

Wissenschaftlicher Beirat - Protokoll -

Freitag, 12. September 2014, 11:00-16:00

Veranstaltungsort

Universität Würzburg
Am Hubland
97974 Würzburg
Zentrales Hörsaalgebäude, Raum 1.013

Agenda

- | | |
|-------------|---|
| 11:00-11:30 | Begrüßung
DARIAH-DE II und DARIAH-EU – ein Update |
| 11:30-13:00 | DARIAH-DE Cluster: erste Ergebnisse <ul style="list-style-type: none">• Cluster 1: Wissenschaftliche Begleitforschung• Cluster 4: Wissenschaftliche Sammlungen• Cluster 5: Big Data in den Geisteswissenschaften• Cluster 6: Fachwissenschaftliche Annotationen |
| 14:00-15:00 | DARIAH-DE Cluster: erste Ergebnisse (Fortsetzung) <ul style="list-style-type: none">• Cluster 2: eInfrastructure• Cluster 3: DARIAH eInfrastructure Service Unit (DeISU)• Aus den AGs:<ul style="list-style-type: none">- AG Research Data Lifecycle- AG Service Lifecycle |
| 15:00-16:00 | Diskussion und Abschluss |

TeilnehmerInnen (jeweils in alphabetischer Reihenfolge)

Beiratsmitglieder

- Prof. Dr. Christian **Bischof** (TU Darmstadt, Leitung Hochschulrechenzentrum)
- Prof. Dr. Thomas **Bürger** (Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden, Generaldirektor)
- Prof. Dr. Elisabeth **Burr** (Universität Leipzig, Französische / frankophone und italienische Sprachwissenschaft)
- Prof. Dr. Thomas **Gloning** (Universität Gießen, Zentrum für Medien und Interaktivität)
- Prof. Dr. Günther **Görz** (Universität Erlangen, Institut für Informatik)
- Dr. Peter **Leinen** (Universität Mannheim, Leitung Rechenzentrum)
- Prof. Dr. Wolfgang **Nagel** (TU Dresden, Direktion Zentrum für Informationsdienste und Hochleistungsrechnen)
- Ao. Univ. Prof. Dipl.-Ing. Dr. techn. Andreas **Rauber** (TU Wien, Information & Software Engineering Group)
- Frank **Scholze** (Karlsruher Institut für Technologie, Direktion Bibliothek)
- Prof. Dr. Angelika **Storrer** (Universität Mannheim, Lehrstuhl für Germanistische Linguistik, Seminar für deutsche Philologie)

Gäste

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

DARIAH-DE Konsortium

- Dr. Mirjam **Blümm** (Niedersächsische Staats- und Universitätsbibliothek Göttingen)
- Prof. Dr. Axel **Horstmann** (ProWiss)
- Heiko **Hütter** (DAASI International)
- Prof. Dr. Fotis **Jannidis** (Universität Würzburg, Lehrstuhl für Computerphilologie und Neuere deutsche Literatur)
- Dr. Heike **Neuroth** (Niedersächsische Staats- und Universitätsbibliothek Göttingen)
- Andreas **Schmidt** (Rechenzentrum Garching)
- Dr. Stefan **Schmunk** (Niedersächsische Staats- und Universitätsbibliothek Göttingen)
- Dr. Ulrich **Schwardmann** (Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen)
- Dr. Juliane **Stiller** (Max-Planck-Institut für Wissenschaftsgeschichte)
- Dr. Rainer **Stotzka** (Karlsruher Institut für Technologie)
- Prof. Dr. Manfred **Thaller** (Universität Köln – Historisch-Kulturwissenschaftliche Informationsverarbeitung)

Protokoll

(zusammengefasst nach Themen und in zeitlicher Abfolge des Diskussionsverlaufs)

1. DARIAH-DE II und DARIAH-EU - ein Update

- DARIAH-DE startete im März 2014 in die zweite Förderphase (Laufzeit bis Februar 2016) mit 20 Partnern; neue Partner sind: DHd-Verband / Uni Hamburg, Uni Leipzig, HAB Wolfenbüttel, MPI-WG, Open Knowledge Foundation;
- DARIAH-EU hat seit 15. August 2014 eine eigene Rechtsform: DARIAH-ERIC mit 15 Gründungsmitgliedern: Belgien, Dänemark, Deutschland, Frankreich (Host-Country), Griechenland, Irland, Kroatien, Luxemburg, Malta, Niederlande, Österreich, Serbien, Slowenien, Zypern; Deutschland ist zusammen mit Frankreich und den Niederlanden seit zehn Jahren in DARIAH engagiert;

2. DARIAH-DE Cluster: erste Ergebnisse

a. Cluster 1 - Wissenschaftliche Begleitforschung

- Fragen zu Usability sollten schon als Komponente bei der Toolentwicklung Berücksichtigung finden und dabei nicht nur auf visuelle Aspekte sondern auch auf weitere Kriterien ausgeweitet werden wie z.B. Effizienz etc.
- Generell sollte bei der Begleitforschung / Usability ein breites Spektrum von Geisteswissenschaften abgedeckt werden, also über die Geschichtswissenschaften hinausgehen; Vorarbeiten der Archäologie sollten in Hinblick auf Usability einbezogen werden; Einbeziehung von Wissenschaften, die mehr auf Bildern (statt auf Texten) basieren (→ Kontakt zu Georg Schelbert);
- Wissenschaftliche Begleitforschung sollte auch Studierende miteinbeziehen;
- Weiterverwendung von Datenaufnahme (z.B. Auszeichnung von Texten in TEI) durch Dritte → Mehraufwand nützt nicht dem Individuum aber der Gemeinschaft: Wie kann man Wissenschaftler dazu animieren aufwendige Datenauszeichnungen zu machen?
- Wiedererkennungswert muss nicht so hoch bewertet werden, es gibt auch Gründe warum nicht alles gleich aussehen muss → Usability sollte auch Individualisierung erlauben;

b. Cluster 4 - Wissenschaftliche Sammlungen

- In der Collection Registry werden nur Sammlungsbeschreibungen verschlagwortet, im DARIAH-Repository auch einzelne digitale Objekte; es sollte geprüft werden, ob ein Thesaurus oder Schnittstellen für kontrollierte Vokabulare angeboten werden können oder das Schlagwortsystem von den Nutzern generiert werden soll;
- Ein Handlungsstrang von DARIAH-DE könnte sein, ein Social Tagging System zu etablieren, da sich Kategorien ständig ändern bzw. erweitern (→ Kontakt zu Hubertus Kohle);
- Bei der Beschreibung wissenschaftlicher Sammlungen muss darauf geachtet werden, dass kein Mapping-Problem in Bezug auf museale Sammlungen entsteht; hier empfiehlt sich die Entwicklungen von Europeana einzubeziehen;

c. Cluster 5 - Big Data in den Geisteswissenschaften

- Der Begriff „Big Data“ scheint angesichts eines Korpus von ca. 2.000 deutschsprachigen Romanen relativ; computergestützte Analysen dauern nur wenige Sekunden;
- Die Tools sollten unter heuristischen Gesichtspunkten nochmal auf größeren Korpora ausgetestet werden, um weitere (Forschungs-)Fragen zu produzieren;
- Eine weitere Idee wäre, Genres nach den Textmerkmalen der Linguistik experimentell zu untersuchen und die Ergebnisse auszuwerten;
- Dem Problem der Mehrsprachigkeit von Texte könnte mit DeKaPro begegnet werden, das verschiedene Sprachen mit linguistischen Tools unterstützt und über eine einheitliche Schnittstelle angesprochen werden kann;

d. Cluster 6 - Fachwissenschaftliche Annotationen

- Der Open Annotation Standard sollte weiter verfolgt werden, da er verspricht zu dem einschlägigen Standard für Annotationen zu werden;
- Projektübergreifende Annotationen wären sehr wünschenswert, die Speicherung sollte intern (im Projekt, in DARIAH) erfolgen;
- Die Polygone, die mit dem Tool Sem to Notes erzeugt werden, sollten TEI-konform sein;

e. Cluster 2 - eInfrastructure

- Bei den Ebenen von Data Curation müssen verschiedene Ebenen berücksichtigt werden, die Sicht der Rechenzentren, die Sicht der Fachwissenschaftler, ...;
- Die Migrationsfähigkeit von Daten sollte gleich im Entstehungsprozess integriert werden („Modell Blackbox“);
- Der Standardisierung von Metadaten kommt bei der Replikation ein hoher Stellenwert zu;

f. Cluster 3 - DARIAH eInfrastructure Service Unit (DeISU)

- Die DeISU ist ein Baustein, um die DARIAH-DE Infrastruktur zu sichern und nachhaltig zur Verfügung zu stellen;
- Dies kann aber nur dann gelingen, wenn entsprechende Vereinbarungen mit Rechenzentren, Förderern, Institutionen und/oder Forschungseinrichtungen geschlossen werden können und vor allem nur dann, wenn es auch gelingt die fachwissenschaftliche Community anzubinden.

g. aus den AGs

AG Research Data Lifecycle

- Die Motivation für die Community ist hoch, wenn von Anfang an konkrete Werkzeuge implementiert werden;
- Vom konkreten Werkzeug als Ausgangspunkt sollte abstrahiert werden, um einen generischen Nutzen sicher zu stellen;

AG Service Lifecycle

- Prozesse für die Aufnahme neuer Tools und Dienste sind definiert. Allerdings muss hierzu auch eine Regelung getroffen werden, wann und wie ein Dienst abgeschaltet wird;
- DARIAH-DE übernimmt Dienste ohne sie selbst weiter zu entwickeln, falls sie nicht mehr betrieben werden können, der Bedarf aus der Community jedoch besteht, muss ein Antrag auf Finanzierung für die Weiterentwicklung gestellt werden;

3. Eindrücke und Empfehlungen des Wissenschaftlichen Beirates

- DARIAH-DE ist auf einem guten Weg, es hat die Projektlaufzeit gut genutzt und ist fachwissenschaftlich sowie technisch gut aufgestellt;
- Die Rolle von digitalen Infrastrukturen für Erkenntnisse in der geisteswissenschaftlichen Forschung sollte stärker herausgestellt werden („Mehrwert“);
- Mögliche Gemeinsamkeiten mit CLARIN z.B. im Bereich Standardisierung sollten eruiert und genutzt werden;
- Generell muss darauf geachtet werden, Arbeit nicht zu duplizieren sondern möglichst kooperativ anzugehen;
- Es muss gelingen, den wissenschaftlichen Nachwuchs stärker einzubinden und Digital Humanities in die Infrastruktur – sowohl als Anforderung, als auch in der Lehre – der Universitäten zu integrieren;
- DARIAH-DE ist immer noch sehr fokussiert auf Text, durch Einbindung der digitalen Kunstgeschichte könnte eine Erweiterung um den Untersuchungsgegenstand Bild erreicht werden;
- Eine Möglichkeit der Verstetigung für DARIAH-DE könnte die Einbindung weiterer Cluster sein;
- Gespräche wie die Verstetigung von DARIAH-DE aussehen kann werden gerade mit Förderern und Stakeholdern durchgeführt;