

Historisches Datenzentrum Sachsen-Anhalt - Kernelemente eines fachspezifischen Datenzentrums

Dr. Katrin Moeller

Institut für Geschichte, Martin-Luther-Universität Halle-Wittenberg

- Fachspezifisches Datenzentrum für Geschichtswissenschaft
- Institutionelle Anbindung in der Fachwissenschaft seit 2008
- Schwerpunktbildung: biografisch-lebenslaufspezifische Daten im weitesten Sinne

**Universitäts- und
Landesbibliothek Halle
(ULB)**

Beratung,
Lehre,
Weiterbildung

Daten-
modellierung

Quellen- und
Datenauf-
bereitung
Datenkuration

Standards,
Normdaten
Ontologien

Publikation
und
Archivierung
von Daten



**Historisches Datenzentrum
Sachsen-Anhalt**

**Universitäts- und
Landesbibliothek Halle**

Beratung,
Lehre,
Weiterbildung

Daten-
modellierung

Quellen- und
Datenauf-
bereitung
Datenkuration

Standards,
Normdaten
Ontologien

Publikation
und
Archivierung
von Daten



**Historisches Datenzentrum
Sachsen-Anhalt**

Datenmodellierung, Forschungsdatenmanagement und digitale Methoden als zentrales Beratungs- und Lehrangebot

- Für alle wissenschaftlichen Projekte
- BA-, MA- und Lehramtsstudiengänge
- Verpflichtende Datenberatung und Speicherung Masterarbeit
- Schulungen für Citizen Science
- Workshops und Schulungen im Rahmen von DH

**Universitäts- und
Landesbibliothek Halle**

Beratung,
Lehre,
Weiterbildung

Daten-
modellierung

Quellen- und
Datenauf-
bereitung
Datenkuration

Standards,
Normdaten
Ontologien

Publikation
und
Archivierung
von Daten

Portal
Zugangs-
beschränkter
Daten



**Historisches Datenzentrum
Sachsen-Anhalt**

- Datenrepository zur langfristigen
Verfügbarmachung von digitalisierten Quellen und
Forschungsdaten durch die Universitäts- und
Landesbibliothek Sachsen-Anhalt/Rechenzentrum
Universität Halle nach FAIR-Prinzipien

- Spezialportal mit angebundenen Repository zur passfähigen Repräsentation von zugangsbeschränkten Daten (Datenschutz, Datenethik, personenbezogene Daten) z.B. Oral History
- Metadatenmodell zur übergreifenden, fachwissenschaftlich orientierten Erschließung von Tonarchiven

**Universitäts- und
Landesbibliothek Halle**

Beratung,
Lehre,
Weiterbildung

Daten-
modellierung

Quellen- und
Datenauf-
bereitung
Datenkuration

Standards,
Normdaten
Ontologien

Publikation
und
Archivierung
von Daten



**Historisches Datenzentrum
Sachsen-Anhalt**

- Quellenaufbereitung für historische Forschungen
 - Quellenaufnahmen
 - OCR-Erkennung mit Ocropus
- Digitalisierung von Quellen aus veralteten technischen Ressourcen (Magnetbänder, Kassetten, Lochkarten, veraltete Software und Datenbanksystemen)
- Recodierung und Wiederaufbereitung von langzeitarchivierten Daten zur Nachnutzung

- Spezialarbeitsplätze zur Quellenaufbereitung (alte und neue Technik)
- Spezialarbeitsplätze mit Spezialsoftware für geisteswissenschaftliche Forschungen, neben virtuellen Forschungsumgebungen vor allem Bild- und Tonanalyse, quantitative und qualitative Datenanalyse, Verarbeitung von Geodaten, Netzwerkanalyse, Bild- und Tonbearbeitung
- Computerpool für DH und geisteswissenschaftliche Forschungsumgebung

- ULB Halle: FID Orientalistik (Nahost-, Nordafrika- und Islamstudien): OCR-Erkennung orientalischer Schriften und Ground-Truth-Daten für OCR-D in mehreren Sprachen

- Historische Datenzentrum organisiert und führt durch Datenaufnahmen (Handschriften, Oral History) für eigene Forschungsprojekte (Datenmodellierung, Schulung Hilfskräfte, Paläographische Standards, Normalisierung und Datenbereinigung, Prüfung Datenqualität)
- Seit 2014 Aufnahme von 300.000 handschriftlichen Datensätzen (mit mehr als 50 Variablen)

**Universitäts- und
Landesbibliothek Halle**

Beratung,
Lehre,
Weiterbildung

Daten-
modellierung

Quellen- und
Datenauf-
bereitung
Datenkuration

Standards,
Normdaten
Ontologien

Publikation
und
Archivierung
von Daten



**Historisches Datenzentrum
Sachsen-Anhalt**

Fachliche Standards und Werkzeuge:

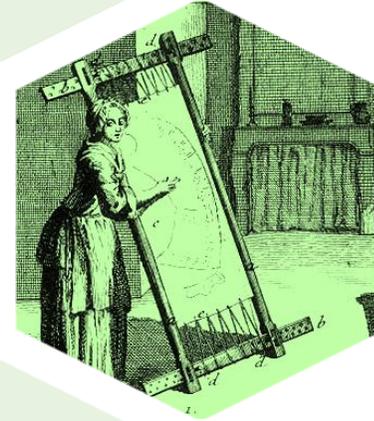
- Konkordanzen/Syntaxen für automatisierte Normalisierung, Lemmatisierung und Annotation für biografisch-lebenslaufspezifische Daten
- Record Linkage und Familienrekonstitution (automatisierte Identifikation von Personen in verschiedenen Daten)
- Ontologie der historischen, deutschsprachigen Amts- und Berufsbezeichnungen

Vom Standard



200.000
Varianten

20.000
Berufe



FAIRe-
Daten



zur Methode
und Datenanalyse

Beispiele eines HISCO-Codes in verschiedenen Dimensionen

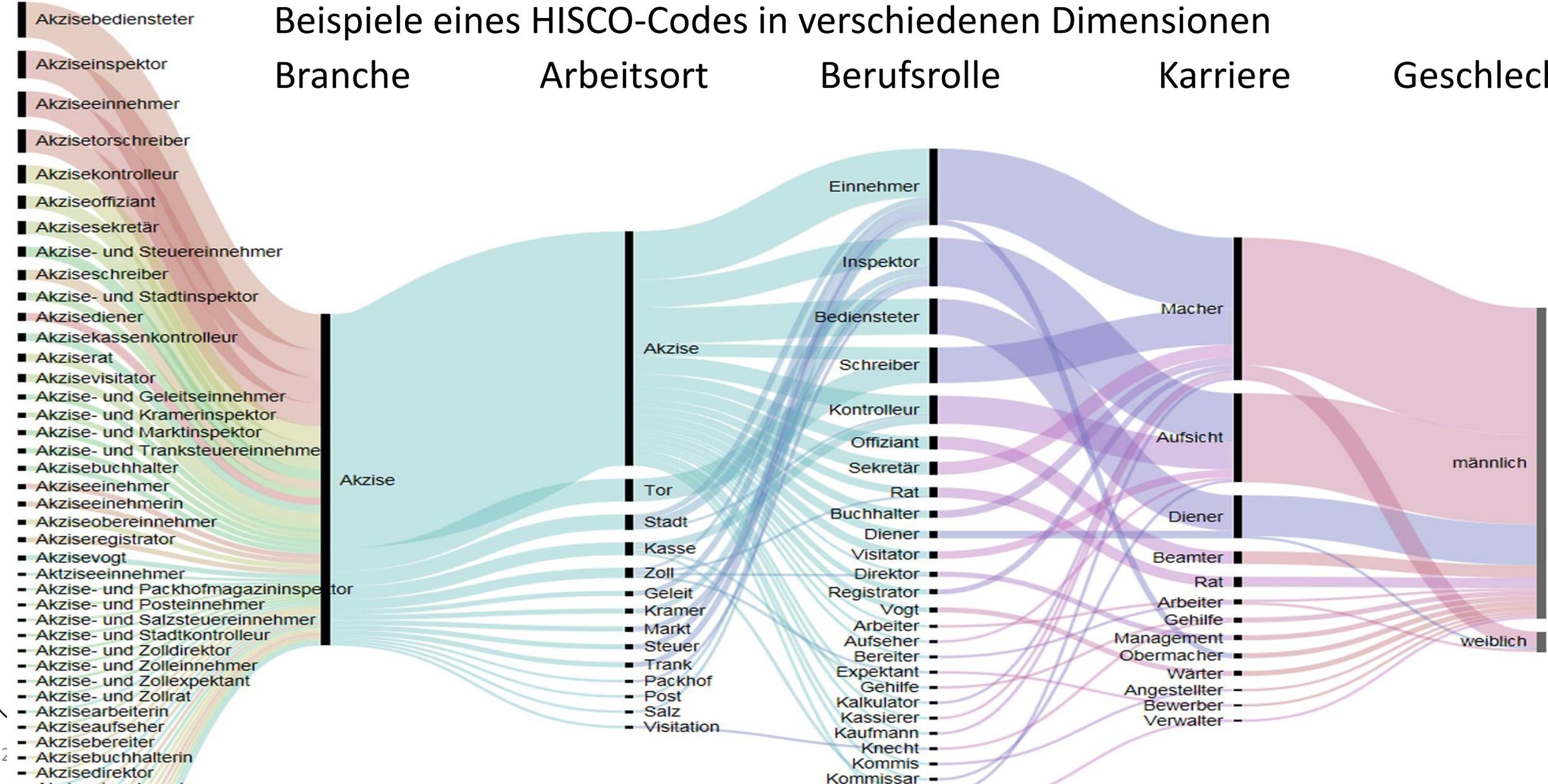
Branche

Arbeitsort

Berufsrolle

Karriere

Geschlecht



Fachliche Standards und Werkzeuge:

- Konkordanzen/Syntaxen für automatisierte Normalisierung, Lemmatisierung und Annotation für biografisch-lebenslaufspezifische Daten
- Record Linkage und Familienrekonstitution (automatisierte Identifikation von Personen in verschiedenen Daten)
- Ontologie der historischen, deutschsprachigen Amts- und Berufsbezeichnungen

