Enhanced Access to Micro Data of Official Statistics - Integrated Data and Meta Data Management System of the German Research Data Centres

Manfred Ehling

Federal Statistical Office of Germany
University of Applied Science Fresenius
German Data Forum

Conference of European Statistics Stakeholders
Budapest, 20–21 October 2016

Content

Background

- German Research Data infrastructure and Research Data Centres (RDC) in official statistics

Problem statement

- Meta data describung micro data is not straightforward

- Available meta data deficient

Solution

- New meta data management system

Conclusion

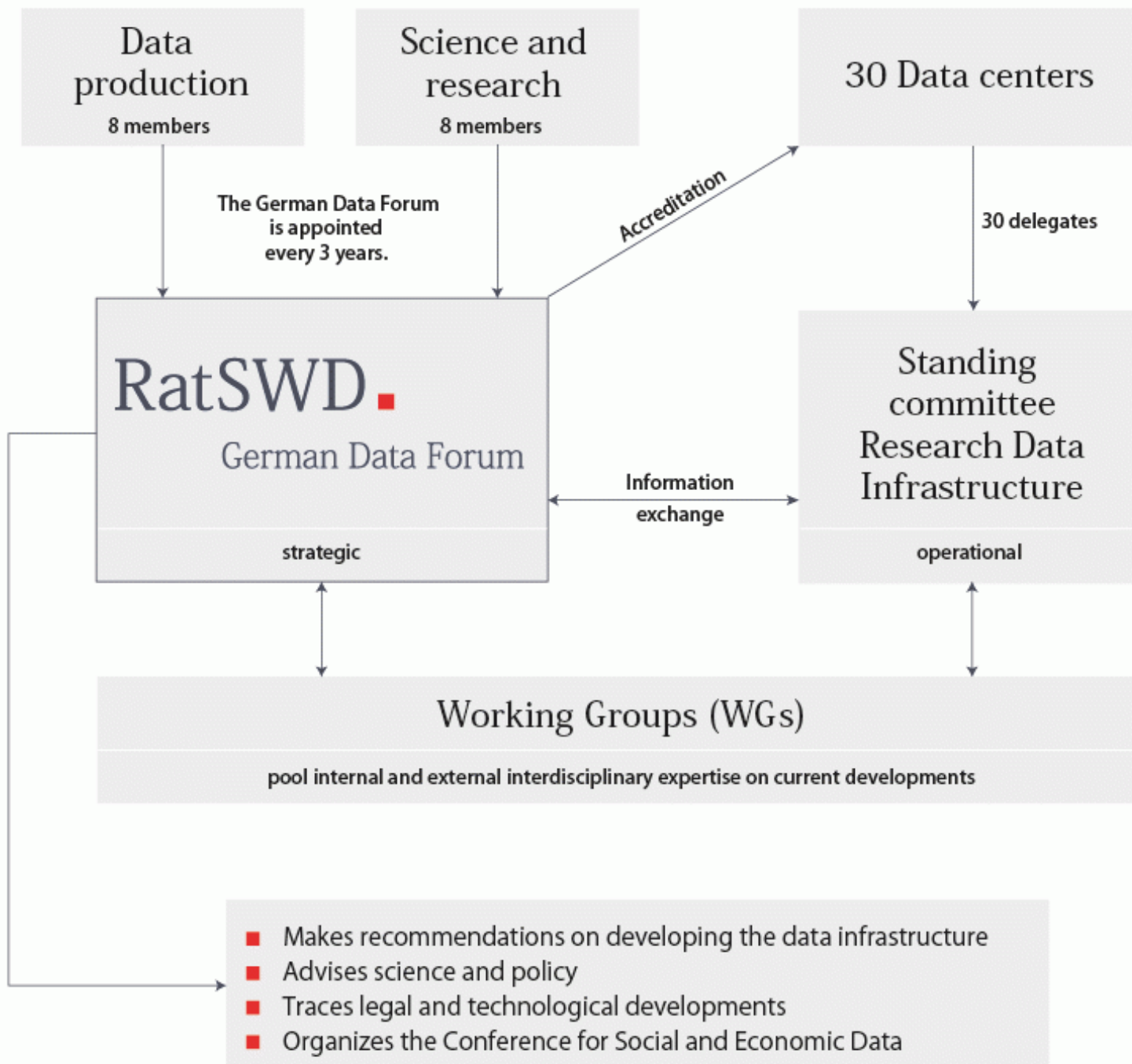- Integration of research data infrastucture

The German Data Forum

- 2004: First appointment of the German Data Forum

- Advisory Council to the Federal Government (8 data producers / 8 data users from research)

- Established a research data infrastructure for the social, behavioral, and economic sciences

- Which consists of 30 research data centres

- Facilitating access to high-quality data

- As a result of independent initiatives from within the scientific community
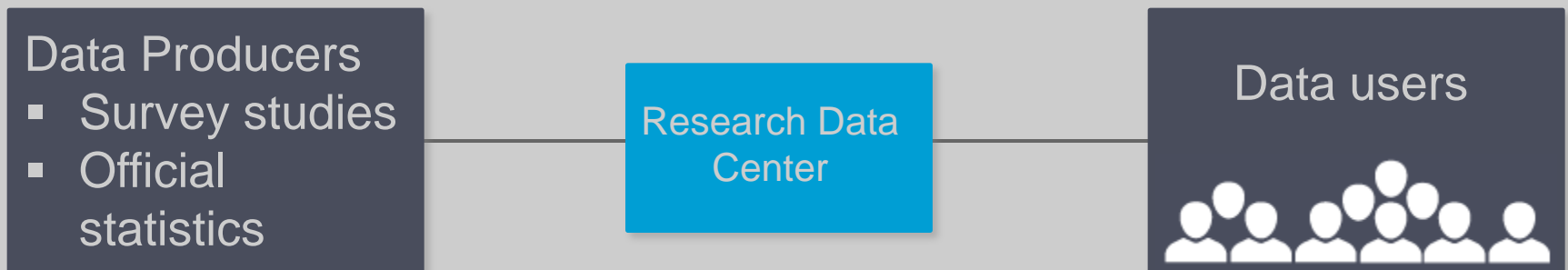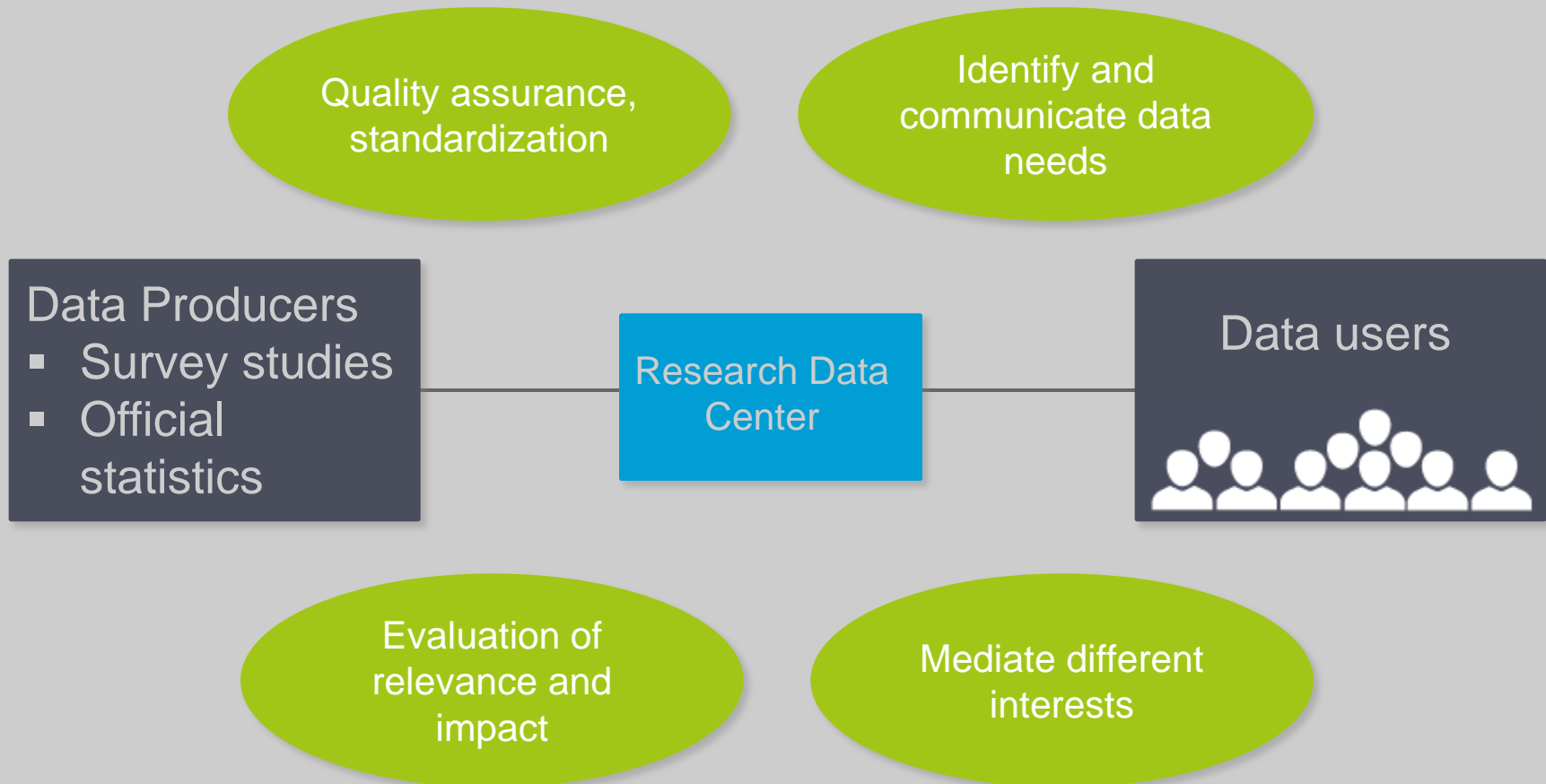
30 RDCs accredited by the RatSWD

Accreditation of RDC

Data Producers
- Survey studies
- Official statistics

Research Data Center

Data users

Accreditation of RDC

Quality assurance, standardization

Identify and communicate data needs

Data Producers
- Survey studies
- Official statistics

Research Data Center

Data users

Evaluation of relevance and impact

Mediate different interests

Current Agenda 2014-2017

Scientific Policy Advising

(e.g. Census 2021, European General Data Protection Regulation)

Data Protection and Research Ethics

Development of the Research Data Infrastructure

(e.g. data management, new types of data)

International Coordination and Cooperation

Harmonising operative tasks for all RDCs

(harmonizing standards, work tasks, documentation and operational procedures)

Metastructure of RDC

(improvement of user support with regard to finding data, design of future structures of the research data infrastructure, meta data portal)

For more detailed information, see www.ratswd.de

Tasks of the RDC in official statistics

Service facility for science

- Producing anonymised data records from the data collected

- Providing appropriate metadata

- Ensuring the data use within the legal framework

RDC - legal basis

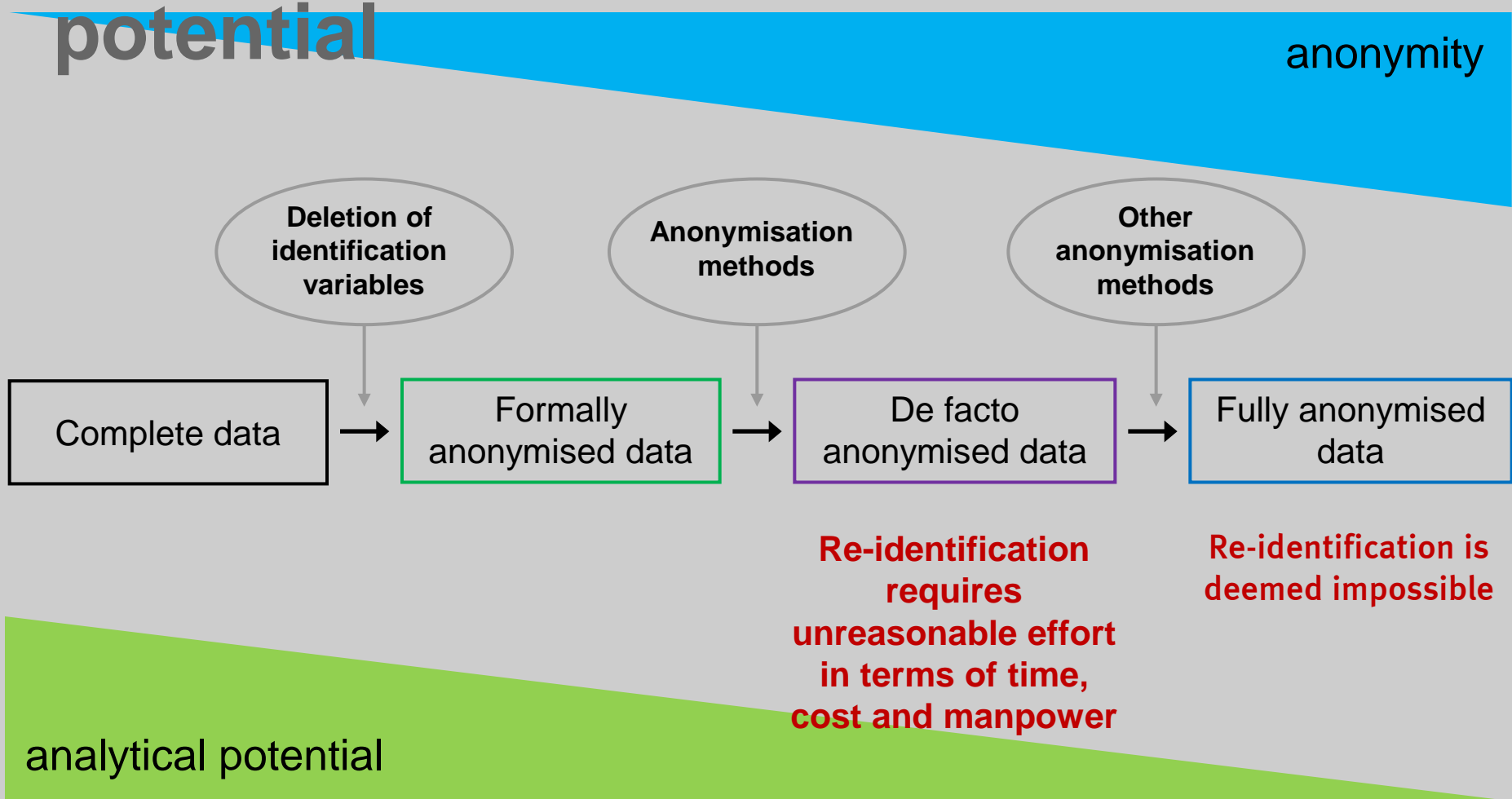**Federal Statistics Law (Article 3)**

- Provision of individual data for scientific purposes as task the Federal Statistical Office
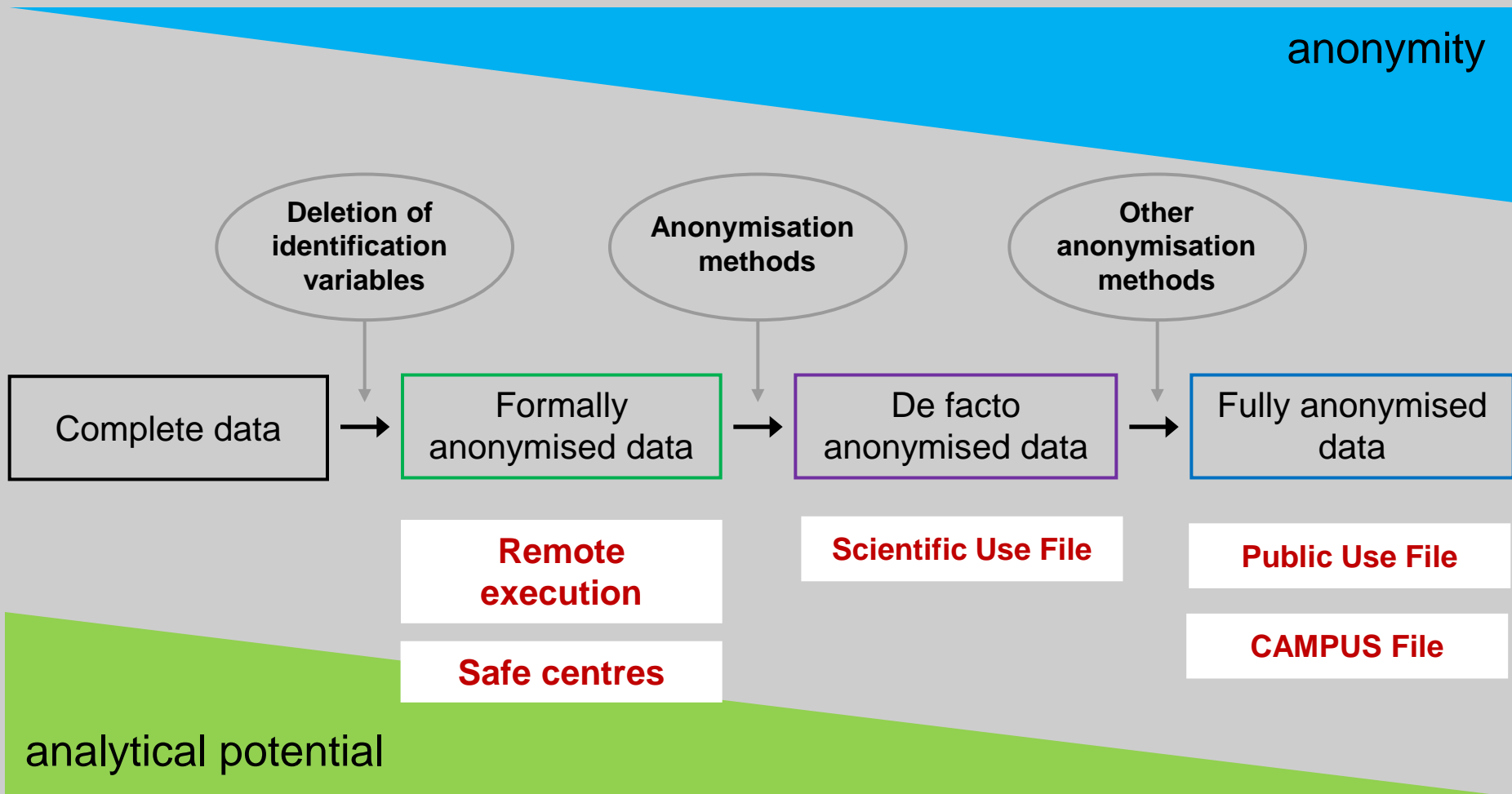
Federal Statistics Law (Article 16)

Individual data will only be passed on

- to institutions of higher education and independent research institutes
- for carrying out scientific projects
- if it requires an unreasonable effort in terms of time, cost and manpower to attribute the data to the statistical units concerned ($\rightarrow$ anonymisation).

Levels of anonymity and data access

anonymity

Deletion of identification variables		Anonymisation methods		Other anonymisation methods	

Complete data → Formally anonymised data → De facto anonymised data → Fully anonymised data

Remote execution

Safe centres

Scientific Use File

Public Use File

CAMPUS File

analytical potential

Demand of Statistics from 2005 to 2015
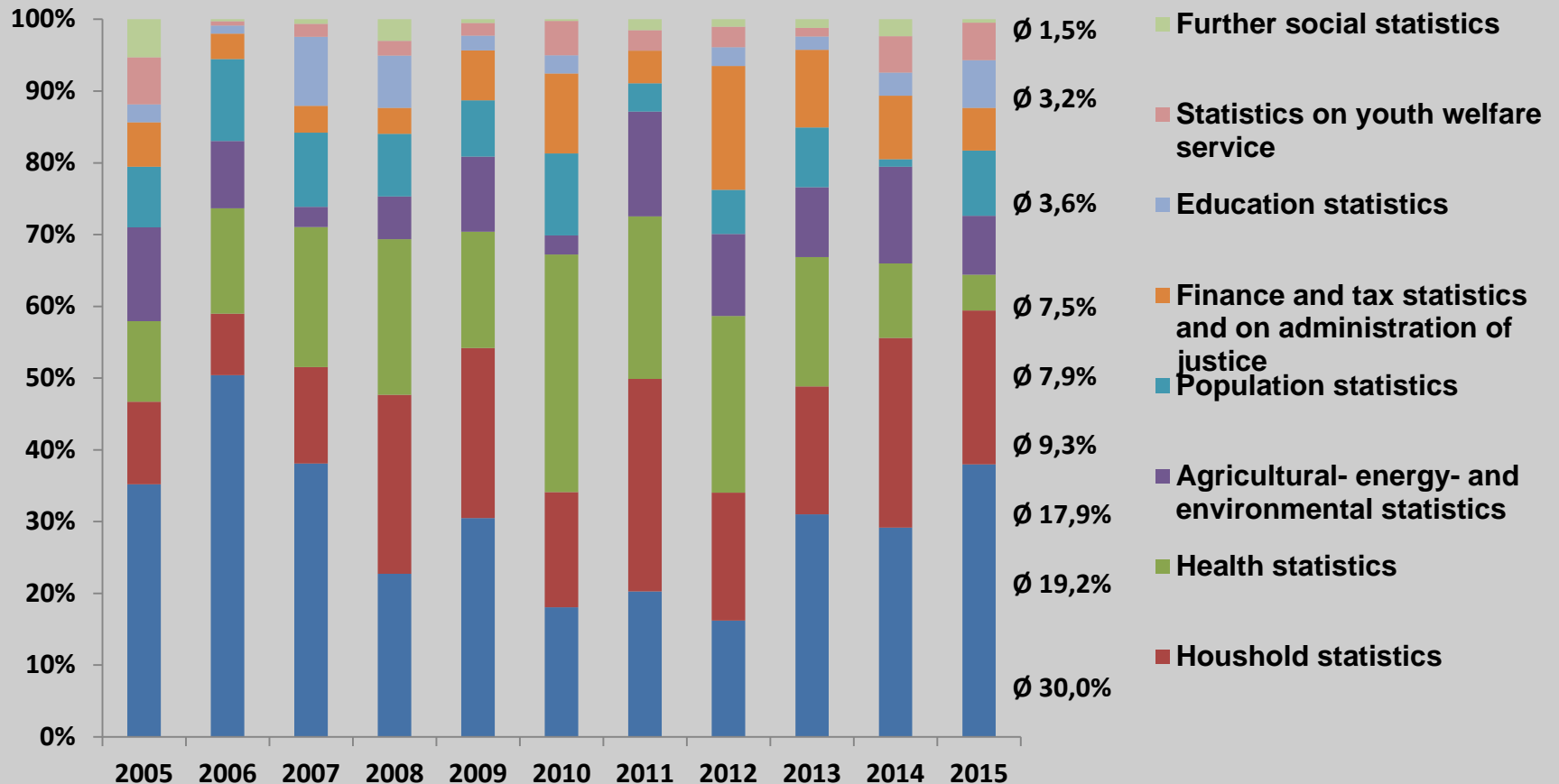
Most Requested Statistics: Top 10 in 2015

TOP	Statistics	Number of applications	Number of statistical products
1	**Micro census**	42	254
2	**Structure of earning survey**	12	30
3	**Official firm data (longitudinal data) - Local units**	8	164
4	**Statistics on youth welfare service**	7	66
5	**Official firm data (longitudinal data) - Enterprises**	6	89
6	**Migration statistics**	4	44
7	**Population census**	4	7
8	**Statistics on income tax**	4	4
9	**Official firm data (longitudinal data) - Products**	3	45
10	**Business register**	3	45

Problem statement

- **Dissatisfaction with the structure and quality of metadata on the RDC website**

- **User support with regard to finding data and the assistance of their analysis to provide accurate results insufficient**

- **Complex and combined micro data of different surveys**

- **Maintenance and administration of the (old) meta data system caused a lot of work in the research data centres**

  - Meta data not up-to-date
  - A lot of manuell work in RDC

Metadata– illustration (1)

31111 – Construction industry

The monthly report provides current data for short-term analyses in the main construction industry and basic information for the calculation of indices of production and new orders. Also, statistics of orders on hand in the main construction industry are compiled on a quarterly basis. Short-term data for the finishing trades and property developers are obtained through the quarterly survey.

The annual structural surveys in the construction industry provide information on the medium-term development and structural changes in the construction industry. The annual survey of investments provides important information on investments made by the enterprises, while the annual cost structure survey supplies detailed information on cost and performance data. The two surveys are complemented by the structural survey of small b... ...ustry.

EVAS: 31111
Periodicity: monthly

📄 Survey questionnaire - main construction industry (pdf, 1 MB)

Editorial work

Statistics database

StaNet (Statistics Portal)

REQUESTED TABLES

🔗 Permits for building construction in Germany, yearl...

🔗 Construction work completed in Germany, year...

🔗 Local units in the field of main construction industry

GENESIS-Online

Regional database

PUBLICATIONS

🔗 GENESIS-Onli...

🔗 Regional database Germany

🔗 Quality report

(dynamic) Link list

LEGAL BASES

🔗 Construction S...

🔗 EU: Regulation

Statistics database

Metadata – illustration (2)

THEME IN BRIEF

Agricultural holdings

In Germany there are almost 300,000 agricultural holdings. Still most of them are sole proprietorships, which means that farmers operate their businesses alone or together with their families.

The agricultural holdings section contains extensive information on the holdings themselves – for example their legal form and whether they are run as a primary or secondary activity, other gainful activities – and on whether holdings have converted to organic farming or operate renewable energy installations. You will also find data on labour, farm succession, ownership and tenancy as well as purchasing values of agricultural real estate.

Use our contact form to request data also on the following subjects: type of farming, grants received in support of rural development, calculation of profits and turnover taxation of holdings, vocational training of the managers of agricultural holdings.

Editorial work

Agricultural holdings

Animals and animal production

Fisheries

Field crops and grassland

Fruit, vegetables, horticulture

Wine

Forests and wood

Production methods

Land use

LINKS

↗ STATmagazin

↗ Census of agriculture 2010

↗ Press pages

Editorial work

DOWNLOADS

📄 Industry classifications, 2008

📄 Goods directory... statistics, 2009

📄 Job classification, 2010

Classification server

DEFINITIONS

…

MDM

SEARCH

List of Municipalities

AGS oder Name   Suchen

GV-ISys

Metadata– Harvesting

Live system

Editorial system

Editor → **Layout** →

DB → Publish, automated when updated

Harvesting

Import | API

Templates for data query → Jobs → **Harvester** → Content ← **Transformation rules for results**

Metadatamanagement

| Source 1 | Source 2 | Source 3 | ... |

solution-sketch...

© Statistisches Bundesamt

Metadata– Harvesting

Live system

Editorial system

Editor → Layout →

Publish, automated when updated → Live system

DB

Harvesting

Import | API

Templates for data query → Jobs → Harvester → Content ← Transformation rules for results

RDC-editorial module
- documents
- online-ressources (DOI)
- contact person
- ...

Metadatamanagement

Source 1 | Source 2 | Source 3 | RDC-data

solution-sketch...

RDC-editor

© Statistisches Bundesamt

Metadata– Data query

Object requirements from the MDM

Templates for data query

```
MDM_READ (<OBJECT>; <LANGUAGE>; <ID>),

    <OBJECT>    {statistic, classification, … }

    <LANGUAGE>  {DE, EN}

    <ID>        {<EVAS-Nr. or structure>}

e.g. MDM_READ (statistic; EN; 12411)
```
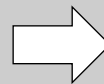
| Attribute	Description	Source
EVAS	EVAS-Nr. of the statistic	statistics database
TITEL	Title of the statistic	statistics database
PZT	Periodicity of the statistic	statistics database
LEGAL_BASIS_NAT_TXT	Text of the national legal basis of the statistic	statistics database
LEGAL_BASIS_NAT_URL	url to the national legal basis of the statistic	statistics database
LEGAL_BASIS_INT_TXT	Text of the international legal basis of the statistic	statistics database
LEGAL_BASIS_INT_URL	url to the international legal basis of the statistic	statistics database
TABLES_TOP_GO	List of the TOP-tables from GENESIS-Online with DeepLink-URL, Download-URL (XLS) name and header, regional granularity	GENESIS-Online
TABLES_TOP_REGDB	List of the TOP-tables from the regional database with DeepLink-URL, Download-URL (XLS) name and header, regional granularity	regional database
TABLES_GO	List of tables from GENESIS-Online with DeepLink-URL, Download-URL (XLS) name and header, regional granularity	GENESIS-Online
TABLES_REGDB	List of tables from the regional database with DeepLink-URL, Download-URL (XLS) name and header, regional granularity	

solution-sketch…

© Statistisches Bundesamt

Metadata– Transformation

Attribute illustration MDM/CMS

MDM Objects		CMS	
Object	**Attribute**	**Portal-content**	**Attribute**
statistic	Name		
	Title	text block	...
	Periodicity	text block	...
	Text of the national legal basis	Box "legal bases" with two links to external content	External content link 1 - Title/Name
	URL to the national legal basis		External content link 1 - URL
	Text of the international legal basis		External content link 2 - Title/Name
	URL to the international legal basis		External content link 2 - URL
	GENESIS-ONLINE TOP-Table 1 -name	Box "requested tables" with two links to external content	./.
	GENESIS-ONLINE TOP-Table 1 -header		External content link 1- Title/Name
	GENESIS-ONLINE TOP-Table 1 -DeepLinkURL		External content link 1- URL
	Regional database TOP-Table 1 -name		./.
	Regional database TOP-Table 1 -header		External content link 2- Title/Name
	Regional database TOP-Table 1 -DeepLinkURL		External content link 2- URL
	...	...	...
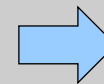
LEGAL BASES

- Construction Statistics Act
- EU: Regulation
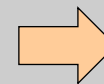
REQUESTED TABLES

- Permits for building construction in Germany, yearly
- Construction work completed Germany, yearly
- Local units in the construction industry

solution-sketch...

Elements of the new meta data system

Meta data reports

- Detailed information on the single statistic
  - General information
  - Methods
  - Quality
  - Publications
- Information on level of the single product out of the statistic
  - Product description (Panel, combination of data sets)
  - Data editing and preparation
  - Anonymization

Digital object identifier (DOI)

- Every used data set will get that as a serial code
- The persistent identification of research data enables researchers to cite the data sets

Conclusion

- Features of the meta data management system of the common statistics portal will be released at the end of 2016

- Meta data management system will provide up-to-date meta data

- Meta data system of RDC with reports for the most used data sets will follow 2017

- After an evaluation of user satisfaction and implementing recommended changes the meta data reports will be expanded to all statistics

- Integration of the German research data infrastructure