

5. Sitzung des Wissenschaftlichen Beirates von DARIAH-DE

- Protokoll -

Freitag, 10. März 2017, 11:00-16:00 Uhr

Veranstaltungsort

Universität Würzburg
Zentrales Hörsaalgebäude
Am Hubland
97974 Würzburg

Agenda

- 11:00-11:10 Begrüßung & Vorstellungsrunde
- 11:10-11:30 DARIAH-DE III - aktueller Stand und Vision
- 11:30-12:00 Anschließende Diskussion
- 12:00-12:30 Wissenschaftlicher Stand DARIAH-DE
 - Wissenschaftliche Begleitforschung
 - Wissenschaftliche Sammlungen
 - Quantitative Datenanalyse
 - Annotieren, analysieren, visualisieren
- 12:30-13:00 Anschließende Diskussion
- 13:00-14:00 Mittagspause
- 14:00-14:30 Umsetzung des technischen und organisatorischen Konzepts
 - Technik
 - Portal
 - Organisation
 - Kooperationen
- 14:30-15:30 Anschließende Diskussion
- 15:30-16:00 Abschluss

Teilnehmende

Beiratsmitglieder

- Prof. Dr. Elisabeth Burr (Universität Leipzig, Französische / frankophone und italienische Sprachwissenschaft)
- Prof. Dr. Ortwin Dally (Abteilung Rom des Deutschen Archäologischen Instituts)
- Prof. Dr. Thomas Gloning (Universität Gießen, Zentrum für Medien und Interaktivität)
- Prof. Dr. Günther Görz (Universität Erlangen, Institut für Informatik)
- Petra Hätscher (Kommunikations-, Informations-, Medienzentrums der Universität Konstanz)
- Dr. Peter Leinen (Deutsche Nationalbibliothek Frankfurt, Leiter Informationstechnik)
- Prof. Dr. Peter Matussek (Universität Siegen, Medienwissenschaftliches Seminar)

- Prof. Dr. Andreas Rauber (Technische Universität Wien, Information & Software Engineering Group)
- Prof. Dr. Robert Sablatnig (Technische Universität Wien, Institut für rechnergestützte Automation)
- Frank Scholze (Karlsruher Institut für Technologie, Direktion Bibliothek)
- Prof. Dr. Manfred Stede (Universität Potsdam, Angewandte Computerlinguistik)

Gäste

- Hans Nerlich (PT-DLR-Geistes-, Sozial- und Kulturwissenschaften für das Bundesministerium für Bildung und Forschung)

DARIAH-DE Konsortium

- Dr. Mirjam Blümm (Staats- und Universitätsbibliothek Göttingen)
- Peter Gietz (DAASI International)
- Prof. Dr. Wolfram Horstmann (Staats- und Universitätsbibliothek Göttingen)
- Prof. Dr. Fotis Jannidis (Universität Würzburg, Lehrstuhl für Computerphilologie und Neuere deutsche Literatur)
- Dr. Tibor Kálmán (Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen)
- Dr. Beata Mache (Staats- und Universitätsbibliothek Göttingen)
- Prof. Dr. Andrea Rapp (TU Darmstadt, Institut für Sprach- und Literaturwissenschaft)
- Dr. Stefan Schmunk (Staats- und Universitätsbibliothek Göttingen)
- Dr. Ulrich Schwarldamm (Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen)
- Prof. Dr. Thomas Stäcker (Herzog August Bibliothek Wolfenbüttel)
- Dr. Rainer Stotzka (Karlsruher Institut für Technologie)
- Dirk Wintergrün (Max-Planck-Institut für Wissenschaftsgeschichte)

Protokoll

1. Begrüßung

- Begrüßung durch Prof. Dr. Horstmann im Namen des DARIAH-DE Konsortiums

2. Änderungen in der Zusammensetzung des Beirats

- neue Beiratsmitglieder:
 - Prof. Dr. Gudrun Gersmann (Universität zu Köln, Historisches Institut)
 - Petra Hätscher (Kommunikations-, Informations- und Medienzentrums (KIM), Universität Konstanz)
 - Prof. Dr. Gudrun Oevel (Zentrum für Informations- und Medientechnologien (IMT), Universität Paderborn)
- ausgeschiedene Beiratsmitglieder:
 - Prof. Dr. Christian Bischof (TU Darmstadt, Leitung Hochschulrechenzentrum)
 - Dr. Gabriele Buschmeier (Akademie der Wissenschaften und der Literatur Mainz)

- Prof. Dr. Dr. Dr. h.c. Ernest W.B. Hess- Lüttich (Universität Bern, Institut für Germanistik)
- Prof. Dr. Reinhold Kliegl (Universität Potsdam, Department Psychologie)
- Dank an alle ausgeschiedenen Beiratsmitglieder für die bisherige Unterstützung und an alle neuen Mitglieder für die Bereitschaft im Wiss. Beirat mitzuwirken

3. DARIAH-DE III - aktueller Stand und Vision

- Zentrales Thema der Sitzung: Nachhaltiger Betrieb von DARIAH-DE als digitale Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften in Deutschland
- Zusammenarbeit mit CLARIN:
 - gemeinsames Interesse an einer nationalen Infrastruktur
 - Wahrung der Heterogenität unter den GKW
 - Bestimmung der Gemeinsamkeiten (Basisdienste in der technischen Infrastruktur, Interoperabilität, Community-Advocacy-Outreach, Institutionalisierung)
- CLARIN-D und DARIAH-DE legten dem BMBF am 1. März 2017 ein gemeinsames Konzept „Nationale Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften (NRIC)“, vor, mit dem Ziel eine umfassende Forschungsinfrastruktur für Geistes- und Kulturwissenschaften in Deutschland in Form eines *National Research Infrastructure Consortium* (NRIC) zu etablieren und dauerhaft ihren Betrieb zu sichern. Das Dokument beschreibt den Status Quo und den Bedarf an Forschungsinfrastrukturen und benennt Betriebs- und Entwicklungskosten für DARIAH-DE und CLARIN-D
- Etablierung einer neuen rechtlichen Struktur: Konsortialstruktur auf der Bundesebene für die nationale Förderung der („personalbasierten“) Betriebskosten (mögliche Betriebsarten: Minimalbetrieb – es wird nur die Geschäftsstelle unterhalten; Regelbetrieb – Geschäftsstelle, Beratung der Community, (Weiter-)Entwicklung der Werkzeuge und Dienste wird sichergestellt; Vollbetrieb – Regelbetrieb und Betreuung der assoziierten Projekte)
- Finanzierung von Betriebskosten als ein Stufenmodell: Finanzierung durch Bund, Länder und Eigenmittel
- Cash-contribution für die Betriebsstelle EU – bisher im Rahmen der Projektförderung; wie es im nachhaltigem Betrieb geregelt werden soll, ist noch nicht endgültig geklärt, das BMBF sucht aber nach einer rechtlichen Regelung, die eine weitere Zahlung durch den Bund ermöglicht
- DARIAH-Vision: Aufbau von nachhaltigen digitalen Angeboten mit dezidiert fachwissenschaftlichen Schwerpunkten auf Forschung und Lehre, Forschungsdaten und auf die technische Infrastruktur(-komponenten); Sicherung der Weiterentwicklung und Integration „externer“ Dienste; Dauerbetrieb von verteilten, föderalen und forschungsorientierten Infrastrukturen

Anschließende Diskussion:

- DARIAH sollte die Verbindung zur EU-Ebene stärken, DARIAH und CLARIN sollten auch auf der EU-Ebene enger zusammenarbeiten.
Grundsätzlich ja, auch wenn das Konzept die nationale Ebene betrifft; NRIC müsse als nationale Konsortialstruktur mit mehreren Partnern zuerst rechtlich gesichert werden, Modelle für Betriebskostenfinanzierung müssen gefunden werden und in der nationalen Forschungslandschaft verankert.

Unsere Ergebnisse fließen als In-Kind-Contribution auf EU-Ebene ein; eine gemeinsame europäische Dateninfrastruktur entsteht - European Open Science Cloud (EOSC), allerdings noch ohne Schnittstellen, DARIAH-DE beobachtet diese Entwicklung. Eine Definition von wissenschaftlichen Workflows ist zusammen mit CLARIN in Arbeit.

- Gibt es Kontakte zu anderen ESFRI-Projekten bzw. ERICs z.B. aus den Sozialwissenschaften?
Kommunikation gibt es, aber es stellt sich die Frage, ob es eher eine Aufgabe von DARIAH-EU ist. Seitens sozialwissenschaftlicher Projekte gibt es Anfragen bezüglich der AAI.
- Geisteswissenschaften sind mehrsprachig, ist DARIAH-EU hier eine mögliche Lösung?
Wir müssen die Heterogenität der Geistes- und Kulturwissenschaften deutlich machen, die Bereiche erweitern, aber wir brauchen bestimmt noch 20 Jahre, um alles zu berücksichtigen. Kleine Fächer werden (zwar noch nicht planmäßig) aber schon in vielen Projekten berücksichtigt.
- „Wildwüchsige“ Entwicklungen sollten als eine Chance und nicht als ein Problem gesehen werden, nicht zentralisieren, sondern Querverbindungen stärken.
Ja, das strebt DARIAH als eine realistische und sinnvolle Vorgehensweise an.
- Zusammenarbeit mit den DFG-geförderten Projekten sollte verstärkt werden, z.B. mit Generic Research Data Infrastructure (GeRDI) und den Projekten der letzten Nachhaltigkeitsausschreibung
Alle arbeiten in die gleiche Richtung, was die Größe des Bedarfs deutlich macht. DFG-Projekte sind toolbasiert; DARIAH-DE könnte als eine die Qualitätsstandards und Nachhaltigkeit sichernde Instanz fungieren, beraten und fortgeschrittene, generische Tools nach dem Abschluss der Projekte in den DARIAH-Katalog aufnehmen. Heterogenität der Geistes- und Kulturwissenschaften muss berücksichtigt werden, generische Lösungen sind unzureichend.
- Gibt es eine Antwort vom RII auf das „Memorandum zur nachhaltigen Bereitstellung digitaler Forschungsinfrastrukturen für die Geistes- und Kulturwissenschaften in Deutschland?“
Keine direkte, da der RII neutral bleiben muss, um in die politischen Diskussionen hineinwirken zu können, aber die Argumente aus dem Memorandum sind im Papier des RIIs sichtbar.
- Der Projektträger ist mit den Arbeiten der letzten Zeit sehr zufrieden: Klare Strukturen und Zukunftsmodelle.

4. Wissenschaftlicher Stand DARIAH-DE

1. Wissenschaftliche Begleitforschung

- die Impactmatrix (ein Katalog für Impactfaktoren und Erfolgskriterien für digitale Infrastrukturen in den Geisteswissenschaften) wurde erfolgreich getestet und von der Community rege diskutiert;
- eine Analyse der Publikationswege wurde durchgeführt, dabei wurde festgestellt, dass die Reputation der neuen Publikationswege und die Sichtbarkeit der digitalen Methoden in den fachwissenschaftlichen Publikationen (auch durch politische Initiativen) erhöht werden soll.

Anschließende Diskussion:

- Digitale Ressourcen sollten als mehr Gewicht bei Begutachtungsverfahren oder Dissertationen erhalten, hierzu müssten die Hochschulordnungen überarbeitet werden. Einbindung der Hochschulleitungen und zuständigen Ministerien
- Ein Problem sind bislang fehlende Bewertungskriterien z.B. für die Arbeit in einem Wiki o.Ä.

2. Wissenschaftliche Sammlungen

- *Collection Registry* in Version 2.0 ist live geschaltet, aktuell sind 135 Sammlungen eingetragen, davon 37 mit einer Schnittstelle; als Use Case werden Bestände des Forschungsverbundes Marbach Weimar Wolfenbüttel (MWW) eigepflegt;
- eine Forschungsdaten-Architektur wird weiterentwickelt; Akzeptanz und Nutzung in den Fachcommunitys wird als ein wichtiges Ziel formuliert.

Anschließende Diskussion:

- Gibt es gemeinsame Ausschreibungen oder Kooperationen mit Museen?
Universitäre Sammlungen werden über das Stakeholdergremium „Wissenschaftliche Sammlungen“ in DARIAH-DE mit einbezogen, die Museumslandschaft ist im Moment noch zu heterogen um sie in DARIAH zu integrieren

3. Quantitative Datenanalyse

- zwei Tools werden (weiter)entwickelt – *CosmoTool* zur automatischen Generierung von Biographien aus strukturierten und unstrukturierten Quellen und *Topic Modeling* zur maschinellen Analyse der semantischen Strukturen in großen Textkorpora.

Anschließende Diskussion:

- Es ist wichtig bei der Entwicklung von den wissenschaftlichen Fragestellungen, nicht von einem Tool auszugehen
Beim Themenkomplex Topic Modelling konnten über die Zeit Veränderungen der Fragestellungen beobachtet werden, z.B. komplementäre Themen in Zeitschriftenkorpora
- Die Brücke von wissenschaftlichen Fragestellungen und Tools herzustellen ist oft eine Gratwanderung
DARIAH versucht die Brücke durch Use Cases, Publikationen in Fachzeitschriften und Python Notizbücher zum Austausch von Daten zu schlagen, dies wird sehr gut rezipiert

4. Annotieren, analysieren, visualisieren

- mehrere Use Cases, die auf Tagungen und in Workshops auf großes Interesse stoßen;
- Berücksichtigung von multimedialen Daten, Einsatz von Linked Open Data.

Anschließende Diskussion:

- Wie kann der Gedanke des Semantic Webs integriert werden?
Es ist eine Frage der Granularität, häufig seien nicht unbedingt große Korpora nötig, zur Beantwortung von Forschungsfragen kommt es v.a. auf die Scala an, die angelegt wird

5. Umsetzung des technischen und organisatorischen Konzepts

- Nachhaltigkeit auch auf internationalen Ebene, Darstellung der Dienste und Serviceangebote sowie Ausbau der Community werden kontinuierlich verbessert.
- Technik: zentrale Nutzerverwaltung, Beratungs- und Supportangebot mit Trackingsystem werden realisiert; Standarddienste (kein Antrag nötig) und Individualdienste (Nutzerantrag erforderlich) beschrieben und angeboten, Authentifizierungsverfahren (Vertrauensketten) für Dienste mit erhöhten Sicherheitsanforderungen (Metahosting, VMs) muss noch definiert werden.
- Portal: das Portal wurde den Empfehlungen entsprechend neu strukturiert, Inhalte in Hinblick auf die heterogenen Zielgruppen überarbeitet; Beschluss über den sofortigen Relaunch wurde vom Konsortium gefasst; eine Design-Agentur wird beauftragt.
- Organisation: Gemeinsames Konzept mit CLARIN-D (s. o.); Gründung des DARIAH-DE Coordinaton Office (Geschäftsstelle und DeISU); weitere Vernetzung mit DARIAH-EU (z.B. Sustainability- Workshop am 27.04.17 in Berlin).

Anschließende Diskussion:

Portal:

- Zustimmung zum sofortigen Relaunch, neue Version unter der alten Webadresse, alte Version soll noch bis zur endgültigen Übertragung der Inhalte online bleiben (unter einer neuen Adresse)
- Kooperation CLARIN-DARIAH auf dem Portal thematisieren
- viele Fallbeispiele anhand des Forschungsprozesses (Forschungsfrage, Lösung, Ergebnis)
- WBR möchte als Tester fungieren – Kommentierung ermöglichen und nach der Umschaltung WBR informieren
- Mehr grafische Darstellungen
- elementare Testverläufe mit echten Nutzern als Usability Tester durchführen (für Sommersemester mit Studierenden geplant)
- Suchfenster testen
- Liste von Studiengängen aktualisieren
- Seiten sprachlich überarbeiten
- Ein grundsätzliches Statement zur DARIAH auf der ersten Seite (Beschreibung der Hauptaufgaben)

Organisatorisches:

- Nachhaltigkeitsdiskurs bei DARIAH-DE wird von anderen DARIAH-EU-Partnern beobachtet; trotz der unterschiedlichen Situation in den einzelnen Partnerländern wird unsere Lösung als ein wichtiger Orientierungspunkt gesehen
- Projektträger geht von der „Ausstrahlung“ der in DARIAH-DE entwickelten Geschäfts- und Organisationsmodelle auf DARIAH-EU-Ebene aus
- eine internationale Diskussion ist notwendig, da die geistes- und kulturwissenschaftliche Forschung international ist, auch die Anforderungen der Philologien könnten auf diesem Wege besser berücksichtigt werden
- Gebührenpraxis schon in der aktuellen Phase entwickeln und testen (bei intensiver Absprache mit dem PT)

- Zusammenarbeit mit Gedächtniseinrichtungen verstärken (z.B. im Bereich Metadaten)
- Gebührenpraxis soll intensiv diskutiert, Kalkulationsgrundlagen beschrieben und getestet werden

Technik:

- Notwendigkeit der Absprache mit Fachwissenschaften zum Thema Langzeitarchivierung der Forschungsdaten: Soll alles archiviert werden? Wer entscheidet? Bisherigen Datenmengen sind nicht so groß, es sollte zuerst alles gesichert werden, Fachcommunity in die Entscheidungsprozesse miteinbeziehen.
- Interoperabilität als zentrales Thema führen.

6. Abschließende Diskussion

Der Wissenschaftliche Beirat bestätigt eine sinnvolle Umsetzung der Monita und lobt die erzielten Fortschritte. Der Projektträger bezeichnet das Konzeptpapier als eine wertvolle (und gut genutzte) Möglichkeit der Selbstreflexion.

Hinweise für weiteres Vorgehen:

Der Frage der Gebührenpraxis soll weiterhin nachgegangen werden. Dabei soll zwischen freien generischen Tools, grundsätzlich kostenpflichtigen Diensten und projektspezifischen Erweiterungen der Werkzeuge, unterschieden werden. Die mögliche Eigenleistung der Partner soll realistisch beschrieben werden (maximal 10 Prozent der zu erwartenden Kosten).