

Processing, Anonymizing and Providing Data from Administrative Processes for Researchers

*How to govern the country better – An example of good
practice for employment policy*

Prague, 11th April 2018

CERGE-EI

Silvina Copestake

IAB

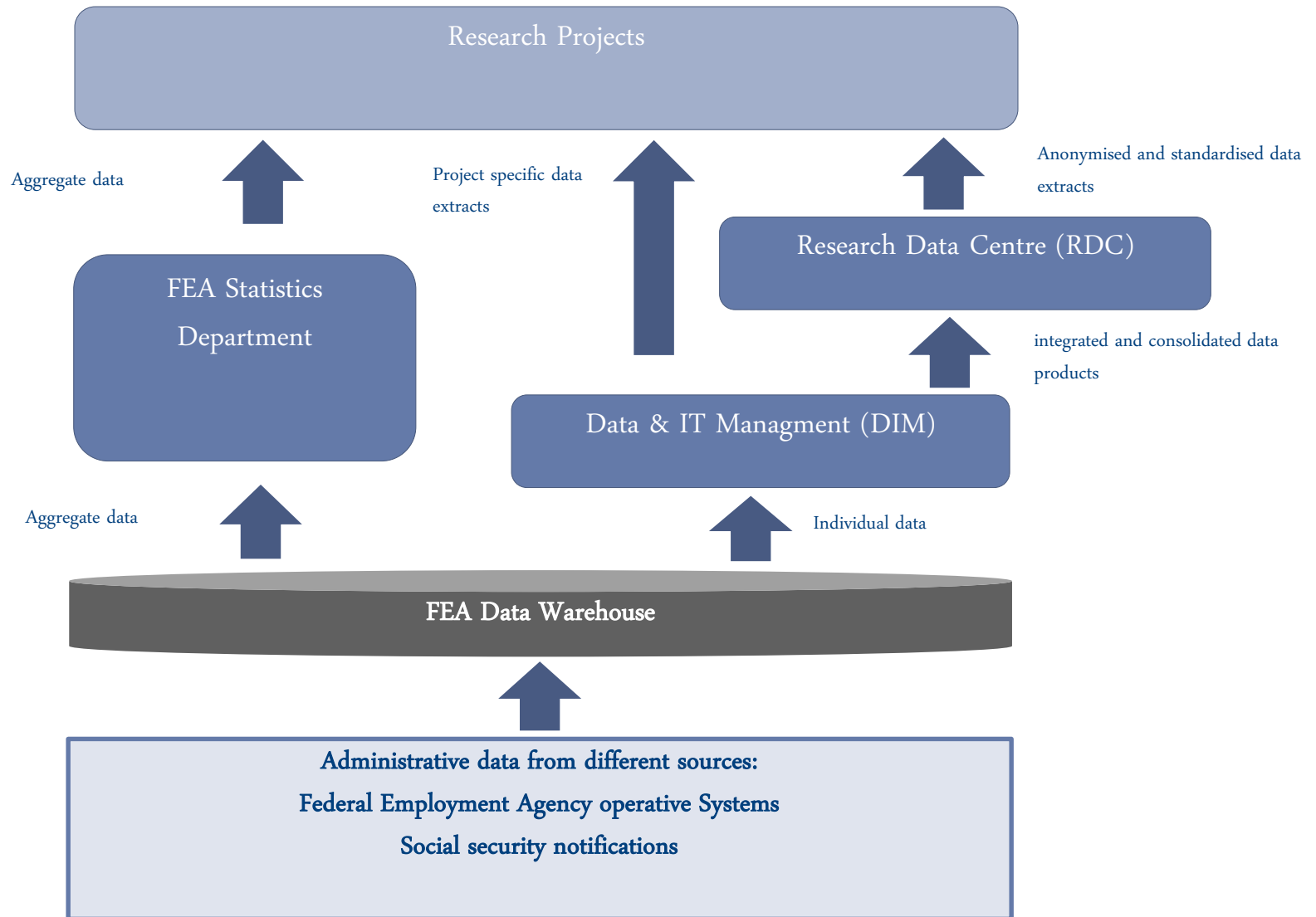


- Rich and unique data base
- Includes survey data and administrative data from the Federal Employment Agency (FEA)
- Can be used for research purposes within the limits of data protection regulations

- Data from administrative processes
- Preparing data for empirical analyses
- Providing data for researchers

Data from Administrative Processes

Data providers at the IAB



Most important data sources



Data from external
IT-Systems

Employment

information from social security notifications
(since 1975)

Unemployment / Job search

information from FEA IT-Systems
(since 1999)

Participation in Active Labour Market Policy Programs

Information from FEA IT-Systems
(since 1999)

Earnings-replacement benefits

Unemployment insurance –
SGBIII (since 1975)

Social assistance

Basic income support for jobseekers – SGB II (since
2005)



Federal Employment Agency
process generated data

What does the data look like?



Person identifier	Begin	End	Data source	State
4753779	01-JAN-02	11-JUN-02	Employment	Regular employment
4753779	12-JUN-02	23-MAY-03	Benefits	Earnings replacement benefits
4753779	12-JUN-02	23-MAY-03	Job Search	Unemployed
4753779	24-MAY-03	31-DEC-03	Employment	Regular employment

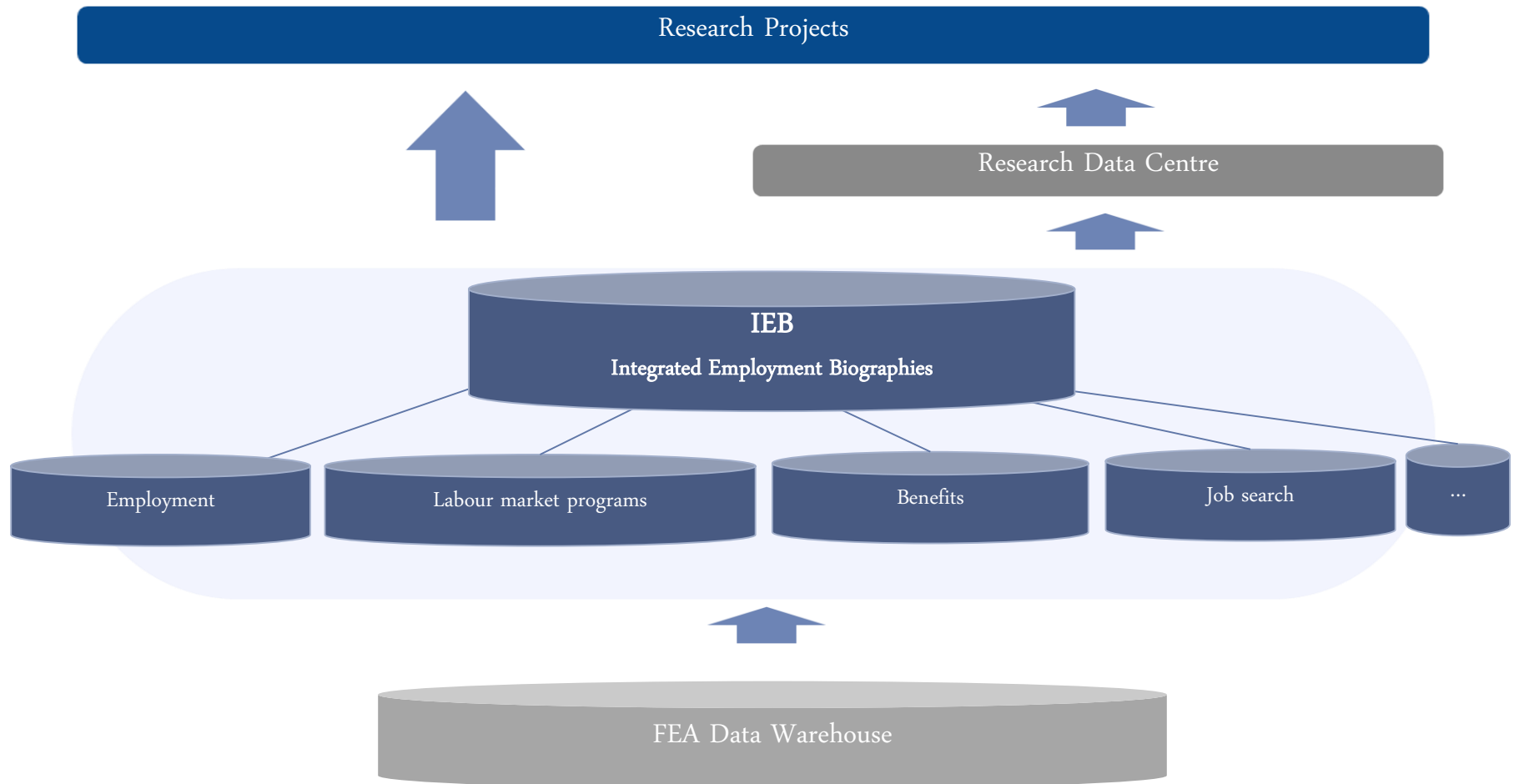
- Individual spell data
- Daily and historical data in long time series
- Different data sources
- Parallel episodes

What information is available?

- Person-Identifier
- Time periods
- Client / employee characteristics
- Different administrative purpose dependent information
- Establishment information

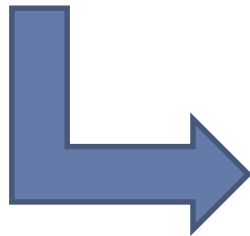
Data collection for administrative purposes – relevance
for research is secondary

Preparing data for empirical analyses



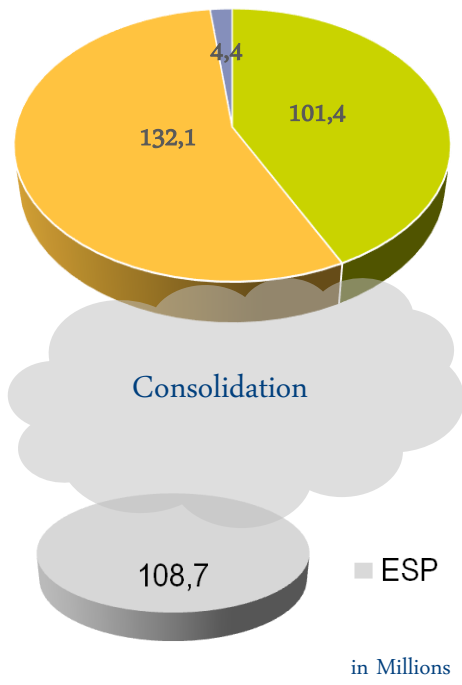
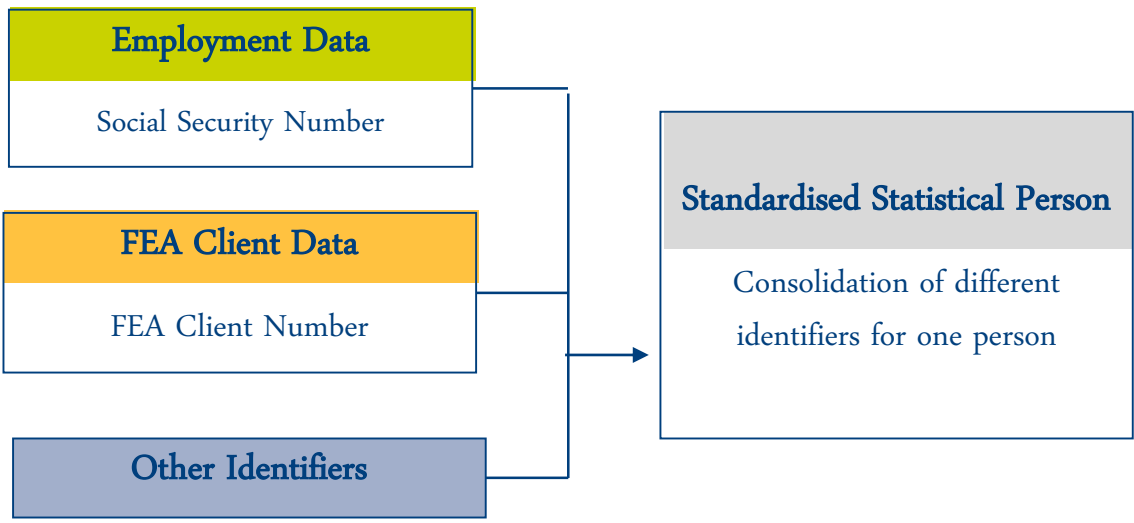
Problems of data generation

- Different identifiers in IT-systems
- Implausible inconsistent information



Solution: data consolidation

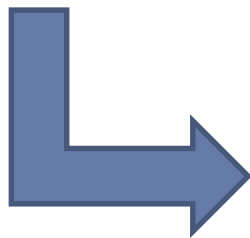
- Comprehensive identification of a person
- Consolidation of overlapping times and inconsistencies



Problems in historical data

- Changes in legislation
- Changes in the operational IT-Systems
- Insufficient documentation of former times

Dataproducts require extensive Metadata and Quality Checks



Possible consolidations:

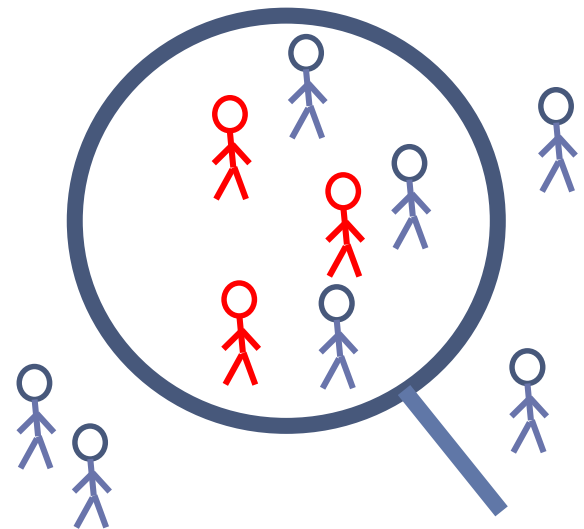
- Exact description of the breach, old and new fact
- Update of regional information
- Mapping of older information in new structures

Providing data for researchers

Data & IT Management (DIM) is an internal service department

- Explore and prepare complex **data sources** in cooperation with researchers
- Develop **data products** and keep them up to date
- Provide **documentation** and advice
- Assure data quality
- IT infrastructure

- Example: Evaluation of an Active Labour Market Policy Program: Promotion of Further Vocational Training (FbW)
- Identify participants in Participants-in-Measures History File
- Identify a relevant control group
- Supply complete Employment Biographies
- Supply addresses for a survey



- Access for external projects is only allowed under strict data protection regulations
- Access to a broader public provided by **Research Data Centre** (RDC):

Samples of data products in
different stages of
anonymization

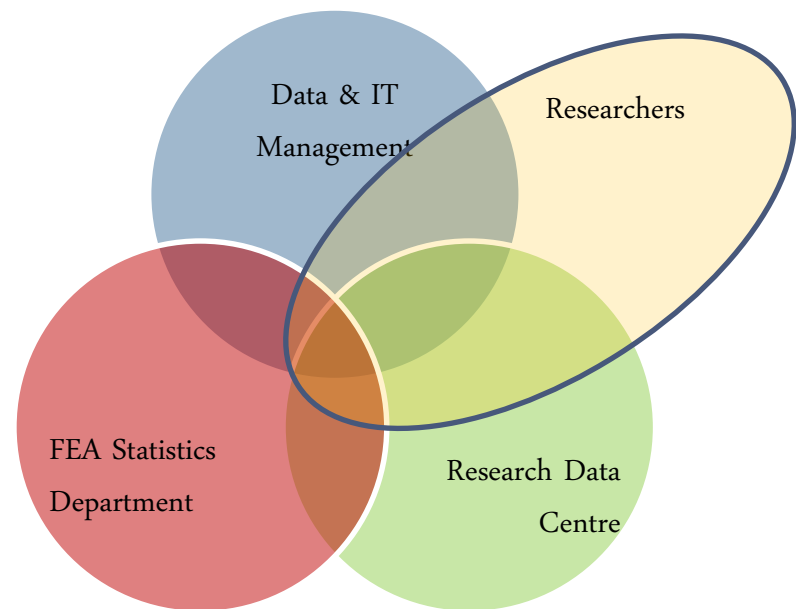
Combination of administrative
data and survey data

Different types of access

Scientific Use Files
Remote Data processing
On-site

Providing a good research data base from administrative data

- Several different actors play together to achieve a good data base for research from process generated data
- Interaction of different skills
- Close feedback between data producers and data users
- Extensive Metadata



Thank you for your attention!

Silvina Copestake

Silvina.Copestake@iab.de