

The IAB metadata portal: Consolidating different data sources in one application

EDDI 2014

London

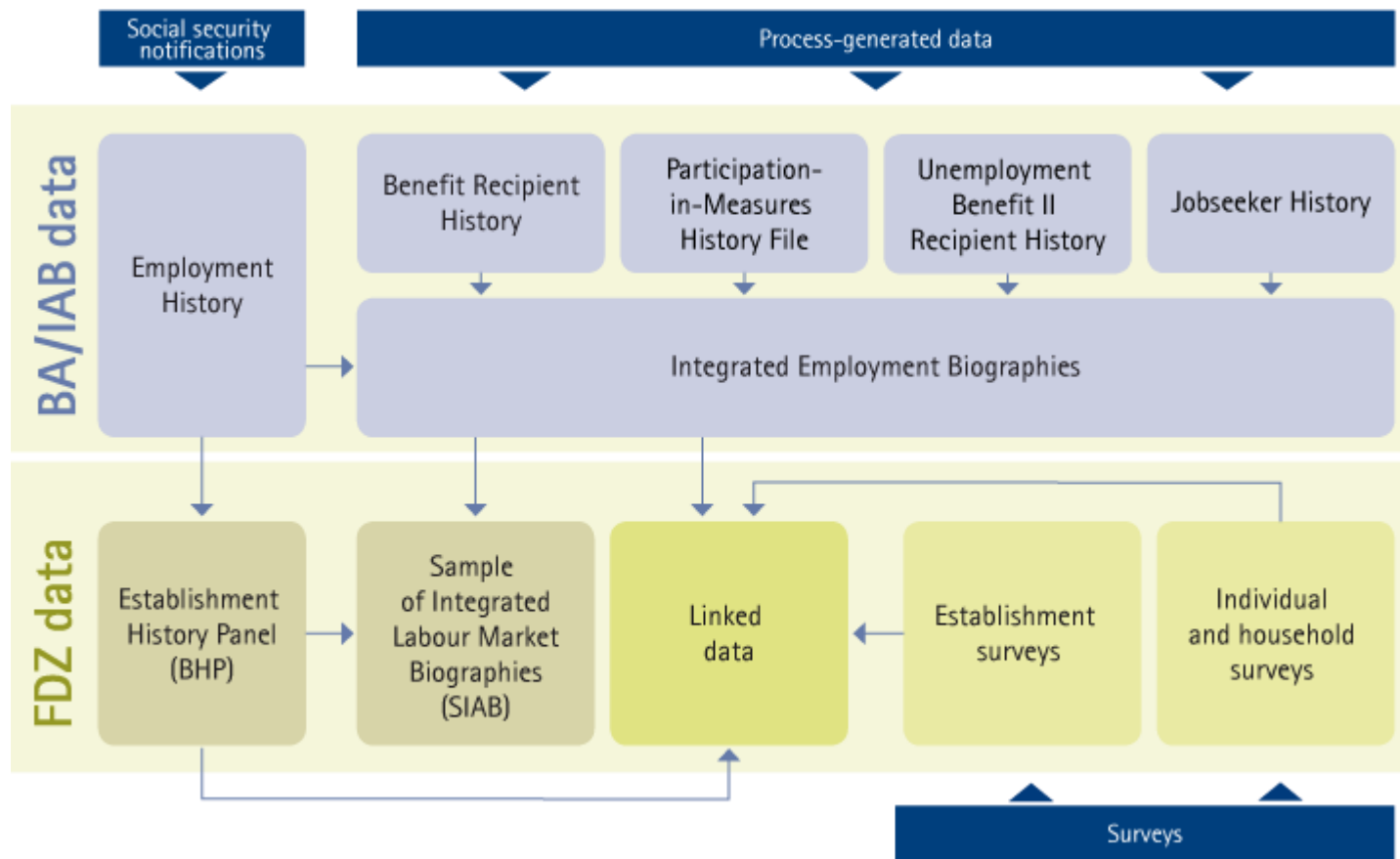
03.12.2014

Iris Dieterich (IAB)
David Schiller (IAB)
Dana Müller (IAB)

Introduction

- member of the DDI Alliance
- while supporting DDI, the standard is not implemented in the internal structures of IAB
- Very heterogeneous documentation in Word, Pdf, Excel, Tex, txt ... files
- Need for structured and standardized database for managing metadata effectively
- Implementation project: IAB Metadata Portal
- Project is carried out by an international consortium DIPF/TBA21 (Germany), Opit Consulting (Hungary)

IAB data sources



IAB data sources

- Different processes of data generation
 - Survey data
 - Register data from administration processes
- Spell data (e.g. income spell, biographical spell)
 - Time period with identical characteristic
- Episode data
 - Generated to put overlapping spells into a fixed time frame (e.g. calendar year)

The future – even more data sources

- IAB offers a powerful source of research data
 - Information about everybody how was employed from 1975 on a daily basis
 - Additional information coming from surveys to enrich process generated data
- Future plans to link register data with more surveys from outside the IAB and other data sources (Open Data, Geo-Data, Big Data)

Referencing Linked Data

Administrative Data

- Source
 - Variables + Answer Schemes

Survey Data

- Wave
 - Instrument + Modules
 - Question Items + Answer Schemes
 - Variables + Answer Schemes

Linked Data

- Identifier for linkage
 - Subset of original source
 - Reference selected sources and variables
 - Add modified variables
- Subset of original source
 - Reference selected waves and variables
 - Add modified variables

Import and handling of multiple sources

- Heterogeneous documentation of process generated data
- Questionnaire based survey data from different survey companies (infas, TNS ...)
- Existing data sets stored in Stata

- No standard import interface possible
- XML as an exchange standard for import

http://n5315022.88/ IAB-Intranet - Meta Data P...

Deutsch English Magyar

Home
New administrative dataproduct
New survey dataproduct
New linked dataproduct
Keyword
Master data
User management
Search
Help
SDSTDietericI

Help ?

Administrative data / Process generated data

- SIAB_de
- SIAB
- Test
- SIAB_en

Survey data products

- IAB-Betriebspanel
- PASS
- IABBP_9312_v2

Linked data products

- LIAB

Home

New administrative dataproduct

New survey dataproduct

New linked dataproduct

Keyword

Master data

User management

Search

Help

SDSTDieterici

Help ?

Administrative data

Sources

Text chapters

State

New

Change state

Name *

SIAB

Version

Test

Long name

Stichprobe der integrierten Arbeitsmarktbiografien

Abstract

Die Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) ist eine 2%-Stichprobe aus der Grundgesamtheit der Integrierten Erwerbsbiographien (IEB) des Instituts für Arbeitsmarkt und Berufsforschung (IAB). Die IEB

Topics

Identifikatoren:
Systemfreie Personennummer
Systemfreie Betriebsnummer
Informationen zur Person:
Geslecht, Geburtsjahr, Staatsangehörigkeit

Data report number

Test

Data unit

Sozialversicherungspflichtig Beschäftigte (ab 1999 auch geringfügig Beschäftigte), Leistungsempfänger, Arbeitssuchende

Number of cases *

1.639.325 Personen

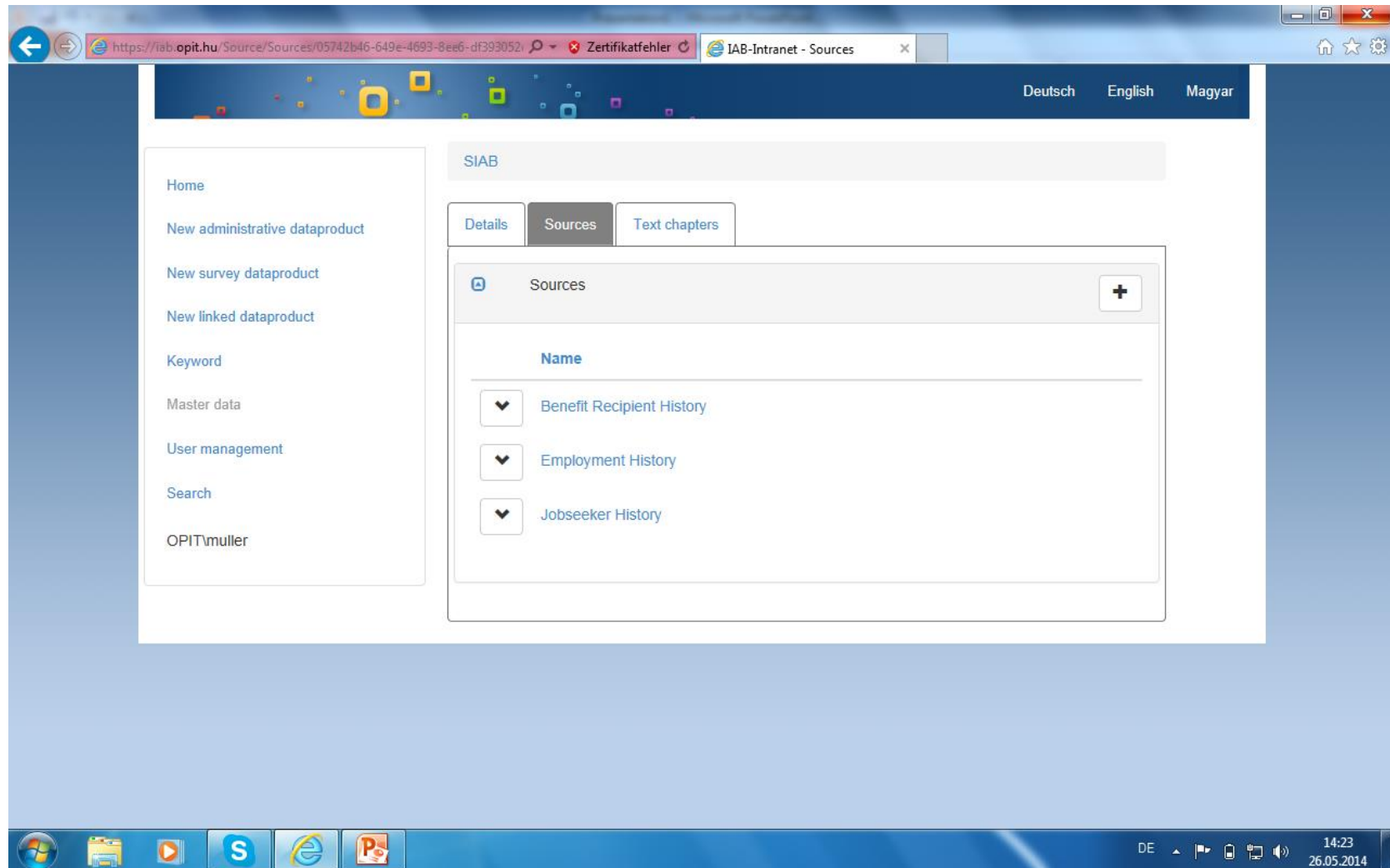
Period

abhängig von der Datenquelle:
Beschäftigung: 1975-2010
Leistungsbezug: 1975-2011
Arbeitssuche: 1997-2011

Time reference

tagesgenaue Erwerbsbiografien

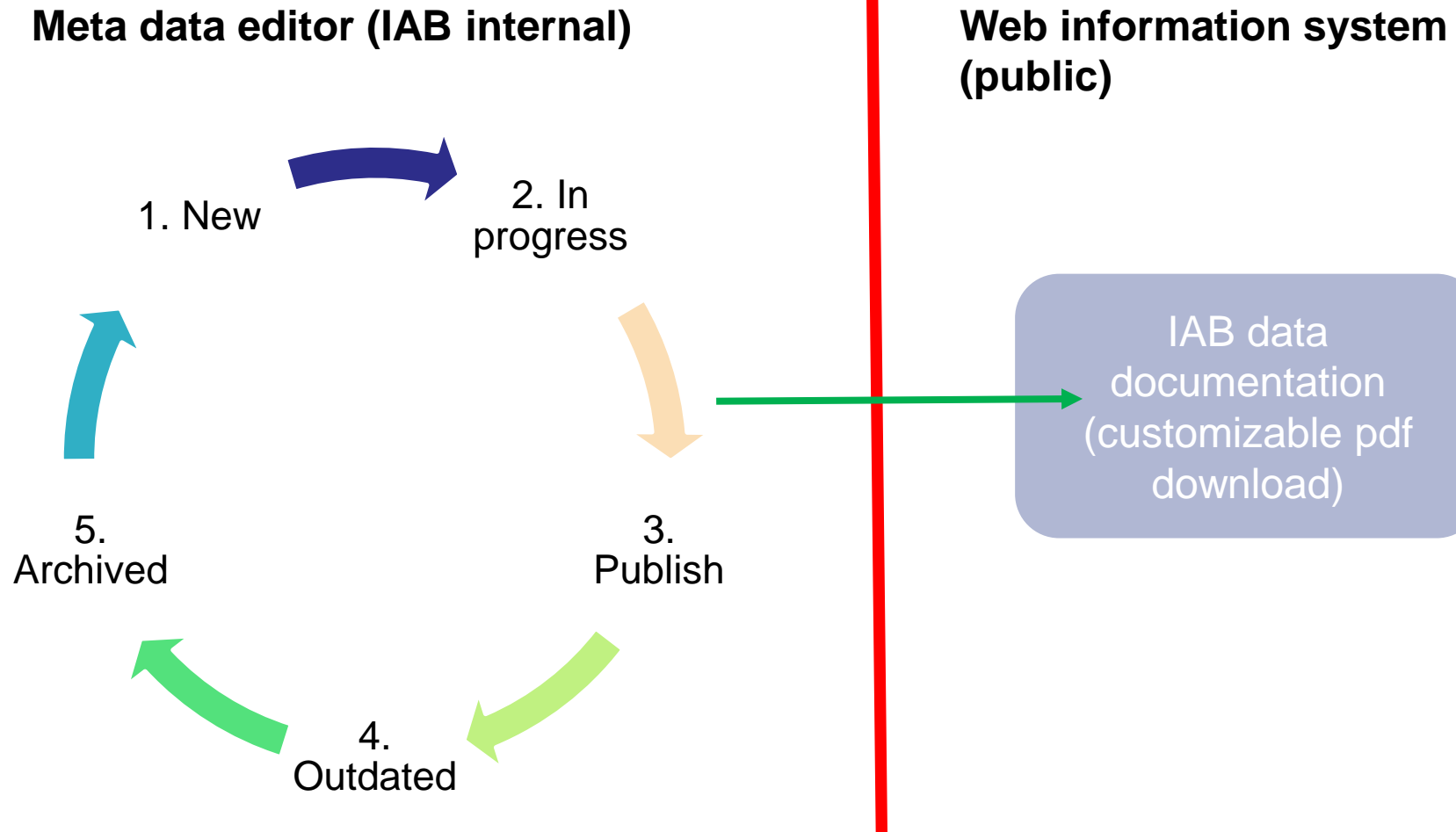
Incorporation of multiple data sources



Structure of IAB data creation process

- A number of departments are involved:
 - ITM – extracting research data out of the data warehouse of the German Federal Employment Agency
 - Different departments (like “education and employment over the life course”) – responsible for single data products
 - KEM – methods department responsible for data editing issues (weighting etc.)
 - FDZ – Research Data Centre quality checks the documentation and gives access to external scientists
- Consequently: need for a tool that supports the data creation process with user roles

Metadata editor and web information system



Outlook

- development and testing still going on
- complicated process of implementation into IT structures of the German Federal Employment Agency
- solution should go live in the 2nd half of 2015 for first data products
- First DDI export applications:
 - CED²AR Cornell University (DDI 2.5)
 - Metadata use in virtual research environments

Thank you for your attention!

iris.dieterich@iab.de
david.schiller@iab.de
dana.mueller@iab.de

<http://fdz.iab.de>