

## Workshop zum Forschungsdatenmanagement in der Romanistik

### **Inhalt**

#### **1. Einleitung**

#### **2. Themenblock 1: Suche nach Forschungsdaten**

- 2.1. Suchinstrumente für romanistische Forschungsdaten (Jan Rohden)
- 2.2. Bisherige Erfahrungen der WissenschaftlerInnen mit der Suche nach Forschungsdaten
- 2.3. Probleme der aktuellen Nachweissituation
- 2.4. Möglichkeiten des neuen Meldesystems - Anforderungen an den Datenbestand
- 2.5. Zusätzliche Beschreibungskategorien für Forschungsdaten - Desiderate

#### **3. Themenkomplex 2: Arbeit mit Forschungsdaten**

- 3.1. Erstellen und Publizieren von Textsammlungen mit [GitHub](#) und [Zenodo](#) (Christof Schöch)
- 3.2. Nachhaltigkeit - aus der Sicht virtueller Forschungsumgebungen (Thomas Krefeld)
- 3.3. Diskussion der Bedarfe

#### **4. Themenkomplex 3: Notwendige Beratungsangebote**

- 4.1. Existierende Informations- und Beratungsangebote (Jan Rohden)
- 4.2. Eruierung der Bedarfe in der Romanistik

#### **5. Fazit**

### **Feedback**

### **Anhang**

## 1. Einleitung

Am 13. März 2017 fand an der Universitäts- und Landesbibliothek (ULB) Bonn ein von der *AG Digitale Romanistik* und dem *Fachinformationsdienst (FID) Romanistik* veranstalteter Workshop statt mit dem Ziel, die spezifisch romanistischen Bedarfe im Bereich Forschungsdatenmanagement zu ermitteln, um auf diese Weise die Grundlage für spätere Unterstützungsmaßnahmen zu schaffen.

In seiner Begrüßung betonte der leitende Direktor der ULB Bonn Ulrich Meyer-Doeringhaus die aktuelle Bedeutung von Forschungsdaten für die heutige Wissenschaftslandschaft und verwies auf zahlreiche Programme, Richtlinien und Positionspapiere, die unter anderem von der *Deutschen Forschungsgemeinschaft (DFG)*, der *Leibniz-Gemeinschaft*, dem *Bundesministerium für Bildung und Forschung*, der *Allianz-Initiative*, dem *Wissenschaftsrat*, der *Hochschulrektorenkonferenz* und dem *Rat für Informationsinfrastrukturen* vorgelegt wurden, sowie auf ein gemeinsam von Bund und Ländern getragenes Programm, das zum Jahresende von der Gemeinsamen Wissenschaftskonferenz (GWK) zu erwarten sei. Die aktuelle Auseinandersetzung der WissenschaftlerInnen mit der Thematik beruhe dabei nicht nur auf den spezifischen Vorgaben der Förderinstitutionen, sondern auch auf dem wachsenden Bewusstsein für die Vorteile, die ein transparenter und nachhaltiger Umgang mit Forschungsdaten für die eigene wissenschaftliche Arbeit biete. Vor diesem Hintergrund verwundere es nicht, dass immer mehr Hochschulen entsprechende eigene Leitlinien und Policies entwickeln.

Zum Einstieg in das Workshop-Programm situierte die FID-Referentin Doris Grüter das Thema im Kontext des FID Romanistik und stellte kurz die Vorhaben zum Forschungsdatenmanagement in der aktuellen Projektphase vor. In diesem Rahmen nannte sie auch das kürzlich fertiggestellte [Meldesystem](#) auf *romanistik.de*, das RomanistInnen die Möglichkeit bietet, der Wissenschaftscommunity Forschungsdaten als publikationsäquivalente Leistung aufzuzeigen.

Anschließend wurden die Bedarfe der romanistischen Fachgemeinschaft anhand von im Vorfeld verschickten Leitfragen in drei Themenblöcken (Suche nach Forschungsdaten, Arbeit mit Forschungsdaten, notwendige Beratungsangebote) diskutiert.

## 2. Themenblock 1: Suche nach Forschungsdaten

### 2.1. Suchinstrumente für romanistische Forschungsdaten (Jan Rohden)

Der erste Themenbereich „Suche nach Forschungsdaten“ wurde mit einem Impulsvortrag von Jan Rohden eröffnet. Nach einer kurzen Definition des Begriffs „Forschungsdaten“ nannte er neben allgemeinen Suchmaschinen wie *Google* eine Reihe von Nachweissystemen, die romanistische Forschungsdaten enthalten: u.a. Portale, die sich eher auf spezifische Sparten wie die Linguistik, die Literatur- und

Kulturwissenschaften oder die (Fach-)Didaktik beziehen wie z.B. *Linguistik.de*, *Clarín VLO*, *TextGrid Repository* und *Forschungsdaten Bildung*, sowie Verzeichnisse, die sich speziell der Romanistik widmen, wie *Textdaten Romanistik* und die Internetquellensammlung des FID Romanistik. Anschließend erläuterte er anhand von zwei einschlägigen Beispielen ([Metashare](#) und [romanistik.de](#)) die unterschiedlichen Qualitäten von Suchinstrumenten hinsichtlich

- der Menge an verzeichneten romanistischen Ressourcen,
- der inhaltlichen Ausrichtung,
- des Servicespektrums (vom bloßen Nachweis über den Zugriff bis zum Hosting),
- der Filter- und Suchmöglichkeiten,
- des Umfangs an Beschreibungskategorien,
- die Provenienz der Nachweise.

## 2.2. Bisherige Erfahrungen der WissenschaftlerInnen mit der Suche nach Forschungsdaten

Die anschließende Diskussion kreiste zunächst um die bisherigen Suchpraktiken der TeilnehmerInnen. Dabei wurden zwei Zielrichtungen unterschieden.

Zum einen ging es um die Zusammenstellung von digitalen Ressourcen, die sich als Ausgangsmaterial für die eigentliche Forschungsarbeit eigneten. Hier wird den Teilnehmenden zufolge in Abhängigkeit von der jeweiligen Fragestellung v.a. in öffentlich zugänglichen digitalen Textsammlungen und Quellen recherchiert. Beispielhaft wurden u.a. die Plattformen *Gallica*, *Ebooks libres et gratuits*, die über den *Karlsruher Virtuellen Katalog (KVK)* durchsuchbaren Portale sowie Archivseiten thematisch einschlägiger Institutionen genannt. Urheberrechtlich geschütztes Material und kommerziell vertriebene Ressourcen blieben hingegen wegen der Zugangs- und Nutzungsbeschränkungen oft unberücksichtigt, obwohl sie inhaltlich durchaus in Frage kämen.

Zum anderen wurde nach Forschungsdaten und -projekten gesucht, die dem eigenen Forschungsbereich nahe stünden und somit für die Konzeption der eigenen Vorhaben relevant sein könnten. Hier würden v.a. Projektwebsites, die sich in der Vergangenheit als erfolgsversprechend erwiesen haben, in regelmäßigen Abständen konsultiert und direkte Kontakte zu den bekannten Forschergruppen gepflegt, um sich auf dem Laufenden zu halten. Entsprechendes gelte für die Suche nach Tools.

Da sich diese auf persönlichen Erfahrungswerten beruhende Vorgehensweise letztlich nur für ExpertInnen bzw. für diejenigen eignet, die bereits lange mit Korpora und anderen digitalen Forschungsdaten arbeiten, stellte sich die Frage nach den Informationsmöglichkeiten für Einsteiger oder Studierende in der Ausbildung. Studierende erhielten bestenfalls eine Einführung durch erfahrene DozentInnen ihrer Hochschule. Ansonsten bekäme man erste Informationen zu relevanten Projekten am ehesten auf Fachtagungen. Für einen systematischen Einstieg eigneten sich auch einige einschlägige Print-Publikationen, die allerdings aufgrund ihres statischen

Charakters schnell veralteten und daher allenfalls als Ausgangspunkt für weitere Recherchen im Schneeballsystem dienen könnten.

### 2.3. Probleme der aktuellen Nachweissituation

Es wurde festgestellt, dass eine systematische Suche nach romanistischen Forschungsdaten derzeit kaum möglich ist. Es gebe zwar eine Reihe von Ansätzen, Initiativen und (wie im Impulsvortrag vorgestellt) diverse Suchinstrumente, allerdings sei keines der existierenden Systeme bislang ausreichend. So seien z.B. Korpora ebenso wie spezielle digitale Wörterbücher nur unzureichend nachgewiesen und dementsprechend schwer zu finden. Außerhalb der Linguistik, z.B. in den Literatur- und Kulturwissenschaften oder in der Didaktik, sei die Nachweissituation angesichts hybrider Forschungsdaten, die aus unterschiedlichen Forschungsumgebungen stammen, oft noch unübersichtlicher.

Als Ursache dafür wurden mehrere Faktoren genannt. Viele Daten werden von den Forschenden

- auf Institutsservern gespeichert, wo sie nicht nur extrem flüchtig, sondern auch nicht systematisch recherchierbar sind,
- quasi privat auf teilweise kommerziellen Plattformen im Internet gespeichert, wo sie für andere gar nicht auffindbar sind,
- in Systemen gespeichert, die zwar öffentlich zugänglich sind, aber keine oder kaum Suchfunktionalitäten anbieten,
- in Systemen gespeichert, die öffentlich zugänglich sind und auch Suchfunktionen anbieten (z.B. *Clarín*), aber nicht mit darüber hinausgehenden Nachweissystemen vernetzt sind, so dass die Daten nur im Rahmen des jeweiligen Systems recherchierbar sind.

In diesem Zusammenhang wurde die Gefahr einer Entwicklung von Parallelstrukturen gesehen und für eine bessere Vernetzung der im Aufbau befindlichen Infrastrukturen mit den traditionellen Nachweissystemen der Bibliotheken plädiert. Ein Suchsystem, das die Informationen der bewährten Kataloge und die Nachweise von digitalen Forschungsdaten vereine, wäre ein Desiderat. Dabei müsse dafür gesorgt werden, dass die Metadaten zu Forschungsdaten systematisch erfasst würden und auf möglichst transparentem Wege in dieses Nachweissystem einfließen. Auch die Standardisierung der Metadaten müsse vorangetrieben werden. Betont wurde in diesem Zusammenhang die Rolle der Bibliotheken, die aufgrund ihrer Erfahrungen für die Aufgabe der Erfassung und Verwaltung von Metadaten prädestiniert seien.

### 2.4. Möglichkeiten des neuen Meldesystems - Anforderungen an den Datenbestand

Die Anforderungen an ein Nachweissystem für Forschungsdaten wurden anschließend v.a. anhand des neuen Meldesystems auf *romanistik.de* diskutiert, das generell als ein Schritt in die richtige Richtung angesehen wurde. Für eine umfassende systematische

Suche käme es allerdings darauf an, den Input deutlich zu erweitern und die Konsistenz der Daten zu garantieren.

Als ein Desiderat zur Erweiterung des Datenbestandes wurde die Berücksichtigung von Forschungsdaten genannt, die im Rahmen von noch nicht abgeschlossenen Projekten anfielen und sich quasi noch in der Entstehungsphase befänden. Auch sollten nicht nur aktuelle, sondern auch ältere Ressourcen einbezogen werden, was eine retrospektive Erfassung notwendig mache.

Um einen in sich konsistenten Datenbestand zu erhalten, der für gezielte Suchanfragen unerlässlich ist, müssten die Metadaten *ausreichende* und *standardisierte* Angaben umfassen und kontinuierlich gepflegt werden. Letzteres sei gerade angesichts der gewünschten Einbeziehung von dynamischen Ressourcen wie Projekten (als potentiellen Datenquellen) wichtig.

Angesichts dieser Herausforderungen wurde die derzeitige Praxis diskutiert, daß Einträge in die Datenbank auf *romanistik.de* ausschließlich durch die Forschenden selbst bzw. die Redaktion erfolgen, weitgehend aus dem deutschsprachigen Raum stammen und sich auf ein Minimum an Angaben (Titel und Link) beschränken können. Es wurden verschiedene Vorschläge zur Anreicherung des Datenmaterials vorgebracht. Zum einen solle die Gestaltung der Eingabemaske, z.B. durch geeignete Auswahlmenüs, den Meldenden ermöglichen, ohne viel Aufwand passgenaue Angaben zu machen. Zum anderen müsse auch die Möglichkeit eingeräumt werden, „fremde“ Ressourcen zu melden und deren Metadaten zu bearbeiten. Damit ergäben sich u.U. auch verschiedene Optionen, den Datenbestand im Rahmen einer Crowd-Beteiligung durch WissenschaftlerInnen, BibliothekarInnen und Studierende komplettieren zu lassen.

Inwieweit es gelingen kann, die Meldenden zu ausführlicheren Angaben und zur Schließung von Lücken zu motivieren, wurde von den Anwesenden unterschiedlich beurteilt.

## 2.5. Zusätzliche Beschreibungskategorien für Forschungsdaten - Desiderate

Schließlich wurden auch Wünsche zu zusätzlichen Beschreibungskategorien geäußert. Diese betrafen:

- die Nutzungsrechte,
- die Unterscheidung von Datentypen und Datenformaten (z.B. einfaches PDF versus strukturiertes Material),
- die Erschließung mittels einer Klassifikation (nach auswählbaren Kategorien) und eines normierten Vokabulars,
- die Möglichkeit zur Bewertung der verzeichneten Datensätze, und zwar sowohl bezüglich der Qualität der Forschungsdaten selbst (beispielsweise Grad und Qualität der Dokumentation, Grad der Einhaltung der Auszeichnungskonventionen und Standards usw.) als auch hinsichtlich der

Qualität der Repräsentation der Daten (z. B. Nutzeroberfläche, Suchmöglichkeiten etc.). Als Vorschlag zur praktischen Umsetzung wurden ein Kommentarsystem oder die Vergabe von Sternen auf einer Skala geäußert. Die entsprechenden Suchmöglichkeiten sollten dabei auch die Option beinhalten, nach bislang noch nicht kommentierten bzw. bewerteten Datensätzen filtern zu können.

### 3. Themenkomplex 2: Arbeit mit Forschungsdaten

Vorgestellt wurden hier zunächst zwei konkrete Vorhaben aus der Literaturwissenschaft und der Linguistik, die als „Best-Practice“-Beispiele Möglichkeiten und Herausforderungen eines wissenschaftsadäquaten Umgangs mit romanistischen Forschungsdaten illustrierten. Anschließend wurden im Plenum die Probleme diskutiert, mit denen sich die WissenschaftlerInnen in der Praxis ihrer Forschungsprojekte konfrontiert sehen.

#### 3.1. Erstellen und Publizieren von Textsammlungen mit [GitHub](#) und [Zenodo](#) (Christof Schöch)

Einleitend schilderte Christof Schöch anhand des [Research Data Lifecycles](#) die einzelnen Arbeitsphasen, die ein wissenschaftlich angemessenes und nachhaltiges Management von Forschungsdaten in der Regel umfasst. Daran anknüpfend stellte er am Beispiel der Projekte der von ihm geleiteten e-Humanities-Nachwuchsgruppe *Computergestützte literarische Gattungsstilistik (ClIGS)* ein Modell für den Prozess der Forschungsdatenverwaltung vor.

Als Datenbasis werden digital vorliegende Texte verwendet. Deren Strukturelemente werden dann mit Hilfe des xml-basierten Standard-Formats TEI kodiert. Damit können u.a. Überschriften, Absätze, Vorworte, eingefügte Gedichte, Sprecher etc. annotiert werden, wobei für die entsprechenden Keywords nach Bedarf verschiedene Sprachen wählbar sind. Auch Metadaten können dem Dokument beigelegt werden, so z.B. für Autor, Titel, Status (der Verfügbarkeit), Herkunft der Datenbasis ebenso wie für verschiedene Verantwortlichkeiten bei kollaborativem Arbeiten. Als Vorteile von TEI wurden nicht nur die vielfältigen Auszeichnungsmöglichkeiten herausgestellt, sondern auch die Möglichkeit, auf komfortablem Wege Dokumente im PDF-Format zu generieren, und nicht zuletzt die Tatsache, dass es als Standardformat den Austausch von Daten erleichtert.

Gespeichert und verwaltet werden die Daten auf der Plattform *GitHub*, die zwar kommerziell betrieben wird, aber kostenfrei nutzbar ist.

Schöch erläuterte die Stärken von *GitHub*:

- die Möglichkeit zur kollaborativen Textbearbeitung,
- die systemeigene automatische Versionierung: Im Github-Repository können komplette Versionen der Dokumente gespeichert und Änderungen automatisch dokumentiert werden,

- die Verknüpfung mit dem Online-Speicherdienst *Zenodo*, wodurch die auf *GitHub* gespeicherten Dokumente mit einem persistenten Identifikator, genauer gesagt einer DOI, versehen werden und so langfristig verfügbar und zitierbar gemacht werden.

Ebenso kam jedoch auch ein Problem zur Sprache, das auf den Themenkomplex „Suche nach Forschungsdaten“ verweist: Die Recherche nach den in *GitHub* und *Zenodo* gespeicherten Daten ist aufgrund der begrenzten Suchfunktionalitäten innerhalb der Plattformen wenig komfortabel, und in (fach)wissenschaftliche Nachweissysteme fließen die Metadaten i.d.R. nicht ein. Auch werden sie kaum über allgemeine Suchmaschinen wie *Google* gefunden.

### 3.2. Nachhaltigkeit - aus der Sicht virtueller Forschungsumgebungen (Thomas Krefeld)

Im zweiten Impulsvortrag beschrieb Krefeld zunächst an mehreren Beispielen die Gefahr, dass wichtige Forschungsdaten für die Fachcommunity oft nicht mehr zugänglich sind, nachdem die Projektförderung ausgelaufen ist oder die damit befassten ForscherInnen in Ruhestand gegangen sind.

Anschließend veranschaulichte er die Herausforderungen einer nachhaltigen Sicherung von Forschungsdaten am Beispiel des DFG-Projekts [Verba Alpina](#), das wesentlich zum Aufbau des Münchener Publikationsportals [Korpus im Text](#) beigetragen hat. Dabei legte er einen umfassenden Begriff von Forschungsdaten zugrunde, der explizit auch digitale *Textpublikationen* einschließt: Objektdaten, Metadaten und digitale Texte stellen zwar inhaltlich verschiedene Typen dar, seien aber in Präsentation und Analyse oft sehr eng miteinander verknüpft und somit in ihrer Gesamtheit als Forschungsdaten zu betrachten.

Angesichts der Komplexität und Vielfalt von Forschungsdaten seien verschiedene Ebenen der Sicherung zu unterscheiden. Die „physische“ Bewahrung der Rohdaten sei dabei relativ unkritisch und könne wie in München vom jeweiligen Rechenzentrum übernommen werden. Problematischer sei die nachhaltige Dokumentation, die garantiere, dass die Daten möglichst unabhängig von Vorkenntnissen oder technischen, zeitlichen und räumlichen Rahmenbedingungen rekonstruierbar sind und verständlich bleiben. Dies sei gerade für die Romanistik in besonderem Maße relevant, da hier nicht nur aktuelle, sondern auch ältere Daten oftmals von großem Wert seien.

Die Voraussetzung dafür sollten die WissenschaftlerInnen selbst schaffen, indem sie schon vor Projektbeginn und während des Projekts für eine nachvollziehbare und lückenlose inhaltliche und technische Dokumentation aller in den Forschungsprozess involvierten Daten sorgten. Dies betreffe Programmcode, digitale Texte und empirisch erhobene Daten, aber auch die verwendete Software sowie die zum Einsatz gekommenen virtuellen Arbeitsumgebungen. Um solche Informationen festzuhalten biete z.B. die virtuelle Forschungsumgebung im Projekt *Verba Alpina* entsprechende Funktionalitäten an.

Darüber hinaus müsse sowohl ein transparenter und standardisierter Mechanismus geschaffen werden, der dafür Sorge, dass die Forschungsdaten mit einem DOI

versehen werden und die Metadaten in bibliothekarische Nachweissysteme, als auch eine institutionelle Absicherung, die bislang noch nicht in ausreichendem Maße gegeben sei. Die Repositorien der neuen Infrastrukturen und virtuellen Forschungsumgebungen wie etwa *DARIAH*, *CLARIN* und *TextGrid* seien derzeit noch nicht ausreichend: Sie unterlägen teilweise selbst einer Projektfinanzierung, ihr Verhältnis untereinander sei nicht sehr transparent und ihre Nutzung oft mit Problemen behaftet. Krefeld betonte hier nochmals die Rolle der Bibliotheken, die dank ihrer Erfahrungen die Verantwortung für Bewahrung, Verfügbarkeit und Pflege von Forschungsdaten übernehmen könnten, unabhängig davon, ob sie auch die entsprechenden Repositoriumsdiensleistungen selbst durchführen würden.

Im Sinne einer nachhaltigen Sicherung der in Projekten anfallenden Forschungsdaten plädierte Krefeld für ein komplexes Zusammenspiel von Wissenschaft, Rechenzentren, Bibliotheken und Förderinstitutionen, das er anhand eines Schaubildes erläuterte ([Abbildung zum nachhaltigen Umgang mit wissenschaftlichen Daten](https://www.kit.gwi.uni-muenchen.de/wp-content/uploads/Nachhaltig_Regelwerk_2.jpg) / [https://www.kit.gwi.uni-muenchen.de/wp-content/uploads/Nachhaltig\\_Regelwerk\\_2.jpg](https://www.kit.gwi.uni-muenchen.de/wp-content/uploads/Nachhaltig_Regelwerk_2.jpg)).

### 3.3. Diskussion der Bedarfe

An den präsentierten Beispielen aus der Praxis konnten einige wesentliche Anforderungen an eine wissenschaftsadäquate Datenverwaltung im Sinne der Empfehlungen des „Research Data Lifecycles“ verdeutlicht werden:

- der Rückgriff auf offene und konventionelle Standard-Formate,
- eine angemessene Dokumentation u.a. mit Metadaten und Versionierung, möglichst schon im Projektverlauf,
- Gewährleistung der langfristigen Verfügbarkeit,
- Zitierbarkeit (DOI-Vergabe),
- nachhaltige Sicherung,
- Auffindbarkeit,
- Nachnutzbarkeit.

Auch wurden bereits zentrale Bedarfe genannt:

- eine institutionelle Absicherung,
- ein *standardisiertes* Verfahren, um Forschungsdaten systematisch mit einem persistenten Identifikator zu versehen,
- ein *standardisiertes* Verfahren zur Integration der Metadaten in einschlägige Nachweissysteme,
- ein koordiniertes Zusammenspiel der verschiedenen Infrastruktureinrichtungen und Förderinstitutionen.

In der anschließenden Diskussion wurden viele der genannten Anforderungen noch einmal unterstrichen. Mit Blick auf die Weiterverarbeitung, auf mögliche



systemunabhängige Kooperationen und auf die problemlose, auch disziplinübergreifende Nachnutzbarkeit sei es von essentieller Bedeutung, möglichst (offene) Standards zu verwenden und dies schon von Anfang an, z.B. bei der Wahl der Plattformen und Formate, zu bedenken.

Gleichzeitig wurde deutlich, dass eine solchen Anforderungen entsprechende Datenverwaltung längst nicht allgemein verbreitet ist. Selbst wenn bereits ein Bewusstsein für deren Notwendigkeit bestehe, fehle es oft an ausreichenden Informationen und Infrastrukturen für eine adäquate Umsetzung. Mehrere Anwesende thematisierten die grundsätzliche Schwierigkeit, geeignete Formate und Plattformen bzw. Virtuelle Forschungsumgebungen zu finden. Pauschale Lösungen gebe es i.d.R. nicht, da die Wahl der Instrumente von der jeweiligen Fragestellung abhängen.

Bereits die bloße Abspeicherung der Forschungsdaten bereite oft Probleme, zumal nicht alle Hochschulen eine Infrastruktur wie die LMU München bereitstellten. Daher sei eine quasi private Ablage, z.T. unter Nutzung kommerzieller Web-Services, vielfach gängige Praxis mit der Folge unzureichender Nachnutzbarkeit, schlechter Auffindbarkeit und der Gefahr eines Verlustes der Daten oder der diesbezüglichen Rechte.

Die Nachhaltigkeitsproblematik wurde auch mit Blick auf kooperative nationale und internationale Projekte diskutiert. Es wurde empfohlen, die dabei auf Repositorien gespeicherten Daten - sofern rechtlich möglich - an verschiedenen Stellen parallel abzulegen und in den Kooperationsvereinbarungen entsprechende Vorkehrungen zu treffen. In diesem Zusammenhang wurde der Wunsch geäußert, für solche Kooperationsverträge allgemein zugängliche Muster bereitzustellen.

Insgesamt belegten die bisherigen Erfahrungen, dass für ein wissenschaftsadäquates Datenmanagement zum einen ein umfassender Informations- und Beratungsbedarf besteht (s. Themenkomplex 3) und zum anderen ein weiterer Auf- und Ausbau von Infrastrukturen für Forschungsdaten notwendig ist.

Mit Blick auf die Schwierigkeit, geeignete Plattformen für Forschungsdaten zu finden, wurden auch die existierenden Angebote wie *DARIAH*, *TextGrid* und *Clarin* diskutiert. Auf der einen Seite wurde bemängelt, dass die entstehenden Infrastrukturen oft zu projektgebunden seien und teilweise in Konkurrenz zueinander stünden. Auf der anderen Seite wurde deren Bemühen um Nachhaltigkeit betont. Sie basierten auf Kooperationen mit etablierten Institutionen wie etwa der SUB Göttingen und seien, in ihrer Organisation als *European Research Infrastructure Consortium - ERIC*, inzwischen mehr als bloße Projekte.

In diesem Zusammenhang wurde auch der Vorschlag geäußert, *romanistik.de* zu einem Publikationsmedium zu erweitern und damit quasi eine eigene Infrastruktur für die Romanistik aufzubauen. Schöch erläuterte dazu, dass man dies mit dem Hinweis auf bereits existierende Alternativen bislang vermieden habe. Beispielhaft genannt wurden hier die Bloggingplattform <http://hypotheses.org/>, die eine vergleichsweise nachhaltige Möglichkeit zur Publikation wissenschaftlicher Texte (auch mit Einbindung

audiovisueller Medien) bereitstelle, sowie die Zeitschriftenplattform [revues.org](http://revues.org) (dort bestehen allerdings hohe Einstiegshürden).

Die Nutzung solcher internationaler Plattformen biete sich auch vor dem Hintergrund an, dass romanistische Forschungsdaten nicht nur für die deutsche Romanistik, sondern auch für ein internationales Publikum relevant seien.

Daran anknüpfend wurden zusätzliche, spezifisch romanistische Herausforderungen für das Datenmanagement genannt: eine internationale Ausrichtung der Nachweissysteme sowie eine möglichst mehrsprachige Metadatenverzeichnung und Dokumentation, auch wenn dies mit einem großen Arbeitsaufwand verbunden sei.

Schließlich wurde die Tatsache problematisiert, dass viele relevante Forschungsdaten gar nicht zugänglich seien. Die Beteiligten plädierten ausdrücklich für eine Bereitstellung im Open Access, so weit dies Urheberrecht und Persönlichkeitsrechte zuließen.

Es bestand Konsens, dass es einer Sensibilisierung der Fachcommunity für die Bedeutung von Forschungsdaten und einer neuen Anerkennungskultur bedarf, zu der auch die Verbände beitragen könnten. Schöch verwies in diesem Zusammenhang auf ein in den DRV-Mitteilungen erscheinendes Positionspapier der *AG Digitale Romanistik*.

#### **4. Themenkomplex 3: Notwendige Beratungsangebote**

##### *4.1. Existierende Informations- und Beratungsangebote (Jan Rohden)*

Den dritten und letzten Themenbereich eröffnete Jan Rohden mit einem Impulsvortrag über bereits existierende Beratungsdienstleistungen. Dazu zählten fachunabhängige Services der Hochschulen und Kompetenzzentren, allgemein geisteswissenschaftliche Angebote (z.B. vom *Data Center for the Humanities* in Köln oder von *DARIAH*), disziplinspezifische Plattformen wie *CLARIN* oder *Forschungsdaten Bildung* und auch fachspezifische Informationsangebote, wie sie z.B. vom *GESIS – Leibniz-Institut für Sozialwissenschaften*, von verschiedenen Fachinformationsdiensten und auch von der *AG Digitale Romanistik* bereitgestellt werden.

Die gesichteten Beratungsangebote unterschieden sich dabei nicht nur in ihrer fachlichen Ausrichtung, sondern auch in der Berücksichtigung der diversen Aspekte des Datenmanagements (Abdeckung des gesamten Lebenszyklus oder Fokussierung auf Datenmanagementpläne, auf Tools und Arbeitsmethoden, auf die Veröffentlichung, die Archivierung oder rechtliche Aspekte etc.), im Spezialisierungsgrad (von allgemeinen Informationen bis zur individuellen, projektbezogenen Beratung) sowie im Format (von Infoseiten mit Linksammlungen im Web über Videos bis zu Online-Kursen, „Training for trainers“ oder Workshops). Dieses Spektrum veranschaulichte Rohden exemplarisch anhand der beiden umfassenden Beratungsangebote des [Data Center for the Humanities](http://DataCenterfortheHumanities.org) in Köln und des Verbunds bildungswissenschaftlicher Kompetenzzentren [Forschungsdaten Bildung](http://ForschungsdatenBildung.de).

#### 4.2. Eruierung der Bedarfe in der Romanistik

Die Frage nach den Erfahrungen der TeilnehmerInnen mit existierenden Beratungsdienstleistungen eröffnete die Diskussion. Dabei wurden zunächst die Schulungen zu Methoden, Tools und Best Practices genannt, die in digital affinen Bereichen wie etwa der Korpuslinguistik bereits seit langem stattfinden. Sie sind i.d.R. innerhalb der einschlägigen Seminare angesiedelt und können auch am besten von den damit arbeitenden WissenschaftlerInnen durchgeführt werden.

Des Weiteren wurde auf die Workshop- und Schulungsangebote der *AG Digitale Romanistik* hingewiesen, die Methodenkompetenz im Umgang mit Forschungsdaten vermitteln. Zunehmend wahrgenommen werden dabei u.a. Beratungsdienstleistungen und Workshops zu allgemeinen Tools und Methoden, zu spezifischen Arbeitstechniken wie dem „Topic Modelling“, zu TEI, zu Software für die quantitative Untersuchung von Texten wie beispielsweise TXM und zu Literaturverwaltungssoftware.

Anschließend wurden darüber hinausgehende Bedarfe nach Informationsangeboten aufgegriffen, die teilweise bereits im Rahmen des zweiten Themenkomplexes aufgekommen waren. Sie betrafen

- die allgemeine Aufklärung über die diversen Aspekte des Datenmanagements im Sinne des [Research Data Lifecycles](#) und über die Bedeutung von Standards,
- spezifische Information, um WissenschaftlerInnen bei der Wahl geeigneter Formate, Tools, Plattformen und Virtueller Forschungsumgebungen zu unterstützen,
- die Bereitstellung von Materialien wie z.B. Formulare zur Verankerung von Nachhaltigkeit in Kooperationsvereinbarungen,
- eine Beratung zu Rechtsfragen, unter anderem zum Urheberrecht, zu Persönlichkeitsrechten und Veröffentlichungsrechten.

Grundlegender Beratungsbedarf bestehe auch oft zum Thema Digitalisierung. Hier könnten nach Auffassung der TeilnehmerInnen v.a. die Bibliotheken, die in diesem Bereich über weitreichende Erfahrung verfügen, entsprechende Services anbieten.

Als Vorschlag, wie ein an die Bedürfnisse der romanistischen Fachgemeinschaft angepasstes Beratungsangebot strukturiert sein könne, wurde ein dreistufiges System ins Gespräch gebracht:

- a) Auf einer ersten Stufe könnten RomanistInnen grundsätzliche Informationen zum Forschungsdatenmanagement in virtueller Form, zum Beispiel durch didaktisch aufbereitete Online-Tutorials oder Lehrvideos verfügbar gemacht werden. (Mögliche Themen: Welche Formatstandards gibt es, welche technischen Tools gibt es?)
- b) Die zweite Stufe sähe die eingehendere Behandlung spezifischer Einzelthemen (z.B. Vorbereitung von TEI-Annotationen) vor, etwa durch Vorträge vor Ort, Schulungen oder Fortbildungen für ExpertInnen, gegebenenfalls auch Schulungen für ProjektmitarbeiterInnen, die ihrerseits später einzustellende Projektkräfte schulen könnten).

- c) Die dritte Stufe würde eine spezifische, auch antragstechnische Projektberatung umfassen und wäre i.d.R. von den Hochschulen der jeweiligen WissenschaftlerInnen zu übernehmen (u.U. in Zusammenarbeit mit *Cologne Center of E-Humanities*).

Die Bereitstellung eines solch umfassenden Informationsangebots wurde als Desiderat formuliert. Ziel wäre es dabei, zum einen die bereits existierenden Dienste für die Romanistik aufzubereiten und Lücken zu schließen, die angesichts spezifisch romanistischer Fragestellungen bisher nicht abgedeckt werden. In diesem Zusammenhang wurde auch der Vorschlag geäußert, das bereits existierende [Workshop- und Schulungsangebot der AG Digitale Romanistik](#) prominenter zu bewerben.

## 5. Fazit

Im Rahmen des Workshops wurden vielfältige Bedarfe im Bereich des Forschungsdatenmanagements in der Romanistik herausgearbeitet. Sie betreffen die derzeitigen Unzulänglichkeiten der Nachweissituation, die Schwierigkeiten, den Anforderungen eines wissenschaftsadäquaten Umgangs mit Forschungsdaten zu genügen, Defizite der zur Verfügung stehenden Infrastrukturen sowie ein breites Spektrum an Informations- und Beratungsbedarf. Ziel wird es nun sein, darauf aufbauende geeignete Strategien und Unterstützungsmaßnahmen zu entwickeln.

## Feedback

Für die weiteren Planungen möchten wir gern die Meinung möglichst vieler RomanistInnen einbeziehen. Wenn Sie zu den Ergebnissen des Workshops Stellung nehmen möchten, Anregungen haben oder zusätzliche Bedarfe melden möchten, können Sie sich per Mail oder telefonisch gern an uns wenden.

## Anhang

Workshop-Programm

Leitfragen des Workshops