



1

Introduction to INEXDA's Metadata Schema

Christian Hirsch
(Deutsche Bundesbank)
July 2018

The views expressed here do not necessarily reflect the opinion of the Deutsche Bundesbank, the INEXDA network, or the Eurosystem.

Motivation

2

- Metadata is data about data, i.e. a description of data
- Examples of existing metadata schema
 - Statistical Data and Metadata Exchange (SDMX)
 - Data Documentation Initiative (DDI)
 - da|ra Metadata Schema
- Important is the purpose
 1. Overview of available and possible comparable granular datasets
 2. Help data users discover and use datasets appropriate for research and analysis by using harmonised metadata.
 3. Provide framework to facilitate a possible harmonisation of datasets in the (near) future.

INEXDA Metadata Database

3

Help Contact Search all shared datasets Logout (data)

INEXDA The Granular Data Network

Searchresult

Search metadata

your search: All fields: (investment)
Active filters [x clear all]: * Keywords: Banks_depository_institutions * Publication Date: 2017

results: 2
Your search returned 2 results in all shared INEXDA datasets.

Monthly Balance Sheet Statistics
Deutsche Bundesbank, Statistics Department (S), Monetary and Financial Statistics (S 1), Monthly balance sheet statistics (S 10)
published 2017-10-10, Version 1
On-site
view metadata in: English
Status: in-house

Statistics of the banks' profit and loss accounts
Deutsche Bundesbank, Statistics Department (S), Monetary and financial statistics (S 1), Cross-sectional tasks and secondary statistics (S 14), Secondary statistics (S 14-3)
published 2017-07-25, Version 1
On-site
view metadata in: English
Status: in-house

Number of (0) datasets: 2

Resource Types
Dataset (2)

Data provider
Deutsche Bundesbank (2)

Creator
Deutsche Bundesbank, Statistics Department (S), Monetary and Financial Statistics (S 1), Monthly balance sheet statistics (S 10), (1)
Deutsche Bundesbank, Statistics Department (S), Monetary and financial statistics (S 1), Cross-sectional tasks and secondary statistics (S 14), Secondary statistics (S 14-3) (1)

Geographic Coverage
Germany (2)

Keywords
Balance sheet (2)
* Banks_depository_institutions

Availability
On-site (2)

Publication Date
* 2017 (2)

- The database serves
 1. as an information tool for INEXDA members, researchers and analysts
 2. as the basis for the harmonisation activities of INEXDA (e.g. item relation)
- Current contributions to the database come from
 - Deutsche Bundesbank 30 (12)
 - European Central Bank (6)
 - Banca d'Italia 26 (6)
 - Banco de España 21 (2)
 - Banco de Portugal 16 (2)

Part 1: Basic Information

4

1	Resource Type
2	Resource Identifier
3	Name of Dataset
4	Creator
5	DOI Proposal
6	URL
7	Language of Resource
8	Publication Date
9	Availability
10	Sampled Universe
11	Sampling
12	Temporal Coverage
13	Time Dimension
14	Collection Mode
15	Unit Descriptions
16	Descriptions
17	Geographical Coverage
18	Keywords
19	Alternative Identifiers
20	Relations
21	Publications

- *Creator* is a mandatory item in da|ra. May be used to provide more granular information on the data compiler
- *URL* refers to the webpage which displays information about the dataset
- *Availability (controlled)* describes the procedure under which the data can be accessed (eg download or on-site)
- *DOI Proposal* provides the suggested DOI name of the dataset. A Digital Object Identifier (DOI) is a permanent, persistent identifier used for citing and tracking datasets

Part 2: Methods

5

1	Resource Type
2	Resource Identifier
3	Name of Dataset
4	Creator
5	DOI Proposal
6	URL
7	Language of Resource
8	Publication Date
9	Availability
10	Sampled Universe
11	Sampling
12	Temporal Coverage
13	Time Dimension
14	Collection Mode
15	Unit Descriptions
16	Descriptions
17	Geographical Coverage
18	Keywords
19	Alternative Identifiers
20	Relations
21	Publications

- *Sampling* displays the type of sample design used to select the observations to present the population
- *Time Dimension* provides information on
 - frequency of observations.
 - whether dataset structure is panel, time-series or cross-sectional
- Structural breaks are defined as major events and revisions that have impacted the dataset
- Examples of structural breaks include:
 - changes to the time frequency with which data is collected
 - changes to the set of collected variables
 - changes in the population or sampling

Part 3: Descriptions

6

1	Resource Type
2	Resource Identifier
3	Name of Dataset
4	Creator
5	DOI Proposal
6	URL
7	Language of Resource
8	Publication Date
9	Availability
10	Sampled Universe
11	Sampling
12	Temporal Coverage
13	Time Dimension
14	Collection Mode
15	Unit Descriptions
16	Descriptions
17	Geographical Coverage
18	Keywords
19	Alternative Identifiers
20	Relations
21	Publications

- *Unit Description* provides information on the entities that are being observed in the dataset
-
- Datasets may contain more than one unit of observation. For example, in a credit register information on the following units are collected:
 - Banks
 - Companies
 - Governments
 - Loans
- *Descriptions* also contains detailed information on structural breaks in the dataset

Part 4: Relations and Publications

7

1	Resource Type
2	Resource Identifier
3	Name of Dataset
4	Creator
5	DOI Proposal
6	URL
7	Language of Resource
8	Publication Date
9	Availability
10	Sampled Universe
11	Sampling
12	Temporal Coverage
13	Time Dimension
14	Collection Mode
15	Unit Descriptions
16	Descriptions
17	Geographical Coverage
18	Keywords
19	Alternative Identifiers
20	Relations
21	Publications

- Describes relations between datasets and databases in the INEXDA metadata database...
 - ... in a given country
 - ... across countries
- Examples of use cases in INEXDA context
 1. Relation between datasets containing similar units (in different countries).
 2. Dataset feeds into a ECB dataset.
- *Publications* provides information on scientific publications related to the dataset.

Conclusion

8

- INEXDA's metadata schema is based on the GESIS DOI registration service da|ra (GESIS is cooperating with DataCite)
<https://www.da-ra.de/en/home>
- Purpose is to foster harmonisation between INEXDA members and broaden metadata sharing within INEXDA and possibly outside
- Name of metadata items closely follows da|ra conventions to enable seamless DOI registration, if desired later in the project
- Using DOI for datasets may help
 - ... facilitate datasets citations
 - ... foster reproducibility of results