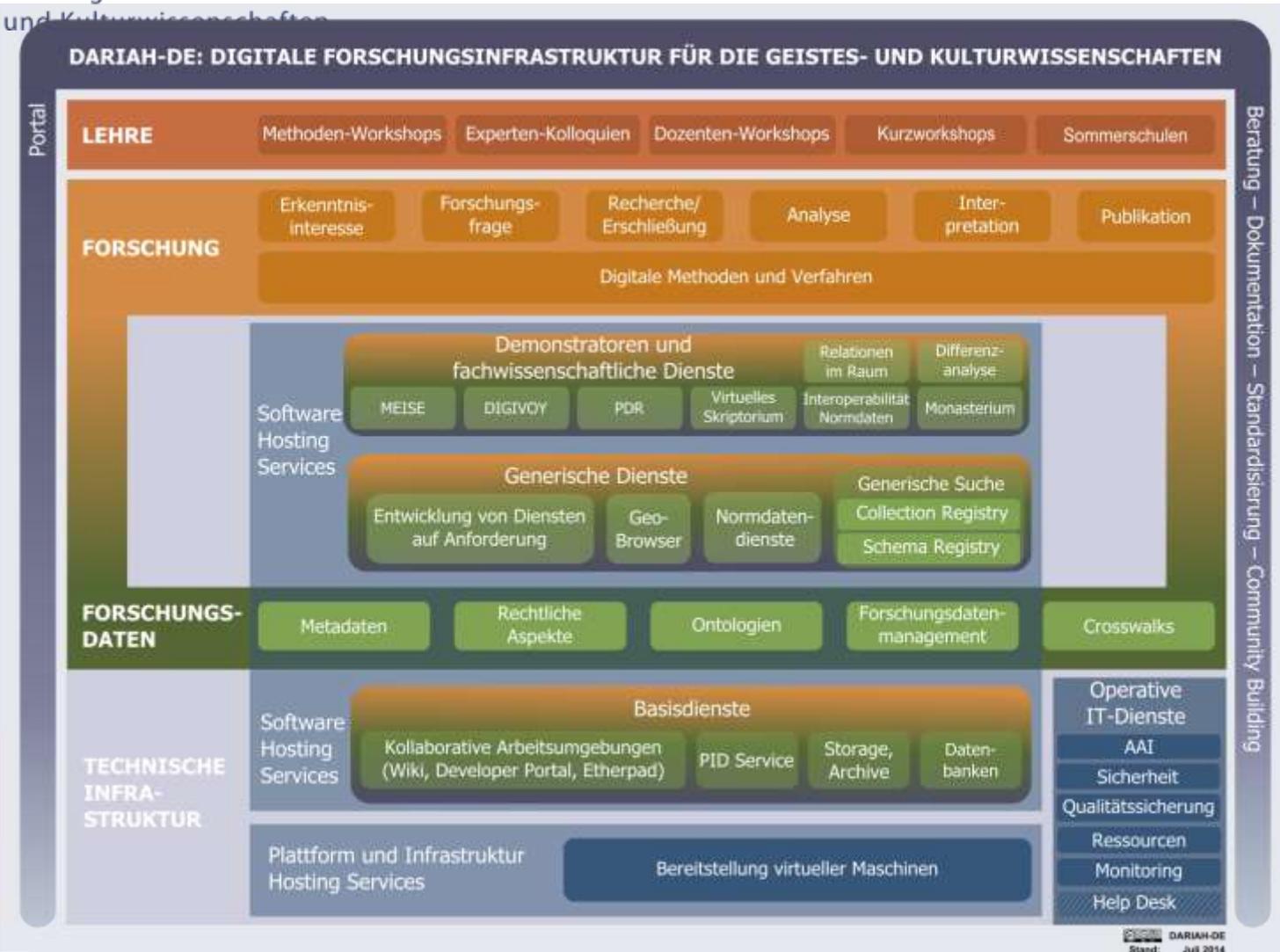


# Fachwissenschaftliche Anforderungen von Geistes- und KulturwissenschaftlerInnen an digitale (offene) Forschungsdaten

Dr. Stefan Schmunk  
DARIAH-DE – SUB Göttingen

# Agenda

1. Einleitung
2. Scholarly Primitives – Methoden und Verfahren der Digital Humanities
3. Anforderungen von FachwissenschaftlerInnen



## Forschungsdaten

Unter digitalen geistes- und kulturwissenschaftlichen Forschungsdaten werden all jene Quellen und Ergebnisse verstanden, die im Kontext einer geistes- und kulturwissenschaftlichen Forschungsfrage gesammelt, beschrieben, ausgewertet und/oder erzeugt wurden und in maschinenlesbarer Form zum Zwecke der Archivierung, Zitierbarkeit und zur weiteren Verarbeitung aufbewahrt werden können.

## Forschungsdaten / Wissenschaftliche Sammlungen

Unter digitalen geistes- und kulturwissenschaftlichen Forschungsdaten werden all jene Quellen und Ergebnisse verstanden, die im Kontext einer geistes- und kulturwissenschaftlichen Forschungsfrage gesammelt, beschrieben, ausgewertet und/oder erzeugt wurden und in maschinenlesbarer Form zum Zwecke der Archivierung, Zitierbarkeit und zur weiteren Verarbeitung publiziert sind.

- Metainformationen / bibliographische Daten
- Digitale und / oder digitalisierte Daten und / oder digitale Repräsentation analoger Daten
  - Volltexte / Transkripte
  - Angereicherte Volltexte
- Normdaten / Kontrollierte Vokabulare

• ...

John Unsworth, Scholarly Primitives: What methods do humanities researchers have in common, and how might our tools reflect this? (2000)

**Discovering / Retrieval** (Sammeln / Zusammenstellen / Sichern)

**Annotating** (Annotation)

**Comparing** (Vergleich / Hermeneutik)

**Referring** (Verweisen / wiss. Diskussion)

**Sampling** (Gruppieren / Statistik)

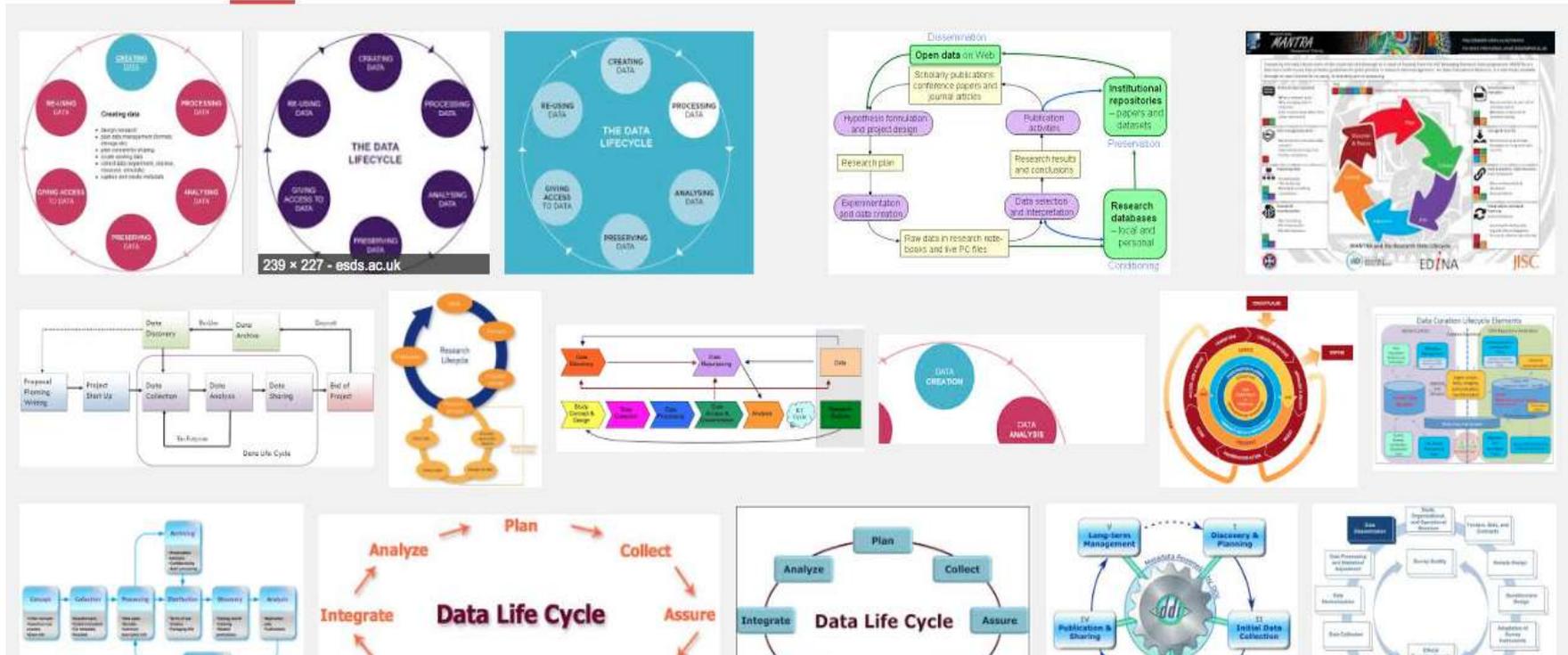
**Illustrating** (Visualisierung / Bildgebung)

**Representing** (Präsentation / Publikation)

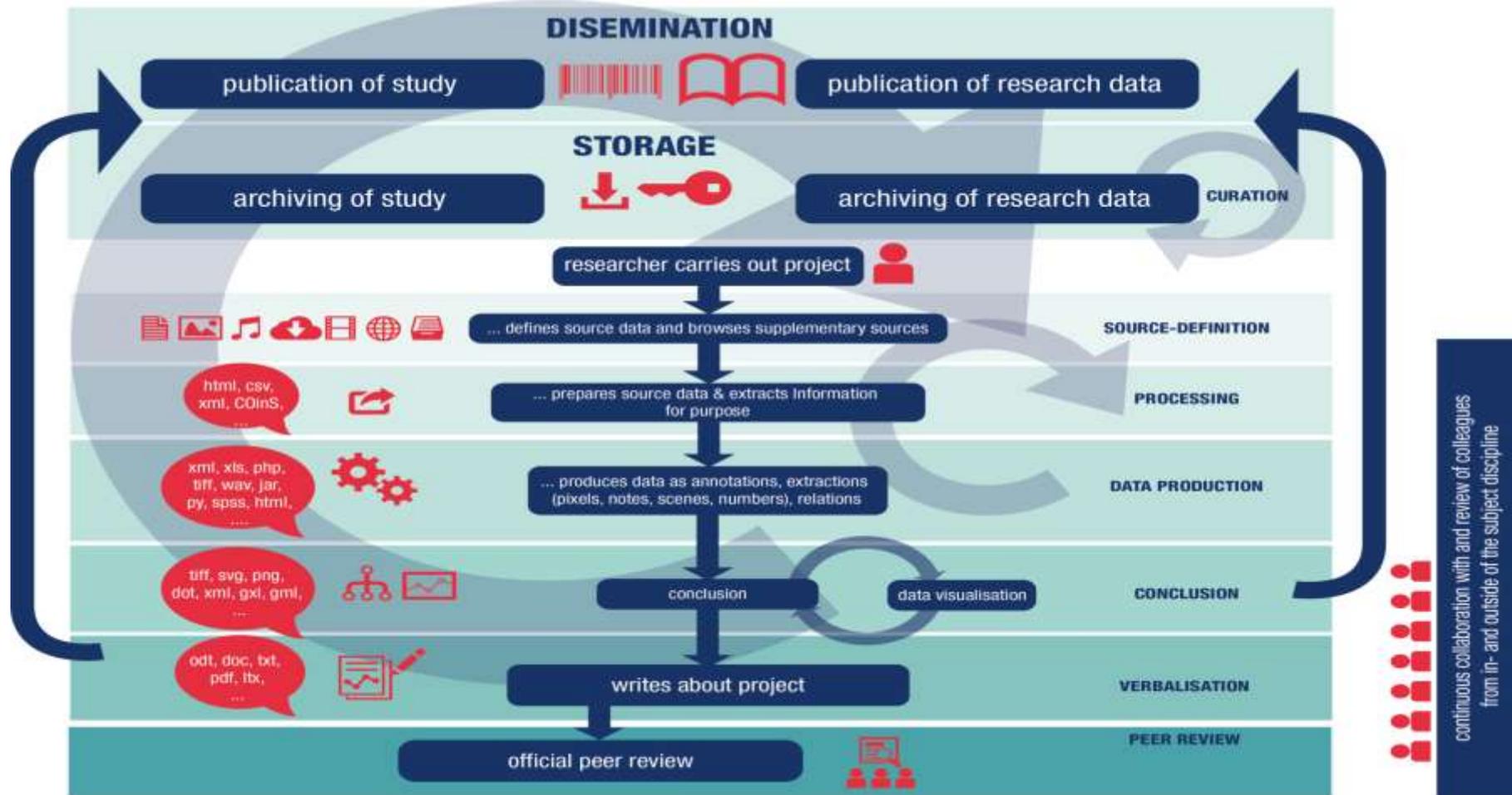
Arbeitsmethoden  
abhängig vom  
Zugang und der  
Qualität der Quellen

Web **Bilder** News Videos Shopping Mehr ▾ Suchoptionen

SafeSearch ▾



Google-Bilder Suche nach Research-Data-Lifecycle



Visualisierung eines Forschungszyklus in den Geistes- und Kulturwissenschaften –

<https://de.dariah.eu/der-dariah-ansatz>





Bibliographische Daten

Metadaten



Digitalisate

Digitale Repräsentation



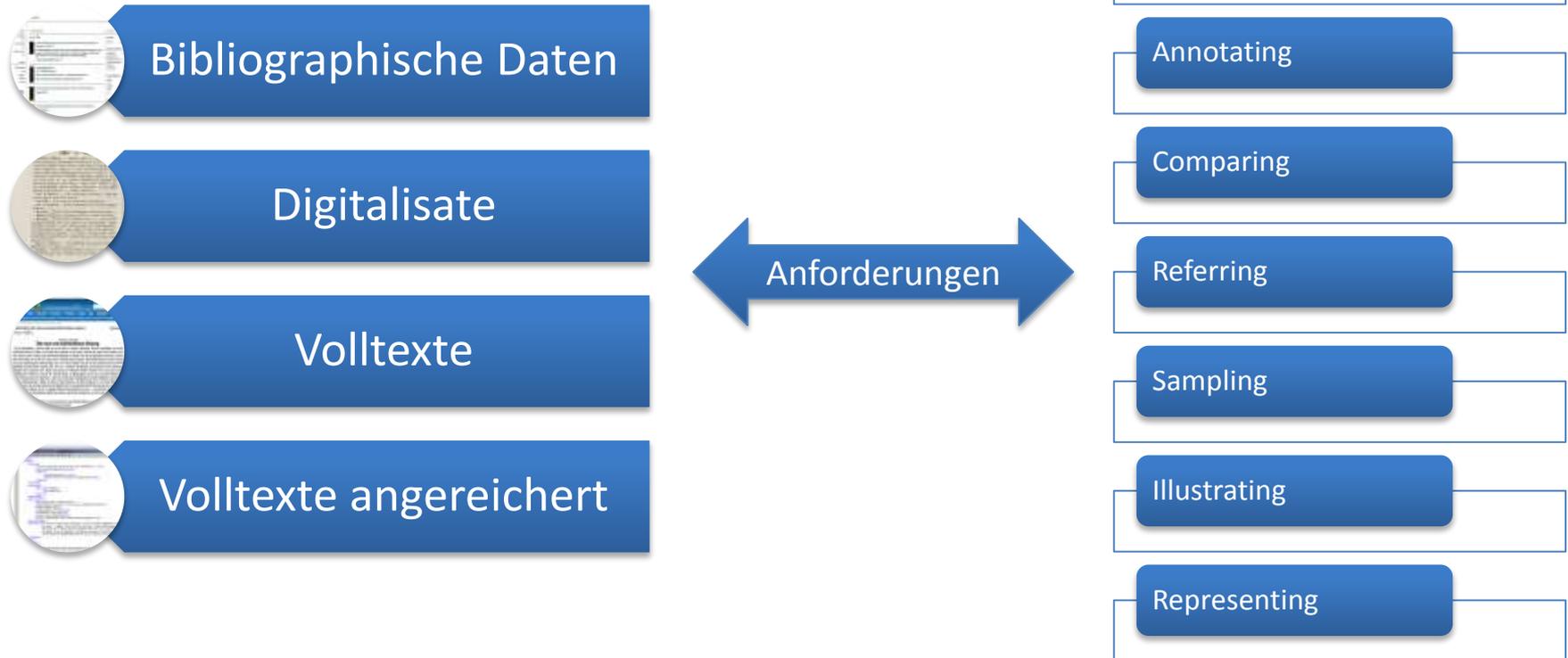
Volltexte

Transkripte

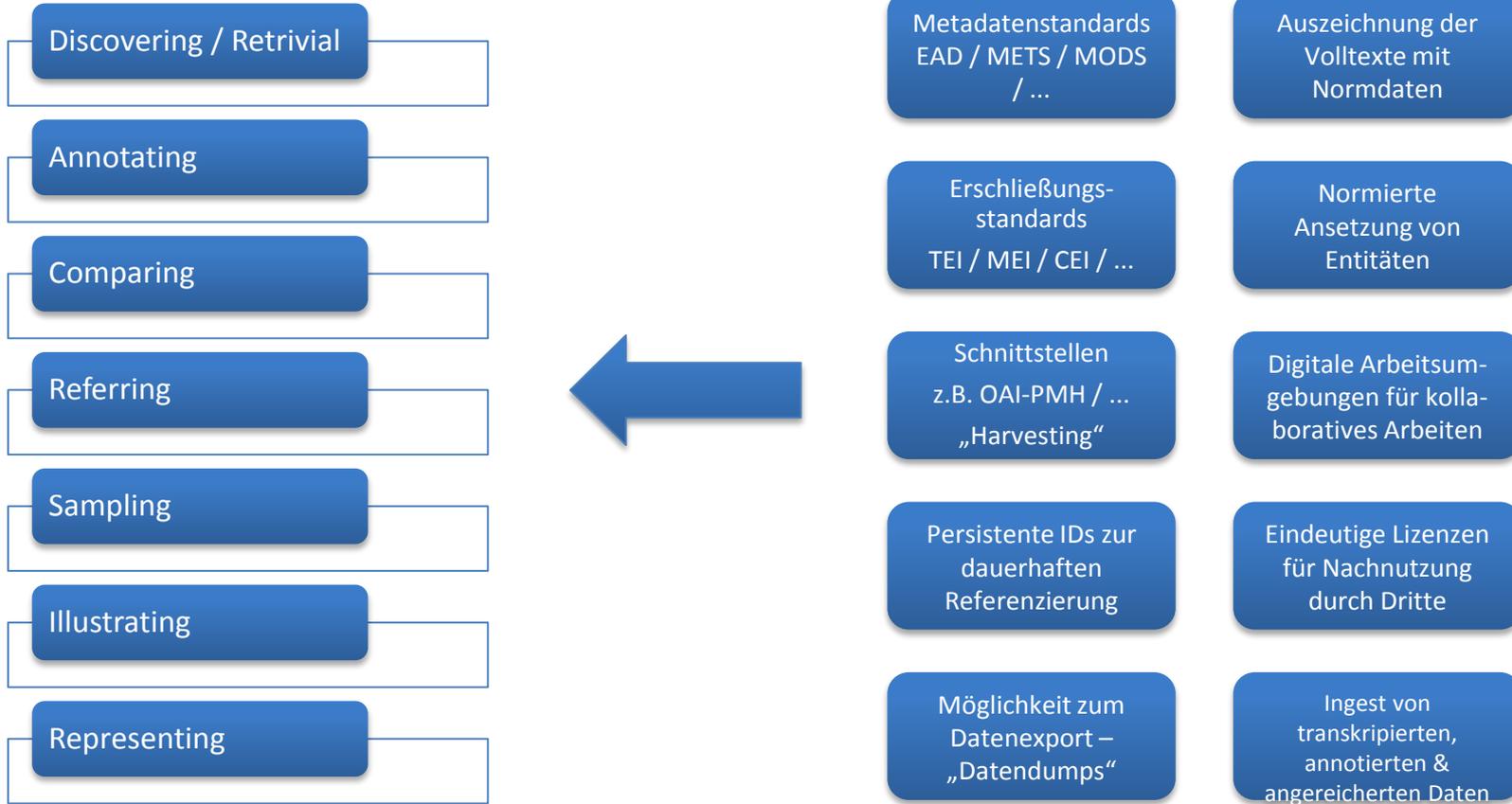


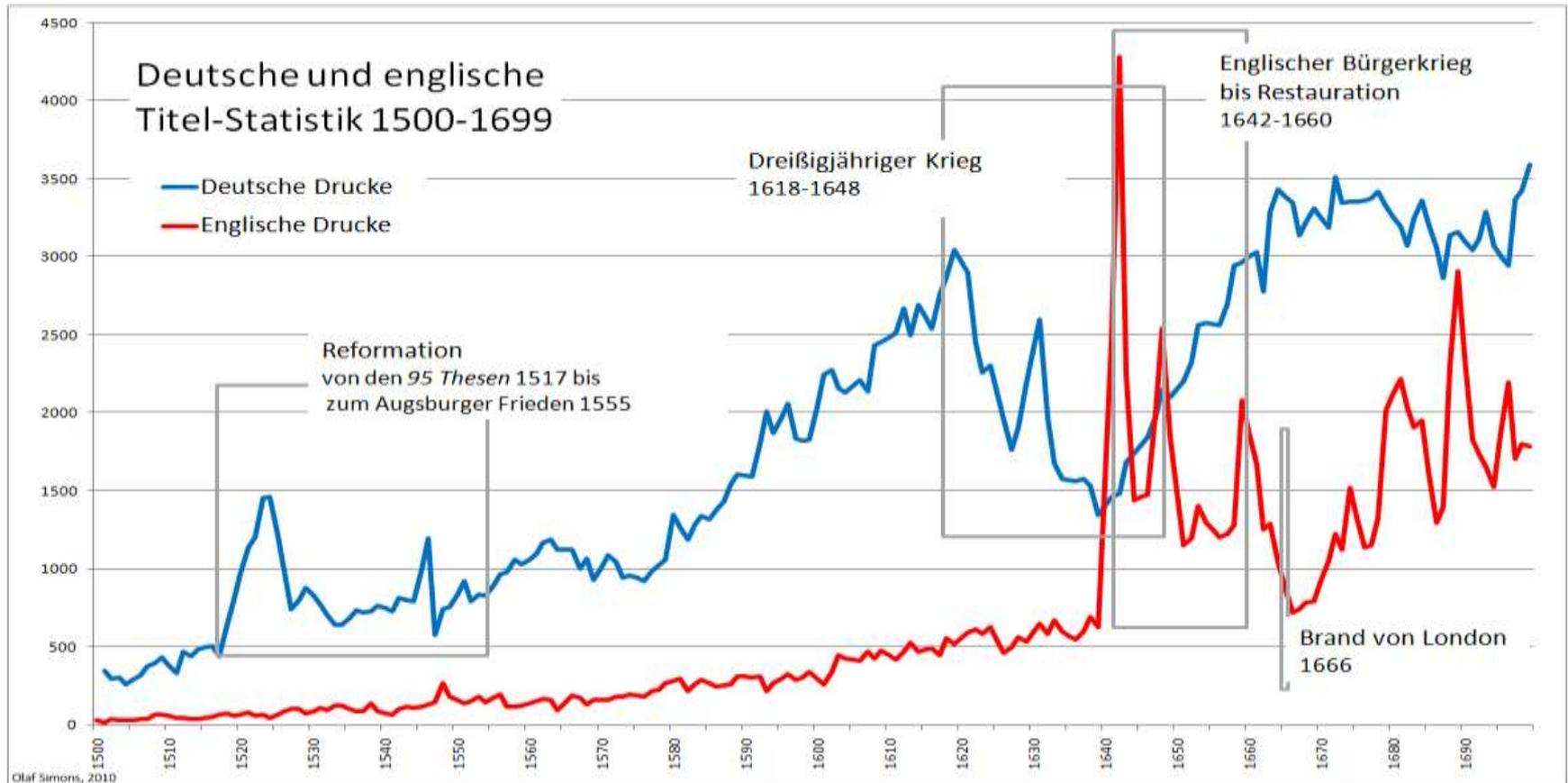
Volltexte angereichert

Annotiert und  
ausgezeichnet



## Anforderungen von FachwissenschaftlerInnen Mindestanforderungen aus Sicht von DH- WissenschaftlerInnen





„Deutsche und englische Titel-Statistik, 1500-1699“, basierend auf VD16, VD17 und English Short Title Catalogue (ESTC) von Olaf Simons.

Lizenziert unter Creative Commons Attribution-Share Alike 3.0 über Wikimedia Commons -

[http://commons.wikimedia.org/wiki/File:Deutsche\\_und\\_englische\\_Titel-Statistik,\\_1500-1699.png#mediaviewer/File:Deutsche\\_und\\_englische\\_Titel-Statistik,\\_1500-1699.png](http://commons.wikimedia.org/wiki/File:Deutsche_und_englische_Titel-Statistik,_1500-1699.png#mediaviewer/File:Deutsche_und_englische_Titel-Statistik,_1500-1699.png)

Access digital: Nutzen und vernetzen – Bundesarchiv, Koblenz – 25. November 2014

## Beispiel Forschungsdatensammlung – epidat

- **Forschungsdaten:** 25.000 Inschriften von über 130 jüdischen Friedhöfen aus Deutschland und den Niederlanden (1050-2000)
- **Enthält:** Transkriptionen, Karten, Digitalisate, Fotografien und sonstige archivalische Quellen
- **Methoden:** Visual Analytics – Relationen im Raum, Named-Entity-Recognition, Raum-Zeit-Visualisierung – Geo-Browser

### \*File and encoding format:

- data format
  - images: jpg png gif tiff eps
  - maps: svg pdf AutoCad kml
  - text
    - extern: x(html)
    - extern: TEI p4 XML (using: epidoc: Epigraphic Documents in TEI XML)
    - extern (wip): TEI p5 XML (using: example)
    - extern: plain/text (no format)
    - extern (on demand): PDF
    - intern: TUSTEP file with epidat structure
- metadata format
  - TEI header
  - DC (simple, example)



## Daten & Quellen müssen auffindbar und digital zu nutzen sein

- Erfassung anhand von **Standards**
- Dauerhafte Referenzierbarkeit durch **persistente IDs**
- **Schnittstellen** zum Export
- Verwendung von **Normdaten / kontrollierten Vokabularen**
- **Maschinenlesbarkeit & Prozessierbarkeit**
- Speicherung / **Langzeitarchivierung der „angereicherten“ Daten / Quellen**
- Eindeutige **Lizenzen**, die maschinenlesbar sind & Nachnutzung ermöglichen
- Anbindung an **digitale Viewer / Werkzeuge**: z.B. DFG-Viewer, Goobi, ...
- **„Qualitätsgrad“** der Transkripte / Volltexte müssen erfasst werden
- **Ingestmöglichkeiten** für transkripierte, annotierte & angereicherte Daten

DARIAH-DE Homepage

<http://de.dariah.eu>

DARIAH-EU Homepage

<http://dariah.eu>

DARIAH-DE Repositorium

<https://de.dariah.eu/forschungsdaten>

DARIAH-DE Dienste

Collection Registry <http://colreg.de.dariah.eu>

Schema Registry <http://dev3.dariah.eu/schereg/>

Generische Suche <http://search.de.dariah.eu>

TextGrid Homepage

<http://www.textgrid.de>

TextGrid Repositorium

<http://textgridrep.de>

EPIC2 PIDs

<http://www.pidconsortium.eu/>

<http://epic.gwdg.de/wiki/index.php/EPIC:API>

Kontakt:

[schmunk@sub.uni-goettingen.de](mailto:schmunk@sub.uni-goettingen.de)

# Herzlichen Dank für Ihre Aufmerksamkeit