FORSbase: Implementing the FAIR principles for social science data archives

IASSIST 2018

May 30, 2018, Montreal
Stefan Buerli
Outline

• What is FORS?
• What is FORSbase and what is its current state?
• Open Data and FAIR data
FORS

FORS is a national centre of expertise in the social sciences. Its primary activities consist of:

• production of survey data, including national and international surveys;
• preservation and dissemination of data for use in secondary analysis;
• research in empirical social sciences, with focus on survey methodology;
• consulting services for researchers in Switzerland and abroad.

FORS collaborates with researchers and research institutes in the social sciences in Switzerland and internationally.
DARIS FORS:
- Research inventory (10,900 project descriptions, 18,800 researchers, 4,500 institutions, 4,760 active users)
- Data catalogue (655 datasets)

Needs:
- Services covering the entire life cycle of research data
- Simplified archiving and publishing
- Simplified access
- Promotion of secondary analysis
- Renewal of the infrastructure
- Integrated workspace for collaborators and researchers
FORSbase Strategic Vision
Technical View

FORS Web

- Inventory & Archive
- Survey Tool (ex SHP)
- Catalogue Tool (Compass)

Common Functions

- Archive Metadata
- Survey Metadata
- ... Metadata
- BFS Metadata

Primary Research Data

International site

Tools

Metadata

Primary Data
Schedule

Version 1.0 Research Inventory
Summer 2013

Version 2.0 Data Archive
February 2016

Version 3.0 Survey Tool
2018
FORSbase characteristics

- A work environment for social scientists and FORS collaborators
- Support for the entire data life cycle
- Integrated data storage for primary data and metadata
- Data preparation and exchange via standards (DDI etc.)
- OAIS compliant data archive
Big Picture Inventory & Archive

Users
- Researcher/Producer
- FORS Collaborator
- Researcher/Consumer

Processes
- Study Elaboration
- Submission
- Quality Assurance
- Publication/Archive
- Search, Browse
- Download

Environment
- Personal Workspace
- FORS Workspace
- Unified Search, Browse, Filter, Access

Global Functions
Communication, Messaging etc.

Modules
- Workspace
- FORS Archive: Ingest, Manage, Access
- Repository
- Compass
- Project Inventory
- NESSTAR Adaptor
- Future Module
Big Picture Inventory & Archive

Users
- Researcher/Producer
- FORS Collaborator
- Researcher/Consumer

Processes
- Study Elaboration
- Submission
- Quality Assurance
- Publication/Archive
- Search, Browse
- Download
- Future Process

Environment
- Personal Workspace
- FORS Workspace
- Unified Search, Browse, Filter, Access
- Future Module

Global Functions
- Communication, Messaging etc.

Modules
- Workspace
- FORS Archive
  - Ingest
  - Manage
  - Access
- Compass
- Project Inventory
- NESSTAR Adaptor
- Future Module
Big Picture Inventory & Archive

Users
- Researcher/Producer
- FORS Collaborator
- Researcher/Consumer

Processes
- Study Elaboration
- Submission
- Quality Assurance
- Publication/Archive
- Search, Browse
- Download
- Future Process

Environment
- Personal Workspace
- FORS Workspace
- Future Module
- Unified Search, Browse, Filter, Access

Global Functions
- Communication, Messaging etc.

Modules
- Workspace
- FORS Archive
  - Ingest
  - Manage
  - Access
- Compass
- Project Inventory
- NESSTAR Adaptor
- Future Module
- Repository
Big Picture Inventory & Archive

**Users**
- Researcher/Producer
- FORS Collaborator
- Researcher/Consumer

**Processes**
- Study Elaboration
- Submission
- Quality Assurance
- Publication/Archive
- Search, Browse
- Download

**Future Process**

**Environment**
- Personal Workspace
- FORS Workspace
- Unified Search, Browse, Filter, Access

**Future Module**

**Global Functions**
- Communication, Messaging etc.

**Modules**
- Workspace
- FORS Archive
- Compass
- Project Inventory
- NESSTAR Adaptor

**Repository**
- Ingest
- Manage
- Access
Big Picture Inventory & Archive

Users
- Researcher/Producer
- FORS Collaborator
- Researcher/Consumer

Processes
- Study Elaboration
- Submission
- Quality Assurance
- Publication/Archive
- Search, Browse
- Download
- Future Process

Environment
- Personal Workspace
- FORS Workspace
- Unified Search, Browse, Filter, Access
- Future Module

Global Functions
- Communication, Messaging etc.

Modules
- Workspace
- FORS Archive
  - Ingest
  - Manage
  - Access
- Repository
- Compass
- Project Inventory
- NESSTAR Adaptor
- Future Module
FORSbase 2.0:

- Online access to all data in our catalogue
- Online data deposit
- Personal messaging system
- Single login
FORSbase 2.0:

Online access to all data in our catalogue:

- One Stop Shop
- Accessing datasets in our catalogue only takes a few mouse clicks
- A dedicated place in your “workspace” allows you to manage all of your data acquisitions
FORSbase 2.0:

Online data deposit:

- Deposit and share your data easily online, at any time, and at your own rhythm!
- Upload your data files and documentation
- Feedback to the researcher on usage of their deposited data
FORSbase 2.0:

Personal messaging system:

- Get connected with other researchers
- Easy communication between FORS collaborators and FORSbase users
- Automated alerts/requests concerning downloads of datasets
FORSbase 2.0:

Single login:

- a single FORSbase account to access our different services
- FORS Nesstar integration
- Easier handling of all data users
- Improved analytics of the FORS data user community
DEMO

https://forsbase.unil.ch/
Open Data:

a great principle
impossible to implement as such in the social sciences
(due to data protection laws)

FORS Open Data endeavors:
Open Research Data Platform Switzerland
Open Data Hackathon
FAIR Data:
Made-to-measure solution
to satisfy the data producer, the data user, and data protection laws
In the social sciences **appropriate infrastructure** is needed to mediate the specific needs of data producers and users in relation to FAIR principles and data protection laws.

FORSbase is an online web application that makes data discovery, deposit, and delivery as convenient and open as possible for data in the social sciences

- from the beginning of the data life cycle to the actual sharing of the data
- made-to-measure options regarding access to the data
- different degrees of openness regarding delivery
FORSbase: FAIR principles

- **FINDABLE:**
  - Each dataset has its own DOI
  - Rich metadata freely available

- **ACCESSIBLE:**
  - retrievable by DOI, study and dataset metadata freely accessible
  - Most data immediately accessible with an email address from an academic institution

- **INTEROPERABLE:**
  - OAIS compliant; DDI standard

- **RE-USABLE:**
  - Clear usage license with possibility of customizing access conditions
FAIR principles: Findable

- Each dataset is assigned a globally unique and persistent identifier (DOI)
- The data are described with rich metadata on the study level and the dataset level
- Some metadata are based on or compatible with international standards (DDI, CESSDA Metadata Standard, Dataverse)
- The dataset’s identifier is explicitly mentioned in the metadata
- The catalog is in the process of being indexed in order to be readable by mainstream search engines
FAIR principles: Accessible

- retrievable by DOI, the study and dataset metadata are freely accessible.
- most longitudinal data are immediately accessible with an email address from an academic institution on our FORS Nesstar server.
- access to data requires a validated user account and a user agreement that is specific for the requested dataset.
- the study metadata is persistent.
FAIR principles: Interoperable

- FORSbase is OAIS compliant.
- The framework to describe and structure data and metadata is based on the DDI standard.
- The actual storage of the data is Fedora Commons, a modular open source repository system, both for access and preservation.
- The DDI standard (DDI Lifecycle 3.2 and DDI Codebook 2.5) used to describe the datasets are freely accessible (https://www.ddialliance.org)
FAIR principles: Re-usable

- Rich metadata related to the nature and the provenance of the data
- Clear usage license for the users
- The possibility of customizing access conditions for the depositor
  - Date of embargo
  - Consent by the data producer before download
## Dataset Details

**Dataset Title**: Meinungsbild zur Sicherheits und Verteidigungspolitik der Schweiz - 2004

**DOI**: 10.23662/FORS-DS-430-1

**Ref dataset**: 430

**Ref study**: 8024

**Dataset language**: German

**Additional information**: Dieser Datensatz ist Teil einer jährlich stattfindenden Erhebung. Die Datensätze seit 1995 sind bei FORS archiviert.

**Publication date**: 07.02.2007

**Isoscat link**: ch.sidos.ddi.430.8024


**Data type**: Quantitative data

**Consent of respondents**: No

**Time method**: Cross-section

**Media**: Digitalized data file

**Data collector**: FORS documentation

**Available document types**: FORS documentation, Other
<table>
<thead>
<tr>
<th><strong>Time method</strong></th>
<th>Cross-section</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Media</strong></td>
<td>• Digitalized data file</td>
</tr>
<tr>
<td><strong>Data collector</strong></td>
<td>FORS documentation</td>
</tr>
<tr>
<td><strong>Available document types</strong></td>
<td>• Other</td>
</tr>
<tr>
<td><strong>Remarks about the documentation</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Restriction</strong></td>
<td>None</td>
</tr>
<tr>
<td><strong>Special order</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Special permission</strong></td>
<td>None</td>
</tr>
<tr>
<td><strong>Status</strong></td>
<td>Published</td>
</tr>
</tbody>
</table>

**Datenfile**

| **Collection date - begin** | 07.01.2004 |
| **Collection date - end** | 09.02.2004 |
| **Analysis unit** | • Individual |
| **Universe** | Stimmberechtigte Schweizer:innen (der deutschen, französischen und italienischen Landesteile) |
| **Sampling description** | Random-Quote (Zufallsschluss von Rufnummern) |
| **Mode of data collection** | • Telephone interview (CATI, etc.) |
| **Data collection instruments** | • Questionnaire |

**Data source**

**Weighting information**


| **Response rate** | 1200 Personen |
| **Data completeness** | |
| **Number of cases** | 1200 |
| **Number of variables** | 142 |
The future: SWISSUbase

SWISSUbase
The future: SWISSUbase

- Separation of the core part from the domain-specific component
- Generic core model (can be used as a OD repository)
- Different domain-specific metadata schemas to plug into on all levels, i.e. study metadata, dataset metadata, contracts and licensing, and researcher metadata
- Plug-ins will allow for the capture and implementation of disciplinary specificities in such a way that each institution using the tool will have an adaptable and customized solution
Any questions?