Open data in Slovenia:

An assessment of accountability among stakeholders

IASSIST’12, June, Washington, DC, USA
National Data Landscapes: Policies, Strategies, and Contrasts
Session-ε, Thu June 07, 15:30-17:30, Location 413-414

Janez Stebe
Content of presentation

• OPEN DATA project description
• Findings and starting points from previous studies, study design
• Results:
  – data planning and creation issues
  – data curation/ documentation
  – access/ data sharing culture in general
• Conclusion
Project description: **Open data project ID-card**

- **Title:** "Open Data – action plan delivery for the open access system of research data financed from public resources in Slovenia"
- **ID number of project:** „CRP with ID V5-1018“
- **Budget:** 60 000 EUR
- **Duration:** 2011-2013
- **Explicit requirement:** to pave ground for **OECD declaration** fulfillment
Project description: **Main goals**

1. **State of the art assessment**
   - Identification of
     - current roles holders,
     - existing data types produced and managed,
     - and nascent data services activities or needs
   - On-going or potential cooperation (national and international)

2. **Mobilization**

3. **Planning**
   - *Empowering*: to build on current emerging services and expertise, existing collaborations and follow/exchange/adapt good international practices
   - *Achieving* consensus among stakeholders regarding strategic goals and their implementation
Phase 1 of the project: An assessment of current situation

Metodology:

22 semi-structured interviews from all disciplinary fields

- founders
- leaders of research institutions
- data creators and users
- data service professionals

Social Science Data Archives
Findings and starting points from previous studies

Expectations:

• to find similar results about the workflows and research data sharing culture as elsewhere

Still, our general intention was:

• to emerge into the field after openly discussing about the research data access with prominent researchers, librarians and institutions’ leaders
Results: Data creation phase

• **High awareness** of data quality criteria in creation phase
  – Realistic immediate quality assessment – it is seen as a compromise in given circumstances
  – Reuse potential is not always obvious (many different data types)

• Awareness that less input in higher quality of data/documentation at creation results in more work/lost information in next phases of Data Curation Lifecycle
"Well, data are so and so, but everything is included in because bad data also say something. If data are missing you are repeatedly trying to discover and reveal why they haven’t been included. Better bad data than no data.” (Miran Hladnik, Slovene literature)

"Some formal criteria do exist, of course. In my opinion formal criteria should contain the level of comparability with metadata requirement.‘, (Head of the research institute’s library)

"There is some general consensus about standards in our scientific field. For example, there is an agreement about what is a clean protein or how to conduct a good quality experiment.‘ (Senior Biologist)

... "documentation was prepared and organised for the scientific analysis, that is why it gets a bit more difficult, some people don’t understand, they say: „You have a certain number of tapes, all you have to do is to digitalize them and put them on web“. Then we say: „No, it is not as easy as it seems“ (Audio archive researcher)
Results: Data types, processing and preservation

What is data?

Standards for data transformations and metadata

Good practices already exist: genomics, meteorology, audio materials, ...

International repositories

Contribution of research librarians?

MOTIVATION  KNOWLEDGE  FINANCING
Detected problems

- There is **no common policy** or rules at the institutional level
- Researchers receive **little incentives** for preparing additional metadata
- **Funding** of long-term data management is **not provided** (short term project work prevails)

Yet

- High demand for tools and standards, guidance and specialized research data management support services
“Such results have very little, or even no value if not properly commented on.” (Sciences Institute Director’s Adviser).

“I think we do not have a unified rule as an institute about what or how every researcher should keep or document his original data. I would say that this is more a matter of methodology that the researcher has got accustomed to during his collaboration with the mentor and the preparation of his doctoral degree and then he probably keeps that in the future. (…). As an institute we do not have unified rules and would also create them with difficulty” (Sciences Institute Director’s Adviser).

“I think data should be maintained by those institutions where the data are produced (…). Of course, we do not always get the project that would provide the updating of some of the data we were working on ten years ago.” (Humanities Institute Director’s Adviser).
Results: Access, use & reuse

- "data owners" – specific culture of data sharing (jealousy, competitiveness, personal engagement)
- low prestige of data related activities
- professionals with specialized data curation knowledge are rare
- unbalanced awareness of data related services landscape – mentoring and guidance potential

Good practices already exist:

- ARKAS
- DEDI
- SISTORY
- EVA
Citations from the interviews

**Needs and priorities**

“As I see at the institute, some of the required preservation or preparation of data for general availability would be considered by our researchers as an extra obligation and additional administration unless they are given additional resources, maybe one month, that would also be paid for to prepare the data in a form and to make them potentially more widely accessible” (Sciences Institute Director’s Adviser).

“Often the general atmosphere is that this is mine and that someone else will find something in the data that I have not seen. It isn’t necessarily the case that this is fine with everyone.” (Andrej Blejec, Biologist).

“Nowhere is this seen as an honourable job, it is slave labour (...)” (Miran Hladnik, Slovene Literature).

**Low prestige of data related activities**

“But if we are to stand behind it, we need to ensure the quality of the data. The data are somewhere in some colleagues’ drawers but, if I do not have enough money to pay them for their work, I cannot require them to properly prepare this material.” (Humanities Institute Director’s Adviser).

**Culture of data owners**

Social Science Data Archives
Conclusion: **Topics for further discussions**

- **Specialists service support / common service**
  - Quality assessment/ specific content and procedures description vs. formal digital preservation criteria/services

- **Ensure researchers „buy-in“:**
  - Incentives/ requirements (Data Cite, OECD principles)
  - Data Management Planning requirements: additional budget
  - Additional administration vs. substantial benefits of high quality research output accessible for a whole scientific community

- **National consensus on priorities** in further data infrastructure services development (ESFRI etc.)

- **Cooperative further development** of nascent disciplinary data centers/ respect for different data types (big/small sciences etc.)