Data Center for the Humanities

Umfragen und Analyse von Beratungsgesprächen als strategischer Wegweiser

DINI-nestor-Workshop „Bedarfserhebungen - Grundlage für passgenaue Infrastrukturen?“, 29. – 30. Januar 2019, Universität Duisburg-Essen

Jonathan Blumtritt, Patrick Helling, Brigitte Mathiak, Katja Metzmacher

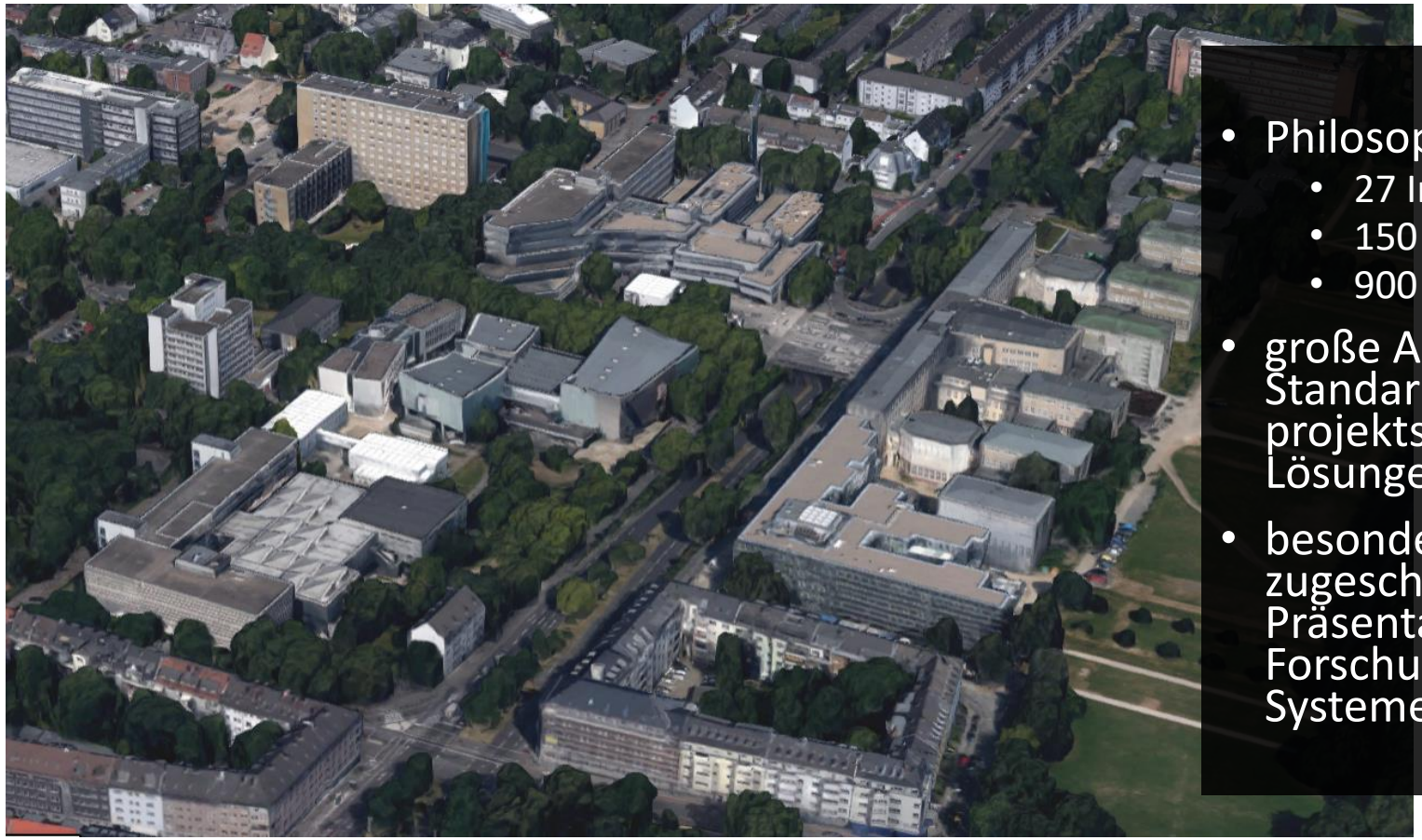
Universität zu Köln
DCH Data Center for the Humanities Kölner Datenzentrum für die Geisteswissenschaften

Data Center for the Humanities (DCH)

Kölner Datenzentrum für die Geisteswissenschaften

- zentrale Serviceeinrichtung der Philosophischen Fakultät der Universität zu Köln
- gegründet 2013, Satzung seit 2018
- unterstützt Forscher*innen an der Fakultät dabei, die dauerhafte Sicherung, Verfügbarkeit und Präsentation von Forschungsdaten und Ergebnissen der geisteswissenschaftlichen Forschung sicherzustellen
- Die Philosophische Fakultät gehört mit zuletzt rund 150 Professor*innen und fast 900 wissenschaftlichen Mitarbeiter*innen zu den größten geisteswissenschaftlichen Einrichtungen Deutschlands
- http://dch.uni-koeln.de

Forschungsdaten in den Geisteswissenschaften



- Philosophische Fakultät, Uni Köln
  - 27 Institute
  - 150 Professoren (Stand 2016/17)
  - 900 wissenschaftliche Beschäftigte
- große Anzahl fachspezifischen Standards/Methoden und projektspezifischen Lösungen/Formaten
- besondere Bedeutung von zugeschnittenen Präsentationssystemen für Forschungsprojekte (aka „lebende Systeme“)

Universität zu Köln
DCH Data Center for the Humanities Kölner Datenzentrum für die Geisteswissenschaften

Strategieentwicklung des DCH

- Bottom-up
- Bedarfserhebungen: Was wollen wir wissen? Was wollen wir erreichen?
  - (fachspezifische) Bedarfe der Forscher*innen
  - Services anbieten, die andere nicht anbieten
  - FDM und verschiedene Forschungstraditionen
  - Cluster-Strategien, wenn möglich
  - blinde Flecken

Universität zu Köln
DCH Data Center for the Humanities Kölner Datenzentrum für die Geisteswissenschaften

Bedarfserhebungen

Universität zu Köln
DCH Data Center for the Humanities
Kölner Datenzentrum für die Geisteswissenschaften

Bedarfserhebungen zu FDM 2016 & 2018

- Bedarfserhebungen zu FDM 2016 und 2018
- **Online-Fragebogen**: 2016 aus 24, 2018 aus 36 geschlossenen Fragen, bei kategorialen Antworten gab es 2018 immer die Möglichkeit, ergänzende Kategorien zu benennen.
- **2016:** Datensatz mit **N=136** Fällen (71% Beendigungsrate, nur vollständige Umfragen werden gezählt). (**30% der Professor*innen der Fakultät**)
- **Phil**: 215 Personen nahmen an der Umfrage teil, 128 beantworteten alle Fragen. Auch hier wurden einige nicht ganz vollständige Fälle in die Analyse miteinbezogen, der Teildatensatz beinhaltet **N=179** Fälle (83%). (knapp **37% der Professor*innen haben die Umfrage ausgefüllt**)
- Bei beiden Umfragen wurden Daten zu Charakterisierung, Speicherung, Aufbewahrung von FD, Kenntnisse in FDM, Position, Fächergruppenzugehörigkeit, Evaluation DCH abgefragt.
- 2018 wurde darüber hinaus noch Stellenwert von sekundärer Datennutzung, Distributionsverhalten und Anforderungen an Archive und Repositorien erhoben.

Universität zu Köln
DCH Data Center for the Humanities Kölner Datenzentrum für die Geisteswissenschaften

Was haben wir aus den Umfragen erfahren und wie haben die Bedarfsumfragen unsere Strategien verändert?

FDM an der Philosophischen Fakultät 2018

- **Textdaten bleiben die häufigsten Forschungsdaten** an der Phil Fak, die meisten Wissenschaftler*innen arbeiten mit Text und zusätzlich mit **multiplen anderen Daten**
- Über **80%** geben **das Gesamtvolumen** ihrer Forschungsdaten mit **bis zu 30TB an, ca. 47%** nur **bis zu 1TB**
- über **50%** der Wissenschaftler*innen arbeitet mit **personenbezogenen Daten**
- Knapp **80%** wollen ihre **Forschungsdaten so lange wie möglich aufbewahren**
- **die drei hauptsächlichen Aufbewahrungsorte von Forschungsdaten sind lokale Rechner (dienstlich/privat) und externe Speichermedien**
- **70%** hatten bereits **Probleme mit veralteten Datenformaten** oder befürchten dies für die Zukunft, **53%** haben **Probleme mit nicht sichergestellter Betreuung von Webressourcen nach Projektende** und **schwer auffindbaren Daten**
- Mehr als **40%** schätzen ihre **FDM Kenntnisse als gering/sehr gering ein, nur 13% als sehr gut/gut**
- Nur knapp **20%** haben einen **Datenmanagementplan**; mehr als **22% wissen nicht, was das ist**
- Die am **häufigsten genannten Probleme** im Umgang mit Forschungsdaten **wären mit gutem FDM vermeidbar gewesen**
- **Alle Serviceleistungen** werden von **mehr als 40% gewünscht, TOP 3: Beratung zu technischen Fragen, Beratung zu rechtlichen Fragen/Bereitstellung Speicherplatz gleich auf, Beratung zu allgemeinem Umgang mit Forschungsdaten**
- Über **2/3 wissen nicht**, dass das **DCH eine Beratung zu FDM anbietet**

Universität zu Köln
DCH Data Center for the Humanities Kölner Datenzentrum für die Geisteswissenschaften

Datennachnutzung und Distribution 2018

- Den **Mehrwert nachnutzbarer Forschungsdaten** halten über **80%** der Wissenschaftler*innen **für ihren Fachbereich** für **etwas hoch/hoch/sehr hoch**.
- **Weitergabe von Forschungsdaten** erfolgt meist jedoch über **persönliche Netzwerke und/oder Kontakte**.
- Nur ca. **1/3** hat bereits **Daten in ein Archiv** abgelegt, sogar **weniger haben Daten aus Archiven nachgenutzt**
- **Hauptgründe**, **Daten nicht abzulegen** waren nach **Unkenntnis** ein **zu hoher Aufwand (Zeit/Geld)** und das **Fehlen eines passgenauen Archivs**.
- **Wichtigste Kriterien für ein Archiv** wären: **Datensicherheit, Professionalität des Archivs** und die **Vertrauenswürdigkeit der Organisation des Archivs**.
- Der **wichtigste Faktor bei der Entscheidung**, wo die eigenen Daten abgelegt werden wollen, **spielt die Empfehlung von Kolleg*innen**.

Universität zu Köln
DCH Data Center for the Humanities Kölner Datenzentrum für die Geisteswissenschaften

Stärken von Online Surveys als Bedarfserhebung fürs DCH

- Viele Wissenschaftler*innen (der Fakultät) werden erreicht
- Auch Wissenschaftler*innen, die sich bisher wenig mit FDM auskennen, werden befragt
- Viele Fragen werden vergleichend abgefragt, Vergleich mit anderen Umfragen möglich (wenn bei item-Bildung bedacht)
- Relativ geringer Aufwand für Datenaufbereitung und –analyse
- Geringer Aufwand für Antwortende, Annahme von weniger sozial erwünschten Antworten als bei persönlichen Befragungen
- Gutes Tool, um Awareness zu messen und gleichzeitig zu erhöhen, Werbung in eigener Sache

Limitierungen von Online Surveys als Bedarfserhebung fürs DCH

- Komplexität von Forschungsdatenbeständen wird kaum oder nur implizit abgebildet
- Selbsteinschätzungen können sehr fehleranfällig sein, besonders bei Nicht-Expert*innen
- Keine Rückkopplungen zu einzelnen Projekten möglich; Fördersituationen bleiben unberücksichtigt (wg. Anonymität)
- Fachspezifität und Fachkulturen können im Sinne der Anonymität nur begrenzt abgefragt werden (bei uns auf Fächergruppenebene)
- Auch politisch/administrative Entscheidung, an welche Emaillisten Einladungen gehen → führt bei uns zu sehr geringer Rücklaufquote bei Promovierenden

Universität zu Köln
DCH Data Center for the Humanities Kölner Datenzentrum für die Geisteswissenschaften

Zwischenfazit – Strategieentwicklung durch unsere Online Surveys

- Online Surveys liefern wichtige Kennzahlen für die Kommunikation mit Dekanat, Unileitung, Rechenzentrum und Co.
- Möglichkeit, Forschungsdatenmanagement ins Bewusstsein der Mitglieder der Fakultät zu heben (→ Awareness)
- Möglichkeit, eigene Institution zu bewerben
- Spezifische FDM Kennzahlen für Geisteswissenschaften waren nicht ohne weiteres aus anderen, generischen Umfragen zu entnehmen
- Problemlagen „unserer“ Fakultät konnten in Projekte/Angebote überführt werden

Universität zu Köln
DCH Data Center for the Humanities Kölner Datenzentrum für die Geisteswissenschaften

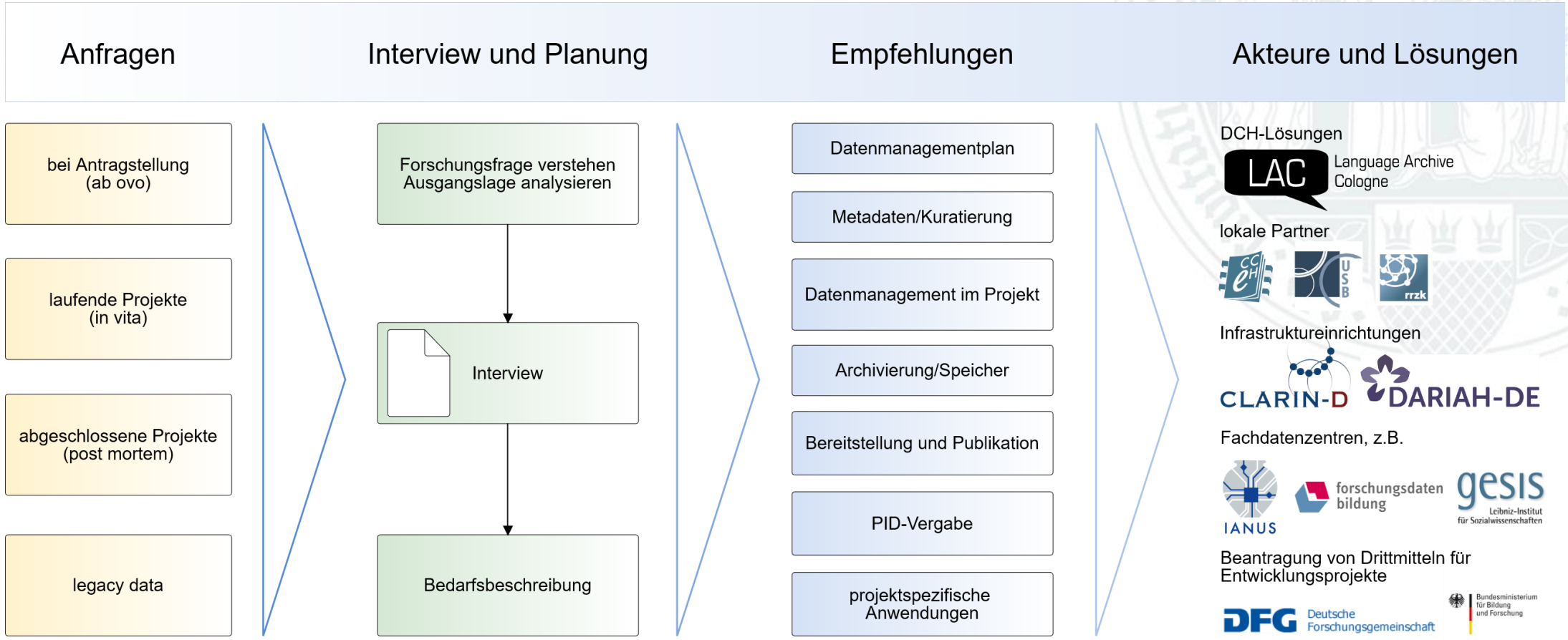
Beratungen am DCH

Beratungsgespräche am DCH

- Beratungsgespräche zu Fragen des Forschungsdatenmanagements sind das Kerngeschäft des DCH
- Sie bilden die Basis für aktives Forschungsdatenmanagement
- Zeitraum: **2015 – 2019**
- ca. **60** Beratungsgespräche mit Forscher*innen an der Philosophischen Fakultät der UzK und darüber hinaus

Bei allen Beratungsgesprächen wurden semistrukturierte Protokolle zur internen Dokumentation und Organisation erstellt

Beratungsworkflow



Beispiele?

„Im DFG-Projekt […] **nähert sich nach drei Jahren das Ende der Förderung** […] und wir überlegen gerade, wie wir die angefallenen Forschungsdaten am besten aufbewahren können. Im Antrag hatten wir die Formulierung "**Archivierung der Daten und Unterlagen**" aufgenommen, wobei es sich in diesem Projekt um typische geisteswissenschaftliche qualitative Daten handelt.

Leider haben wir zu Beginn des Projekts versäumt, Standards festzulegen, deshalb habe ich drei Fragen: […]

- Liegt nach dem Projektende das Urheberrecht - zum Beispiel für Präsentationen - beim Projekt oder bei mir als individueller Wissenschaftlerin?
- Wie sieht es mit dem **Nutzungsrecht von digitalen Fotos** aus, die ich in Archiven nach Registrierung eines persönlichen Accounts für den eigenen wissenschaftlichen Gebrauch gemacht habe, sowie bei Digitalisaten, die ich als Wissenschaftlerin für bestimmte Publikationen mit Projektgeldern erworben habe?
- Wie lässt sich nach dem Ende des Projekts die betreffende […] **Datenbank** am besten archivieren?“

Vorgänge 2018

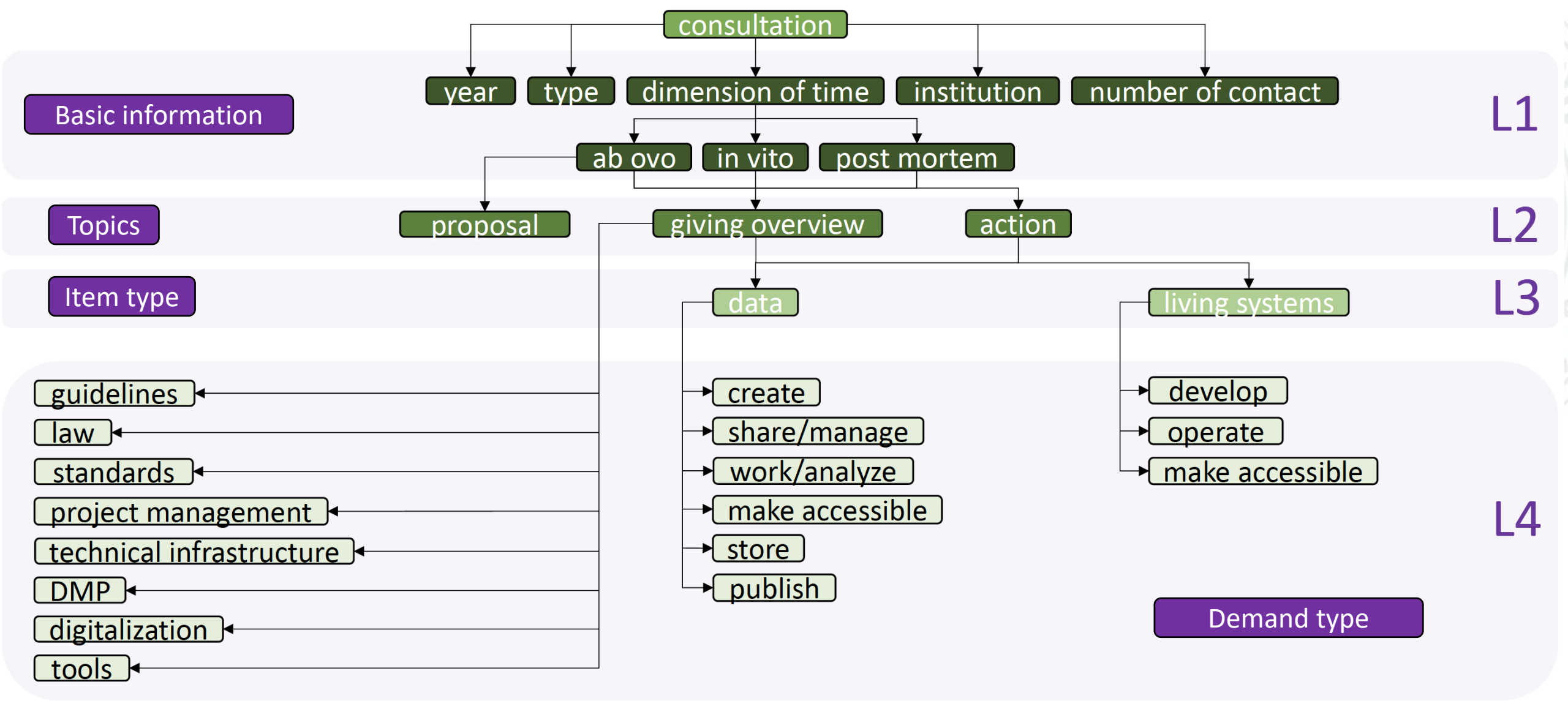
Name	Projekt	Institut	Projektstatus	Anfrage
**********	Computational Musicology	Musikwissenschaftliches Institut	Antragsstellung	Beteiligung am Antrag und Übernahme Workshop zu FDM im Projekt mit Geldbeteiligung
**********	Digital Pynchonology	Institut für Digital Humanities, Englisches Seminar I	Antragsstellung	Antragsberatung DMP mit Geldbeteiligung
**********	Regieanweisungen in Dramen der frühen Neuzeit	Institut für Medienkultur und Theater	Antragsstellung	Datenmanagement zu Projektstart mit Beratung zu Analysetool
**********	Verarbeitung von kinem. und akust. Daten von Essentiellen Tremor Patienten	Institut für Linguistik	Antragsstellung	Antragsberatung DMP
*******	Zorostrian Middle Persian: Digital Corpus and Dictionary (MPCD)	Institut für Digital Humanities	Antragsstellung	Datenmanagentplan, Archivierung, Publikation mit Geldbeteiligung
**********	GSSC Next Cloud	GSSC	ausgelaufen	Datenmanagement nach Ende Exzellenzgeld
**********	Early Cinema	Institut für Medienkultur und Theater	ausgelaufen	Wiederherstellung und dauerhafte Verfügbarkeit Webressource
**********	Digitalisierung Berichte Stasi	Institut für Deutsche Sprache und Literatur II	beginnend	Beratung zu Digitalisierung, Spracherkennung und Auswertung
********	eXist-Application für die Darstellung von Referenzgrammatik des Àbèsàbèsì	Institut für Linguistik	beginnend	Beratung zu langfristigen Verfügbarkeit einer zu entwickelnen eXist-Apllikation
**********	Merian Centre Latin America	GSSC/Inst. Für iberoamerikanische Geschichte	beginnend	Beratung zu Datenmanagement
**********	Rezeption grafischer Literatur durch Kinder und Jugendliche	Institut für Deutsche Sprache und Literatur II	beginnend	Beratung zu technischen Möglichkeiten
**********, **********	SagIchs	Department für Heilpädagogik und Rehabilitation, HF	beginnend	Beratung zur technischen Umsetzung eines Online Entscheidungstools
**********	Your Study	Department für Erziehungs- und Sozialwissenschaften, HF	beginnend	Datenmanagement zu Projektstart
**********	Inkunabeln	Historisches Institut	endend	Veröffentlichung der Forschungsdaten
**********	NRW80+	ceres	endend	Datenspeicherung nach Projektende
**********	Musikalische Preisausschreiben	CCeH	laufend	Betreuung der Webressource
*********	Sicherung Digitalisate Theaterwissenschaftliche Sammlung	Institut für Medienkultur und Theater	laufend	Bakup Strategie für Digitalisate
**********	Ausstellung im LWL Museum für Naturkunde Münster			Bereitstellung von Audiodaten
**********		Department für Erziehungs- und Sozialwissenschaften, HF		allgemeine Beratung zu Forschungsdatenmanagement

Beratungsanalyse

Beratungsanalyse am DCH

- Entwicklung einer Taxonomie zur Beschreibung von FDM-Vorgängen
  - Ausrichtung an bestehenden Modellen und Glossaren
  - Ausrichtung an bereits bestehenden Workflows im DCH

- Aktueller Stand der Formalisierung: Erfassung und erste Auswertungen der Daten aus dem Zeitraum **Mai 2016 – 2018**
  - Zunächst 31 Beratungsgespräche ausgewertet



Aktueller Stand 29. Januar 2019

Universität zu Köln
DCH Data Center for the Humanities
Kölner Datenzentrum für die Geisteswissenschaften

Beratungsanalyse - Grundlegendes

In welcher Projektphase wird das DCH konsultiert?

- Knapp **die Hälfte** der Forscher*innen kommen bei **laufender Projektphase** auf das DCH zu, **ein Drittel** bereits **vor Beginn** einer Projektphase
- Nur **jeder Fünfte** meldet sich erst **nach Ende** der Förderphase

Was ist der grundlegend nachgefragte Bedarf?

- Die **große Mehrheit** benötigt eine **allgemeine Übersicht** über FDM-Themen, **11%** fragen **Unterstützung bei einem Antrag** nach
- **Aktives Datenmanagement** durch das DCH wurde in **gut 1/5 der Fälle** benötigt

Details Beratung

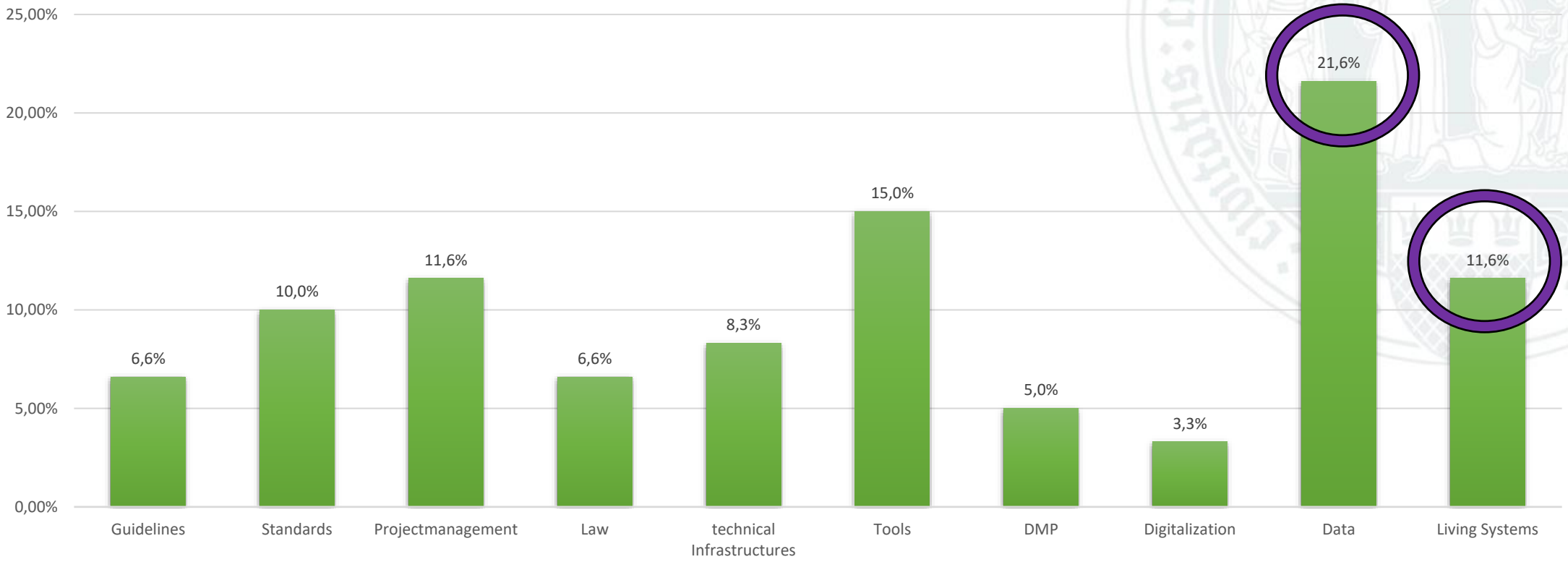


Abb. 5: Beratungsbedarfe in der Übersicht. (N=60)

Universität zu Köln
DCH Data Center for the Humanities Kölner Datenzentrum für die Geisteswissenschaften

Beratungsanalyse – Kategorie Daten

Allgemeine Übersicht

- Beim Bedürfnis eine allgemeine Übersicht beim Thema Daten zu erhalten, ging es in den meisten Fällen um die **Sicherung**, **Publikation von Daten** und die **Zugänglichkeit** von eigenen Daten

Aktives Datenmanagement

- Ähnlich verhält es sich, wenn aktives Datenmanagement seitens des DCH nachgefragt wurde: in **allen Fällen** wurden Daten durch das DCH **gesichert/archiviert** und in **3 von 4** Anfragen wurden Daten **zugänglich** gemacht oder **publiziert**

Beratungsanalyse – Kategorie Living Systems

Allgemeine Übersicht

- Wenn es um sich um den Bedarf nach einer allgemeinen Übersicht beim Thema Living Systems handelte, ging es in **über der Hälfte der Fälle** um die **Entwicklung**, **den Betrieb** oder **die Bewahrung der Zugänglichkeit** zu Lebenden Systemen.

Aktives Datenmanagement

- Ähnlich verhält es sich, wenn aktives Management seitens des DCH nachgefragt wurde: in **allen Fällen** wurde die **Zugänglichkeit** zu Lebenden Systemen gewährleistet, in **der Hälfte der Fälle** musste auch der **Betrieb** von Lebenden Systemen gesichert werden

Zwischenergebnisse der Beratungsanalyse

- Vorhandener Beratungsbedarf bei **Promotionsstudent*innen**
- nur knapp **jede fünfte** Beratung findet **„post mortem“** statt
- Fachspezifische **Beratung** bleibt weiterhin ein **Kernbereich** des FDM
- Häufige Bedarfsbereiche allgemein:
  - **Data, Living Systems, Tools und Project-Management**
- Häufige Bedarfsbereiche bei Daten:
  - **Speicherung, Verfügbarkeit und Publikation von Daten**
- Häufige Bedarfsbereiche bei lebenden Systemen:
  - **Zugänglichkeit von lebenden Systemen**

Stärken der Beratungsanalyse

Die Analyse der Beratungsgespräche ermöglicht…

- quantitative und inhaltliche Messbarkeit eigener Arbeit
- verlässliche Darstellung des Ist-Zustandes
- Abbildung von Entwicklungen/Erstellung von Prognosen
- Umfassende Analyse und Darstellung von FDM innerhalb einzelner Projekte/Fachbereiche
- Entwicklung zielorientierter Strategien und Workflows
- mehr Transparenz nach innen und nach außen

Für die Analyse der Beratungsgespräche ist keine weitere Datenerhebung notwendig.

Limitierung der Beratungsanalyse

- Keine (repräsentative) Erfassung des Stands von FDM an der Fakultät
- Wir modellieren unsere Sicht auf das Thema Forschungsdatenmanagement (Frage der Vergleichbarkeit zu anderen FDM-Institutionen)
- Keine Awareness-Maßnahme im Sinne einer Umfrage
- Aufwendiger Prozess

Fazit – to be continued….

- Online Surveys und die Analyse von Beratungsgesprächen liefern ganz unterschiedliche Ergebnisse auf verschiedenen Ebenen
- Online Surveys:
  - Grober Überblick
  - Erreicht viele Wissenschaftler
  - Eigene Institution bewerben
  - Awareness
  - Kennzahlen
- Situation in den einzelnen Fächern sehr relevant für FDM!!
  - Surveys zu Nutzungspraktiken in den Fachwissenschaften?

Fazit

- Analyse der Beratungsgespräche bieten sehr tiefen Einblick in konkrete Anfragen mit allen technischen Details und in Fachkulturen
- Für Verbesserung des Beratungs-/Arbeitsworkflow ist die Analyse der Beratungsgespräche sehr erfolgsversprechend
  - Hier auch tiefer Einblick in einzelne Projekte
  - Bessere Einblicke in die Fächer durch intensiven, persönlichen Kontakt
- Begonnen: Follow-up/Feedbacks-Interviews der Beratungsgespräche

→ mixed-methods Ansätze

Quellen

- Witt, A.; Blumtritt, J.; Helling, P.; Mathiak, B.; Rau, F.: Forschungsdatenmanagement in den Geisteswissenschaften an der Universität zu Köln. In: o-bib. Das offene Bibliotheksjournal 5 (3), 2018, S. 104-117, doi:10.5282/o-bib/2018H3S104-117.
- Simukovic, E.; Kindling, M.; Schirmbacher, P.: Umfrage zum Umgang mit digitalen Forschungsdaten an der Humboldt-Universität zu Köln, 2013, doi:10.18452/13568.
- Heinrich, M.; Jahn, S.; Schäfer, D.: Stakeholderanalyse 2013 zu Forschungsdaten in den Altertumswissenschaften, 2014, doi:10.13149/000.jah37w-g.
- Helling, P.; Blumtritt, J.; Mathiak, B.: Der Beratungsworkflow des Data Center for the Humanities an der Universität zu Köln. In: o-bib. Das offene Bibliotheksjournal 5 (4), 2018, S. 248-275, doi:10.5282/o-bib/2018H4S248-261.
- Blumtritt, J.; Mathiak, B.: Consulting Workflow for Humanities Research Data, Forschungsdaten in den Geisteswissenschaften (FORGE 2016), Hamburg 2016, https://www.fdm.uni-hamburg.de/ueber-uns/a-nachrichten/aktivitaeten/forge16/praesentationen/programmheft.pdf#page=21, Stand. 21.01.2019
- Tenopir, C. et al. 2011. Data Sharing by Scientists: Practices and Perceptions. PLoS ONE 6(6): e21101. doi: 10.1371/journal.pone.0021101.
- Tenopir, C. et al. 2015. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists worldwide. PLoS ONE 10(8): e0134826. doi:10.1371/journal.pone.0134826.
- Fecher, B., Friesike, S., and Hebing, M. (2015). What drives academic data sharing?. PloS one, 10(2), e0118053.