Cataloguing editions and other resources in one unified system

The case of the Text+ Registry

Tobias Gradl
Media Informatics Group, University of Bamberg

Daniela Schulz
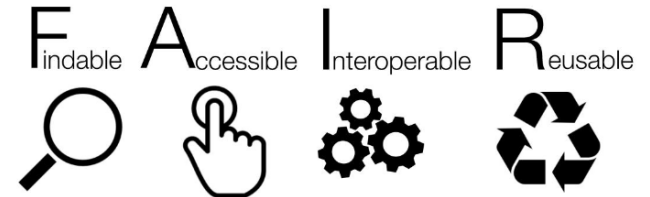Herzog August Bibliothek Wolfenbüttel

HERZOG
AUGUST
BIBLIO
THEK

The bigger picture: NFDI / NFDI e.V.

- Recommendation by RfII in 2016

- Joint Science Conference 2018: establishing a National Research Data Infrastructure (NFDI)

- Objective: To develop and secure data stocks in the long term
  - along the FAIR principles
  - without disciplinary boundaries
  - via a consortial structure

Findable Accessible Interoperable Reusable

Graphics: SangyaPundir - Own work, CC BY-SA 4.0,
https://commons.wikimedia.org/w/index.php?curid=53414062

Text+

2

Cataloguing editions and other resources in one unified system – The case of the Text+ registry (DHC Sheffield 04/05 Sept 2024)

The bigger picture: NFDI / NFDI e.V.

Cataloguing editions and other resources in one unified system – The case of the Text+ registry (DHC Sheffield 04/05 Sept 2024)
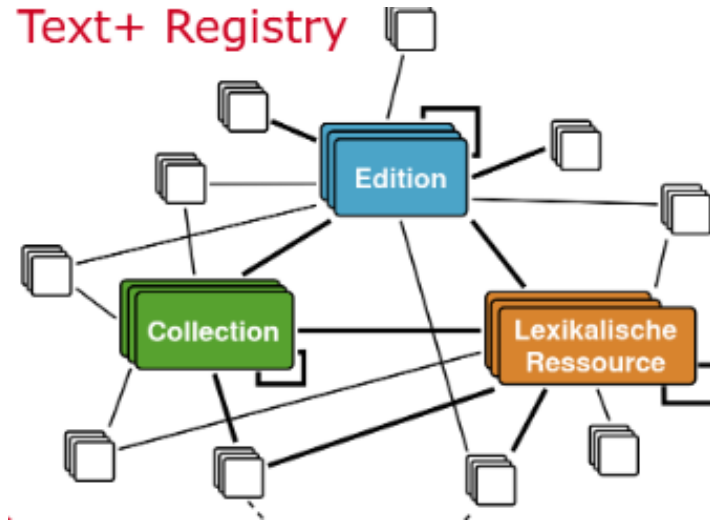
Text+: Aim and Scope

- Focused on speech and textual data
- Structured along the data domains (Task Areas) collections, lexical resources, and editions
- Aims at making textual and speech-based data available and preserve it

Collections

Editionen

Lexikalische Ressourcen

Text+

Cataloguing editions and other resources in one unified system – The case of the Text+ registry (DHC Sheffield 04/05 Sept 2024)
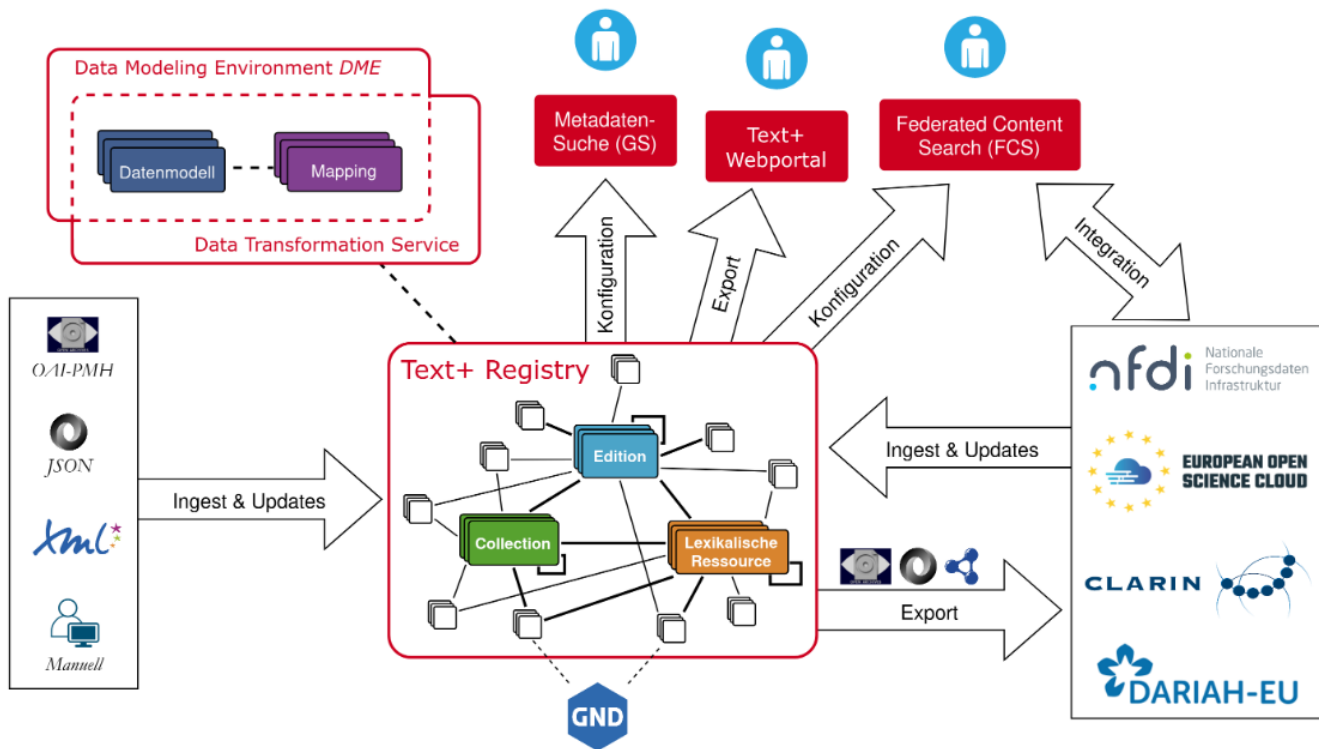
The Text+ Registry: features

- Central tool for registering, researching, and accessing data from the different domains

- Provides APIs to connect to other knowledge bases

- Under consideration of the FAIR principles

- Including interconnected meta / authoritative data (LOD) and a Federated Context Search (FCS)

Text+ Registry

The Text+ Registry: ingest & output

Cataloguing editions and other resources in one unified system – The case of the Text+ registry (DHC Sheffield 04/05 Sept 2024)

Import sources

Institutions

Persons

Works

Text+ Registry

The Text+ Registry is a central, cross-domain research and information tool that records, correlates and links ~~~~~ ces, collections and editions. The Registry is not intended to replace existing directories and catalogs, but to complement them by offering a res~~~~~ e'. Data is enriched in the registry and made available again via interfaces, which enables the connectivity of the registry to overarching structure ~~~~~ whole (e.g. knowledge graphs).

Resources and Entities

Textual collections 118

Services 35

Editions 1096

Lexical Resources 20

Repositories 18

https://registry.text-plus.org/

Cataloguing editions and other resources in one unified system – The case of the Text+ registry (DHC Sheffield 04/05 Sept 2024)

Resources and entities ▾

Further Entities ▾   🌐 EN ▾   🚪 Login   ❓ ▾

Text+ Registry

The Text+ Registry is a central, cross-domain research and information tool that records, correlates and links resources such as lexical resources, collections and editions. The Registry is not intended to replace existing directories and catalogs, but to complement them by offering a research-supporting 'one stop store'. Data is enriched in the registry and made available again via interfaces, which enables the connectivity of the registry to overarching structures and offerings of the NFDI as a whole (e.g. knowledge graphs).

Resources and Entities

Textual collections 118	**Services** 35	**Editions** 1096
Lexical Resources 20	**Repositories** 18	

Text+ —— ⑧

Cataloguing editions and other resources in one unified system – The case of the Text+ registry (DHC Sheffield 04/05 Sept 2024)

Desiderata and Challenges (Editions)

- Discoverability of editions: print vs. digital

- Integration of information from various sources with varying degrees of informational depth and granularity

- Data model to describe all different kinds of editions

- Catering for various purposes and research questions

- Controlled vocabularies vs free text

- identification and disambiguation of the actors (via authority files)

Hierachy	Level	Filed label	Function	Selection	Help	DataCite reference	Comment
0	0	Administrative Information					This section contains information about the record itself (e.g. its provenance, information about the time it was created or relationships to other records).
0.0	1	resource type	1	lexical resource, collection, edition, service	Please indicate which resource type you are dealing with.	10. ResourceType — DataCite Metadata Schema 4.5 documentation (datacite-metadata-schema.readthedocs.io)	
0.1	1	Text+-ID	1				
0.2	1	provenance	1	manual, AGATE, GEPRIS, CSE, CDSE, curated	The data source or provenance from which the entry originated can be specified here. In the case of "manual entry", this entry itself is the provenance.		
0.3	1	timestamp	1				
0.4	1	versions	0..n		The versioning of the registry entry is displayed here. Previous/subsequent edition versions of the edition itself can be specified under "Relations".		
0.4.1	2	versions	0..n				
0.4.2	2	type of relation	1				
0.5	1	part of the Text+ portfolio	1	[x]			
1	0	output type					This section contains information on the output type(s) in which the edition is or will be available.
1.1	1	type	1	book version (print/ebook), digital edition (portal, webpresentation), hybrid, research data	Please select the output form of the edition. If there are different forms of edition, you can enter these in separate entries and then link these entries via "Relations".		
1.2	1	citation	0..n		You can enter a recommended citation here.		
1.3	1	language of output type	0..n		Enter the language(s) used in the output form here. The languages that occur within the edition itself are recorded under "Information on the edition object(s)".		
1.4	1	licences	0..n			16. Rights — DataCite Metadata Schema 4.5 documentation (datacite-metadata-schema.readthedocs.io)	
1.4.1	2	information on accessibility	0..n	free, free for academic purposes, restricted, not accessible	Specify whether the resource is free or only accessible with restrictions.		
1.4.2	2	licence	0..n	CC, free text	Select a licence from the list or use the blank text field.		
2	0	basic information					This section contains basic information about the edition or project, such as title, publication year(s), existing IDs

basic information

main title   short title   IDs

year(s) of publication   links

APIs   discipline(s) (DFG)

Notes on disciplines

funding period

output type   type   citation

language of output type

information on accessibility

licence

relations   series

series number

Relations to other editions

abstract

actors   person   institution

editorial realisation

type of edition

components of the edition

edendum or edenda

medium   specification   period

language(s)   writing system

originator(s)   work

technologies

data formats (in creation, presentation, and provision)

software

further technological features

additional information

tags, keywords   free text

Review available (e.g. in RIDE)?

Administrative

Information

provenance

part of the Text+ portfolio

Registry Metadata

Resource (latest version)

Displayed version

Version timestamp

Creator of the version   Versions

Resource created

Creator of the resource

Resource layers

FAIR & RDM

RDM information provided?

Use of PIDs?

Use of common standards?

Use of authority files?

Availability of research data via a certified repository?

Availability of research data (via download)?

Accessibility of research data via APIs?

Licence allows extensive reuse?

Website is in accordance with current web accessibility standards?

Cataloguing editions and other resources in one unified system – The case of the Text+ registry (DHC Sheffield 04/05 Sept 2024)

Model-driven architecture

2	0	basic information	
2.1	1	IDs	0..n
2.2	1	title	1
2.2.1	2	main title	1
2.2.2	2	short title	0..1
2.3	1	year(s) of publication	1..n

```
entity: edition
blocks:
  - name: basic_information
    properties:
    - field: title
      type: text
      mandatory: true
      multiple: false
      multilang: true
```

basic information

main title   short title   IDs

year(s) of publication   links

APIs   discipline(s) (DFG)

Notes on disciplines

funding period

basic information

main title	Edition of the Frankish Capitularies eng
	Edition der fränkischen Herrschererlasse deu
	AGATE — Edition der fränkischen Herrschererlasse eng / deu
short title	Capitularia lat
IDs Modify	PR.130
	II.D.40
	AGATE PR.130 — II.D.40
year(s) of publication	2015-
links	http://capitularia.uni-koeln.de/ Edition Präsentationsschicht
	https://capitularia.uni-koeln.de/project/tech/ Dokumentation
	https://capitularia.uni-koeln.de/resources/downloads/ Forschungsdaten Dokumentation Editionsrichtlinien Schemadefinition
	https://cceh.github.io/capitularia/index.html Dokumentation Schemadefinition

Text+

basic information

main title
Edition of the Frankish Capitularies `eng`

Edition der fränkischen Herrschererlasse `deu`

`AGATE` — Edition der fränkischen Herrschererlass `deu`

short title
Capitularia `lat`

IDs `Modify`
PR.130

II.D.40

`AGATE` PR.130 — II.D.40

year(s) of publication
2015-

links
http://capitularia.uni-koeln.de/ Edition Präsentationsschicht

https://capitularia.uni-koeln.de/project/tech/ Dokumentation

https://capitularia.uni-koeln.de/resources/downloads/ Forschungsdat Dokumentation Editionsrichtlinien Schemadefinition

https://cceh.github.io/capitularia/index.html Dokumentation Schemadefinition

https://agate.academy/id/PR130 Projektbeschreibung

https://www.awk.nrw/fraenkische-herrschererlasse-was-der-kaiser-schrieb-und-

output type

type
Digitale Edition (Portal, Webpräsentation)

citation
Capitularia. Edition der fränkischen Herrsch bearb. von Karl Ubl und Mitarb., Köln 2014 https://capitularia.uni-koeln.de

language of output type
Deutsch (deu)

Englisch (eng)

information on accessibility
Frei

licence `Modify`
CC BY-SA

`AGATE` CC BY-SA

relations

series

series number

Relations to other editions

abstract `Modify`
The decrees of the Frankish rulers are know capitularies because of their subdivision in capitula). They are amongst the most impo for the history of the Frankish ...

Die fränkischen Herrscherlasse, nach der Kapitel (lat. capitula) auch Kapitularien ger

actors

person

Ubl; Karl; https://d-nb.info/gnd/122233581 — Leiter (led)

Mischke; Britta; Diss., Uni Bonn, 2013; https://d-nb.info/gnd/1033598992; Mischke Forschungsteammitglied (rtm)

Kaschke; Sören; https://d-nb.info/gnd/1203231466; Kaschke — Forschungsteammitglied (rtm)

Trump; Dominik; https://d-nb.info/gnd/1166880516; Trump — Forschungsteammitglied (rtm)

Schulz; Daniela; 08/2010 – 09/2010: Mita im Projekt „Rubens online" (Wolfenbütteler Digitale Bibliothek) an der Herzog August Bibliothek, Wolfenbüttel ; 05/2011 – 08/2011 Mitarbeit im „Saint Patrick's Confessio Hyper Stack Project" an der R... — Forschungsteammitglied (rtm) Datenmodellierung, XML-Prozessierung, Encoding

Gregorii Borisov — Forschungsteammitglied (rtm)

Marcello Perathoner — Andere (oth) Developer

Nils Geißler — Andere (oth) Developer XML Processing

Martina Gödel — Schemadefinition

Henny-Krahmer; Ulrike; Ulrike Henny-Krah ist Mitglied der Nachwuchsgruppe „Computergestützte Literarische Gattungsstilistik" an der Universität Würzbur und promoviert an der Graduate School of th Humanities zu „Topic and Style in Subgenres

information of the edendum or edenda

medium
Text

specification
Druck

Handschrift

period
Fränkisches Reich

Europa

Frühmittelalter

500 v. Chr. bis 950 n. Chr.

Mittelalter

`AGATE` Fränkisches Reich — Europa — Frühmittelalte

language(s)
Latein (lat)

writing system
Latein (215) (Latn)

originator(s)

work

editorial realisation

type of edition
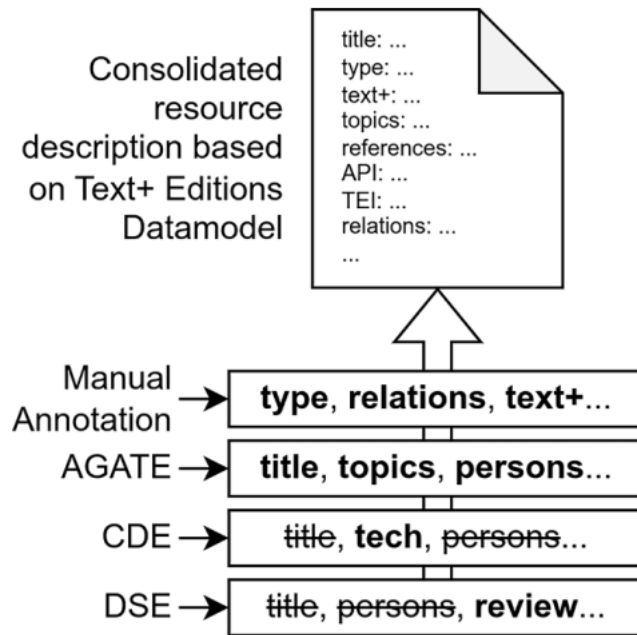Diplomatische Edition

Gesamtausgabe

Minimal entry

Very few mandatory categories

1.1	1	type		1	book version (print/ebook), digital edition (portal, webpresentation), hybrid, research data	Please select the output form of the edition. If there are different forms of edition, you can enter these in separate entries and then link these entries via "Relations".	
2.2	1	title		1			
2.2.1	2		main title	1		Enter the language for the main title. If the main title is available in several languages, make an entry in each language.	3. Title — DataCite Metadata Schema 4.5 documentation (datacite-metadata-schema.readthedocs.io)
2.3	1	year(s) of publication		1..n		For print editions, enter the year of publication; for multi-volume works, enter the period. For digital editions, enter the period from the publication of the first version to the (planned) completion of the edition project. If the edition described here consists of several parts, enter the time span for all parts. Enter the year in the format YYYY, a time period in the format YYYY-YYYY. If the edition project has not yet been completed, you can enter YYYY-	5. PublicationYear — DataCite Metadata Schema 4.5 documentation (datacite-metadata-schema.readthedocs.io) bzw. 8. Date — DataCite Metadata Schema 4.5 documentation (datacite-metadata-schema.readthedocs.io)
2.5	1	discipline(s)		1..n	based on DFG subject classification ("Fachebene"), DDC, RVK, free text	Please select the subject area(s) to which the edition relates. If the default selection is not sufficient, you can also enter your own subject term.	
4.1.1	2		selection	1..n	audio, image, film, musical notation, text		
4.5	1	period		1..n		Please enter the period in which the objects of the edition were created (if possible in the format YYYY or YYYY-YYYY).	8. Date — DataCite Metadata Schema 4.5 documentation (datacite-metadata-schema.readthedocs.io)

Information layering

	AGATE	CDE	DSE	WeGA
Title	+	+	++	++
Abbreviation, Acronym	+	−	(+)	(+)
Description	++	−	+	++
Actors	++	+	(+)	++
Institutions	+	++	(+)	(+)
Date (Range) of Publication	+	−	+	+
Reviews	−	−	+	−
Catalogue ID s	−	++	+	−
URLs	++	+	+	+
API available	−	+	(+)	(+)
Contact Info	−	−	+	+

Consolidated resource description based on Text+ Editions Datamodel

title: ...
type: ...
text+: ...
topics: ...
references: ...
API: ...
TEI: ...
relations: ...
...

Manual Annotation → **type, relations, text+**...

AGATE → **title, topics, persons**...

CDE → ~~title~~, **tech**, ~~persons~~...

DSE → ~~title~~, ~~persons~~, **review**...

Challenges, approaches, future

- Identification of relevant resources remains an issue

- Time and effort

- Various models and degrees of granularity → layering/metamodelling

- sustainability

- Engaging the community → Edit-a-thons

- Ideas?

Thank you!

„Edit" (Mascot of Task Area Editions)

17

Cataloguing editions and other resources in one unified system – The case of the Text+ registry (DHC Sheffield 04/05 Sept 2024)