), ISSDA (Ireland), DANS (Netherlands), NSD (Norway), ADP (Slovenia), SND (Sweden), and FORS (Switzerland).

The workshop began with a presentation of the results of the recent surveys carried out by GESIS. Expert interviews were held with representatives of DANS, GESIS and SND and a quantitative survey addressed all CESSDA Service Providers. After that, five users of persistent identification systems presented their experiences using PIDs and the challenges that they faced during the implementation phase. Those were quite variegated: ČSDA had to deal with the difficulty that no DOI registration agency exists in the Czech Republic, while SND struggled with the decision of whether or not to introduce the PID on study or dataset level and DANS faced challenges linked to the introduction of a second PID system (DOI). Meanwhile, GESIS needed to find a unique way of procedure to integrate data into its catalogue. They explained which PID systems they had chosen and the reasons behind their decisions.

Next, five further data service providers which were not using PID systems explained their experiences and concerns with the application of a persistent identification system. Those concerns were about very different issues: ADP is very interested in the application of a PID system, but needs technical support,

CESSDA AS
Eleanor Smith
Senior Communications Officer

Parkveien 20
5007 Bergen
NORWAY

phone: +47 55 58 36 47
e-mail: eleanor.smith@cessda.net
www.cessda.net

while FORS is introducing a new archiving repository system and first needs to rise to the challenge of defining its workflows. FSD, as a small archive, has to cope with the problem of funding and for ISSDA the barrier consists in not having a DataCite service provider in Ireland. NSD's concerns mainly affect the persistent identification of different versions of datasets.

The role of CESSDA regarding the use of persistent identification was also debated. All archives, especially the PID non-users, would like CESSDA to give recommendations on the use of PIDs and information about their assets and requirements. It was agreed that CESSDA should provide guidance in this area to its Service Providers.

On the second day of the workshop representatives of the PID Service Providers DANS, SND and GESIS da|ra introduced their institutions and services and explained particular challenges as a service provider for PIDs as well as the benefits and possible improvements of the PID systems (DOI and URN) they offer. All representatives had the chance to give feedback, ask questions and share their concerns in discussion rounds which took place after each session. The main issues which came up focused on technical and organisational challenges, such as insufficient technical support and the integration of the PID assignment to the existing workflows, or on the level of granularity and how to deal with versioning. Further points raised were the provision of access points to data through landing pages that contain metadata as well as the handling of duplicates.

There was a common understanding that most of the discussed technical and organisational issues are related to general principles of data handling (within CESSDA), therefore the participants expressed the need for a general data policy within CESSDA.

The main outcome of the workshop was to concretely set the ground for the construction of a CESSDA PID Policy Framework. It was thus agreed that the PID policy would consist of three different parts: general principles, implementation guidelines and best practice examples. It also became clear that a CESSDA PID policy would not be enough on its own and that it would need to be embedded within a wider framework of policies (for example, a "data policy for CESSDA"). Furthermore, the aim of a common PID policy was defined as: to introduce the use of PIDs in all CESSDA Service Providers in order to be able to locate, reference, identify and cite data in a non-ambiguous way. In order to achieve this aim the CESSDA PID Task Force consisting of GESIS (lead), DANS and SND was established.

An article by Brigitte Hausstein, Kerrin Borschewski and Anna Münch from GESIS, edited by Eleanor Smith from CESSDA, 06 April 2016.