

# Criteria for Recording and Categorizing Scholarly Digital Editions

Daniela Schulz, Herzog August Bibliothek Wolfenbüttel  
Simon Sandler, Goettingen State and University Library

[www.clariah.de](http://www.clariah.de) @CLARIAHde

Scholarly Primitives - DARIAH Annual Event 2020 (Session: Tools and Workflows)

# Structure

- On Data Sustainability in Scholarly Digital Editions
- Categorizing Scholarly Digital Editions
- Implementing a Survey
- Conclusion

# On Sustainability of Data in Scholarly Digital Editions

Expressivity of markup and interoperability can compete.

## Expressivity

- Scholars desire a native data model that is able to represent their source as precisely as possible
- Multiple markup options for one phenomenon

*These Guidelines apply to texts in any natural language, of any date, in any literary genre or text type, without restriction on form or content. They treat both continuous materials ('running text') and discontinuous materials such as dictionaries and linguistic corpora.*

**Source:** P5: Guidelines for Electronic Text Encoding and Interchange. iv. About These Guidelines.

## Interoperability

- As the complexity of a data model/markup increases, possibilities of interoperability decrease
- Interoperability is crucial for sustainability and re-use of research data

N.B.: Interoperability  $\neq$  Interchangability

# A Compromise Between Expressivity and Interoperability

Provide research data in a **pivot format**.

- Identify and markup a core body of data
- Make data available to pipelines and tools
- Increased visibility of the edition through integration in corpora

# Categorizing Scholarly Digital Editions

## *Guiding Questions:*

- Are there features that regularly impact conversion complexity?
- Can we establish a workflow for these conversions?
- Can we support the community in converting SDEs?



*Aim:*

Provide a measure of comparability between different native formats of SDEs and aid scholars in converting data to a pivot formats.

*Pivot Format of Choice:*

**Base Format of the German Text Archive**

1. Editorial model
2. Markup
3. Discipline
4. Epoch of base material
5. Type/Transmission of base material
6. Language of base material
7. Number of documents to be encoded

# 1. Editorial model

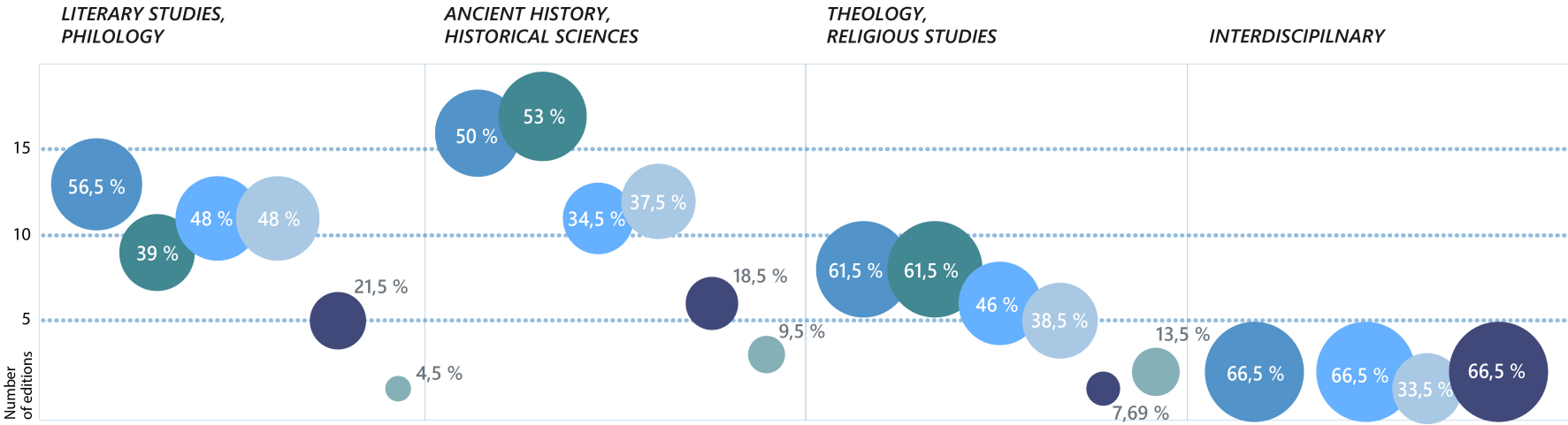
- Historical-Critical Edition
- (Hyper-)Diplomatic Edition
- Genetic Edition
- Reader's Edition
- Transcription
- (partial) Translation

1. Editorial model
2. Markup
3. Discipline
4. Epoch of base material
5. Type/Transmission of base material
6. Language of base material
7. Number of documents to be encoded

## Implementing a Survey

- Estimate the complexity of transformation into pivot format
- Provide basic recommendations regarding transformation
- Cross-reference projects of similar scope

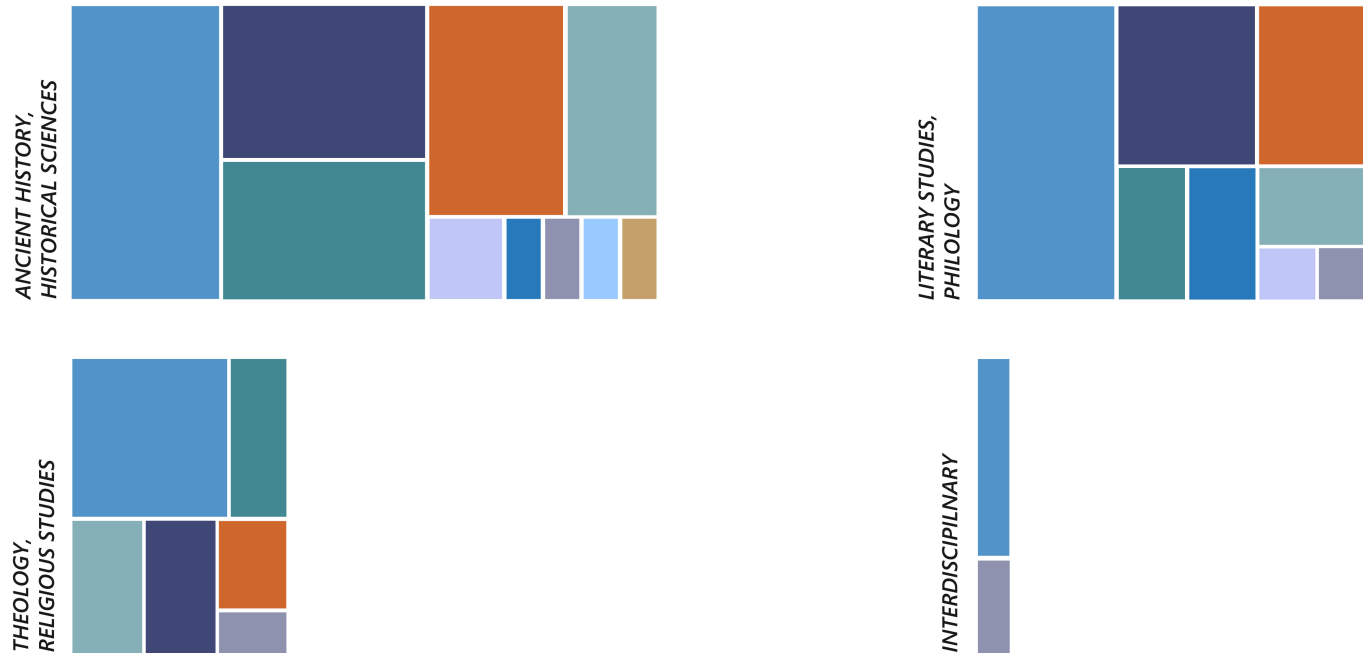
# EDITORIAL MODELS BY DISCIPLINE (BASED ON WDB EDITIONS)



Total number of editions included in the evaluation: 41

- TRANSCRIPT
- HISTORICAL-CRITICAL EDITION
- READING EDITION
- (HYPER-)DIPLOMATIC EDITION
- (PARTIAL) TRANSLATION
- „GENETIC“ EDITION

# SOURCE MATERIAL USE BY DISCIPLINE (BASED ON WDB EDITIONS)



Total number of editions included in the evaluation: 41

■ SINGLE MANUSCRIPT  
 ■ MANUSCRIPTS  
 ■ LETTERS  
 ■ INSCRIPTIONS  
 ■ HISTORICAL RECORDS  
 ■ SINGLE PRINT  
 ■ PRINTS  
 ■ DIARIES  
 ■ PAMPHLETS  
 ■ CONVOLUTE

## Conclusion

- Provision of data in a pivot format as well as the native format as a compromise between expressiveness and interoperability
- Create recommendations for "pivot-friendly native formats"



# References

- Syd Bauman: “Interchange vs. Interoperability.” In: Proceedings of Balisage: The Markup Conference 2011. Balisage Series on Markup Technologies, Volume 7 (2011).  
<https://doi.org/10.4242/BalisageVol7.Bauman01>
- Martin Holmes: "Whatever happened to interchange?" In: Digital Scholarship in the Humanities, Volume 32, Issue suppl\_1, April 2017, Pages i63–i68.  
<https://doi.org/10.1093/lhc/fqw048>
- Daniela Schulz: Überlegungen zu Harmonisierung und Standardisierung im Rahmen des Projekts CLARIAH-DE – oder: Aus der „Suppenküche“ von Arbeitspaket 1: Forschungsdaten, Standards und Verfahren. In: DHd-Blog, 10.07.2020. <https://dhd-blog.org/?p=14064> (11.10.2020).
- FAIR Principles: <https://www.force11.org/group/fairgroup/fairprinciples>.
- P5: Guidelines for Electronic Text Encoding and Interchange. iv. About These Guidelines: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/AB.html> (11.10.2020).

*Graphics courtesy of Susann Schwaß, Wolfenbüttel*



# Criteria for Recording and Categorizing Scholarly Digital Editions

Daniela Schulz, Herzog August Bibliothek Wolfenbüttel  
Simon Sendler, Goettingen State and University Library

[www.clariah.de](http://www.clariah.de) @CLARIAHde

Scholarly Primitives - DARIAH Annual Event 2020 (Session: Tools and Workflows)