



IAB-ITM

Finding the Right Tags in DDI 3.0: A Beginner's Experience

IASSIST 2009, 27th-30th Mai 2009

Topics



- Data of the IAB
- DDI 3.0 and IAB Documentation
- Features of DDI 3.0 Used
- Where we had to use notes
- Summary

Administrative Data in the Federal Employment Agency

Two different sources:

- notification procedure of the Social Security System (health-, pension-, unemployment- and long-term care insurance)
- administrative data of the Federal Employment Agency

Notification procedure of the Social Security System

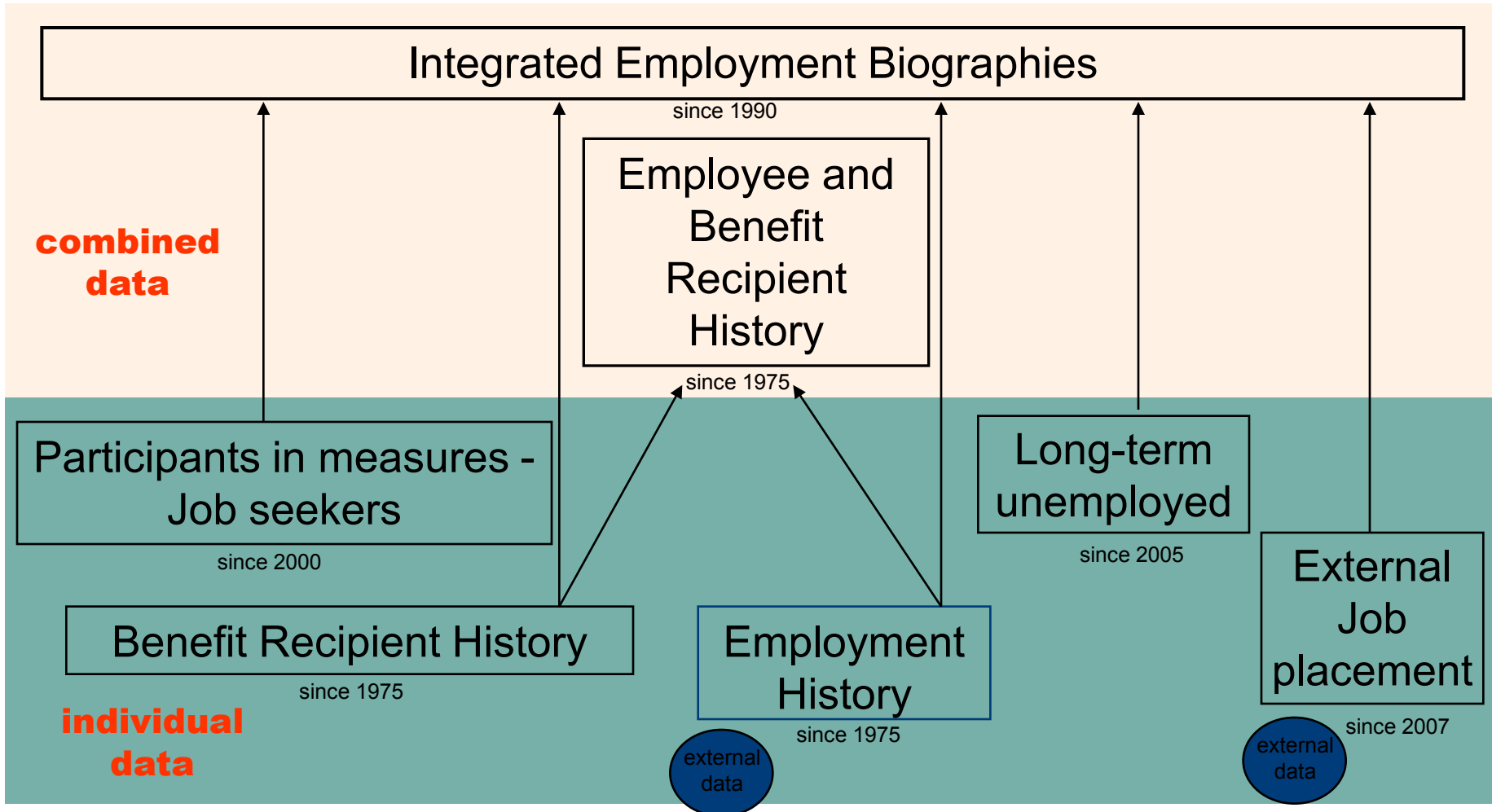
Intention of data collection:

- calculation of social security contribution and entitlement to benefits
- generation of statistics

Scope and coverage:

- every employee and marginal worker covered by social security system (notification requirement)
- identification: social security number and establishment number
- annual, begin or end of employment, employment interruption, change of health insurance, period of exclusion from benefit

Administrative data: Individual data - Combined Data



DDI 3.0 and IAB Documentation

A database has been designed to hold all of the IAB documentation

- Original documentation was stored in Word documents
- Documentation was fairly well-structured and consistent

The database has been mapped to DDI 3.0 and will allow for import and export of DDI 3.0 XML

The DDI 3.0 will be used to present documentation in many different ways in the new Knowledge management system

Process for Analysis

The most complex data set (IEB) was chosen as a subject for analysis

- The documentation was used for the analysis
- This was mapped against the DDI 3.0

To make sure it was complete

- The documentation database was also mapped against the DDI 3.0

Features of DDI 3.0 Used

Studies are “simple” in DDI 3.0 terms

- They do not use grouping or comparison features
- Each data set is documented individually
- The data sets are very large however, with many variables and lots of complexity

StudyUnit module

is used for abstracts and descriptive material formatted in XHTML

DataCollection module

was used to describe methodology (sampling) and data cleaning process

LogicalProduct module

is used to capture categories, codelists, and variables

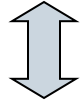
Archive module

is used to record information about the individuals, organizations, and their roles, as well as some storage information for files

Notes were widely used

Issue: Register vs. Survey Data

- DDI 3.0 is well-suited to describe survey data



- It is not as well suited for register data

- Documentation about *using* the data did not have a natural place in DDI 3.0

Which variables can be used together, etc.

- Documentation about *data quality* did not have a natural place in DDI 3.0
- This documentation went into Notes



For survey data, this documentation isn't needed (researchers can look at the collection instrument, etc.)

Issue: Custom Data Sets

The data sets used by researchers at IAB are processed out of larger data files

- The structure and location of data within the large files is held in a database system which can be queried when customized sets are assembled

The DDI requires logical and physical data products, and one or more physical instances

- These could be created when data sets are assembled, but not before

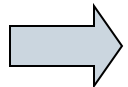
Much of the DDI 3.0 structure is useful to archives, to document static data sets

- Not as useful to organizations like the IAB which focus on research
- These modules were not used

Issue: Codes in Integrated Data Sets

In the IEB, variables from different sources may contain duplicate codes with different meanings

- You need to know the source as well as the code to know what it means
- This is not technically allowed in DDI 3.0
- It is not possible to re-code the source variables

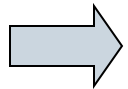


The solution – do nothing! It was possible to represent the duplicate codes in DDI 3.0, even though this is not actually supported

Issue: Security Fields in Database

The IAB maintains some critical security information about their data sets

- This is used when displaying the data from the XML
- There was no place in DDI 3.0 to store this information



Placed in Notes

Making Decisions

Sometimes it was difficult to decide where the metadata fit into DDI 3.0

- We decided to only use fields that were a good match
- Other metadata went into Notes
- Within some fields, the XHTML <div> element was used to describe specific sections
- All formatting uses XHTML

Summary

- All DDI issues have been reported to the DDI Alliance TIC Committee



This may improve the standard for use with register/administrative data in the future

- It was possible to capture all needed metadata in the DDI 3.0
- Had to use Notes in many places, however



contact information:

Claudia Lehnert

IT Services and Information Management

mail: claudia.lehnert@iab.de

Phone: +49 911 1793675

ITM 28.05.2009