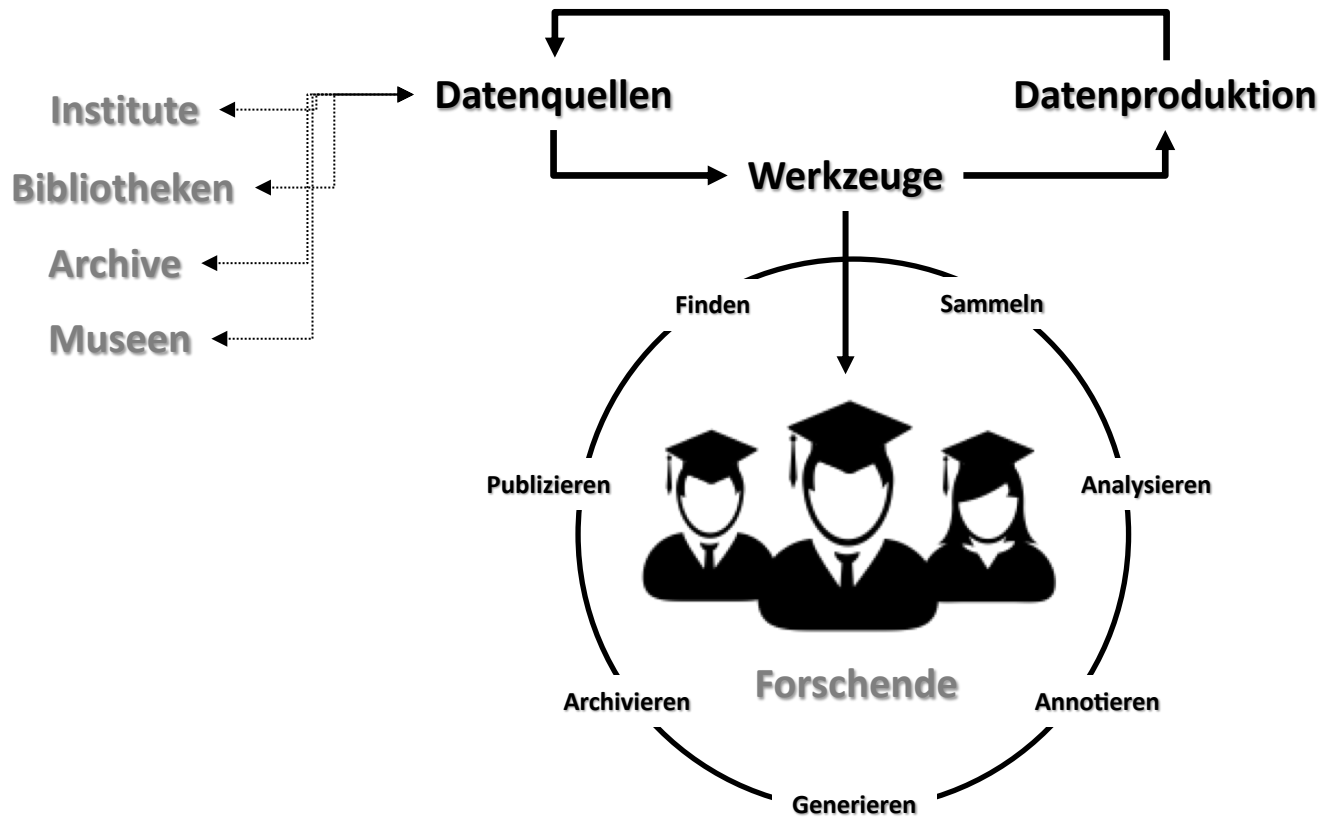


Geisteswissenschaftliche Forschungsdaten aus Bibliotheken

Wissenschaftsgeleitete
Forschungsinfrastrukturen für die Geistes- und
Kulturwissenschaften in Deutschland

15. Juni 2018, Berlin

Forschungsdateninfrastruktur für die Geistes- und Kulturwissenschaften



Forschungsdateninfrastruktur für die Geistes- und Kulturwissenschaften

Bibliotheksbestände

Literatur
Spezielsammlungen
Digitalisate

Lizenzierte Daten

Nationallizenzen
Allianzlizenz
FID-Lizenzen

Fachinformationsdienste

Beispiel Geschichte
Beispiel Kunst
Beispiel Karten

Institute

Bibliotheken

Archive

Museen

Datenquellen

Werkzeuge

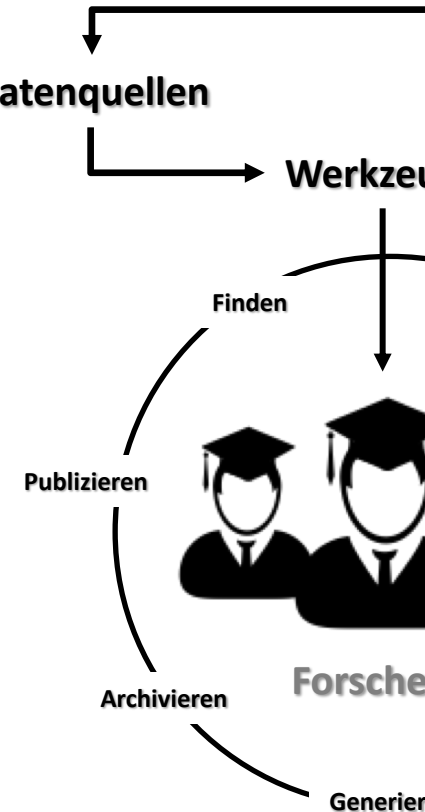
Finden

Publizieren

Archivieren

Forscher

Generieren



Bibliotheken als Infrastruktur für die Geistes- und Kulturwissenschaften

- Rd. 9100 öffentliche und rd. 260 wissenschaftliche Bibliotheken in Deutschland ([dbv](#))
- Wissenschaftlich: Länder und Bund mit c. 379 Mio. Medien
 - Bücher, Spezialsammlungen, Patente etc., o. Zeitschriften ([dbs](#))
- Davon bieten Hochschulen 2,9 Mio. Studierenden 185 Mio. Medien an ([Bibliotheksportal](#))
- Digital und besonders forschungsrelevant in den Geistes- und Kulturwissenschaften und
 - Verzeichnisse deutscher Drucke, inkl. OCR
 - Digitale Lizenzen

Die Verzeichnisse Deutscher Drucke

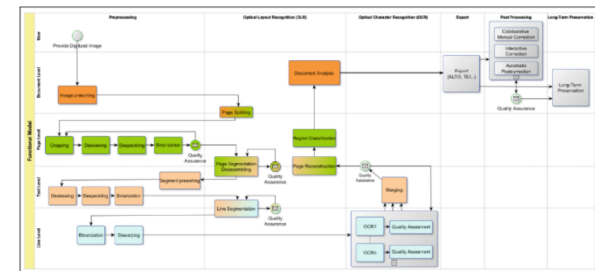
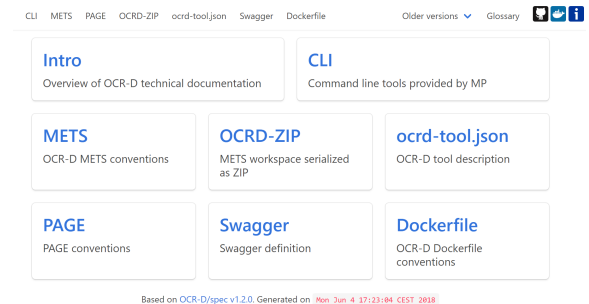
Projekt	Gesamtumfang	Bearbeitete Menge (Stand Mai 2018)
VD 16 (seit 1969, Digitalisierung seit 2006)	ca. 120.000 Drucke	Katalogisiert: 105.000 Drucke Digitalisiert: 60.000 Drucke
VD 17 (seit 1982, Digitalisierung seit 2006)	ca. 320.000 Drucke	Katalogisiert: 300.000 Drucke Digitalisiert: 130.000 Drucke
VD 18 (seit 2009)	ca. 585.000 Drucke	Katalogisiert und digitalisiert: 220.000 Drucke

Die insgesamt **410.000 Digitalisate** mit ca. **50 Millionen Seiten** aus VD 16, VD17 und VD 18 sind nicht nur in den VD-Datenbanken und den deutschen Verbundkatalogen nachgewiesen, sondern können auch im **zvdd**, der **DDB** und **europena** recherchiert werden. Hinzu kommen **Handschriften in Kalliope**.

Im Rahmen der laufenden VD-Projekte, aber auch darüber hinaus, werden zudem sukzessive die **125.000 Titel** digitalisiert, die über die **Arbeitsgemeinschaft Sammlung Deutscher Drucke** (AD SDD, www.ag-sdd.de) seit 1989 erworben wurden.

OCR-D: Koordinierte OCR für VD-Projekte

- Hintergrund
 - Volltexterschließung der digitalisierten Bestände bisher uneinheitlich
 - Massive Fortschritte bei OCR dank neuronaler Netze
- Phase 1 (2015-2018):
 - Entwicklung eines Funktionsmodells für die Domäne OCR
 - Bedarfsermittlung und Vorbereitung der Modulprojekte
- Phase 2 (2018-2020):
 - Betreuung der Modulprojekte und Integration der Ergebnisse
 - Prototypische Implementierung und Workflows
 - Weiterentwicklung der DFG-Digitalisierungsrichtlinien
- Stand 5. Juni 2018:
 - Modulprojekte haben Arbeit aufgenommen (Kaiserslautern, Mannheim, Leipzig, Göttingen, Mainz, Erlangen, München)
 - Spezifikationen mit Fokus auf Massendigitalisierung und LZA (<https://ocr-d.github.io>)
 - Agile, kooperative Entwicklung über GitHub (<https://github.com/OCR-D>)



Lizenzen in Deutschland

- Lokale Lizenzen / Konsortiallizenzen

mit DFG-Förderung

- **Modell Nationallizenzen**

- Zugriff für Nutzer/innen an allen wissenschaftlichen Einrichtungen in D
- 100% DFG-finanziert / Förderung bis 2010, d.h. Zugriff auf bis dahin erworbene Lizenzen, aber keine neuen Nationallizenzen

- **Allianzlizenzen**, zielen auf den Grundbedarf

- Opt-in Modell / Zugriff für Einrichtungen, die sich an der Finanzierung beteiligen
- 25% DFG-finanziert / 75% durch beteiligte Einrichtungen
- **Mehrwertkomponenten: Moving-Wall, OA-Komponente, Hosting- und TDM-Rechte**

- **FID-Lizenzen**, zielen auf den spezialisierten Bedarf

- Zugriff in unterschiedlichen Modellen (Nutzerkreis-definiert, Campus-Lizenz, Nationallizenz)
- 85% DFG-finanziert, 15% durch FID-Bibliothek
- nach Möglichkeit mit Mehrwertkomponenten wie bei den Allianz-Lizenzen

Fachinformationsdienste

Ziel der Förderung und Aufgaben

- Überregionalen, direkten und schnellen Zugriff zu Forschungsliteratur/Fachinformationen als Ergänzung zur Grundversorgung
- Fachbezogener Service zur Bereitstellung von Medien aller Art und mit modernen Nachweis-/Suchsysteme
- Vorausschauendes Engagement in der Entwicklung fachspezifischer Informationsdienstleistung
- Verbesserung des Zugangs zu digitalen Medien

Fördergrundsätze

- (1) Forschungsinteressen im Mittelpunkt, systematische Rückkopplung mit den Zielgruppen
- (2) Mehrwert der Angebote für die Forschung, Abgrenzung von Grundaufgaben

Projekte

- aktuell 37 Fachinformationsdienste in Förderung, davon 31 in den Geistes- und Sozialwissenschaften
- Fördervolumen: durchschnittlich ca. 15 Millionen € p.a.

Fachinformationsdienste **exemplarisch**

- Altertumswissenschaften
- Anglo-American Culture
- Benelux
- CrossAsia – Asien
- Finno-uralische Sprachen
- Germanistik
- **Geschichtswissenschaft**
- Jüdische Studien
- **Kartographie**
- **Kunst, Fotografie, Design**
- Literaturwissenschaft
- Linguistik
- Musikwissenschaft
- Nahost, Nordafrika, Islam
- Philosophie
- Ost-Ostmittel-Südeuropa
- Romanistik
- Slawistik
- Sozial & Kulturanthropol.
- Theologie

FID Geschichtswissenschaft

- **Formatunabhängige** Sammlungsorientierung, digitale Forschungsdaten gewinnen an Bedeutung
- Digitale Sammlungen: **2,3 Mio. Titel**, u. a. vergriffene Werke des 20. Jhds., 1 Mio. PDF-Downloads
- Forschungsdatenservice: „Daten für die Forschung“ (**DaFo**), 11.262 Aufträge
- Zugang zu Forschungsdaten: Nachweis in Recherche-instrumenten des FID
- **Infrastruktur für Elektronisches Publizieren:** z.B. recensio.net/regio, 33.000 Rezensionen, 50.000 Besuche

FID Geschichtswissenschaft

- In Vorbereitung: Dokumentenserver, dynamisches Publizieren = „**Kontextualisierungs**“-Modell für Forschungsdaten
- Aktive Beteiligung an der **NFDI-Diskussion** im Fach (u.a. VHD-Gremien)
- Eintritt der BSB in ein vom VHD koordiniertes **Konsortium Geschichtswissenschaft der NFDI**
 - Forschungsmassendaten für Quantitative Analysen
 - Elektronisches Publizieren und koordinierte Kuratierung
 - Erschließungskompetenzen und Standardisierung
 - Recherche und Sichtbarkeit der Daten
 - Langzeitarchivierung

FID Kunst, Fotografie, Design

- **Community-Orientierung:** 3.300 im VDK organisierte Kunsthistoriker, 77 kunsthist. Institute, 250 Dissertationen/Jahr, 678 Kunstmuseen ...
- **Community-Service:** Aggregieren von Datenbeständen aus Bibliothekskatalogen, Museen und Archiven, aus Bibliographien und Open Access Repositories, aus Datenbanken, E-Books, E-Journals und Internetquellen sowie aus Bilddatenbanken in **arthistoricum.net**.

Volumen: > 10 Mio. digitale Quellen

- **Spezialisierung auf bild- und materialbezogene Forschungsdaten:** Digitale Repräsentation, Dokumentation und analytische Auswertung von Kulturerbe

FID Kunst, Fotografie, Design

- **arthistoricum.net: Kompetenzen und Angebote für ein NFDI-Konsortium materielle und bildliche Kulturen**
 - Etablierte Netzwerke: Bibliotheken (AKMB), Bildarchive (AKBF), Arbeitskreise Kunstarchive, Digitale Kunstgeschichte, Graphik vernetzt
 - Recherche und Sichtbarkeit von Forschungsdaten
Qualifizierter digitaler Bildsuchraum, perspektivisch digitale Bildanalyse
 - Infrastruktur für Elektronisches Publizieren im Open Access (E-Books, E-Journals, Articles, Digitale Editionen, Digitale Werkverzeichnisse u.v.m.)
 - Normdaten und Standardisierung
 - Zertifizierte Langzeitarchivierung

FID Kartographie

- Starke **Vernetzung** mit der Community
- **Trans- und Interdisziplinarität** als Fakt und Prinzip
- **Kooperationsnetze** (FID, Forschungsprojekte, Staatl. Stellen)
- **Cockpitfunktion**
 - thematisch fokussierte **Digitalisierung**
 - projektbezogene **Erwerbung** von digitalen und analogen Karten & Lit.
 - Ausbau und Anreicherung der **Bibliographia Cartographica** mit Volltexten
 - Aufbau / Etablierung eines **Fachrepositoriums**

FID Kartographie

- **1,2 Mio Karten**, intensive Nutzung vor Ort
- **Daten:** ca. 3500 Orders / Jahr
- Beratung und Vermittlung von **Geodaten** aus aller Welt (Geobasisdaten, Geofachdaten)
- **NFDI:** Teilnahme an Konsortiumsgesprächen

Leitfragen des Workshops / Diskussion

- (1) Welche konkreten Angebote – abseits von Basisdiensten wie E-Mail, Webzugang, Storage und Computing-Leistungen – können wir für die geisteswissenschaftliche Forschung und Lehre bieten? (2) Welche konkreten Angebote bezogen auf die für die Geisteswissenschaften besonders wichtigen Datentypen kann meine Einrichtung für eine einzurichtende nationale digitale Forschungsinfrastruktur in den Geisteswissenschaften beitragen? (4) Welche konkreten Anforderungen der Geisteswissenschaften können die Angebote meiner Einrichtung abdecken? (6) Welche Angebote kann meine Einrichtung in einem nationalen Kontext über meinen Standort hinaus einbringen?
 - Datenquellen über Bestände, Lizenzen und FID, Zielort der Forschungsdatenproduktion
- (3) Welchen Einfluss haben meine digitalen Forschungsinfrastrukturangebote für die Geisteswissenschaften auf Lehre und Weiterbildung?
 - Verwendung über die Forschenden, aber auch die Öffentlichkeit
- (5) Welche Auswirkung haben die Angebote meiner Einrichtung für die Geisteswissenschaften für die Begutachtung und Antragstellung von Projekten?
 - Eine nachhaltige Forschungsbasis