

About the Social Science Data Archive in Slovenia by Janez Štebe, Head of Arhiv družboslovnih podatkov / Social Sciences Data Archives



Q: Depositing and reusing data

Social Science Data Archives (Arhiv družboslovnih podatkov, ADP) are a part of the University of Ljubljana, the Faculty of Social Sciences, which is the leading research and teaching institution in the social sciences field in Slovenia. Yet, we are not an institutional repository as such. Instead we position ourselves as the national disciplinary repository open to any scientifically important research data in our topical range. The main criteria for accepting research data into ADP is its research potential for secondary analysis (demonstrated high methodological quality and conceptual design that makes it interesting for the users). We require good quality of documentation that contains the topical and methodological description of research project that the data springs out of. We are also open to accept some commercial surveys data, and work closely with the Statistical Office of the Republic of Slovenia to facilitate the research and educational use of micro data in their possession. In the future we wish to include more of the data from qualitative style of social sciences research [[1](#)] (e.g. anonymised interviews transcribes), as well as data from some of the disciplines that are less represented, such as economy and public health. We are members of international data service organisation CESSDA, which itself is in the process of becoming a new formal legal entity under the ESFRI Roadmap.

Q: The clear need for a service

Indeed what we do are service activities of adding value and, in case of digital preservation activities, of sustaining value for the future generations, of data we accept in our custody. We see large part of our service activities also in relation to our users. These are of two types: data creators who deposit their work, and data users who further analyse the data. Both constituencies need active support.

Q: How do you contact researchers and encourage them to deposit and share data?

We are still largely dependent on good will and altruism of the researchers who decided to share their research data sooner or later after the project is finished. Periodically we trace the on-going projects funded by the Slovenian Research Agency [[2](#)] and send invitations to PIs to consider depositing the data. We also react to users' recommendations or queries about the particular research data instance, and – not very systematically – follow research projects public presentations and journal publications for the underlying data. There is still much to be done in this respect! We are really anxious to see first the articulation and then the implementation of decisive journals and funders' policies that will contain as a mandate for depositing data into well-established disciplinary data archive or similar places to make them accessible. This will be a stick. Of course, we encourage depositors with the set of positive stimuli too, a carrot. We explain all the benefits for the depositor if the data are stored and made accessible by us. Nonetheless, there are occasions when someone loses his own copy of the data after a few years, and seek if it's available at the data archive. The motivation for active researchers is their reputation that builds up, when the deposited research data under their authorship are reused and cited in scientific reports of secondary users. Just as it is the case of citing other scientific reports in one's own publication. For that sake, a recent innovative building of robust persistent identifiers infrastructure (URNs or DOIs or similar) and recommended citation format for research data that includes relevant identifiers is going to help tracing the publications using a particular data.

We try to understand the hesitation the authors of the data sets have and adapt to their needs. It's legitimate to claim a certain reasonable period of exclusive use of the data for their own research purposes, before the data are finally openly available. We also offer a full support for data creators, giving them advice and tools and services, making it easier for them to prepare archive ready data. It is of critical importance that planning for and preparation of data and documentation is on-going in all phases of research. In our opinion the recent innovative research founders' policy approach of requiring the data management plans at the research project application is of immense value in this respect.

Q: How do you provide user support and training?

It is also critical to us! We would like to equip more users with the skills necessary for realising a full analytic potential that lies in data. It is not only students and lay public, there are a lot of academics that are not well trained in data and statistical literacy, needed for sometimes complex analytical approaches that are possible with the data available. The data are still underused just because of low analytical skill level.

Here we see our role in collaboration with others as we are relatively small. We have prepared guidance and manuals for navigating through our pages, about search, registration, download and online analysis tools. We offer occasional workshops [[3](#)] for users and depositors. Still the bulk of resources that are relevant here are available in our partners organisations around the world, so what we do is to try to follow what is new in the environment and to inform our users, where on the web to find relevant guidance, user group, or training material for specific problem or type of data they are interested in. There are great summer schools organized, such as ECPSR [[7](#)] f. e., that we actively promote.

Some technical details

We rely on DDI [[4](#)] standard for data and study description that includes a native format for bibliographic information, and relevant sections of method description, and of details of a data format. It's also an integrator of all related documents that accompany a data set, typically this are original survey questionnaires. The access is through web, written in Django environment that uses the study package, integrated through DDI xml contained information. Parallel to this is NESSTAR, the application that enables documentation viewing, online analysis of raw data, or download facility for registered users.

Q: How do you handle the diversity of data types, formats, technical standards and so on?

We are lucky in a sense that a majority of digital objects are of similar type, mainly these are simple data matrix, and related documents that we can convert to uniform textual/graphical format, e.g. PDF/A for preservation and access. We require from depositors to prepare the materials in one of the ADP preferred format. We will encounter more of the diversity of other media when other, non-quantitative studies, will start to come in mass.

Q: How are the data licensed?

The depositor signs the agreement granting reuse of data and related materials by Creative Commons licences CC BY by default, or CC BY NC on request. We didn't encounter any problems with applying that reasonably simple licence model to research data. One can opt for an embargo period that we allow, to encourage early submissions.

Q: Popularity of Social Science Data Archives

Our most frequent users are students (pre - and post graduates), following by researchers. For this reason usage of our services is highest in the spring months when student are finishing their papers. Users' statistics are as follows. We have between 1000 and 2000 monthly visitors to our web page, and we give out between 50 and 100 new usernames to access NESSTAR. We distribute all of our materials via Internet but users can come also to our office and receive help. Even if this numbers are not so low, we still see the data as underutilised and try to attract more usage, as explained earlier.

Q: You mention on the website that data are often used as background materials for teaching and practical exercises with students and in some cases even specially adapted modules for teaching are available. Can you please tell more about this?

This again is something that we use inspiration from great examples in our partners data archives [[6](#)]. We actively promote the "re-use" of teaching modules with real data examples that are created elsewhere.

Locally we are often invited to deliver a lecture for various social sciences study programs, so that students can use the data for their assignments easily. And yes, there are few modules of the undergraduate introductory methodology course, using data from ADP and other sources, and for political science postgraduate method and multivariate statistics course concentrated on CSES data analysis.

Q: You mention on the website that archived data are regarded as equivalent to scientific publications of Slovenian Research Agency criteria (Rules about indicators and criteria of scientific and professional efficiency, Article 5, paragraph 3. E.). Can you please tell more about this?

This area is still very new. In this respect we are having a role of publisher, which is more and more, as we perceive ourselves, the data publisher. When a data creator goes through a selection process and fulfils all the criteria of methodological and scientific relevance of the data set, properly documented, and when we publish the data in our catalogue, this then qualifies to enter into bibliography of a researcher under the category mentioned. Bibliographies of researchers are under the auspices of the Slovenian Research Agency (the supervision is carried out by disciplinary Central Specialised Information Centres). Official bibliographic lists and scores attached are available in bibliographic information system, either following the researchers ID in Slovenian CRIS (SICRIS), or in the Slovenian union catalogue (COBIB.SI).

A concern might be that the formal 'peer reviews' are not in place. Here we look for models that would satisfy the requirements. When data accompanies the 'peer reviewed' publication, the decision about 'scientific' relevance is easier. Otherwise we know that the same criteria as for traditional scientific publication could not be applied to data, as data can be used for different purposes and it's not necessary that only the most excellent data is useful. Sometimes there are certain historical or geographic location related data that are unique, and standards about sampling and methodological quality might be less stringent.

All in all, this is the carrot we mentioned before, and we rarely reject the data offered by the researcher, willing to go through the necessary steps in formal preparation of deposit package. The effect of the measure is visible when the dates of the next research project calls approach and some researchers are eager to get through the study processing faster to obtain additional scores for the next project evaluation. Though still we would suggest that measure to be accompanied with the funders mandate to consider archiving and prepare the Data Management Plan in project application. This is then the project obligation and the data publication record is yet an additional reward for following the mandate.

Q: What is the relationship of data archiving to green and gold Open Access for journal articles?

Our main concern regarding journal publishers' policies is that these include, as a formal rule,

the obligation that data related to publication is publicly available through a disciplinary data archive or similar reliable place. We believe that there are a lot of advantages if such disciplinary data centres are well established, especially in the aspect of accumulating specialist professional knowledge of data scientists, assisting and supporting research data management in all phases of data life cycle, which helps to build up a professional profile in the future [5].

Journal editors and reviewers might monitor more closely the appropriate use of citation format for research data to be included in the references section. Finally, when these two conditions are satisfied, it is seamless Open Access from data to journal articles and back that gradually makes the scientific transparency and cumulative character of knowledge more feasible.

Q: How do you think a data archive should measure success?

Users' needs and satisfaction are the main point in any service delivery. However, there are caveats to this, as we need to guess the needs of future users, sometimes this future could be distant. In the long run, topics, theories and problems that are popular today, will change. And as we mentioned, users sometimes do not know what could be done with the data we offer. That's why yes, it's the efficiency of the job we are doing, but measured more broadly in the environment of open scientific inquiry. Therefore the ultimate measure of success is new and more relevant discovery that becomes possible by open use of rare data resources inside scientific community. Here is the role for funders to realise that potential and facilitate its realisation by supporting the intermediate infrastructure information services.

Q: You launched [a community wide discussion about open science](#) – can you please tell more about this?

Data archives were in existence long before the advent of open access movement. Some have already celebrated their 50th anniversary (ICPSR). Yet we find ourselves in the natural coalition with the open access advocates and activists, as we share the same principles and objectives. There is a need for a cultural revolution in minds and habits of some members of the scientific community to 'buy' the ideals of open access. The problem is that current research rewards and the esteem system based on only the traditional publishing model is not a stimulus for that to occur. The reality is that the spontaneous response to our invitation to share did not happen. It's not enough to open space for discussion. We need to put more effort to actively engage with researchers. We started with a formal project recently, which aims to deliver an action plan for the Slovenian Ministry of Science. Inside this project we are trying to facilitate information

exchange about needs and responsibilities among stakeholders in open research data access for all disciplines. It is based on our experiences, but with the awareness that many questions and solutions are specific to a particular type of data and a research discipline. A part of these efforts is also introducing the OpenAccess.Si information portal, where we continue to update the section about research data sharing among others in collaboration with research librarians.

REFERENCES:

- [1] <http://www.iassistdata.org/iq/issue/34/3> [2] <http://sicris.izum.si/> [3] http://adp.fdv.uni-lj.si/za_uporabnike/delavnice2011/
- [4] <http://www.ddialliance.org/>
- [5] <http://www.jisc.ac.uk/publications/reports/2008/dataskillscareersfinalreport.aspx>
- [6] <http://www.icpsr.umich.edu/icpsrweb/ICPSR/training/index.jsp>
- [7] http://new.ecprnet.eu/methodschools/2012_Ljubljana/CourseList.aspx