

Quality of Open Data and Metadata

An Introduction to the NQDM - Project



Lina Bruns
Fraunhofer FOKUS
Prague, 2nd November 2018

Agenda

- Fraunhofer FOKUS
- Open Data in Germany
- Quality of Open Data
- NQDM - Project



3



2.3

Billion euros Research Volume

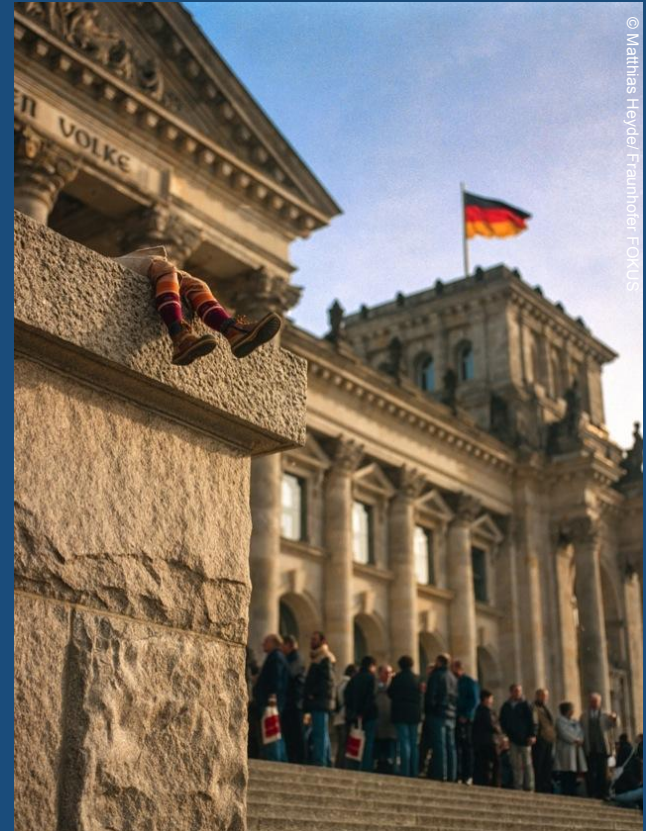
Fraunhofer FOKUS



Digital Public Services – DPS

- eGovernment infrastructures and processes
- Open Government
- Data and service platforms
- Usability of specialist applications
- Electronic file – selection and launch
- Digitization and society
- Certification assistance in CC certification processes

DPS bundles the activities encompassing eGovernment and public IT and supports politics, administration and industry in the digitization in the public space and in public administration.



Open Data in Germany



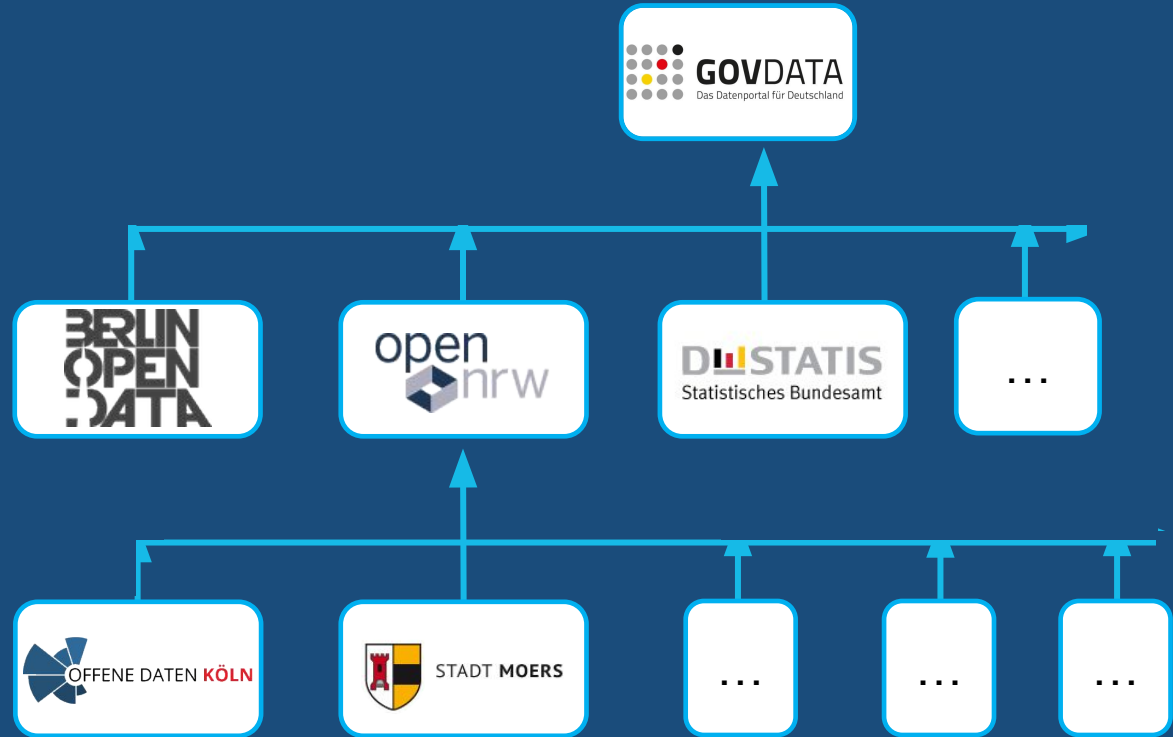
Structure of Open Data in Germany

Overall Data
Portal

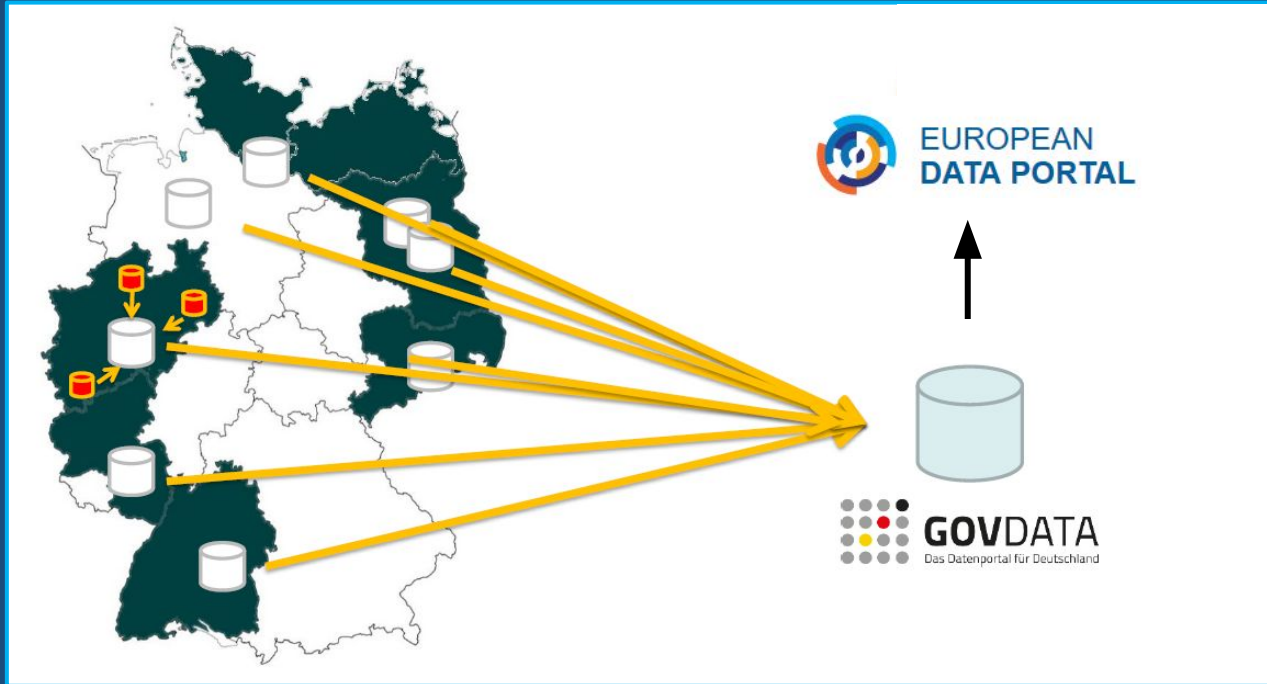
State / Specialist
Portals

Municipal
Portals

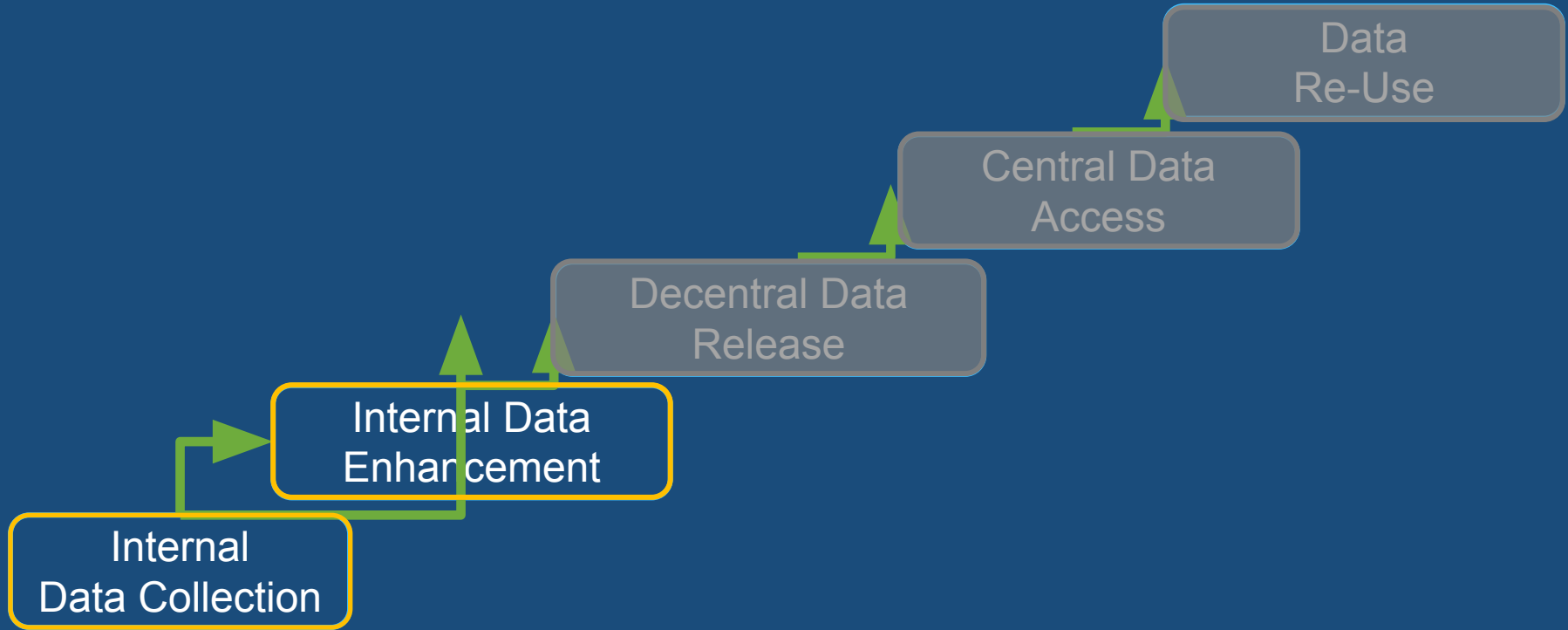
7 Portals



Structure of Open Data in Germany



Publishing Open Data



Quality of Open Data



Open Data Quality Issues I

```
EU;Badestelle;Code;Übermittlung an die EU;Datum Probenentnahme;Entnahmestelle;Proben-ID
true;Zeetze;DENI_PR_TK26_2731_01;;04.05.2015;Badestelle Zeetze;2015M0400413
false;Naturbad Bardowick;;true;07.05.2012;;
true;Barumer See;;true;04.05.2015;Badestelle Barumer See;
true;Inselsee- DLRG Station;;true;04.05.2015;Badestelle Inselsee;
true;Reihersee Lüdershausen;;true;04.05.2015;Reihersee Lüdershausen;
true;Sumter See, Badestelle Haar;;true;04.05.2015;Sumter See, Badestelle Haar;
;;;;;;
false;Naturbad Westergellersen;;false;15.08.2016;Naturbad Westergellersen;2016M0400707
false;Naturbad Kirchgellersen;;false;15.08.2016;Naturbad Kirchgellersen;2016M0400706
false;Naturbad Bardowick;;false;15.08.2016;Naturbad Bardowick;2016M0400708
true;Sumter See, Badestelle Haar;DENI_PR_TK25_2731_02;true;18.07.2016;Badestelle Haar;2016M0400589
true;Badestelle Zeetze;DENI_PR_TK25_2731_01;true;18.07.2016;Badestelle Zeetze;2016M0400590
true;Inselsee;DENI_PR_TK25_2628_02;true;18.07.2016;DLRG station;2016M0400586
true;Reihersee;DENI_PR_TK25_2628_03;true;18.07.2016;Campingplatz Kloodt;2016M0400587
true;Barumer See;DENI_PR_TK25_2628_01;true;18.07.2016;Barumer See;2016M0400588
```

Open Data Quality Issues II

	A	B	C
1	Bezeichnung	Schulform	Telefon
2	Grundschule Betzendorf	Grundschule	04138/510918
3	Grundschule Handorf	Grundschule	04133 6325
4	Grundschule Fürstenwall-Schule	Grundschule	05851 7859
5	Grundschule Reppenstedt	Grundschule	04131 672741
6	Grundschule Westergellersen	Grundschule	04135/8702-77/-78
7	Grundschule Embsen	Grundschule	Tel 04134-916610
8	Gymnasium Lüneburger Heide	Schule in freier Trägerschaft	04134 91500
9	Grundschule Barendorf	Grundschule	04137/7831
10	Grundschule Artlenburg	Grundschule	04139 7008
11	Montessori-Schule (Grundschule)	Schule in freier Trägerschaft	04131 - 22 05 15 6
12	Geschwister-Scholl-Realschule Embsen	Realschule	04134-916650
13	Oberschule Neuhaus/Elbe	Oberschule	038841 21003 oder 21004
14	Realschule Bleckede	Realschule	0 58 52 - 97 80 20
15	Schäfer-Ast-Grundschule	Grundschule	04178 400
16	Christianischule Oberschule am Kreideberg	Realschule	04131/20769-0
17	Grundschule Bardowick	Grundschule	0 41 31 - 12 17 38
18	Jörg-Immendorff-Schule Bleckede	Hauptschule	05852 / 97 80 10
19	Gymnasium Bleckede	Gymnasium	05852 - 95 11 200
20	Grundschule Neuhaus/Elbe	Grundschule	038841 21300
21	Grundschule Tripkau	Grundschule	038841 21300
22	Grundschule Melbeck	Grundschule	Tel. 04134 - 7005

Importance of Data Quality



Data User

- Processing of low-quality data is difficult and often requires
- Manual data enhancing



Data Publisher

- High quality data can be re-used by public administration
- Data quality prevents misinterpretation

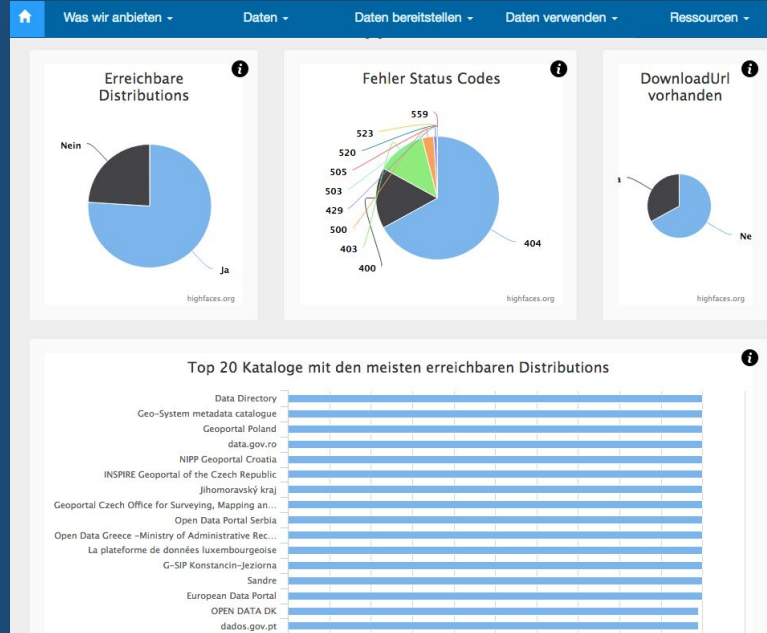
How
can

Increasing Quality – Approaches

Tools for Data Enhancement



e.g. anonymization tools,
automated quality check of
metadata



Increasing Quality – Approaches

Tools for Data Enhancement



e.g. anonymization tools,
automated quality check of
metadata

Improving Publishing Process



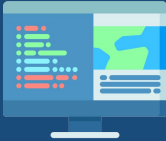
e.g.: Standardization of
publishing process



<https://www.oeffentliche-it.de/>

Increasing Quality – Approaches

Tools for Data Enhancement



e.g. anonymization tools,
automated quality check of
metadata

Improving Publishing Process



e.g.: Standardization of
publishing process

Employee Training



NQDM

Normentwurf für Qualitative Daten und Metadaten

(Draft Standard for Qualitative Data and Metadata)





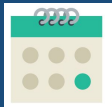
Lack of quality of published data and metadata



Create a guide that helps data publishers to publish high quality open data and metadata



Research project funded by Federal Ministry of Economic Affairs and Energy



01.09.2017 - 31.08.2019

Definitions
of relevant terms



Decision
Support



Rating schemes for
quality



Evaluation
matrix for
metadata



Quality Characteristic



File Format
specific
recommendation





Employees of public administration responsible for publishing data with

- Little knowledge of open data
- Little or no knowledge of data quality aspects
- Little or no knowledge of file formats



Questions addressed:

- How to get started with publishing high quality data?
- What is data / metadata quality?
- What are suitable formats?
- Which tools support publication?



Expert Panel

Evaluation and Support

~20 Members

- Data publisher
- Data users
- Scientists
- Politics



Prototypical Implementation

Use Case – Approach

Data Blending for

- identifying quality issues and
- testing recommendations from NQDM

Done

- First Draft created
- Content Coordination with Expert Panel
- First Feedback Iteration running



To Be Done

- Fine-Tuning with Expert Panel in November
- Extension of Content
- Finalization of Draft
- Preparation of Publishing as Specification or Draft Standard

Contact

Lina Bruns

Phone +49 (30) 34 63 -7283
lina.bruns@fokus.fraunhofer.de

Fraunhofer FOKUS

Institute for Open Communication Systems

Kaiserin-Augusta-Allee 31

10589 Berlin, Germany

info@fokus.fraunhofer.de

www.fokus.fraunhofer.de