

# Modellprojekt „eHumanities - interdisziplinär“

Forschungsdatenmanagement im Rahmen  
des „Digitalen Campus Bayern“

Sonja Kümmer, Universitätsbibliothek der Ludwig-Maximilians-Universität München  
Jürgen Rohrwild, Universitätsbibliothek der Friedrich-Alexander-Universität Erlangen-Nürnberg

## eHumanities – interdisziplinär

### Förderung:

- Bayerisches Staatsministerium für Wissenschaft und Kunst
- Laufzeit Januar 2018 – Dezember 2020



- Universitätsbibliothek der Friedrich-Alexander-Universität Erlangen-Nürnberg (**UB FAU**) [*federführend*]
- Universitätsbibliothek der Ludwig-Maximilians-Universität München (**UB LMU**)
- IT-Gruppe Geisteswissenschaften (**ITG**) der Ludwig-Maximilians-Universität München

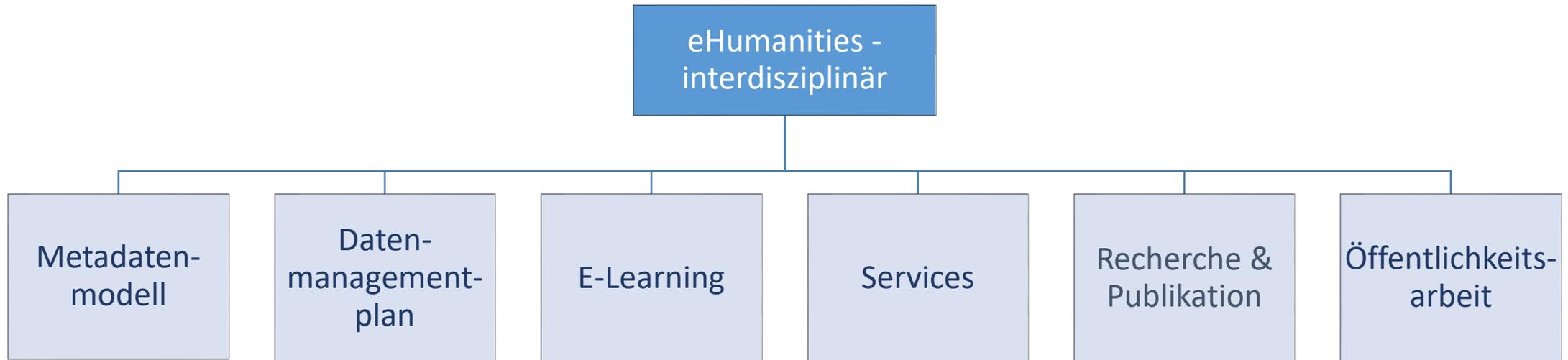
# IT-Gruppe Geisteswissenschaften (ITG)



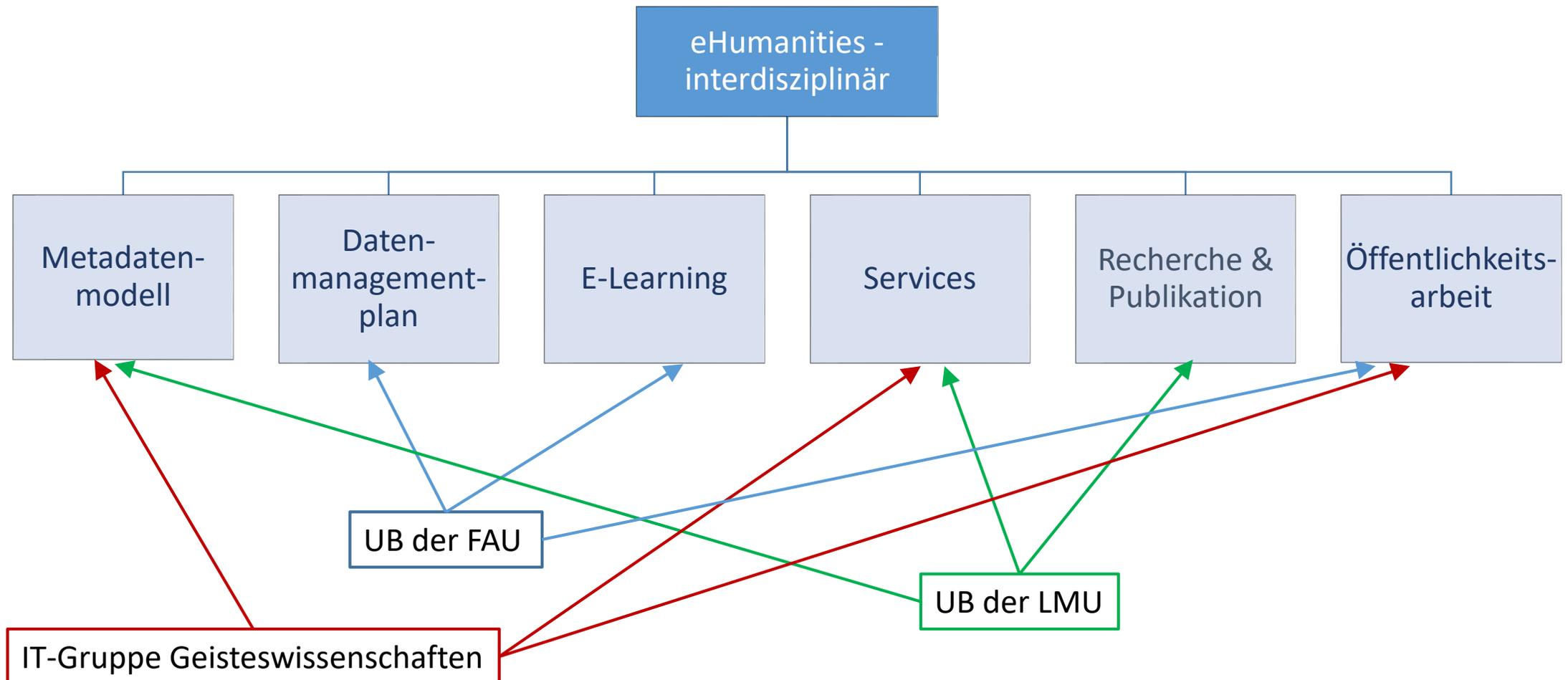
- Betrieb der IT-Infrastruktur an den geisteswissenschaftlichen Fakultäten
- Forschung & Lehre digital: Lehrangebote in den Digital Humanities und Projektpartner in DH-Projekten ab ovo
- Forschungsdatenmanagement: Aufbau domänenspezifischer Services im Rahmen des Modellvorhabens

IT-Gruppe Geisteswissenschaften  
**LMU Center for Digital Humanities**





# Verteilung der Arbeitspakete



Es soll ein auf den **FAIR-Prinzipien**<sup>1</sup> basierendes Metadatenmodell entwickelt werden, mit dem sich Forschungsdaten aus den Digital Humanities **einheitlich** beschreiben lassen.

Um die Auffindbarkeit und Nachnutzung von Forschungsdaten bestmöglich zu unterstützen, soll das Metadatenmodell neben **formalen** auch **inhaltliche Metadaten** umfassen, um eine möglichst tiefe inhaltliche Erschließung der Daten zu ermöglichen.

1) <https://www.go-fair.org/fair-principles/>

# DataCite – Beispieldatensatz

```

<?xml version="1.0" encoding="UTF-8"?>
<resource xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://datacite.org/schema/kernel-4" xsi:schemaLocation="
http://datacite.org/schema/kernel-4 http://schema.datacite.org/meta/kernel-4.1/metadata.xsd">
  <identifier identifierType="DOI">10.5072/D3P26Q35R-Test</identifier>
  <creators>
    <creator>
      <creatorName nameType="Personal">Fosmire, Michael</creatorName>
      <givenName>Michael</givenName>
      <familyName>Fosmire</familyName>
    </creator>
    <creator>
      <creatorName nameType="Personal">Wertz, Ruth</creatorName>
      <givenName>Ruth</givenName>
      <familyName>Wertz</familyName>
    </creator>
  </creators>
  <titles>
    <title xml:lang="en">Critical Engineering Literacy Test (CELT)</title>
  </titles>
  <publisher>Purdue University Research Repository (PURR)</publisher>
  <publicationYear>2013</publicationYear>
  <subjects>
    <subject xml:lang="en">Assessment</subject>
    <subject xml:lang="en">Information Literacy</subject>
  </subjects>
  <language>en</language>
  <resourceType resourceTypeGeneral="Dataset">Dataset</resourceType>
  <version>1.0</version>
  <descriptions>
    <description xml:lang="en" descriptionType="Abstract">
      We developed an instrument, Critical Engineering Literacy Test (CELT), which is a multiple choice instrument designed to measure
      undergraduate students' scientific and information literacy skills. It requires students to first read a technical memo
      and, based on the memo's arguments, answer eight multiple choice and six open-ended response questions. We collected data from 143
      first-year engineering students and conducted an item analysis. The KR-20 reliability of the instrument was .39. Item
      difficulties ranged between .17 to .83. The results indicate low reliability index but acceptable levels of item difficulties and item
      discrimination indices. Students were most challenged when answering items measuring scientific and mathematical
      literacy (i.e., identifying incorrect information).
    </description>
  </descriptions>
</resource>

```

# DataCite – Kritik

```

<?xml version="1.0" encoding="UTF-8"?>
<resource xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://datacite.org/schema/kernel-4" xsi:schemaLocation="
http://datacite.org/schema/kernel-4 http://schema.datacite.org/meta/kernel-4.1/metadata.xsd">
  <identifier identifierType="DOI">10.5072/D3P26Q35R-Test</identifier>
  <creators>
    <creator>
      <creatorName nameType="Personal">Fosmire, Michael</creatorName>
      <givenName>Michael</givenName>
      <familyName>Fosmire</familyName>
    </creator>
    <creator>
      <creatorName nameType="Personal">Wertz, Ruth</creatorName>
      <givenName>Ruth</givenName>
      <familyName>Wertz</familyName>
    </creator>
  </creators>
  <titles>
    <title xml:lang="en">Critical Engineering Literacy Test (CELT)</title>
  </titles>
  <publisher>Purdue University Research Repository (PURR)</publisher>
  <publicationYear>2013</publicationYear>
  <subjects>
    <subject xml:lang="en">Assessment</subject>
    <subject xml:lang="en">Information Literacy</subject>
  </subjects>
  <language>en</language>
  <resourceType resourceTypeGeneral="Dataset">Dataset</resourceType>
  <version>1.0</version>
  <descriptions>
    <description xml:lang="en" descriptionType="Abstract">
      We developed an instrument, Critical Engineering Literacy Test (CELT), which is a multiple choice instrument designed to measure
      undergraduate students' scientific and information literacy skills. It requires students to first read a technical memo
      and, based on the memo's arguments, answer eight multiple choice and six open-ended response questions. We collected data from 143
      first-year engineering students and conducted an item analysis. The KR-20 reliability of the instrument was .39. Item
      difficulties ranged between .17 to .83. The results indicate low reliability index but acceptable levels of item difficulties and item
      discrimination indices. Students were most challenged when answering items measuring scientific and mathematical
      literacy (i.e., identifying incorrect information).
    </description>
  </descriptions>
</resource>

```

✓ Gute Abdeckung der formalen Metadaten

✓ „Quasi-Standard“

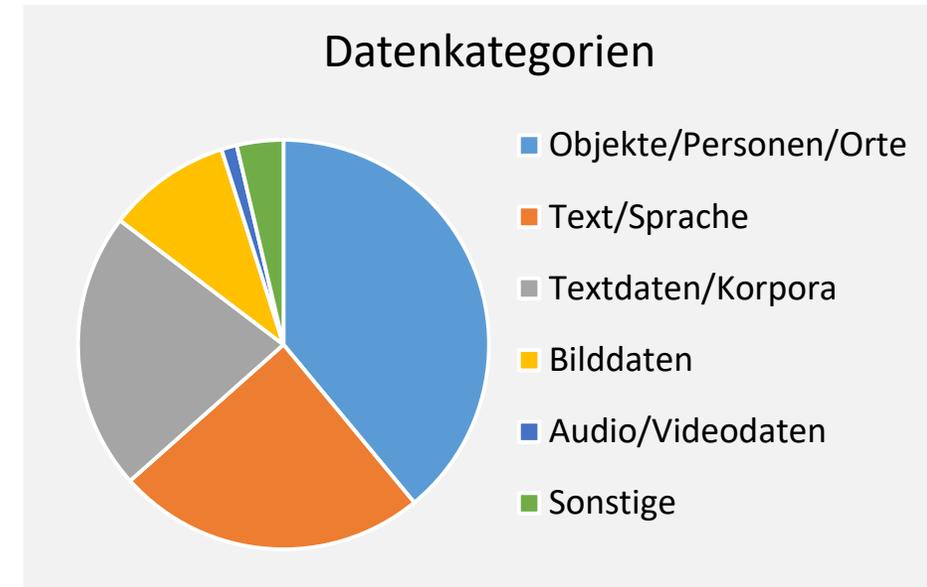
– Nur zwei Felder (subject, description) für die inhaltliche Beschreibung

```
<?xml version="1.0" encoding="UTF-8"?>
<resource xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://datacite.org/schema/kernel-4" xsi:schemaLocation="
http://datacite.org/schema/kernel-4 http://schema.datacite.org/meta/kernel-4.1/metadata.xsd">
  <identifier identifierType="DOI">10.5072/D3P26Q35R-Test</identifier>
  <creators>
    <creator>
      <creatorName nameType="Personal">Fosmire, Michael</creatorName>
      <givenName>Michael</givenName>
      <familyName>Fosmire</familyName>
    </creator>
    <creator>
      <creatorName nameType="Personal">Wertz, Ruth</creatorName>
      <givenName>Ruth</givenName>
      <familyName>Wertz</familyName>
    </creator>
  </creators>
  <titles>
    <title xml:lang="en">Critical Engineering Literacy Test (CELT)</title>
  </titles>
  <publisher>Purdue University Research Repository (PURR)</publisher>
  <publicationYear>2013</publicationYear>
  <subjects>
    <subject xml:lang="en">Assessment</subject>
    <subject xml:lang="en">Information Literacy</subject>
  </subjects>
  <language>en</language>
  <resourceType resourceTypeGeneral="Dataset">Dataset</resourceType>
  <version>1.0</version>
  <descriptions>
    <description xml:lang="en" descriptionType="Abstract">
      We developed an instrument, Critical Engineering Literacy Test (CELT), which is a multiple choice instrument designed to measure
      undergraduate students' scientific and information literacy skills. It requires students to first read a technical memo
      and, based on the memo's arguments, answer eight multiple choice and six open-ended response questions. We collected data from 143
      first-year engineering students and conducted an item analysis. The KR-20 reliability of the instrument was .39. Item
      difficulties ranged between .17 to .83. The results indicate low reliability index but acceptable levels of item difficulties and item
      discrimination indices. Students were most challenged when answering items measuring scientific and mathematical
      literacy (i.e., identifying incorrect information).
    </description>
  </descriptions>
</resource>
```

Idee: Erweiterung von DataCite um ein inhaltserschließendes Metadatenmodell.

Ca. **85 ITG-Projekte**, die für eine detailliertere Untersuchung in Frage kommen.

Diese basieren überwiegend auf **relationalen Datenbanken**.



## Beispielprojekt: **Verba Alpina**

(<https://www.verba-alpina.gwi.uni-muenchen.de/>)

- Ziel: Sprach- und Kulturraum des gesamten Alpengebiets multidimensional zugänglich machen
- Methode: Crowdsourcing

# Forschungsdaten in Bibliothekskatalogen?



**Umfrage unter Forschenden in den digitalen Geisteswissenschaften**

# Mapping DataCiteToMARC – Zwischenfazit

**Grundsätzlich lassen sich die meisten Informationen aus DataCite gut in das wesentlich umfangreichere MARC übertragen; Ausnahmen/Schwierigkeiten:**

- Nur ein kleiner Teil der contributorType-Auswahlliste wird von den Beziehungskennzeichen abgedeckt
- Lassen sich in MARC Informationen zur Projektfinanzierung (funding) unterbringen?
- Aktuell ist noch fraglich, inwieweit sich die z.T. komplexen Beziehungen zwischen Forschungsdaten in MARC abbilden lassen.

	A	B	C	D	E	F	G
1	DataCite (= Quellformat)			MARC (= Zielformat)			
2	Properties	Häufigkeit	DataCite-Beispiel	Felder	Indikator	Unterfeld	Feldinhalt
3	resource						
4	identifier	1	10.5072/example-full	024		a	
5	identifierType	1	DOI	024	Ind.1=7, Ind.2=#	2	doi
6	creators						
7	creator	1-n		100		4	cre
8	creatorName	1	Miller, Elizabeth	100		a	
9	nameType	0-1	Personal				
10	givenName	0-1	Elizabeth	245		c	[100], [100], [100]; [700], [700]
11	familyName	0-1	Miller	245		c	
12	nameIdentifier	0-n	0000-0001-5000-0007	100		0	
13	schemeURI	0-1	<a href="http://orcid.org/">http://orcid.org/</a>				
14	nameIdentifierScheme	1	ORCID				
15	affiliation	0-n	DataCite				
16	titles						
17	title	1-n	Full DataCite XML Example	245		a	
18	xml:lang		en-US				
19	titleType	0-1	Subtitle	245		b	
20	titleType	0-1	AlternativeTitle	246	Ind.1=1, Ind.2=3	a	
21	titleType	0-1	TranslatedTitle	246	Ind.1=1, Ind.2=1	a	
22	titleType	0-1	Other	246	Ind.1=1, Ind.2=3	a	
23	publisher	1	DataCite	264		b	
24	publicationYear	1	2014	264		c	
25	subjects						
26	subject	0-n	000 computer science	082	Ind.1=0 Ind.2=#	a	000
27	xml:lang		en-US				
28	schemeURI	0-1	<a href="http://dewey.info/">http://dewey.info/</a>				
29	subjectScheme	0-1	dewey				
30	subject	0-n	LB 60015	084	Ind.1=# Ind.2=#	a	LB 60015



## Projekte (Auswahl)



20.09.2018



Universitätsbibliothek  
**OPEN DATA LMU**

„Datenaggregator“



# GeRDI

„Datenbroker“

- Ziel: virtuelle Verknüpfung von Forschungsdatenspeichern in ganz Deutschland
- Projektpartner: ZBW, LRZ, TU Dresden, DFN-Verein, CAU Kiel
- Gefördert durch die DFG
- Projektlaufzeit: 2016-2019
- [www.gerdi-project.de](http://www.gerdi-project.de)

# Datenfluss



Projekte (Auswahl)

Pilotprojekt



20.09.2018



Universitätsbibliothek  
**OPEN DATA LMU**

„Datenaggregator“

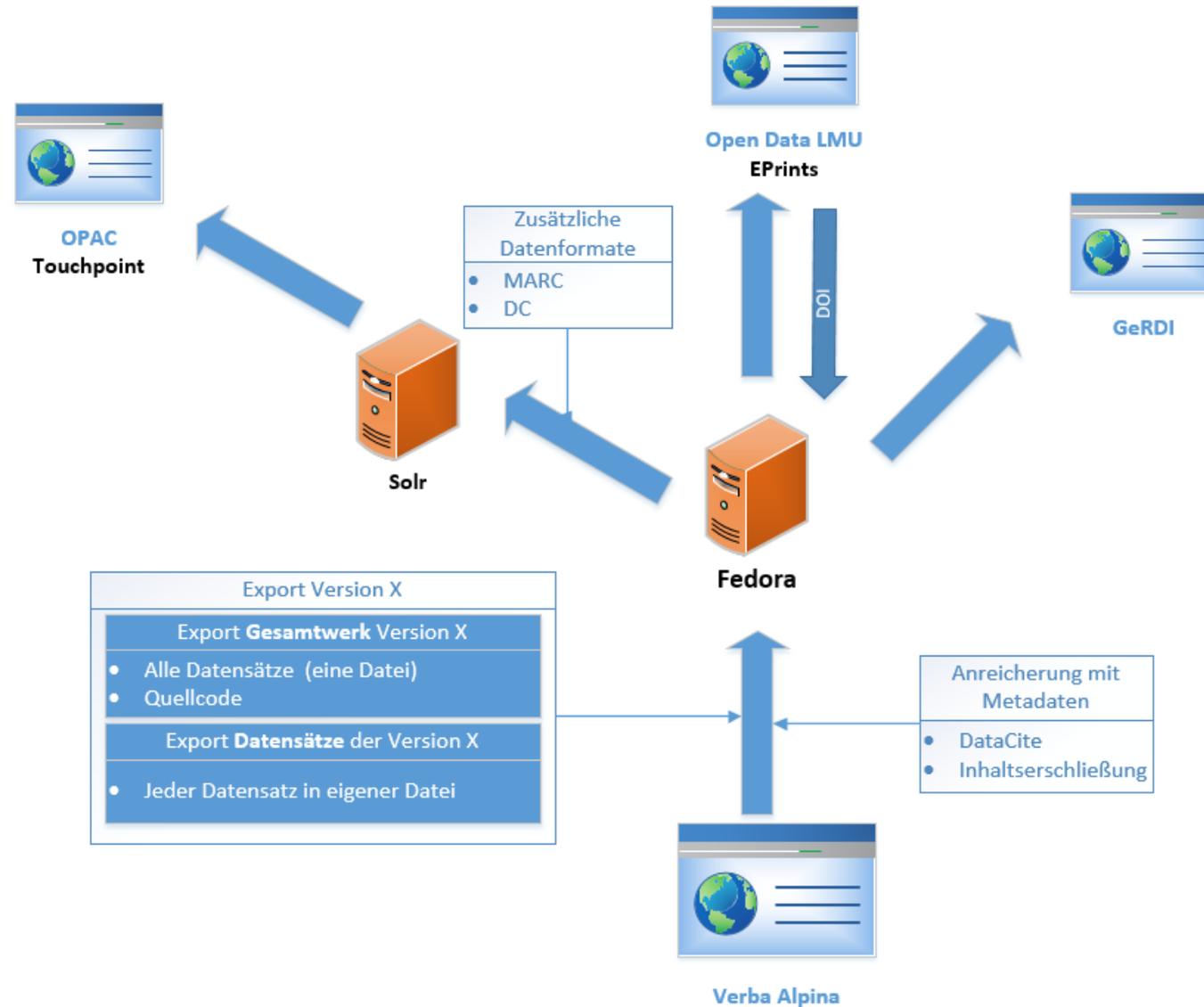


# GeRDI

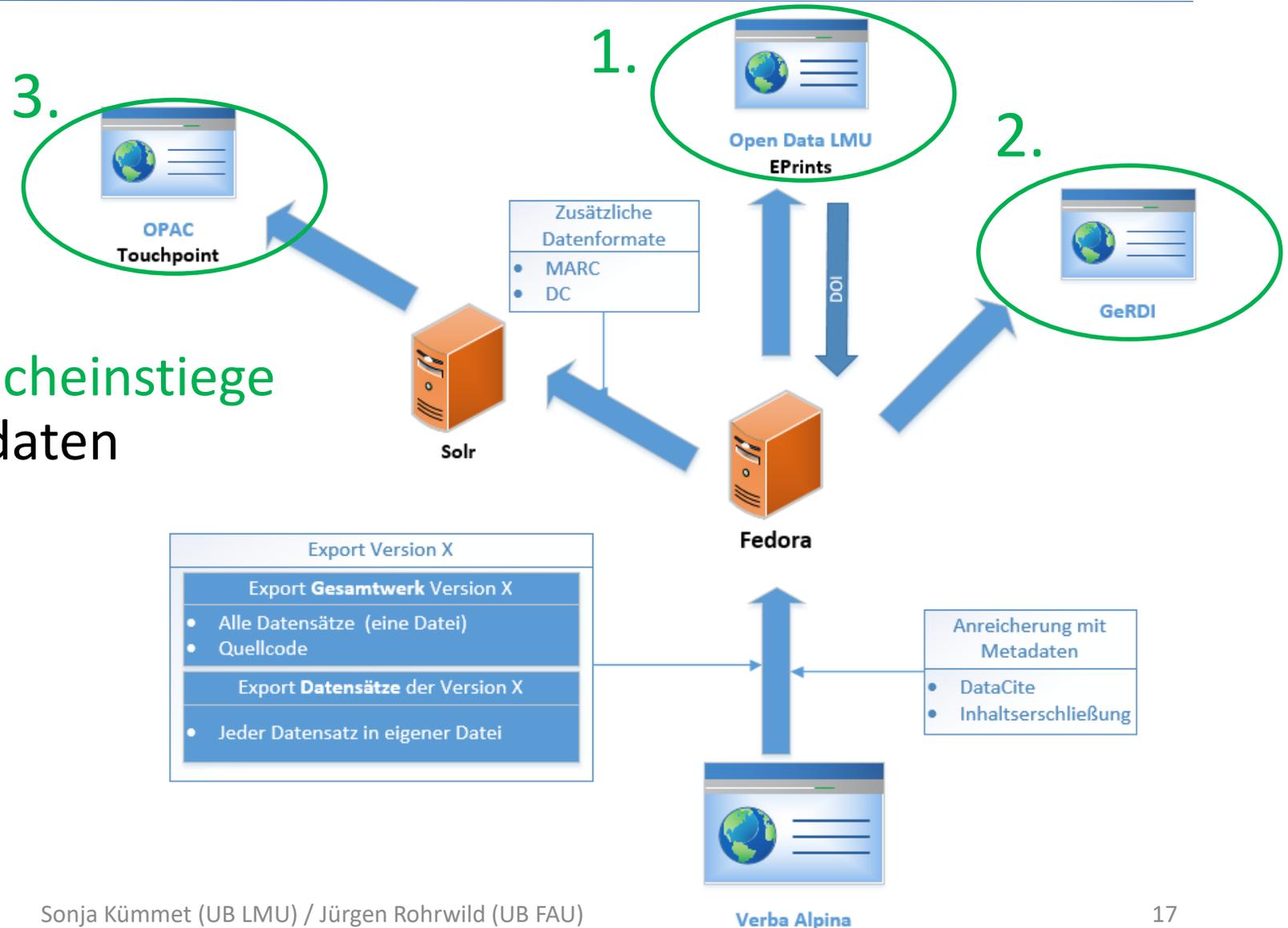
„Datenbroker“

- Ziel: virtuelle Verknüpfung von Forschungsdatenspeichern in ganz Deutschland
- Projektpartner: ZBW, LRZ, TU Dresden, DFN-Verein, CAU Kiel
- Gefördert durch die DFG
- Projektlaufzeit: 2016-2019
- [www.gerdi-project.de](http://www.gerdi-project.de)

# Services für das FDM - Architekturskizze



# Tools für Recherche und Publikation



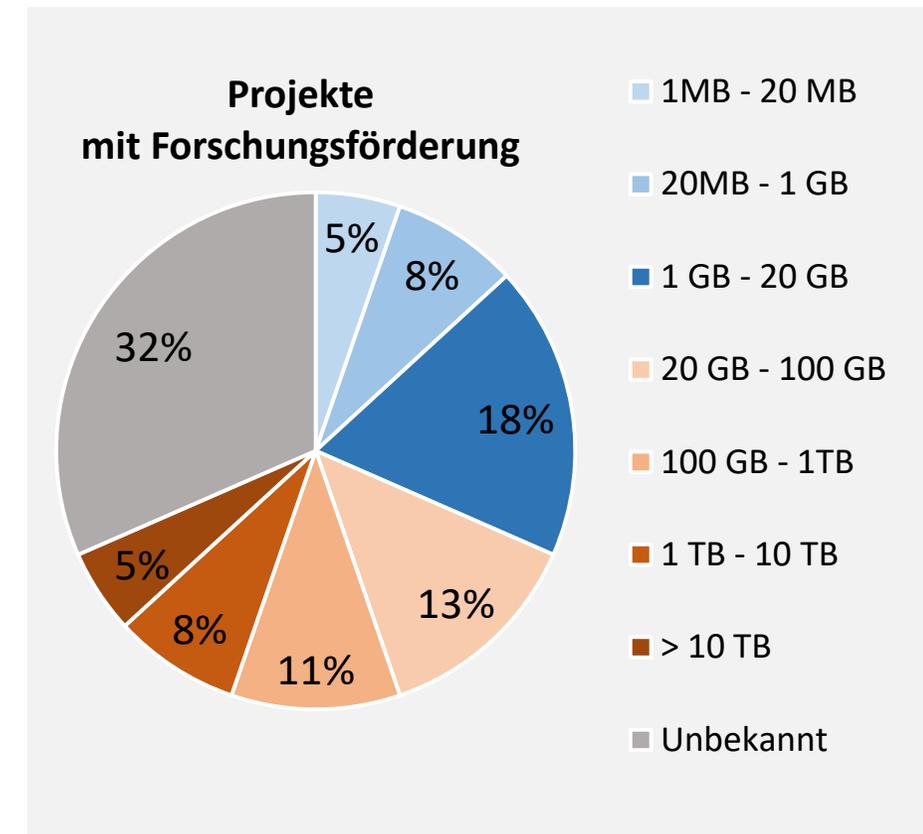
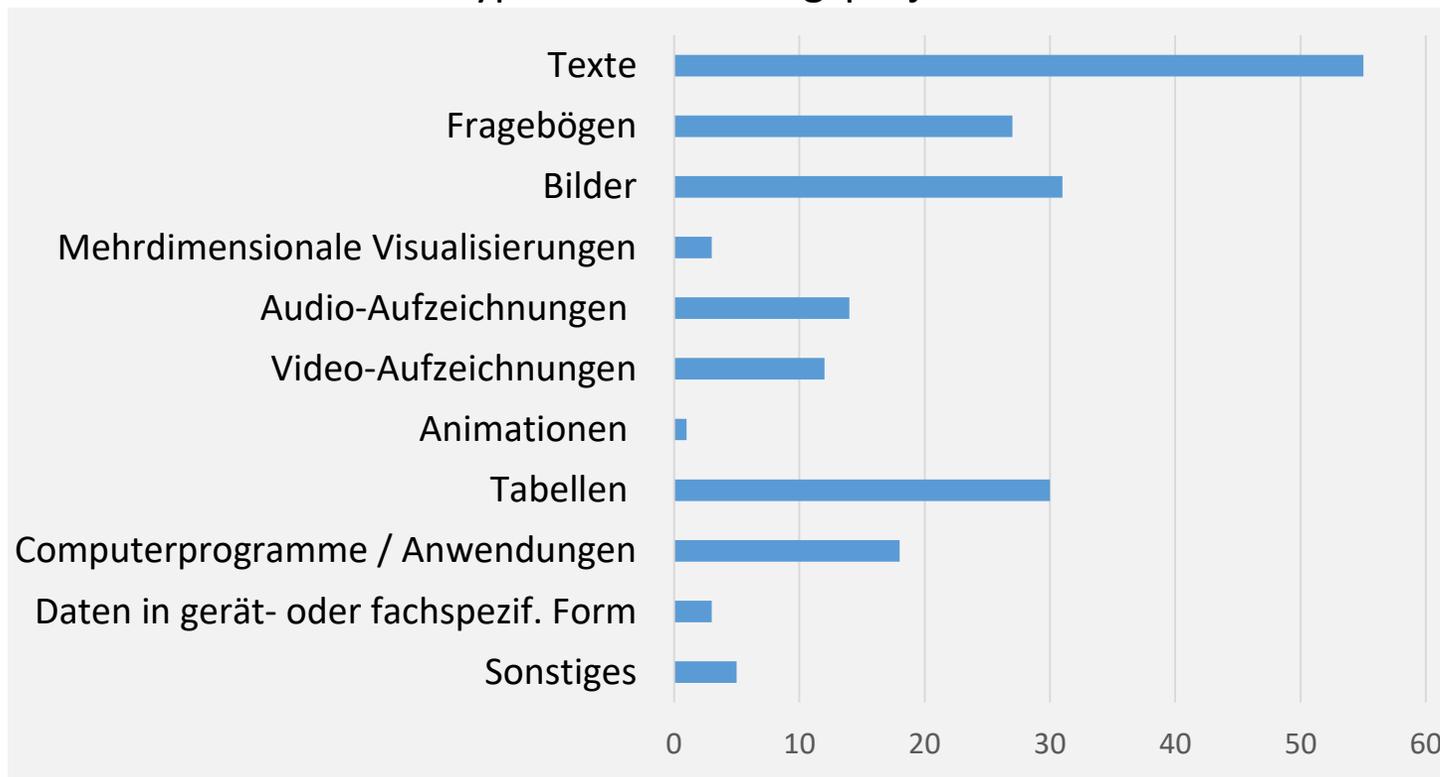
Aktuell planen wir 3 **Sucheinstiege** für unsere Forschungsdaten anzubieten:

1. Open Data LMU
2. GeRDI
3. OPAC

# Vorarbeiten an der FAU

- Umfrage unter Forschenden in den digitalen Geisteswissenschaften
- Interviews mit Dozenten des Studiengangs

Datentypen in Forschungsprojekten



## Datenmanagementpläne:

Werkzeug zur Planung des Datenhandling über den gesamten Forschungsprozess

- frühzeitiges Erkennen von Hindernissen (Datenschutzauflagen, Speicherkapazitäten, Personalmittel,...)
- vereinfachte Zusammenarbeit der Forscher (gleiche Strukturen, dokumentierte Bezeichnungen, ...)
- Nachhaltigkeit

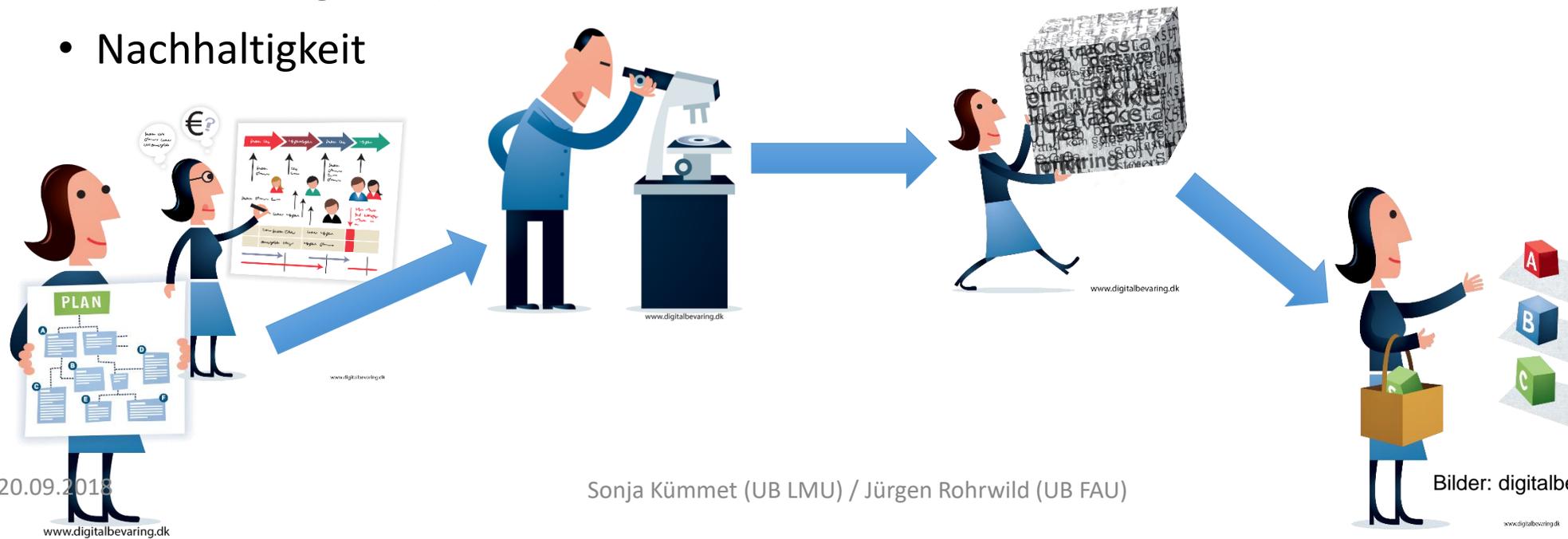


# Transferierbare Datenmanagementpläne

## Datenmanagementpläne:

Werkzeug zur Planung des Datenhandling über den gesamten Forschungsprozess

- frühzeitiges Erkennen von Hindernissen (Datenschutzauflagen, Speicherkapazitäten, Personalmittel,...)
- vereinfachte Zusammenarbeit der Forscher (gleiche Strukturen, dokumentierte Bezeichnungen, ...)
- Nachhaltigkeit



# Transferierbare Datenmanagementpläne

## Datenmanagementpläne:

Reagieren auf steigende Anforderungen von Forschungsförderern



# DMP - Anforderungen

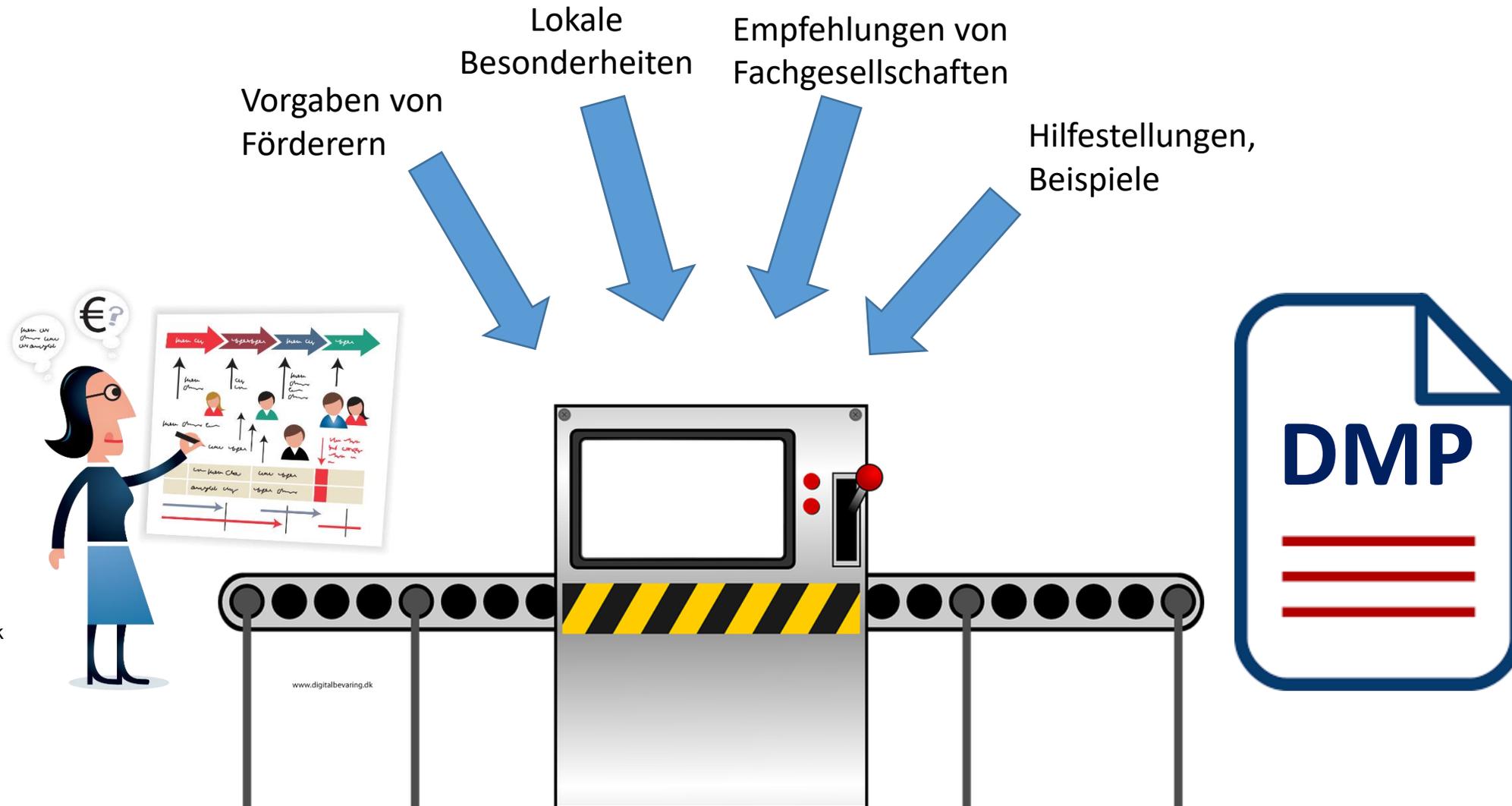


Bild:  
digitalbevaring.dk  
CC BY 2.5 Denmark

## Evaluierung bestehender Softwaretools

### Anforderungen Wissenschaftler:

- 90% kooperatives Arbeiten
- 85% Ansprechpartner vor Ort
- 65% auf Server der eigenen Universität
- ...

### Anforderungen Bibliothek:

- Export / Import-Funktionen (Schnittstellen)
- schnelles Anpassen / Erstellen von Vorlagen
- Softwarelizenz
- multilingual
- ...

Knapp 40 (gewichtete)  
Kriterien

## Zwei Gruppen von Werkzeugen:

- **spezialisierte Tools einzelner Datenzentren**, wie der Clarin-D Wizard oder GFBio DMPT, gut für sehr spezielle Anwendung, aber wenig flexibel
- **Allgemeine Tools**, davon zwei genauer untersucht: [roadmap](#) und [RDMO](#)
  - vergleichbarer Funktionalitätsumfang
  - Unterschiede oft begründet in der unterschiedlichen „Philosophie“
  - *roadmap* erlaubt sehr schnelles Erstellen von Förderer-spezifischen Vorlagen
  - *RDMO* scheint besser für komplexere Pläne geeignet (jenseits von Förderbedingungen)

roadmap

RDMO

# DMP - Ergebnisse (bisher)

- **Einbau des Default-RDMO-Plans** in roadmap und anders herum
- **deutsche und englischsprachige Hilfstexte** zu den 14 *DCC themes* (*data format, intellectual property rights, ...*), die Basis für schnelle Erstellung von Vorlagen in roadmap (zunächst mit Spezialisierung auf die FAU)

## FAU

- └  History, History of Art, Archeology and Classical & Ancient Studies
- └  FAU General Guidance
- └  Social and Economic Sciences
- └  Educational Sciences and Psychology
- └  Philologies
- └  Allgemeine FAU Hilfestellungen

# DMP - Ergebnisse (bisher)

## Fischereiwesen in der Steinzeit

Project Details Plan overview Write Plan Share Download

expand all | collapse all

2/13 answered

Data Collection (2 / 2)

Documentation and Metadata (0 / 1)

What documentation and metadata will accompany the data?

**B** *I*    

Save

Guidance

Comments

FAU General Guid

Social and Econo

Educational Scie

Philologies

Metadata & Documentation

Common metadata standards are

- The XML-based [TEI P5 standard](#) is the most common format for the detailed semantic markup of texts, esp. for editions.
- [MEL \(Music Encoding Initiative\)](#) is the

# DMP - Ergebnisse (bisher)

## Fachspezifische Anpassung des RDMO-Plans:

- fachliche Formate / Policies / Zentren
- Beispiele aus den Geisteswissenschaften
- Bis jetzt in vier „Klassen“ (wie roadmap):
  - Sprach- und Literaturwissenschaften
  - Psychologie & Erziehungswissenschaften
  - Sozialwissenschaften
  - Altertumswissenschaften/(Kunst-) Geschichte
- erstes Feedback zeigt aber: noch zu grob

## Fragebogen für Projekt *eHumanities* - ein Testplan

### Allgemein / Weitere Anforderungen

Gibt es von weiteren Seiten (z. B. von der Fachcommunity) Anforderungen an das Datenmanagement, die beachtet werden müssen?

Forschungsethik und Datenschutzes werden in späteren Fragen noch genauer thematisiert. Es bietet sich aber an entsprechende Leitlinien (beispielsweise der verschiedenen Fachgesellschaften) bereits hier zu erwähnen.

Einige Beispiele für fachspezifische Empfehlungen und Richtlinien:

Sprachen & Literaturwiss. Sozialwissenschaften Psychologie/Erziehungswiss.

Geschichte/Kunstgeschichte/Archäologie

Sowohl für die Psychologie als auch für die Erziehungswissenschaften hat die DFG gesonderte, zusätzlich Förderbedingungen bzw. Empfehlungen:

- Memorandum des Fachkollegiums „Erziehungswissenschaft“ der DFG
- Empfehlungen zu Umgang mit Forschungsdaten in der Psychologie (DGP Vorstand) und der Kommentar des DFG Fachkollegs "Psychologie" zu diesen Empfehlungen
- DFG Richtlinien für Replikationsstudien in der Psychologie

Einzelne Förderlinien haben manchmal eigene, strengere Auflagen, die die allgemeinen Anforderungen der Policies konkretisieren oder erweitern. Dies ist häufig der Fall bei BMBF-Programmen, wie beispielsweise dem Programm "Bildung durch Sprache und Schrift (BISS)".

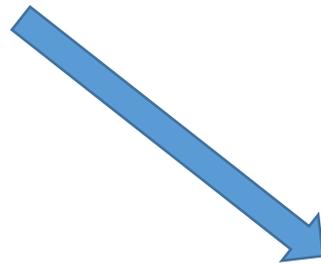
Der Berufsverband Deutscher Psychologinnen und Psychologen und die Deutschen Gesellschaft für Psychologie haben darüber hinaus [Richtlinien zur \(Berufs-\)Ethik und zu Persönlichkeitsrechten](#). Die American Psychological Association verfügt über einen ähnlichen [Ethikkodex](#).

Die Deutsche Gesellschaft für Erziehungswissenschaft (DGfE) verfügt ebenfalls über einen [Ethik-Kodex](#).

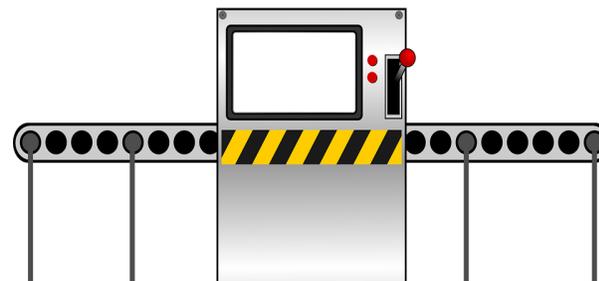
Ja

# Nächste Schritte I

## Schnittstellen:



XML-Format



**Nachnutzen der (Meta-) Informationen aus dem Datenmanagementplan**

→ Export von Metadaten in erweitertem DataCite-Format

→ kann beispielsweise als Grundlage für Metadaten im Datenrepositorium dienen

**Ziel:**

Doppelarbeit für Forscher vermeiden

# Nächste Schritte II

---

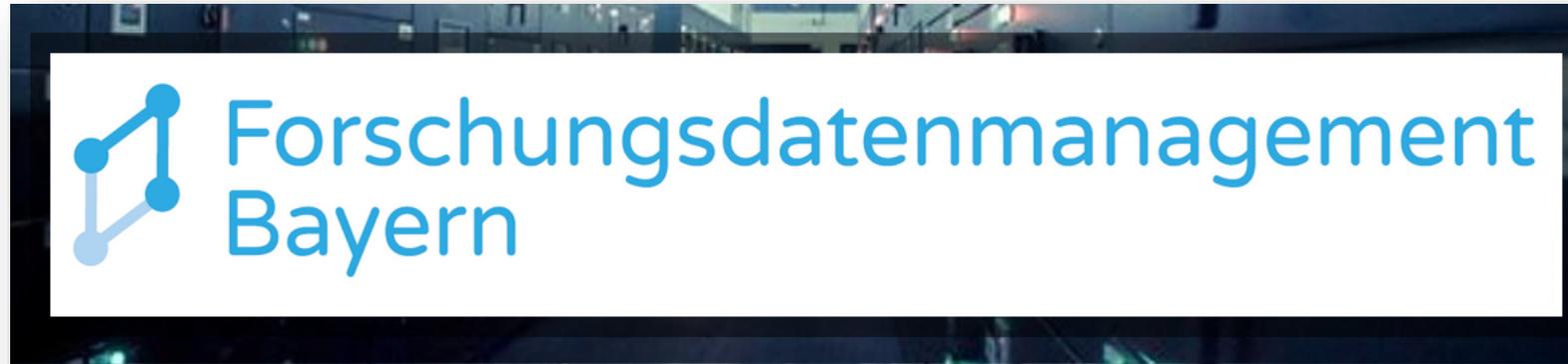
- Mapping-Tool für Templates, um einrichtungsspezifische Informationen automatisch einbauen zu können, etwa
  - Datenschutzbeauftragter mit Kontaktdaten
  - Patentstelle
  - Repository
  - Daten-Policy
  - ...
  
- Am wichtigsten: Nutzertests mit den Tools für weiteres Feedback!

## Steht noch am Anfang:

- Anforderungen der Dozenten des DH-Studiengangs
- Kooperation mit dem Institut für Lern-Innovation (ILI) der FAU

## Ziel: OER für BAB, VHB, ...

- fachübergreifende Module (Datenstrukturierung, ...)
- DH-spezifische Module (z.B. „Einsatz/Erzeugen von Ontologien/kontrollierten Vokabular“, ...)
- Module zu Werkzeugen/Standards aus den anderen AP



Forschungsdatenmanagement

eHumanities – interdisziplinär

Tools

Policies

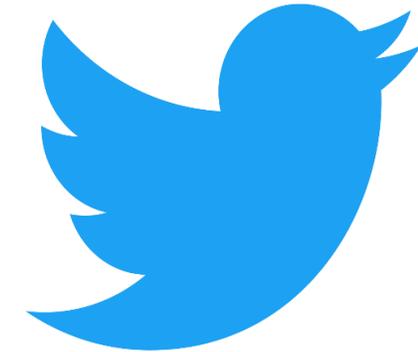
## Startseite

Im Rahmen des vom [Bayerischen Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst](#) geförderten Projektes eHumanities – interdisziplinär entsteht [Forschungsdatenmanagement-Bayern.org](#).

Die Webseite soll nicht nur dem Projekt [eHumanities – interdisziplinär](#) dienen, sondern auch anderen bayerischen und überregionalen Projekten im Bereich Forschungsdatenmanagement eine Plattform bieten und auf bestehende Initiativen und Ansprechpartner in bayerischen Einrichtungen hinweisen.

Die bayerischen Universitäten und Forschungseinrichtungen bieten ihren Wissenschaftlerinnen und Wissenschaftlern mehrheitlich bereits [Dienstleistungen](#) im Bereich Forschungsdatenmanagement an und sind an verschiedenen [Pilotprojekten](#) beteiligt. Die dabei gewonnenen Erfahrungen, Ergebnisse und [Tools](#) können auch für andere Institutionen von Interesse sein. [Forschungsdatenmanagement-Bayern.org](#) ist nur eine von mehreren [Initiativen](#) in Deutschland, die anstreben, den Informationsfluss und eine nachhaltige Vernetzung der verschiedenen Akteure zu verbessern.

Diese „horizontalen“ Strukturen sind auch aus Sicht des Rates für Informationsinfrastruktur (RfII) einer der Wege, eine effiziente Nationale Forschungsdateninfrastruktur (NFDI) aufzubauen. Gegenseitiger Informationsaustausch ist hierfür eines der Werkzeuge, um unnötige Doppelarbeit zu verhindern, noch vorhandene Lücken zu identifizieren und die [Policies](#) von Förderorganisationen und Hochschulen effektiv erfüllen zu können.



@fdm\_bayern

[www.fdm-bayern.org](http://www.fdm-bayern.org)

## Forschungsdatenmanagement

### Forschungsdatenmanagement an bayerischen Einrichtungen

Universität Augsburg

Universität Bayreuth

Friedrich-Alexander-Universität  
Erlangen-Nürnberg

Ludwig-Maximilians-Universität  
München

Technische Universität München

Universität Regensburg

Julius-Maximilians-Universität  
Würzburg

FDM-Projekte in Bayern

## Ludwig-Maximilians-Universität München

### IT-Gruppe Geisteswissenschaften - LMU Center for Digital Humanities ▲

Die IT-Gruppe Geisteswissenschaften (ITG) ist eine den geisteswissenschaftlichen Fakultäten der LMU zugeordnete Einrichtung, die auf der Grundlage eines fach- und fakultätsübergreifenden Konzeptes umfassend in den Bereichen Service (Investitions-, IT-, Projekt- und Wissensmanagement) sowie Forschung und Lehre in den Digital Humanities tätig ist und sich auf diesem Gebiet als Kompetenzzentrum versteht.

#### Dienstleistungen:

- Langjährige Erfahrung in den Bereichen Projektberatung (Konzeption, Antragstellung) und -begleitung (technische Realisierung)
- Beratung in allen Fragen zu zeitgemäßer Publikation, Nachhaltigkeit und Nachnutzbarkeit von Forschungsergebnissen
- Lehre in den Digital Humanities, auch zum Thema Nachhaltigkeit von Forschungsdaten ([DHVLab](#) Datenrepositorium)

#### Homepage:

<http://www.itg.uni-muenchen.de>

#### Kontakt:

[julian.schulz@lmu.de](mailto:julian.schulz@lmu.de)

- gemeinsames Modellprojekt von Bibliotheken und einem IT-Zentrum
- Universitäten mit unterschiedlicher Aufgabenverteilung und Vernetzung der Infrastruktureinrichtungen
- fachnahe technische Lösungen und Dienstleistungen, aber nicht „fachexklusiv“
- Bei Fragen oder Ideen jederzeit Kontakt aufnehmen.

**Vielen Dank für Ihre  
Aufmerksamkeit!**