



Nutzen und Grenzen der Erfassung und
Kategorisierung digitaler Editionen.
Oder: Fifty Shades of Grey?

Daniela Schulz (Herzog August Bibliothek Wolfenbüttel)

Virtuelles DH-Kolloquium BBAW

09.04.2021

clariah.de

 [@CLARIAHde](https://twitter.com/CLARIAHde)

Gliederung

1. Projektkontext
2. Entstehung und Zweck von EdMa
3. Kategorien von EdMa
4. Probleme und Schief lagen
5. EdMa als „Tool“
6. Ergebnisse – aktueller Stand – Auswertungen



1. Projektkontext

1. Projektkontext



CLARIAH-DE als Zusammenführung von CLARIN-D und DARIAH-DE

Förderphase: 04/2019-03/2021

Verlängerung: 04/-09/2021

Ziele:

- Herstellung von Interoperabilität
- Etablierung einheitlicher Workflows & einer gemeinsamen Organisationsstruktur

GEFÖRDERT VOM

1. Projektkontext



AP 1- „Forschungsdaten, Standards und Verfahren“

Als Task- und Co-Taskleader verantwortlich

- Alexander Geyken (Berlin-Brandenburgische Akademie der Wissenschaften)
- Andreas Henrich (Lehrstuhl für Medieninformatik, Otto-Friedrich-Universität Bamberg)



Matthias Boenig
Marius Hug



Tobias Gradl
Robin Jegan

IDS

LEIBNIZ-INSTITUT FÜR
DEUTSCHE SPRACHE

Bernhard Fisseni



NIEDERSÄCHSISCHE STAATS- UND
UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN | SUB

Simon Sendler

H E R Z O G
A U G U S T
B I B L I O
T H E K

Daniela Schulz

GEFÖRDERT VOM



1. Projektkontext

AP 1- „Forschungsdaten, Standards und Verfahren“

AP	Aufgabe	KPI
1.1.1	Prüfung und Aufwandsabschätzung für die Zusammenführung und Homogenisierung digitaler Datenbestände	Konzept zur Konvertierung von TEI-Ressourcen des TextGridRep in das DTABf; Evaluation des DTABf für Editionen
1.1.2	Zusammenführung und Homogenisierung digitaler Datenbestände	Konvertierung ausgewählter TEI-Ressourcen des TextGridRep in das DTABf und deren Integration in die DTA-/CLARIN-D-Infrastruktur
1.2	Aufbereitung von Beständen für bestehende Rechercheumgebungen	Erarbeitung von drei Beispielszenarien
1.3	Erstellung von Handreichungen und Empfehlungen	Zusammenführung bzw. Neu-Erstellung von Handreichungen
1.4	Kooperation mit Standardisierungsgremien	Fortführen der bisherigen Gremienarbeit
1.5	Integration von Angeboten aus der Community	Erarbeitung eines Workflows zur Integration in CLARIAH-DE
1.6	Dokumentation	Fünf Blog-Artikel, drei Schritt-für-Schritt-Anleitungen der Tools

GEFÖRDERT VOM

1. Projektkontext

AP 1- „Forschungsdaten, Standards und Verfahren“

AP	Aufgabe	KPI
1.1.1	Prüfung und Aufwandsabschätzung für die Zusammenführung und Homogenisierung digitaler Datenbestände	Konzept zur Konvertierung von TEI-Ressourcen des TextGridRep in das DTABf; Evaluation des DTABf für Editionen
1.1.2	Zusammenführung und Homogenisierung digitaler Datenbestände	Konvertierung ausgewählter TEI-Ressourcen des TextGridRep in das DTABf und deren Integration in die DTA-/CLARIN-D-Infrastruktur
1.2	Aufbereitung von Beständen für bestehende Rechercheumgebungen	Erarbeitung von drei Beispielszenarien
1.3	Erstellung von Handreichungen und Empfehlungen	Zusammenführung bzw. Neu-Erstellung von Handreichungen
1.4	Kooperation mit Standardisierungsgremien	Fortführen der bisherigen Gremienarbeit
1.5	Integration von Angeboten aus der Community	Erarbeitung eines Workflows zur Integration in CLARIAH-DE
1.6	Dokumentation	Fünf Blog-Artikel, drei Schritt-für-Schritt-Anleitungen der Tools

GEFÖRDERT VOM



2. Entstehung und Zweck von EdMa

Wofür brauchen wir / braucht man überhaupt Kriterien
zur Erfassung und Kategorisierung digitaler Editionen?

2. Entstehung und Zweck von EdMa



Kategorisierung und Auswahl der digitalen Editionen für die drei Beispielszenarien (AP 1.2)

- Problem der Diversität der Daten bzw. Textauszeichnung steht bei Digitalen Editionen in Bezug zu
 - Materialgrundlage
 - Editionsziel
 - editorischem Vorgehen
 - ...
- Notwendigkeit zur Schaffung von Transparenz in der Auswahl „geeigneter“ digitaler Editionen
- Entwicklung von Kriterien, um die Komplexität digitaler Editionen festzustellen
 - Editionsmatrix (EdMa) als Erfassungssystem

GEFÖRDERT VOM

2. Entstehung und Zweck von EdMA

- Projektinterne Grundlage für Auswahl der Fallbeispiele
- Ableitung von Empfehlungen



- Identifizierung objektiver Kriterien zur Erfassung und Kategorisierung Digitaler Editionen allgemein
- Zusammenhang zwischen Kriterien und Komplexität der Überführung in ein Pivot-Format (DTABf), welches Interoperabilität und Nachnutzbarkeit der Daten befördert



3. Kategorien von EdMa

3. Kategorien von EdMa



Editionsmatrix (EdMa)

0. Erhebung allgemeiner Angaben zur Edition
1. Editorisches Grundmodell / „Ziel“ der Edition
2. Art und Tiefe der Textauszeichnung nach TEI-P5
3. Disziplinäre Zuordnung
4. Epochale Zuordnung
5. Ausgangs-/Quellenmaterial
6. Sprache(n)
7. Breite / Menge an Dokumenten

GEFÖRDERT VOM

3. Kategorien von EdMa



Editionsmatrix (EdMa)

- 0. Erhebung allgemeiner Angaben zur Edition
- 1. Editionstyp / Editorisches Grundmodell**
- 2. Art und Tiefe der Textauszeichnung nach TEI-P5
- 3. Disziplinäre Zuordnung
- 4. Epochale Zuordnung
- 5. Ausgangs-/Quellenmaterial
- 6. Sprache(n)
- 7. Breite / Menge an Dokumenten

GEFÖRDERT VOM

3. Kategorien von EdMa



Editionsmatrix (EdMa)

0. Erhebung allgemeiner Angaben zur Edition
1. Editionstyp / Editorisches Grundmodell
- 2. Art und Tiefe der Textauszeichnung nach TEI-P5**
3. Disziplinäre Zuordnung
4. Epochale Zuordnung
5. Ausgangs-/Quellenmaterial
6. Sprache(n)
7. Breite / Menge an Dokumenten

GEFÖRDERT VOM

3. Kategorien von EdMa



Editionsmatrix (EdMa)

0. Erhebung allgemeiner Angaben zur Edition
1. Editionstyp / Editorisches Grundmodell
2. Art und Tiefe der Textauszeichnung nach TEI-P5
- 3. Disziplinäre Zuordnung**
- 4. Epochale Zuordnung**
5. Ausgangs-/Quellenmaterial
6. Sprache(n)
7. Breite / Menge an Dokumenten

GEFÖRDERT VOM

3. Kategorien von EdMa



Editionsmatrix (EdMa)

0. Erhebung allgemeiner Angaben zur Edition
1. Editionstyp / Editorisches Grundmodell
2. Art und Tiefe der Textauszeichnung nach TEI-P5
3. Disziplinäre Zuordnung
4. Epochale Zuordnung
- 5. Ausgangs-/Quellenmaterial**
6. Sprache(n)
7. Breite / Menge an Dokumenten

GEFÖRDERT VOM

3. Kategorien von EdMa



Editionsmatrix (EdMa)

0. Erhebung allgemeiner Angaben zur Edition
1. Editionstyp / Editorisches Grundmodell
2. Art und Tiefe der Textauszeichnung nach TEI-P5
3. Disziplinäre Zuordnung
4. Epochale Zuordnung
5. Ausgangs-/Quellenmaterial
- 6. Sprache(n)**
7. Breite / Menge an Dokumenten

GEFÖRDERT VOM

3. Kategorien von EdMa



Editionsmatrix (EdMa)

0. Erhebung allgemeiner Angaben zur Edition
1. Editionstyp / Editorisches Grundmodell
2. Art und Tiefe der Textauszeichnung nach TEI-P5
3. Disziplinäre Zuordnung
4. Epochale Zuordnung
5. Ausgangs-/Quellenmaterial
6. Sprache(n)
- 7. Breite / Menge an Dokumenten**

GEFÖRDERT VOM



4. Probleme und Schiefungen

4. Probleme und Schief lagen

1. Editorisches Modell bzw. Editionstyp
Mehrfachnennungen möglich, da sich diese gerade im Bereich der Digitalen Editionen nicht immer trennscharf unterschieden lassen bzw. Editionsprojekte verschiedene "Ausgaben" ermöglichen bzw. bündeln

(Hyper-)Diplomatische bzw. dokumentarische Edition
dokumentennah, auch materielle Aspekte berücksichtigend

Transkription meist eher basal ausgezeichnete Wiedergabe eines Textbestandes; kann auf graphematische Genauigkeit abzielen

Leseausgabe den Textbestand mit dem Ziel einer Leseerleichterung anpassend wiedergebend z.B. graphematisch, mit Auflösung von Abkürzungen, Korrektur von Schreibfehlern etc. --> nicht immer trennschaft von einer Transkription abzugrenzen, aber mit einem anderen Impetus

(Text-)Genetische Edition Fokus auf Genese eines Dokumentes bzw. Textes

(Historisch-)Kritische Edition
textkritischer Zugang zum Text, idealtypisch mit Variantenapparat etc.

(Teil-)Übersetzung(en)

Sonstiges

2. Auszeichnungstiefe bzw. Verwendung bestimmter TEI-Elemente bzw. Module
Mehrfachnennungen möglich

Erfassung der Textstruktur z.B. Kodierung von Abschnitten wie Kapitel, vgl. TEI-P5-Guidelines c. 4

Erfassung physischer Merkmale der Originalobjekte z.B. visuelle Einschnitte, vgl. TEI-P5-Guidelines c. 11

Semantische und/oder syntaktische Analyse vgl. TEI-P5-Guidelines cc. 17, 18

Auszeichnung von Entitäten
Personen, Orte etc., vgl. TEI-P5-Guidelines c. 13

Textkritische Annotation z.B. TEI-Elemente app, rdg oder auch sic, corr etc., vgl. TEI-P5-Guidelines cc. 11, 12

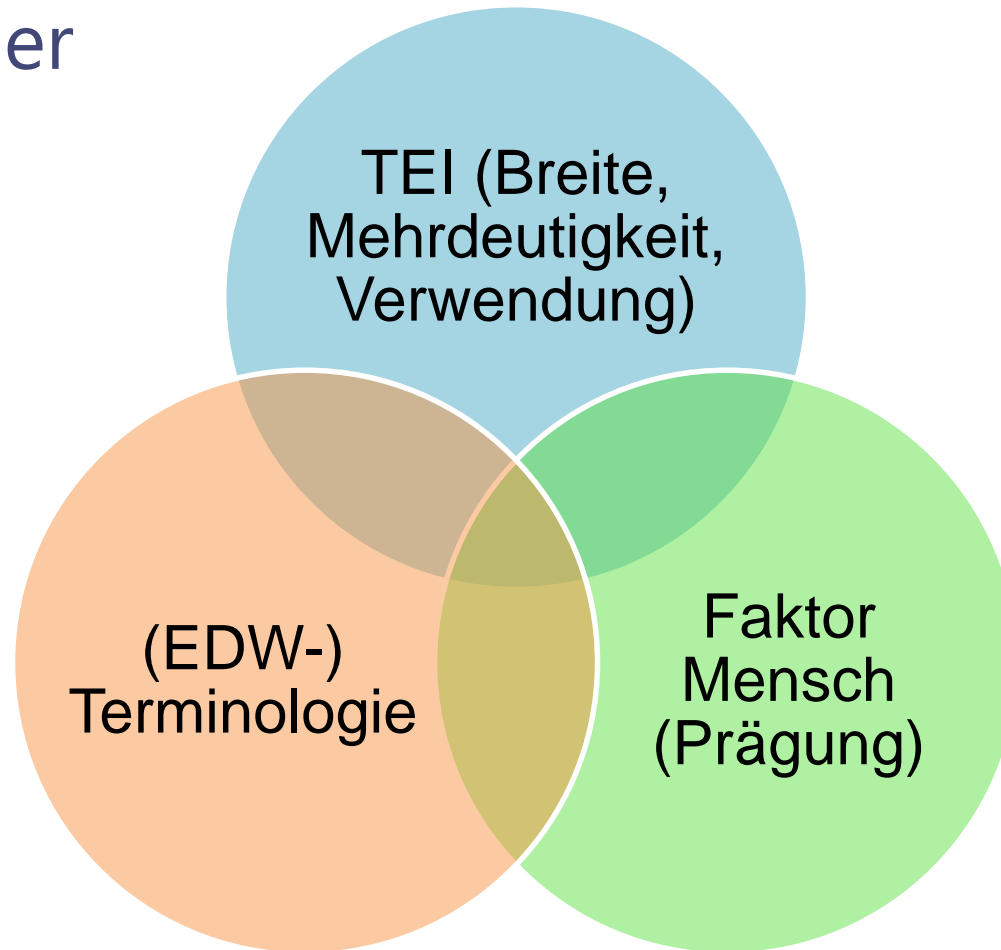
Auszeichnung von mündlicher Rede, Aufführungstexten oder Verstexten
vgl. TEI-P5-Guidelines cc. 6-8

Kodierung von Tabellen, Graphiken oder Bildern vgl. TEI-P5-Guidelines c. 14

Linking & Standoff-Annotation vgl. TEI-P5-Guidelines c. 16

4. Probleme und Schiefungen

Problemfelder



4. Probleme und Schief lagen



- „Jeder editorische Gegenstand hat seine eigenen Anforderungen und Probleme und führt zu ganz bestimmten Lösungen.“
- „Die zunehmende Differenzierung der Entwicklungen in ein Raster einzufügen scheint aber ein allgemeines Bedürfnis zu sein, wobei die erstaunliche Breite der Typendebatte auch Ausdruck des insgesamt hohen methodologischen Niveaus der (Editions-)Philologien sein mag.“
- „Die Praxis der Edition bringt heute mehrheitlich Zwischenpositionen hervor, die aber immer noch stärker an der Idealität der historisch-kritischen als an der Materialität der dokumentierenden Edition orientiert sind.“

(Sahle, Digitale Editionsformen)

4. Probleme und Schiefhlagen

(mögliche) Ansätze:

- Pragmatisches Vorgehen → sich nicht in Details verlieren
- Anwendung eher breiter und inkludierender Definitionen
- Kurze Erklärungen
- Großzügigkeit gegenüber „Widersprüchen“
- Vergleich mehrerer Erfassungen durch unterschiedliche Personen



5. EdMa als „Tool“

5. EdMa als „Tool“



Editionsmatrix (EdMa) zur Bestimmung des Komplexitätsgrades einer Konvertierung zum DTABf

Dieses Formular ist im Kontext von Arbeitspaket (AP) 1 "Forschungsdaten, Standards und Verfahren" des CLARIAH-DE Projekts (2019-2021) entstanden. Es dient dazu, anhand objektiver Kriterien abzuschätzen, wie komplex sich die Überführung einer digitalen Edition in das Basisformat des Deutschen Textarchivs (DTABf) in etwa gestaltet. Es werden insgesamt sieben Kriterien abgefragt. Nutzerinnen und Nutzer erhalten eine Rückmeldung zum Komplexitätsgrad sowie weitere Hinweise.

Herzlich Willkommen zum EdMa-Formular zur Bestimmung des Komplexitätsgrades der Überführung in das DTABf.

In dieser Umfrage sind 7 Fragen enthalten.

Um die Umfrage zu öffnen, akzeptieren Sie bitte unsere Datenschutzerklärung.

[Datenschutzerklärung anzeigen](#)



Weiter

Ziele:

1. Bereitstellung einer Survey auf clariah.de für Nutzerinnen und Nutzer
2. Nachhaltige und erweiterbare Übersicht digitaler Editionen ebenda

GEFÖRDERT VOM

5. EdMa als „Tool“

1. Editorisches Modell bzw. Editionstyp

Bitte wählen Sie die zutreffenden Antworten aus:

- (Hyper-)Diplomatische bzw. dokumentarische Edition** dokumentennah, auch materielle Aspekte berücksichtigend
- Transkription** meist eher basal ausgezeichnete Wiedergabe eines Textbestandes; kann auf graphematische Genauigkeit abzielen
- Leseausgabe** den Textbestand mit dem Ziel einer Leseerleichterung anpassend wiedergebend z.B. graphematisch, mit Auflösung von Abkürzungen, Korrektur von Schreibfehlern etc. → nicht immer trennschaft von einer Transkription abzugrenzen, aber mit einem anderen Impetus
- (Text-)Genetische Edition** Fokus auf Genese eines Dokumentes bzw. Textes
- (Historisch-)Kritische Edition** textkritischer Zugang zum Text, idealtypisch mit Variantenapparat etc.
- (Teil-)Übersetzung(en)**
- Sonstiges

Mehrfachnennungen möglich, da sich diese gerade im Bereich der Digitalen Editionen nicht immer trennscharf unterschieden lassen bzw. Editionsprojekte verschiedene "Ausgaben" ermöglichen bzw. bündeln

5. EdMa als „Tool“

2. Auszeichnungstiefe bzw. Verwendung bestimmter TEI-Elemente bzw. Module

! Bitte wählen Sie die zutreffenden Antworten aus:

- Erfassung der Textstruktur** z.B. Kodierung von Abschnitten wie Kapitel, vgl. [TEI-P5-Guidelines c. 4](#)
- Erfassung physischer Merkmale der Originalobjekte** z.B. visuelle Einschnitte, vgl. [TEI-P5-Guidelines c. 11](#)
- Semantische und/oder syntaktische Analyse** vgl. [TEI-P5-Guidelines cc. 17, 18](#)
- Auszeichnung von Entitäten** Personen, Orte etc., vgl. [TEI-P5-Guidelines c. 13](#)
- Textkritische Annotation** z.B. TEI-Elemente `app`, `rdg` oder auch `sic`, `corr` etc., vgl. [TEI-P5-Guidelines cc. 11, 12](#)
- Auszeichnung von mündlicher Rede, Aufführungstexten oder Verstexten** vgl. [TEI-P5-Guidelines cc. 6, 7, 8](#)
- Kodierung von Tabellen, Graphiken oder Bildern** vgl. [TEI-P5-Guidelines c. 14](#)
- Linking & Standoff-Annotation** vgl. [TEI-P5-Guidelines c. 16](#)

? Mehrfachnennungen möglich

GEFÖRDERT VOM

Konversionshinweise

An dieser Stelle werden, falls zutreffend, Hinweise zu den erhobenen Merkmalen Ihrer Edition, die möglicherweise die Komplexität einer Konvertierung der Editionsdaten aus dem nativen Format in das DTABf erhöhen.

(Hyper-)Diplomatische bzw. dokumentarische Edition

Das DTABf bietet nur eine eingeschränkte Möglichkeit, materielle Aspekte eines Objektes wie Blattformate, Beschädigungen oder Textpositionen wiederzugeben. Teile dieser Informationen können bei einer Konversion nicht berücksichtigt werden.

Vgl. [Leitlinien für die Weiterentwicklung des DTABf](#).

Erfassung physischer Merkmale der Originalobjekte

Das DTABf bietet nur eine eingeschränkte Zahl an Elementen des TEI-Moduls zur Erfassung der Originalobjekte an.

Vgl. [TEI-P5-Guidelines c. 11](#), Neologie, Hainhofer.

Semantische und/oder syntaktische Analyse

Das DTABf sieht für die Analyse der ausgezeichneten Texte standoff-Annotation vor. TEI-Elemente wie `w` sind dementsprechend nicht erlaubt.

Vgl. [TEI-P5 c. 17](#), [TEI-P5 c. 18](#), [DTABf: Übersicht über die Elemente im Textbereich](#), Mantel.

Überlieferungszeitraum vor 1600

Bitte beachten Sie, dass der Fokus des DTABf auf dem Zeitraum von ca. 1600 bis 1980 liegt. Eine Auszeichnung von Drucken und handschriftlichen Texten, die außerhalb dieses Zeitraums entstanden sind, ist grundsätzlich möglich. Unter Umständen können nicht alle Feinheiten der epochenspezifische Schriftkultur abgebildet werden.

Vgl. [Leitlinien für die Weiterentwicklung des DTABf](#).

Mehrere Handschriften / Manuskripte

Hier sei besonders auf die Evaluationsberichte zum Diarium sowie zu Hainhofer verwiesen, da diese Projekte handschriftlich überlieferte Texte edieren.





CLARIAH-DE



6. Ergebnisse – aktueller Stand –
Auswertungen etc.

6. Ergebnisse – aktueller Stand – Auswertungen



„Produkte“

- Formular zur Bestimmung des Komplexitätsgrades
- u.a. detailliertere Anleitung zu EdMa, einen allgemeinen Bericht, drei Evaluationsberichte (in Veröffentlichung)
- Übersicht digitaler Editionen (aktuell 76 Editionen erfasst nach EdMa; bald auf clariah.de)

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Titel der Edition		Acta Pacis Westphalicae	Algermann: Prachtstammbaum	Andreas de Escobar: Modus Confitendi. Digitale Edition der mittelniederdeutschen Handschrift.	Annibale Pocaterra Two dialogues on shame. English translation of: Annibale Pocaterra Due Dialogi della Vergogna Ferrara 1592	Arthur Schnitzler digital. Digitale historisch-kritische Edition (Werke 1905 bis 1931)	Athanasius Kircher an Herzog August den Jüngerem : lateinische Briefe der Jahre 1650-1666 aus den Sammlungen der Herzog August Bibliothek Wolfenbüttel	Balthasar Staindl: Ein sehr künstlichs und nutzlichs Kochbuch	Bibliothek der Neologie. Kommentierte kritische Auswahledition in zehn Bänden	Briefe der Fruchtbringende Gesellschaft und Beilagen. Die Zeitschriften des Fürst Ludwigs von Anhalt-Köthen 1617-1650. 1. Band 1617-1626
Informationen	Link zum Editionsprojekt	https://apw.digitale-sammlungen.de/apw/static.html	http://diglib.hab.de/edoc/ed000222/start.htm	http://diglib.hab.de/edoc/ed000236/start.htm	http://diglib.hab.de/edoc/ed000237/start.htm	https://www.arthur-schnitzler.de/	http://diglib.hab.de/edoc/ed000005/start.htm	http://diglib.hab.de/edoc/ed000241/start.htm	https://bdn-edition.de	http://diglib.hab.de/edoc/ed000213/start.htm
	Status (sofern/wie angegeben)		fertig	fertig	fertig	laufend	fertig	in Arbeit	beta-Version	fertig
	Institution(en)	Friedrich-Wilhelms-Universität Bonn Bayerischen Staatsbibliothek München	HAB	HAB	HAB	BU Wuppertal, University of Cambridge, University College London, University of Bristol (NRWAWK)	HAB	HAB	SUB, WWU Münster	HAB, SAW
	Laufzeitbeginn	2008	2013	2015	2013	2012	2006	2016	-	2014
	Laufzeitende	2014	2013	2018	2013		2006		-	2016
	Anmerkungen				PDF	XML nicht einsehbar			XML nur in Teilen einsehbar	
	1. Editorisches Modell bzw. Editionstyp Mehrfachnennungen möglich, da sich diese gerade im Bereich der Digitalen Editionen nicht immer trennscharf unterscheiden lassen bzw. Editionsprojekte verschiedene "Ausgaben" ermöglichen bzw. bündeln	(Hyper-)Diplomatische bzw. dokumentarische Edition dokumentennah, auch materielle Aspekte berücksichtigend			X		X			
Transkription meist eher basal ausgezeichnete Wiedergabe eines Textbestandes; kann auf graphematische Genauigkeit abzielen							X	X		
Leseausgabe den Textbestand mit dem Ziel einer Lesereleichterung anpassend wiedergebend z.B. graphematisch, mit Auflösung von Abkürzungen, Korrektur von Schreibfehlern etc. --> nicht immer trennscharf von einer Transkription abzugrenzen, aber mit				X		X		X		

6. Ergebnisse – aktueller Stand – Auswertungen



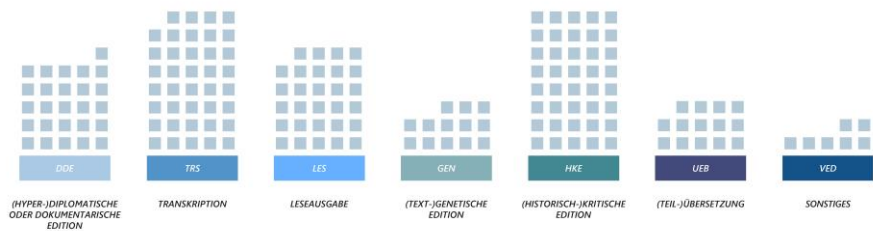
„Produkte“

- Formular zur Bestimmung des Komplexitätsgrades
- u.a. detailliertere Anleitung zu EdMa, einen allgemeinen Bericht, drei Evaluationsberichte (in Veröffentlichung)
- Übersicht digitaler Editionen (aktuell 76 Editionen erfasst nach EdMa; bald auf clariah.de)
- Formular zur Erfassung digitaler Editionen → Beiträge zur Übersicht

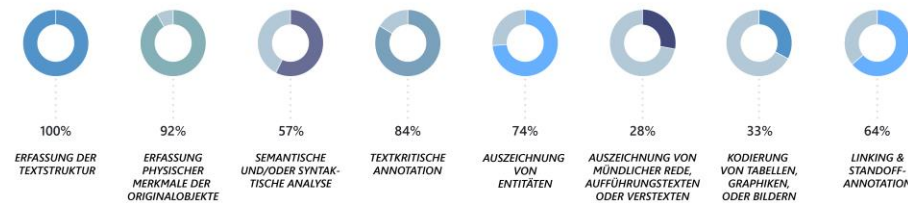
GEFÖRDERT VOM

6. Ergebnisse – aktueller Stand – Auswertungen

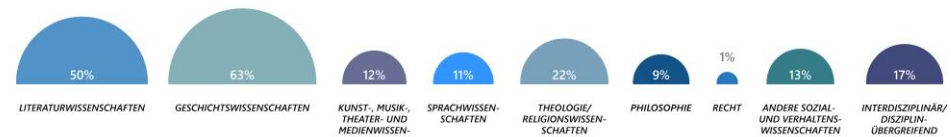
Visualisierungen einzelner Kategorien (von Susann Schwaß, HAB)



Gesamtzahl der in die Auswertung aufgenommenen Editionen: 76



Gesamtzahl der in die Auswertung aufgenommenen Editionen: 61

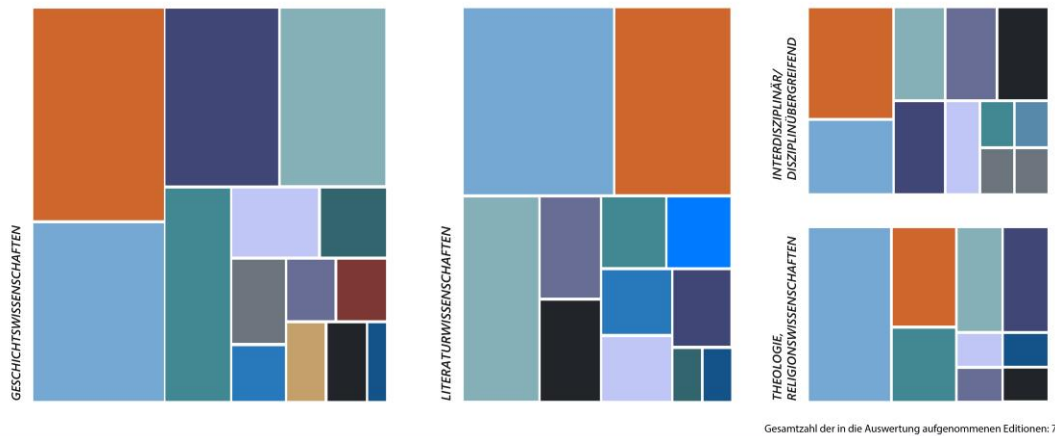


Gesamtzahl der in die Auswertung aufgenommenen Editionen: 76

GEFÖRDERT VOM

6. Ergebnisse – aktueller Stand – Auswertungen

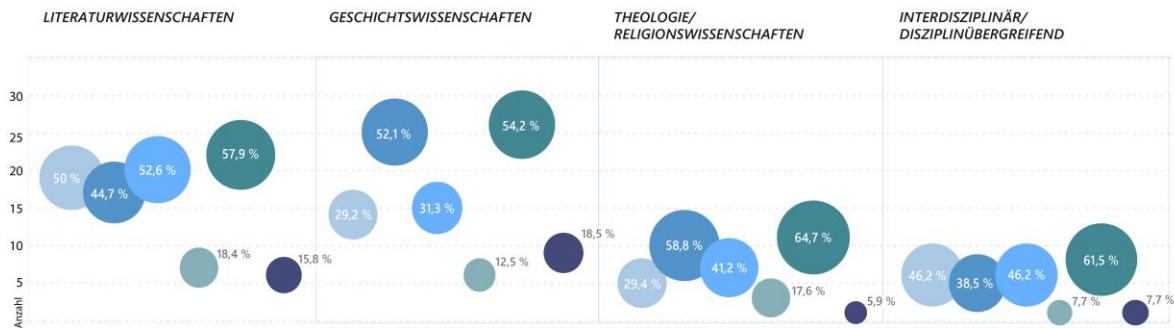
Visualisierungen von Zusammenhängen (von Susann Schwaß, HAB)



Quellenmaterial nach Disziplin



Editionstyp nach Disziplin



6. Ergebnisse – aktueller Stand – Auswertungen



Ausblick

- To do: Erprobung und Schärfung der Kriterien durch praktischen Einsatz
- Offen: Pflege und Weiterführung der Übersicht

GEFÖRDERT VOM



Danke für die Aufmerksamkeit!

Daniela Schulz

Herzog August Bibliothek Wolfenbüttel



clariah.de

 [@CLARIAHde](https://twitter.com/CLARIAHde)

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

FÖRDERKENNZEICHEN
01UG1910 A bis I