Haute école de gestion de Genève
Filière Information documentaire

Digitale Forschungsinfrastrukturen für die Geistes- und Geschichtswissenschaften

Jasmin Hügi (jasmin.hugi@hesge.ch),
René Schneider (rene.schneider@hesge.ch)
Haute école de gestion de Genève (HEG-GE)

Studie im Auftrag von:
infoclio.ch
Enrico Natale
Bern

Genf, 24. Januar 2013

Zitiervorschlag:

HÜGI, Jasmin, SCHNEIDER, René. *Digitale Forschungsinfrastrukturen in den Geistes- und Geschichtswissenschaften*. Genf : Haute école de gestion de Genève, 2013.

Executive Summary - Deutsch

Auftrag

Der Fachbereich "Information documentaire" der Haute école de gestion de Genève (Fachhochschule für Wirtschaft in Genf) hat von infoclio.ch, dem Fachportal für die Geschichtswissenschaften der Schweiz, den Auftrag erhalten, eine Studie zum Thema digitale Forschungsinfrastrukturen für die Geistes- bzw. Geschichtswissenschaften in der Schweiz zu erstellen. Infoclio.ch ist in einer Kommission der Schweizerischen Akademie für Geistes- und Sozialwissenschaften vertreten, welche ein Pilotprojekt bezüglich der dauerhaften Sicherung von Forschungsdaten anhand eines Datenrepositoriums sowie der Vernetzung bereits existierender Infrastrukturen organisiert.

Kontext

Forschungsfördernde Institutionen stellen immer stärker die Anforderung, Daten von Forschungsprojekten, die mit öffentlichen Mitteln finanziert werden, der Allgemeinheit öffentlich zugänglich zu machen. Gleichzeitig entsteht ein immer grösser werdendes Bedürfnis, digitale Daten zu sichern und für eine spätere Nachnutzung auch langfristig zu erhalten. In diesen Bereichen besteht in der Schweiz in den Geisteswissenschaften ein grosser Nachholbedarf, da keine Infrastruktur bereitgestellt ist, um geisteswissenschaftliche Forschungsdaten aufzunehmen.

Digitale Forschungsinfrastruktur

In dieser Studie wird ausschliesslich der Begriff "digitale Forschungsinfrastruktur" verwendet. Dabei stützen wir uns auf die allgemeine Definition einer Informationsinfrastruktur, d.h. eines technischen, sozialen und politischen Rahmens, welcher Menschen, Technologien, Werkzeuge und Dienstleistungen vereint.

Forschungsdaten

Es stellt sich die Frage, was denn Forschungsdaten in den Geisteswissenschaften sind. Wird die engere Definition des Begriffs, so wie er in den Naturwissenschaften angewendet wird, als Grundlage genommen, so würden in den Geisteswissenschaften nur diejenigen Projekte Forschungsdaten produzieren, welche eine natur- oder sozialwissenschaftliche Methode benutzt haben, um quantitative Daten zu generieren.

Doch die Geisteswissenschaften unterscheiden sich von den Naturwissenschaften durch eine grosse Methodenvielfalt, die u.U. von Forschenden zu Forschenden variiert, was eine grosse Heterogenität der erzeugten Daten zur Folge hat. Eine weitere Besonderheit besteht darin, dass wissenschaftliche Monographien nach wie vor die wichtigste Publikationsform darstellen. Die Geschichtswissenschaften gestalten sich ähnlich wie die Geisteswissenschaften, da diese Disziplin ebenfalls durch eine Vielzahl von Teilfächern besteht. Auch hier stellen Monographien die häufigste Publikationsform dar. Eine Spezifität dieser Disziplin ist, dass deutlich mehr Wissen kumuliert als aktualisiert wird.

Aufgrund dieser Unterschiede kann es als zu einschränkend und zu einfach gesehen werden, den Forschungsdatenbegriff so wie er in den Natur- und Sozialwissenschaften angewendet wird, auf die Geisteswissenschaften zu übertragen. Es stellt sich die Frage, ob nicht vielmehr von Forschungsprodukten gesprochen werden sollte, welche sowohl Datensätze aus Datenbanken, wie auch Primär- und Sekundärquellen, Digitalisate oder Hilfsmittel umschliessen würden.

Digitale Forschungsinfrastruktur - organisatorische Aspekte

Eine digitale Forschungsinfrastruktur kann sehr unterschiedlichen Zwecken dienen, welche in dieser Studie aufgelistet werden. Dabei wird klar, dass einige dieser Zwecke je nach Disziplin eine unterschiedliche Wichtigkeit haben. Gilt beispielsweise das Argument, Forschungsergebnisse überprüfen zu wollen, für qualitative Daten genauso wie für quantitative Daten?

Eine digitale Forschungsinfrastruktur, welche auch einfach als Repositorium gesehen werden kann, kann verschiede Charakteristiken aufweisen. Die einzelnen Aspekte werden in dieser Studie aufgezeigt.

Die zeitlich begrenzte bzw. unbegrenzte Aufbewahrung stellt immer wieder grosse Herausforderungen an digitale Infrastrukturen. Dabei ist festzuhalten, dass unterschiedliche Zwecke einer Infrastruktur eine andere Art von Aufbewahrung verlangen. Geht es beispielsweise um den Erhalt der Integrität von Forschungsdaten, damit Forschungsergebnisse überprüft werden können, müssen die Daten in ihrem ursprünglichen Format gesichert werden. Für eine dauerhafte Archivierung müssen aber Datenobjekte in offene Formate konvertiert werden. Es stellt sich die Frage, welche Daten wann, wie lange und in welchem Format gespeichert werden sollen.

Eine Abschätzung der Kosten einer eventuellen Forschungsinfrastruktur stellt sich als äusserst schwierig dar. Dies ist u.a. darauf zurückzuführen, dass es eine Vielzahl von höchst unterschiedlichen Beteiligten gibt, eine Infrastruktur unterschiedlichen Zwecken dienen kann und dass es schwer antizipierbar ist, wie sehr eine Infrastruktur schlussendlich auch benutzt werden wird.

Digitale Forschungsinfrastruktur - technische Aspekte

Im Zusammenhang mit digitalen Forschungsinfrastrukturen gibt es Standards hinsichtlich Klassifikationssystemen, kontrollierter Vokabulare, Normen für die Archivierung und für die Vertrauenswürdigkeit sowie Metadaten, welche für die Beschreibung von unterschiedlichen Aspekten digitaler Objekte verwendet werden können. Diese Standards können sowohl fachspezifisch wie auch fachübergreifend sein. Welche Standards in einer Forschungsinfrastruktur berücksichtigt werden sollen, hängt davon ab, welchen Zweck die Infrastruktur verfolgt, welche Ausrichtung sie hat und welche Disziplinen damit abgedeckt werden sollen.

Um die Nachnutzung von Forschungsdaten zu vereinfachen, ist es empfohlen, persistente Identifikatoren zu vergeben. Die bekanntesten Identifikatoren werden in der Studie vorgestellt. Der Gebrauch von persistenten Identifikatoren ist aber auch mit Nachteilen behaftet: so ist er einerseits nicht immer kostenfrei, andererseits wird eine Abhängigkeit mit dem Identifikator-Anbieter, als auch mit dem Anbieter des Resolvers eingegangen.

Auch die Frage der Lizenzierung von Forschungsdaten und -produkten muss geklärt sein. Für Forschungsprodukte in Dokumentform können die Creative Commons-Lizenzen erweiterte Nutzungsrechte garantieren. Bezüglich Daten und Datenbanken können die Open Data Commons-Lizenzen in Frage kommen. Dabei stellt sich generell die Frage der Granularität, da einerseits die Daten, die Datensätze und die Datenbanken jeweils mit einer Lizenz versehen werden können. Dass Typen von Lizenzierungen die Nachnutzung von Forschungsdaten auch beeinträchtigen können, wird in der Studie näher erklärt.

Wie schon die Kosten ist auch das Datenvolumen, welches eine digitale Forschungsinfrastruktur zu verwalten hat, schwer abschätzbar. Aus einer Umfrage mit Forschenden der Geisteswissenschaften in der Schweiz geht jedoch hervor, dass mit ca. 570 Terabytes gerechnet werden kann. Gehören hochauflösende Digitalisate auch dazu, dann wird sich das Volumen im Petabyte-Bereich befinden.

Digitale Forschungsinfrastruktur - menschliche Aspekte

Damit eine digitale Forschungsinfrastruktur auch benutzt werden wird, müssen für die Forschenden persönliche Anreize geschaffen werden. Das in den Naturwissenschaften gültige Argument, dass durch Bereitstellung von Forschungsdaten die Forschenden häufiger zitiert werden, scheint in den Geisteswissenschaften weniger tragfähig zu sein, weil Zitierungen in diesen Disziplinen eine geringere Bedeutung beigemessen wird. Eine Infrastruktur sollte deshalb sehr gut den Bedürfnissen seines Publikums entsprechen.

Mögliche Rollen bezüglich des Forschungsdatenmanagements werden in der Studie vorgestellt. Da aber die Zwecke und Funktionen einer Infrastruktur sehr unterschiedlich ausfallen können, lassen sich notwendige Kompetenzen zuletzt nur aus dem Pflichtenheft der zu entwickelnden oder entwickelten Infrastruktur ableiten.

Qualitative Fallstudie

Im Rahmen des Auftrags wurden 10 Historiker in der Schweiz zum Thema der digitalen Forschungsinfrastrukturen befragt. Dafür wurde ein semi-strukturiertes Interview von einer durchschnittlichen Dauer von 90 Minuten durchgeführt. Als erstes war festzustellen, dass sehr unterschiedliche Inhalte als Forschungsdatum in den Geschichtswissenschaften gelten können. Dabei wurden sowohl etliche Datenformen und Inhalte aufgezählt, während auch die Meinung vertreten wurde, dass es in den Geschichtswissenschaften gar keine Forschungsdaten gäbe. Im Bezug zu digitalen Forschungsinfrastrukturen gingen die Meinungen, Bedürfnisse und Erwartungen weit auseinander. Ein Konsens konnte dabei nicht entdeckt werden. Dies zeigt auf, dass in den Geschichtswissenschaften zunächst einerseits der Begriff der Forschungsdaten bzw. -produkte definiert werden muss, um danach die Bedürfnisse der Fachgemeinschaft bezüglich einer Infrastruktur festzulegen.

Fazit

Es ist wichtig festzustellen, dass aufgrund der hohen Diskrepanzen zwischen den Natur- und den Geisteswissenschaften die in den Naturwissenschaften erreichten Ergebnisse und Erfolge bezüglich Forschungsinfrastrukturen nicht auf einfache Art und Weise auf die Geisteswissenschaften übertragen werden können. Die Definition von Forschungsdaten, sowie die noch zu verfolgenden Zwecke einer Forschungsinfrastruktur und deren Ausrichtung sind Fragen, die für die Geistes- und Geschichtswissenschaften noch nicht zufriedenstellend zu beantwortet sind. Der diesbezügliche Aufwand muss von der Fachgemeinschaft geleistet werden, um mit einer gemeinsamen Basis den logischen Aufbau von einem Netzwerk von Forschungsinfrastruktur zu fördern.

Executive Summary - Français

Mandat

La filière "Information documentaire" de la Haute école de gestion de Genève a été mandatée par infoclio.ch, le portail professionnel des sciences historiques en Suisse, pour réaliser une étude concernant les infrastructures de recherche numériques pour les sciences humaines, notamment historiques en Suisse. Infoclio.ch est représenté au sein du comité de l'Académie suisse des sciences humaines et sociales qui organise un projet pilote pour l'archivage à long terme des données de recherche, grâce à un dépôt numérique ainsi qu'à la mise en réseau des infrastructures existantes.

Contexte

Les institutions qui financent des projets de recherche insistent de plus en plus sur la publication des données de recherche issues des projets qui ont été financés par des deniers publics. En même temps, le besoin de sauvegarder des données numériques et de les préserver pour une réutilisation ultérieure augmente. Concernant ces domaines en Suisse, il existe un grand retard à combler puisqu'aucune infrastructure n'est mise à disposition pour héberger des données de recherche issues du domaine des sciences humaines.

Infrastructures de recherche numériques

Dans cette étude, le terme d' « infrastructures de recherche numériques » est retenu. Il se base sur la définition générale d'une infrastructure d'information, c'est-à-dire qu'elle représente un cadre technique, social et politique qui unit des personnes, des technologies, des outils et des services.

Données de recherche

La définition de ce que sont les données de recherche pour les sciences humaines est moins évidente. En effet, si le terme est compris selon la définition dans les sciences dures, qui part du principe que des données quantitatives peuvent être considérées comme des données de recherche, seul une petite part des projets de recherche dans les sciences humaines produiront des données de recherche selon cette définition.

Mais les sciences humaines se distinguent des sciences naturelles par une grande diversité des méthodes de recherche, qui mène à une grande hétérogénéité des données générées. Une autre spécificité présente le fait que les monographies sont toujours la forme de publication la plus importante. Les sciences historiques ressemblent aux sciences humaines puisque ces disciplines consistent en une multitude de sous-disciplines. En outre, les monographies représentent également la forme de publication la plus utilisée. Une autre particularité de ces disciplines est que l'accumulation des connaissances est prépondérante par rapport à l'actualisation de celles-ci.

En raison de ces différences, il est trop limitant et trop simple de calquer sur les sciences humaines le terme des données des recherche tel qu'il est usité dans les sciences naturelles et sociales. La question s'impose de savoir si on ne devrait pas plutôt parler des produits de recherche qui incluent à la fois les ensembles des données issus des bases de données, les sources primaires et secondaires, ainsi que les images numériques ou les aides diverses.

Infrastructures de recherche numérique - aspects organisationnels

Une infrastructure de recherche numérique peut avoir des buts très différents qui sont énumérés dans cette étude. Il s'avère que quelques-uns de ces buts ont une importance différente selon la discipline concernée. Par exemple, est-ce que l'argument de la vérification des résultats d'une recherche à la même pertinence pour des données qualitatives que pour les données quantitatives?

Une infrastructure de recherche numérique qui peut aussi être vue comme un dépôt numérique peut avoir plusieurs caractéristiques. Les aspects spécifiques sont illustrés dans cette étude.

L'archivage à durée limitée ou illimitée représente un grand défi pour les infrastructures numériques. Nous sommes à même de constater que les différents buts d'une infrastructure nécessitent une autre forme de préservation. S'il s'agit de la préservation de l'intégrité des données de recherche pour permettre une vérification des résultats, les données doivent être sauvegardées dans leur format initial. Un archivage à long terme demande par contre une conversion des objets de données dans des formats ouverts. Se posent alors les questions suivantes : Quelles données doivent être enregistrées ? Quand ? Pour combien de temps et dans quel format ?

Une estimation des coûts d'une infrastructure de recherche éventuelle semble très difficile. Ceci est lié à la multitude des utilisateurs potentiels, aux buts différents auxquels une infrastructure peut service ainsi qu'à la difficulté d'anticiper une future utilisation.

Infrastructures de recherche numériques - aspects techniques

Dans le contexte des infrastructures de recherche numérique, il existe des standards concernant les systèmes de classifications, les vocabulaires contrôlés, les normes pour l'archivage et pour la fiabilité ainsi que les métadonnées qui peuvent être utilisées pour la description des aspects différents d'objets numériques. Ces standards peuvent être spécifiques à une discipline ou interdisciplinaires. Le choix des standards à respecter dans une infrastructure de recherche dépend du but que l'infrastructure poursuit, son orientation et les disciplines couvertes.

Afin de faciliter la réutilisation des données de recherche, il est recommandé d'attribuer des identificateurs persistants. Les identificateurs les plus connus sont présentés dans l'étude. L'utilisation des identificateurs persistants peut aussi avoir des inconvénients: d'un côté, elle n'est pas toujours sans frais, de l'autre côté, une relation de dépendance vis-à-vis du fournisseur de l'identificateur ainsi que du résolveur est créée.

En outre, la question concernant l'attribution d'une licence pour les données de recherche doit être clarifiée. Pour des produits de recherche sous forme de document, les licences Creative Commons peuvent être appliquées pour garantir des droits d'utilisation élargis. Concernant les données et les bases de données, les licences Open Data Commons peuvent être envisagées. Nous expliquerons plus en avant que le type de licence peut avoir des conséquences négatives pour la réutilisation des données de recherche.

Tout comme les coûts, le volume des données qui devrait être géré par une infrastructure de recherche numérique est difficile à estimer. Les chiffres issus d'un sondage auprès des chercheurs des sciences humaines en Suisse montrent qu'on peut compter avec un volume d'environ 570 téraoctets. Si des images numérisées font aussi parti des objets à gérer, le volume se trouvera à l'échelle des pétaoctets.

Infrastructures de recherche numériques - aspects humaines

Pour qu'une infrastructure de recherche numérique soit utilisée, il faut répondre aux intérêts personnels des chercheurs. L'argument de l'augmentation du nombre de citations grâce à la mise à disposition des données de recherche, valable dans les sciences naturelles, parait être moins solide dans les sciences humaines, puisque les citations ont une signification moins importante. Pour cette raison, une infrastructure devra, pour être utilisée, fortement correspondre aux besoins de son public.

Des rôles possibles pour la gestion des données de recherche sont présentés dans l'étude. Puisque les buts et fonctionnalités d'une infrastructure peuvent varier, les compétences nécessaires se laissent uniquement définir à partir du cahier des charges de l'infrastructure à développer.

Étude de cas qualitative

Dans le cadre de ce mandat, 10 historiens suisses ont été interviewés sur le sujet des infrastructures de recherche numérique. Pour ce faire, un entretien semi-structuré d'une durée moyenne de 90 minutes a été mené. Nous pouvons constater que des contenus très divers peuvent être considérés comme des données de recherche des sciences historiques. Des nombreux formes et contenus ont été énumérés, alors même que quelqu'un a estimé qu'il n'y avait pas de données de recherche en sciences historiques. Par rapport aux infrastructures de recherche numériques, les avis, les besoins et les attentes exprimés divergent. Un consensus n’a pas pu être détecté. Ceci indique que, dans les sciences historiques, il faudra d'abord définir le terme des données de recherche pour ensuite établir les besoins de la communauté scientifique concernant une infrastructure.

Conclusion

Il est important de souligner qu'en raison de la grande divergence entre les sciences naturelles et humaines, les résultats et succès atteints dans les sciences naturelles ne peuvent pas être calqués de manière simple sur les sciences humaines. La définition des données de recherche ainsi que les buts et l'orientation d'une infrastructure de recherche représentent des questions qui n'ont pas encore été résolues de manière satisfaisante pour les sciences humaines et historiques. Un effort considérable devra être fourni par la communauté scientifique afin de créer une base commune qui permettrait, dans un second temps, le développement d'un réseau d'infrastructures de recherche.

INHALTSVERZEICHNIS

Abbildungsverzeichnis

Tabellenverzeichnis

1. Einleitung

1.1. Auftrag

Der Fachbereich "Information documentaire" der Haute école de gestion de Genève (Fachhochschule für Wirtschaft in Genf) hat von infoclio.ch, dem Fachportal für die Geschichtswissenschaften der Schweiz, den Auftrag erhalten, eine Studie zum Thema digitale Forschungsinfrastrukturen in den Geistes- bzw. Geschichtswissenschaften in der Schweiz zu erstellen.

Die Diskussion zum Thema der Forschungsdaten in den Geisteswissenschaften in der Schweiz findet mit dieser Studie nicht zum ersten Mal statt. Im Folgenden soll kurz die diesbezügliche Entwicklung zusammengefasst werden.

Aufgrund des vom Staatssekretariat für Bildung und Forschung (SBF, ab 1.1.2013 Staatssekretariat für Bildung, Forschung und Innovation SBFI) erstellten Zwischenbericht "Schweizer Roadmap für Forschungsinfrastrukturen", welcher schliesslich 2011 veröffentlicht wurde (SBF, 2011), hat die Schweizerische Akademie für Geistes- und Sozialwissenschaften (SAGW) 2008 eine Arbeitsgruppe gegründet mit dem Ziel, die Erstellung eines Datenzentrums für die Geisteswissenschaften abzuklären (Zimmermann, Pfister, 2008 (1):4). Infolgedessen hat die SAGW die Hochschule für Technik und Wirtschaft Chur (HTW Chur) damit beauftragt, eine Bedarfsanalyse bezüglich der digitalen Langzeitarchivierung in geisteswissenschaftlichen Institutionen der Schweiz durchzuführen. Diese Bedarfsanalyse diente wiederum als Grundlage für den Bericht "Digitale Infrastrukturinitiative für die Geisteswissenschaften" (Immenhauser, 2009), welcher dem Staatssekretariat für Bildung und Forschung vorgelegt wurde. Daraufhin hat das Staatssekretariat das Bundesarchiv mandatiert, seine Einbringung in diesem Bereich darzulegen. Die Schlussfolgerung wurde gezogen, dass die Ausgangslage komplex sei, einerseits bezüglich der Datenarten, andererseits bezüglich der unterschiedlichen Verantwortlichkeiten (BBI 2012 3099, 2012:3204). Das SBF beauftragte schliesslich die SAGW damit, eine Kommission mit dem Auftrag, ein Pilotprojekt durchzuführen, zu bilden. Dabei sollen "Fragen bezüglich der Definition von Standards, der Organisation (zentral, dezentral) und der Finanzierung einer entsprechenden Fachstelle zu bearbeiten" (BBI 2012 3099, 2012:3204) beantwortet werden.

Seitens der geisteswissenschaftlichen Fachgemeinschaft in der Schweiz existiert ein grosses Interesse und auch ein grosser Bedarf an einer Forschungsinfrastruktur für die nachhaltige Sicherung und Nachnutzung von Forschungsdaten, nur fehlt bis dato eine dafür zuständige Infrastruktur. Zudem gibt es in der Schweiz keinen zentralen Knoten, welcher die Zusammenarbeit mit europäischen Grossprojekten wie beispielsweise DARIAH koordinieren könnte (SBF, 2011:44). Die SAGW sieht diesbezüglich zwei grosse Massnahmenbereiche: ein Angebot für die dauerhafte Sicherung von Forschungsdaten mit der Entwicklung eines Datenrepositoriums sowie die Vernetzung bereits existierender Infrastrukturen. Des Weiteren sollen weitere Fachportale für die einzelnen Disziplinen in den Geisteswissenschaften, ähnlich desjenigen von infoclio.ch für die Geschichtswissenschaften, gefördert werden.

Das Fachportal infoclio.ch ist in der Kommission, welche das Pilotprojekt durchführen soll, vertreten. Der Fachbereich "Information documentaire" der Haute école de gestion in Genf wurde von infoclio.ch beauftragt, das Thema der digitalen Forschungsinfrastrukturen aus der Perspektive der Geistes- und Geschichtswissenschaften zu analysieren. In diesem Zusammenhang wurde des als förderlich erachtet, dass zusätzlich zu einer Literaturauswertung qualitative Interviews mit

Historikern durchgeführt werden sollen. Dabei sollten vor allem Einstellungen, Erwartungen und Ansichten der Historiker bezüglich ihrer Forschung, Forschungsdaten und Forschungsinfrastrukturen erfragt werden.

1.2. Problematik

Aufgrund der Erhebung von immer grösser werdenden Datenmengen in den Naturwissenschaften wurde es nötig, dafür geeignete Dateninfrastrukturen zu entwickeln, die diese Daten aufnehmen und verarbeiten können. Das Gross-Experiment des CERN mit dem Large Hadron Collider illustriert auf eindrückliche Weise das Ausmass, welche eine solche Datenerhebung annehmen kann: es sollen um die 15 Petabytes pro Jahr an Daten produziert werden (siehe Worldwide LHC Computing Grid, Cern [online], 2008). In diesem Zusammenhang wird auch oft von Big Data gesprochen. Das Kernproblem von Big Data besteht darin, dass Sammlungen von Datensätzen so gross und komplex werden, dass traditionelle Datenbankprogramme nicht mehr ausreichen, um die Daten zu verwalten. Zu den Herausforderungen gehören dabei u.a. die Erhebung, Aufbewahrung, Verarbeitung und Visualisierung der Daten (Big Data, Wikipedia, 2013).

Da immer mehr Daten erhoben werden, stehen auch immer mehr Daten zur Verfügung, zu viele, um von den bisherigen Dateninfrastrukturen angemessen verarbeitet werden zu können, weshalb immer häufiger auch vom "Data Deluge" gesprochen wird. Gemeint ist damit die immer grösser werdende Zahl vielleicht relevanter Daten, die frei zugänglich sind. Für Forschende wird es immer schwieriger, eine Übersicht über das zu analysierende Material zu erhalten (Borgman, 2007:212-214). Daneben stellt sich das Problem, dass im Rahmen einer späteren Verwendung immer nur wenige konkrete Daten von Interesse sind. Neben Big Data stellt sich von daher auch das Problem von Smart Data, also sehr kleinen Datensätzen, die in ihren jeweils sehr spezifischen Kontexten für Nachnutzung eine Rolle spielen.

Gleichzeitig entwickelt sich ein Bestreben seitens der forschungsfördernden Institutionen, Daten von Forschungsprojekten, die mit öffentlichen Mitteln finanziert werden, der Allgemeinheit öffentlich zugänglich zu machen. Dieses Vorhaben wird von der OECD (Organisation für wirtschaftliche Zusammenarbeit und Entwicklung) unterstützt, welche 2007 einen Bericht veröffentlicht hat (OECD, 2007), in welchem betont wird, dass frei zugängliche Forschungsdaten von gesellschaftlichem Nutzen sind. Auf EU-Ebene hat die European Science Foundation ebenfalls die freie Zugänglichkeit von öffentlich finanzierten Forschungsdaten in einem Bericht (Jamieson, 2000) festgeschrieben.

In Deutschland haben die forschungsunterstützenden Institutionen Richtlinien zur guten wissenschaftlichen Praxis herausgegeben. So sollen beispielsweise Primärdaten für zehn Jahre aufbewahrt werden oder Publikationen wenn möglich als Open Access bereitgestellt werden (Pampel et al., 2009:2). Auch in der Schweiz haben Forschungsinstitutionen und Organe zur Forschungsförderung Vorschriften verfasst, welche unter anderem die sichere und dokumentierte Aufbewahrung von Primärdaten auch nach Abschluss des Forschungsprojekts verlangen (Keller-Marxer, 2008:1).

Mit diesen Forderungen gehen aber auch Probleme einher. So stellt sich beispielsweise die Frage, wer eine Infrastruktur zur Verfügung stellen soll, damit die Daten bereitgestellt werden können, wer diese Infrastruktur finanziert und wer die Verantwortung dafür trägt.

Ein weiteres Problem, welches mit Daten einhergeht, beruht auf ihrer digitalen Natur. Viele Daten gehen verloren, weil sie nicht richtig gespeichert werden, weil kein Backup gemacht wird bzw. weil das Datenformat veraltet ist und keine Software die Datei mehr lesen kann. Da die Daten in der

Verantwortung des einzelnen Forschenden liegen, werden sie überall anders gehandhabt und anders verwaltet. Es kommt also vor, dass Wissenschaftler selbst nicht mehr auf ihre eigenen Daten zugreifen können. In der Schweiz besteht besonders bezüglich der nachhaltigen Sicherung von Forschungsdaten vor allem in den Geisteswissenschaften ein immenser Nachholbedarf (Immenhauser, 2009:13).

Doch wie sieht diese Problematik der Forschungsdaten für die Geisteswissenschaften, respektive für die Geschichtswissenschaften aus? A priori werden in diesen Disziplinen eher selten quantitative Daten erhoben. Wird über den Begriff des Datums einmal hinweggesehen, stellt sich die Frage, ob es in diesen Fachrichtungen dennoch Informationen oder Dokumente gibt, die nachhaltig gesichert werden sollten, bzw. die für Dritte von Nutzen sein könnten.

Dieser Bericht erörtert die Thematik der Forschungsdaten in Hinblick auf die Geistes- sowie die Geschichtswissenschaften. Dabei kann dieser Bericht keine eindeutigen Antworten liefern, er dient dazu, Fragen aufzuwerfen, die von der Community beantwortet werden müssen.

1.3. Ausblick

Im zweiten Kapitel werden grundlegende Konzepte erläutert, welche im Zusammenhang mit Forschungsdaten allgemein eine Rolle spielen. Es wird näher auf die Open Access-Bewegung eingegangen und die Begriffe Open Data und Linked Data werden erklärt. Des Weiteren werden die verschiedenen Begriffe bezüglich einer Forschungsinfrastruktur definiert.

Im dritten Kapitel wird die Problematik an die Geistes- und die Geschichtswissenschaften herangetragen, indem die Spezifitäten dieser Disziplinen gegenüber den Naturwissenschaften herausgearbeitet werden und der Frage nach der Rolle der Forschungsdaten in diesen Wissenschaften nachgegangen wird.

Das vierte Kapitel ist den organisatorischen, technischen und menschlichen Aspekten gewidmet, welche bei der Entwicklung einer digitalen Forschungsdateninfrastruktur geregelt sein müssen.

Es folgt im fünften Kapitel der Beschrieb einer qualitativen Studie, die ihm Rahmen des Auftrags von infoclio.ch durchgeführt worden ist. Die Resultate der Interviews mit Historikern zum Thema von Forschungsdaten werden vorgestellt.

Im sechsten Kapitel wird auf die Gründe eingegangen, welche die bisherige Abwesenheit eines Forschungsinfrastruktur-Frameworks in den Geisteswissenschaften erklären können. Des Weiteren werden die Zuständigkeiten und möglichen Stakeholder im Bezug zu digitalen Forschungsinfrastrukturen in den Geisteswissenschaften in der Schweiz aufgelistet. Darauf folgend wird ein möglicher Aufbau skizziert und die jeweilige Verantwortlichkeiten diskutiert. Als letztes wird auf mögliche Risiken bei der Einführung einer digitalen Forschungsinfrastruktur hingewiesen.

2. Kontext

2.1. Open Access

Die Open Access Bewegung hat im grossen Masse den Grundstein für die heutige Forschungsdatendiskussion gelegt. Es geht dabei darum, "[…] sofortigen, permanenten, freien, kostenlosen und elektronischen Zugang zu wissenschaftlichen Publikationen" (siehe Definitionen - Open Access, SAGW [online]) zu haben.

Die Open Access Bewegung fand ihren Beginn 1991 mit arXiv, einem Dokumentenserver, der zur Archivierung von Preprints in der Physik eingerichtet wurde und die Dokumente frei zugänglich machte. Die Aussicht von überall auf frei verfügbare wissenschaftliche Fachliteratur Zugriff zu haben führte im Jahr 2001 zur Budapest Open Access Initiative, welche Richtlinien für die Unterstützung der Open Access Bewegung in der Wissenschaft erstellte und das Ziel hat, Forschungsergebnisse öffentlich frei verfügbar zu machen (Geschichte, Informationsplattform Open Access [online], 2012).

Im Jahr 2003 folgen das Bethesda Statement on Open Access Publishing und die Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen. Das Statement bietet eine Definition des Konzepts "Open Access" und einzuhaltende Prinzipien, während die Erklärung seine Unterzeichnenden verpflichtet, die Bewegung des Open Access zu unterstützen, weiterzuentwickeln und ihre Forscher dazu aufzufordern, unter Open Access zu veröffentlichen (Geschichte, Informationsplattform Open Access [online], 2012).

Die Open Access Bewegung wird auch von Schweizer Universitäten, Bibliotheken und anderen wissenschaftlichen Institutionen unterstützt. Als Beispiel sei hier der Schweizerische Nationalfonds angeführt, welcher 2006 die Berliner Erklärung unterzeichnet hat. Seit 2007 werden von diesem Organ geförderte Projekte dazu aufgefordert, ihre Publikationen als Open Access zu veröffentlichen (SNF, 2013).

2.2. Open Data & Linked Data

2.2.1. Open Data

Im Zusammenhang mit Forschungsdaten wird häufig auch von Open Data gesprochen. Open Data oder "Offene Daten" meinen nicht nur, dass Daten online zur Verfügung stehen. Damit Daten "offen" sind, müssen sie technischen, juristischen und ökonomischen Kriterien entsprechen. Damit Daten zu Open Data werden, müssen sie folgende Voraussetzungen erfüllen (Chignard, 2012:13-14):

- Die Daten müssen in einem möglichst offenen Format vorhanden sein, das die (informatik-orientierte bzw. maschinelle) Wiederbenutzung erleichtert, und das Format darf den Gebrauch von einer proprietären Software nicht verlangen.
- Die Daten müssen unter einer freien Lizenz veröffentlicht werden und verfügen über keine oder wenige Einschränkungen für die Wiederbenutzung.
- Die Gebühren für die Wiederbenutzung sollten so weit wie möglich limitiert werden.

Dabei ist wichtig zu beachten, dass die Wiederbenutzung von Daten unter einer freien Lizenz nicht unbedingt unentgeltlich ist, was im dritten Punkt ausgedrückt wird.

Die Notwendigkeit, Daten zu publik zu machen, entstand Mitte der 90er Jahre in den Bereichen der Geophysik und Umweltwissenschaften, da in diesen Disziplinen der Zugriff auf weltweit vorhandene

Daten notwendig ist, um globale Phänomene studieren zu können (Chignard, 2012:26). Es folgte die Angst der Privatisierung der Forschung bzw. Forschungsergebnisse im Bereich der Medizin und Pharma-Industrie durch die Möglichkeit der Patentierung von beispielsweise DNA-Sequenzen vom menschlichen Genom (Gene patents in the United States bzw. Human Genom Project, Wikipedia, 2012).

Obwohl der Begriff "Open Data" im Bereich der wissenschaftlichen Forschung seinen Ursprung fand, wird er heutzutage mit Public Open Data, d.h. mit Daten von Regierungen und öffentlichen Verwaltungen gleichgesetzt (Chignard, 2012:25; siehe beispielsweise opendata.ch).

2.2.2. Linked Data

Laut Tim Berners-Lee (2009) sind Linked Data weder ein Standard noch ein Format, sondern ein erwünschtes Verhalten (Coyle, 2012:12). Dazu müssen nach ihm folgende Regeln befolgt werden (Berners-Lee, 2009):

1. "Use URIs as names for things.
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL).
4. Include links to other URIs, so that they can discover more things."

Diese Regeln beinhalten die Hauptaussage, dass Identifikatoren sowohl für die Dinge, die mit Metadaten beschrieben werden, benutzt werden sollen, als auch für die Metadaten selber zu verwenden sind. Da diese Identifikatoren mit "http://" beginnen, ist es möglich, auf der entsprechenden Seite nützliche Informationen zu hinterlegen (Coyle, 2012:12).

Linked Data sind nicht zwangsläufig offen. Wenn sie aber unter einer freien Lizenz stehen (was nicht heisst, dass die Wiederbenutzung gratis ist), werden die Daten zu sogenannten Linked Open Data.

Des Weiteren hat Tim Berners-Lee (2009) ein 5-Sterne-System kreiert, um zu evaluieren, wie weit entfernt die eigenen Daten vom Linked Open Data-Status sind.

Die Skala sieht wie folgt aus:

* = Die Daten befinden sich online unter einer freien Lizenz.

** = Die Daten sind in einer strukturierten Form und in einem maschinenlesbaren Format.

*** = Die Daten sind in einem nicht proprietären Format vorhanden.

**** = Offene Standards wie RDF und SPARQL werden benutzt, um Dinge zu identifizieren.

***** = Die Daten werden mit Daten von anderen verlinkt, um die Daten in einen gewissen Kontext zu integrieren.

Offene Datensätze werden in der sogenannten Linked Data Cloud hinzugefügt und miteinander verlinkt. Doch bisher existiert noch keine übergreifende Suchmaschine, die das Abfragen dieser Datensätze ermöglichen würde (Coyle, 2012:13). Im Zusammenhang mit Forschungsdaten bietet dieses erwünschte Verhalten ein grosses Potential, um die Datensätze miteinander zu verlinken und sie in einen Kontext zu integrieren. Diesbezüglich existieren bereits Initiativen wie beispielsweise das Projekt data.bnf.fr[1] der französischen Nationalbibliothek, welche ihre Daten und Metadaten als

[1] http://data.bnf.fr/

Linked Open Data zur Verfügung stellt, oder die Initiative LODLAM (Linked Open Data in Libraries, Archives and Museums)[2], welche das Thema der Linked Open Data im kulturellen Bereich analysiert.

2.3. Forschungsumgebung, e-Science & cyberinfrastructure

Ganz allgemein kann unter dem Begriff der Informationsinfrastruktur ein technischer, sozialer und politischer Rahmen verstanden werden, um Menschen, Technologien, Werkzeuge und Dienstleistungen zu vereinen. Eine Informationsinfrastruktur sollte zudem eine dezentralisierte und kollaborative Nutzung von Inhalten ermöglichen (Borgman, 2007:19). Solche Informationsinfrastrukturen für die Forschung gibt es in analoger Form schon lange. Dazu zählen laut der ACLS Commission on Cyberinfrastructure (ACLS Commission on Cyberinfrastructure):

- Bibliotheken, Archive und Museen, welche Informationen aufbewahren;
- Bibliographien und Hilfsmittel, welche Informationen auffindbar machen;
- Zeitschriften und Universitätspressen, welche die Informationen verbreiten;
- Editoren, Bibliothekare, Archivare und Kuratoren, welche den Betrieb dieser Struktur mit den Wissenschaftlern verbinden.

All diese Infrastrukturen haben Erweiterungen und Gegenstücke in der digitalen Welt.

Um digitale Informationsinfrastrukturen zu beschreiben benutzen Forschungsprogramme tendenziell ein "e"-Präfix wie zum Beispiel e-Science oder e-Research. Das "e"-Präfix steht dabei aber nicht für "electronic" wie zum Beispiel in E-Mail, sondern für "enhanced". Vor allem in Europa, Asien und Australien findet das "e"-Präfix seinen Gebrauch. In den USA wird bevorzugt das Präfix "cyber" benutzt, beispielsweise in "cyberinfrastructure". Im Zusammenhang mit Infrastrukturen für informationsintensive Forschung sind die beiden Präfixe als äquivalent zu betrachten (Borgman, 2007:20).

Ein weiterer diese Thematik betreffender Begriff ist derjenige der virtuellen Forschungsumgebung (Virtual Research Environment). Er wird meistens synonym zu cyber- bzw. e-Infrastruktur verwendet. Wird ein Unterschied gemacht, dann beschreibt die virtuelle Forschungsumgebung einen Forschungskontexts aus einer holistischen Sicht, während e-Infrastruktur auf Kerndienstleistungen fokussieren, über welche eine virtuelle Forschungsumgebung laufen soll (Fraser, 2005).

Die Begriffe e-Science oder Cyberinfrastructure können mehr oder weniger breit aufgefasst werden. So stellt eine Infrastruktur für die einen vor allem einen technischen und operationellen Rahmen dar, anhand dessen kollaboratives Arbeiten sowie das Teilen von Daten und Resultaten ermöglicht werden. Für andere bedeutet eine Infrastruktur eher den Zugang zu Inhalten und weniger technische Hilfsmittel. Für wieder andere muss eine Infrastruktur beides beinhalten (Moulin et al., 2011:4).

Laut Rockwell (2010:7) ist eine Cyberinfrastruktur dazu gedacht, Forschung zu fördern. Eine gute Infrastruktur sollte zu reduzierten Kosten und zu mehr Forschung von jedem Forschenden führen. In diesem Zusammenhang bedeutet e-Science eine verbesserte Wissenschaftspraxis. Diese verbesserte Forschung soll dadurch ermöglicht werden, dass unter anderem Primärdaten überhaupt verfügbar sind und zusätzlich dazu Werkzeuge zu deren Verarbeitung zur Verfügung stehen sowie die Ergebnisse direkt in die Forschungsumgebung integriert werden können. Das Bereitstellen von digitalen Ressourcen stellt aber noch keine Forschungsumgebung im Sinne der e-Science dar. Ausschliesslich der bewusste Aufbau einer Umgebung, welche eine technische Infrastruktur,

[2] http://lodlam.net/

Ressourcen und Werkzeuge auf einer einzigen Plattform integriert, kann als e-Science bezeichnet werden (Sahle, 2008:64).

Für die Lektüre von englischsprachigen Roadmaps, Forschungsprojekten und Infrastrukturinitiativen soll auf Folgendes aufmerksam gemacht werden: In der englischen Sprache wird klar zwischen "science", "social sciences" und "humanities" unterschieden. Wie das Oxford English Dictionary spezifiziert, wird der Begriff „science“ im modernen Sprachgebrauch häufig mit Naturwissenschaften und Physik gleichgesetzt.[3] Der Begriff "science" sollte deshalb in der anglophonen Literatur allererst als Naturwissenschaften verstanden werden, denn meistens wurden für die anderen Wissenschaftsbereiche entsprechende Begriffe wie “e-Humanities” oder “e-Social Sciences” eingeführt. Bei e-Science-Frameworks wird es sich folglich um eine Infrastruktur für die Naturwissenschaften handeln.

In dieser Studie wird ausschliesslich der Begriff "digitale Forschungsinfrastruktur" verwendet. Dabei basieren wir uns auf die allgemeine Definition einer Informationsinfrastruktur, d.h. es ist ein technischer, sozialer und politischer Rahmen, welcher Menschen, Technologien, Werkzeuge und Dienstleistungen vereint.

Im Folgenden werden drei Modelle erläutert, welche im Zusammenhang mit der Verwaltung von Forschungsdaten von hoher Wichtigkeit sind. Als erstes wird das OAIS-Referenzmodell vorgestellt, welches einen allgemeinen Rahmen für die Langzeitarchivierung von digitalen Dokumenten vorschlägt. Darauf wird das Data Continuum Model erklärt, welches die Forschungsarbeit und die damit entstehenden Daten in drei Domänen unterteilt. Als letztes wird das Modell des Digital Curation Cycle präsentiert, welches als Grundlage für die Diskussion über die Zuständigkeiten bezüglich technischer und kuratierender Aspekte dienen soll.

2.4. Modelle

2.4.1. OAIS-Referenzmodell

Das Referenzmodell für ein offenes Archiv-Informationssystem (OAIS) ist ein Modell für die digitale Archivierung. Es ist ein strikt logisches Modell und deshalb unabhängig von jeder Implementierung, weshalb es sich weltweit durchgesetzt hat (Locher, 2011:4). Es geht über eine simple Speicherung der Bits (*bitstream preservation*) hinaus und definiert ein System, welches für die langfristige Erhaltung von Information und deren Zugang verantwortlich ist (CCSDS, 2012:2-1). Das vom Referenzmodell aufgezeichnete Archiv befindet sich in einem Umfeld, welches die Rollen der Informationsproduzenten, Informationskonsumenten und Verwalter des Systems beinhaltet. Der Begriff der Information wird dabei so verstanden, dass es sich dabei um Wissen handelt, welches ausgetauscht werden kann und immer in Form von Daten ausgedrückt bzw. repräsentiert wird. Die Information benötigt immer auch eine Repräsentationsinformation, damit sie verstanden werden kann (CCSDS, 2012:2-3). Im Falle einer digitalen Datei kann es sich bei der Repräsentationsinformation beispielsweise um Hard- und Software handeln. Ein Datenobjekt wird folglich anhand seiner Repräsentationsinformation interpretiert und ergibt damit ein Informationsobjekt (CCSDS, 2012:2-4). Damit ein Informationsobjekt langfristig erhalten werden kann, müssen das dahinterliegende Datenobjekt (die Bits) und die Repräsentationsinformation klar

[3] "In modern use, [science is] often treated as synonymous with ‘Natural and Physical Science’ […]". # "science, n.". OED Online. September 2012. Oxford University Press. http://www.oed.com/view/Entry/172672?redirectedFrom=science& (accessed September 24, 2012).

identifiziert werden, worin die grosse Herausforderung besteht. Denn die Repräsentationsinformation ist rekursiver Natur und besteht meist ihrerseits aus eigenen Datenobjekten mit eigenen Repräsentationsinformationen.

Das Referenzmodell definiert des Weiteren das Konzept des Informationspakets. Dieses ist ein Container für zwei Informationstypen: *Content Information*, welche das Datenobjekt und die Repräsentationsinformation enthält, und *Preservation Description Information*. Das Paket als Ganzes erhält beschreibende Informationen, um es auffindbar zu machen.

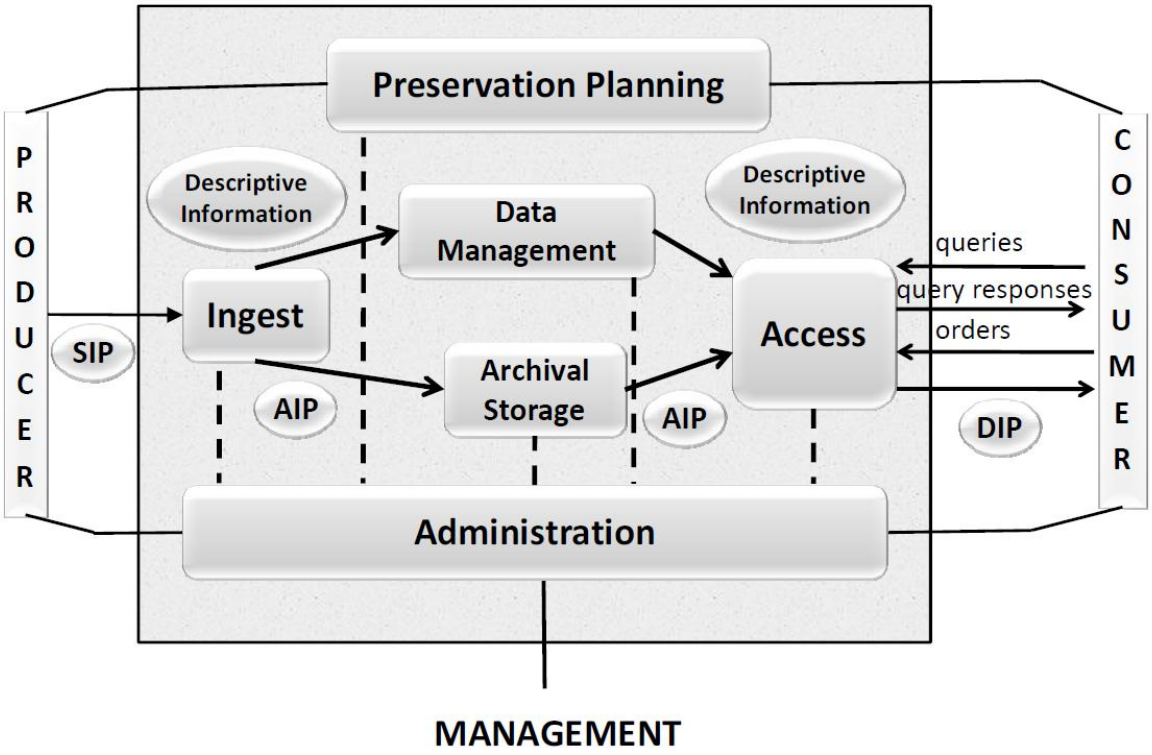


Abbildung 1: Funktionale Entitäten des OAIS-Modells (CCDS, 2012:4-1)

Das OAIS-Referenzmodell unterscheidet zwischen sechs Hauptfunktionen, welche hier zusammenfassend dargestellt sind (siehe Abbildung 1):

- **Ingest (Datenübernahme):** Hierbei handelt es sich um die Übernahme von vom Informationsproduzenten erzeugten Informationspaketen, welche überprüft und in eine archivierbare Form gebracht werden.
- **Archival Storage (Archivspeicher):** Bei dieser Funktion geht es darum, die Informationspakete zu erhalten, indem regelmässig die Datenintegrität überprüft und ein Backup erstellt wird.
- **Access (Abfrage):** Diese Funktion ermöglicht es, anhand einer Schnittstelle die im Archiv vorhandenen Informationspakete zu durchsuchen und dabei auf die Zugriffsberechtigungen zu achten.
- **Administration:** Zu dieser Funktion gehören die Konfiguration von Hard- und Software sowie die Steuerung der Gesamtabläufe.
- **Data Management (Datenverwaltung):** Diese Funktion beinhaltet die Verwaltung der beschreibenden Informationen und Metadaten.

- **Preservation Planning (Archivierungsplanung):** Hierbei handelt es sich um die Verfolgung eventueller Weiterentwicklungen von Technologien und Standards und um die daraus resultierende Datenmigration.

Eine deutsche Übersetzung einer älteren Version des OAIS-Referenzmodells, welche im Jahr 2002 erschienen ist, wird vom Kompetenznetzwerk zur digitalen Langzeitarchivierung nestor zur Verfügung gestellt (siehe Büchler et al., 2012). Eine Übersetzung derselben Version aus dem Jahr 2002 ist ebenfalls auf Französisch bereitgestellt (siehe CCSDS, 2005).

2.4.2. Data Continuum Model

Das Data Continuum Model entstand aufgrund der Erfahrungen, welche an der Monash University in Australien mit Repositorien gesammelt wurden. Laut Treloar, Groenewegen und Harboe-Ree (2007) begann 2003 ein Projekt (ARROW), welches ein *single repository* für alle Aktivitäten der Universität entwickeln sollte, u.a. Forschung, Lehre und Administration. Während der Umsetzung des Repositoriums wurde klar, dass ein *single repository* nicht die richtige Lösung sein kann. Die Gründe dafür waren einerseits, dass die Arten der aufzunehmenden digitalen Objekte stark variieren. Andererseits unterschieden sich die Managementeigenschaften und der Zugang zu den digitalen Objekten je nach Funktion. Diese Differenzen können laut den Autoren nicht auf eine einfache Art und Weise in eine gemeinsame Infrastruktur übertragen werden. Dies trifft besonders auf Repositorien zu, welche neben Publikationen und diskreten Objekten zusätzlich noch durch e-research generierte Daten aufnehmen müssen (Treloar, Groenewegen, Harbor-Ree, 2007). Diese Erfahrung führte die Autoren dazu – wie sie es in einem späteren Artikel (Treloar & Harboe-Ree, 2008:9) klar ausdrücken – sich für *mehrere, wenngleich interoperable*, Repositorien auszusprechen.

Es stand also fest, dass in der Monash University ein Repositorium nicht alles aufnehmen kann und dafür mindestens zwei Infrastrukturen benötigt werden. Um zu entscheiden, wie sich diese Repositorien voneinander unterscheiden sollen und welche Objekte in welche Infrastruktur gehören, wurden Informationsmanagementdimensionen bestimmt. Anhand von Benutzeranforderungen, Fallstudien und der Sichtung des aktuellen Forschungsstands erschien es als schwierig, spezifische Werte innerhalb einer Dimension zu definieren. Stattdessen wurde entschieden, sogenannte Continua festzulegen, welche zwischen zwei Endpunkten einer Dimension fluktuieren (siehe Abbildung 2; Treloar, Groenewegen & Harboe-Ree, 2007).

Object:	Less Metadata	⟷	More Metadata
	More Items	⟷	Fewer Items
	Larger Objects	⟷	Smaller Objects
	Objects continually updated	⟷	Objects static/derived snapshots
Management:	Researcher Manages	⟷	Organisation Manages
	Less Preservation	⟷	More Preservation
Access:	Mostly Closed Access	⟷	Mostly Open Access
	Less Exposure	⟷	More Exposure

Abbildung 2: Continua (Treloar, Groenewegen, Harboe-Ree, 2007)

In der Abbildung 2 werden die Dimensionen drei Aspekten zugeteilt: Objekt, Management und Zugang.

Zu den Dimensionen gehören:

- die Metadaten;
- die Anzahl aufzunehmender Objekte;
- die Grösse der Objekte;
- die (statische oder aktualisierte) "Fixierung" der Objekte;
- die Managementverantwortung;
- die Langzeitarchivierung;
- der Zugang;
- die externe Auffindbarkeit (Exposure).

Die Auflistung ist so zu verstehen, dass es einerseits Repositorien gibt, die weniger oder mehr Metadaten verwalten müssen. Andererseits gibt es Repositorien, die eher mehr oder eher weniger Objekte aufnehmen müssen. Um das Ausmass einer Infrastruktur zu beschreiben, sollte bewusst festgelegt werden, auf welcher Seite der Dimensionen sich das Repositorium am ehesten befinden soll.

Aufgrund dieser Dimensionen kam die Monash University 2007 zum Schluss, zwei unterschiedliche Repositorien anbieten zu wollen: eine Infrastruktur, die der Kollaboration dient und eine Infrastruktur, die der Publikation dient (Treloar, Groenewegen, Harbor-Ree, 2007).

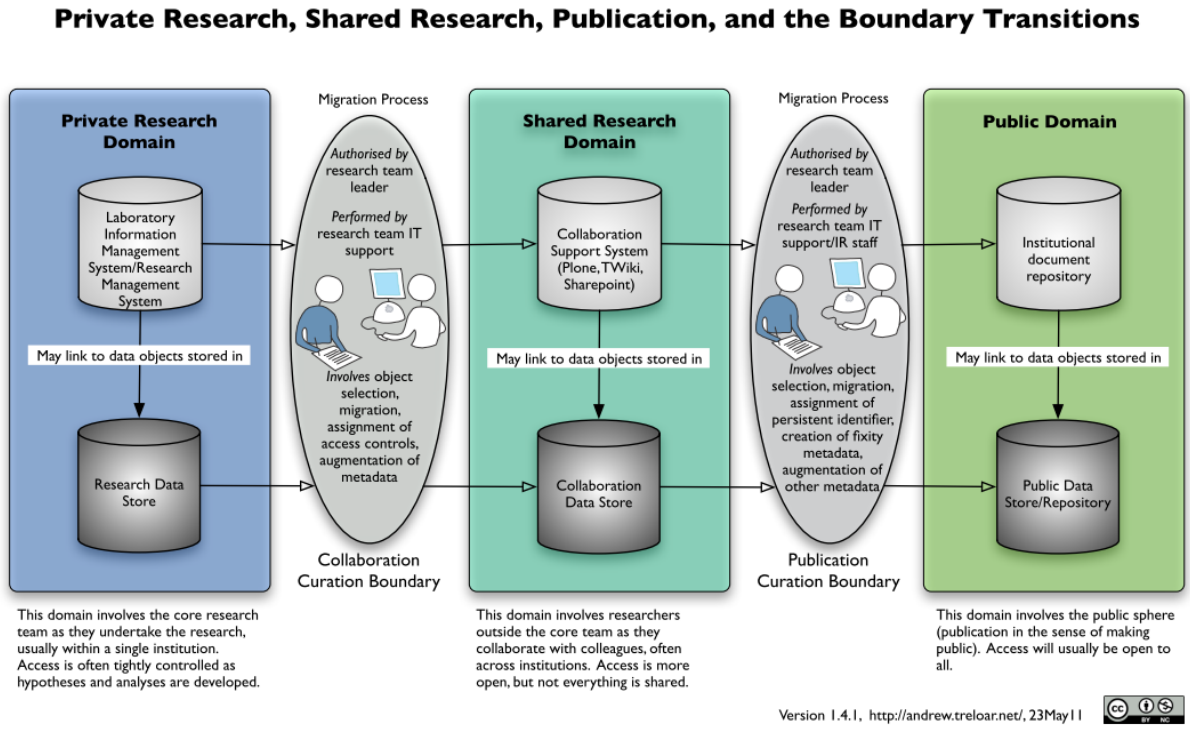


Abbildung 3: The Data Continuum Model (Treloar, 2011)

Im Jahr 2008 haben die Autoren Treloar und Harboe-Ree (2008) den zwei Repositorien ein drittes hinzugefügt und alle drei Infrastrukturen in drei Domänen unterteilt. Die erste Domäne ist der private Forschungsbereich. Hier arbeitet das Forschungsteam im Alltag und erstellt Daten und Resultate. Die zweite Domäne ist der kollaborative Forschungsbereich. Hier wird ein Teil der

Forschungsresultate anderen Forschenden zugänglich gemacht. Die dritte Domäne ist der öffentliche Bereich. An diesem Punkt sind die Resultate in einer "fertigen" Form vorhanden und für die Öffentlichkeit einsehbar. Diese drei Domänen wurden von den Autoren in das Data Continuum Model übersetzt (siehe Abbildung 3).

Die von den Autoren festgelegten Eigenschaften für die drei definierten Domänen sind hier in tabellarischer Form wiedergegeben (siehe Tabelle 1).

Dimensionen	**Privater Forschungsbereich**	**Kollaborativer Forschungsbereich**	**Öffentlicher Bereich**
Metadaten	Wenig Metadaten	Mehr Metadaten	Noch mehr Metadaten
Anzahl Objekte	Viele Objekte	Weniger Objekte	Noch weniger Objekte
Grösse der Objekte	Grössere Objekte	Kleinere Objekte	Noch kleinere Objekte
Fixierung der Objekte	Häufige Aktualisierung	Eher statisch	Statisch
Managementverantwortung	Verwaltung durch den Forscher	Verwaltung durch den Forscher	Verwaltung durch die Organisation
Langzeitarchivierung	Wenig Langzeitarchivierung	Wahrscheinlich mehr Langzeitarchivierung	Mehr Langzeitarchivierung
Zugang	Beschränkter Zugriff	Weniger beschränkter Zugriff	Open Access
Externe Auffindbarkeit	Keine externe Auffindbarkeit	Keine externe Auffindbarkeit	Externe Auffindbarkeit

Tabelle 1: Charakteristiken der drei Forschungsbereiche (Treloar & Harboe-Ree, 2008:5-7)

2.4.3. Digital Curation Cycle

Ein weiteres Modell, welches im Zusammenhang mit einer Cyberinfrastruktur häufig zitiert wird, ist der Digital Curation Cycle, welcher vom Digital Curation Center (DCC) entwickelt worden ist. Das DCC ist ein Kompetenzzentrum in Grossbritannien für die Kuration von digitaler Information (siehe About the DCC, Digital Curation Center [online]).

Das Modell des Digital Curation Cycle wurde in Bezug auf die Kuration und die Erhaltung von digitalen Objekten entwickelt (Higgins, 2008). Digitales Material ist aufgrund seiner Natur technischen Veränderungen ausgesetzt. Aktionen, die während des Lebenszyklus von digitalem Material unternommen oder ausgelassen werden, haben einen Einfluss auf dessen Nachhaltigkeit. Die Autoren vertreten deshalb eine Herangehensweise, die auf dem Lebenszyklus von digitalem Material basiert, um einzelne Etappen wie auch nötige Aktionen darzustellen und zu planen. Mit diesem Vorgehen soll die Authentizität, die Vertrauenswürdigkeit, die Integrität und die Benutzerfreundlichkeit von digitalem Material erhalten bleiben (Higgins, 2008:135).

Das DCC hat dafür ein Modell entwickelt, welches generischer Natur ist (siehe DCC Curation Lifecycle Model, Digital Curation Centre [online]; siehe Abbildung 4). Im Zentrum dieses Modells stehen die

Daten, das heisst digitale Objekte sowie Datenbanken. Diese Daten sind in einen Lebenszyklus integriert, der in vier grosse Etappen unterteilt ist (Higgins, 2008:135-137):

- die Beschreibung und Präsentation der Information;
- die Planung der Datenerhaltung;
- das Monitoring bezüglich Standards, Tools und Software;
- die Förderung von Kuration und Datenerhaltung.

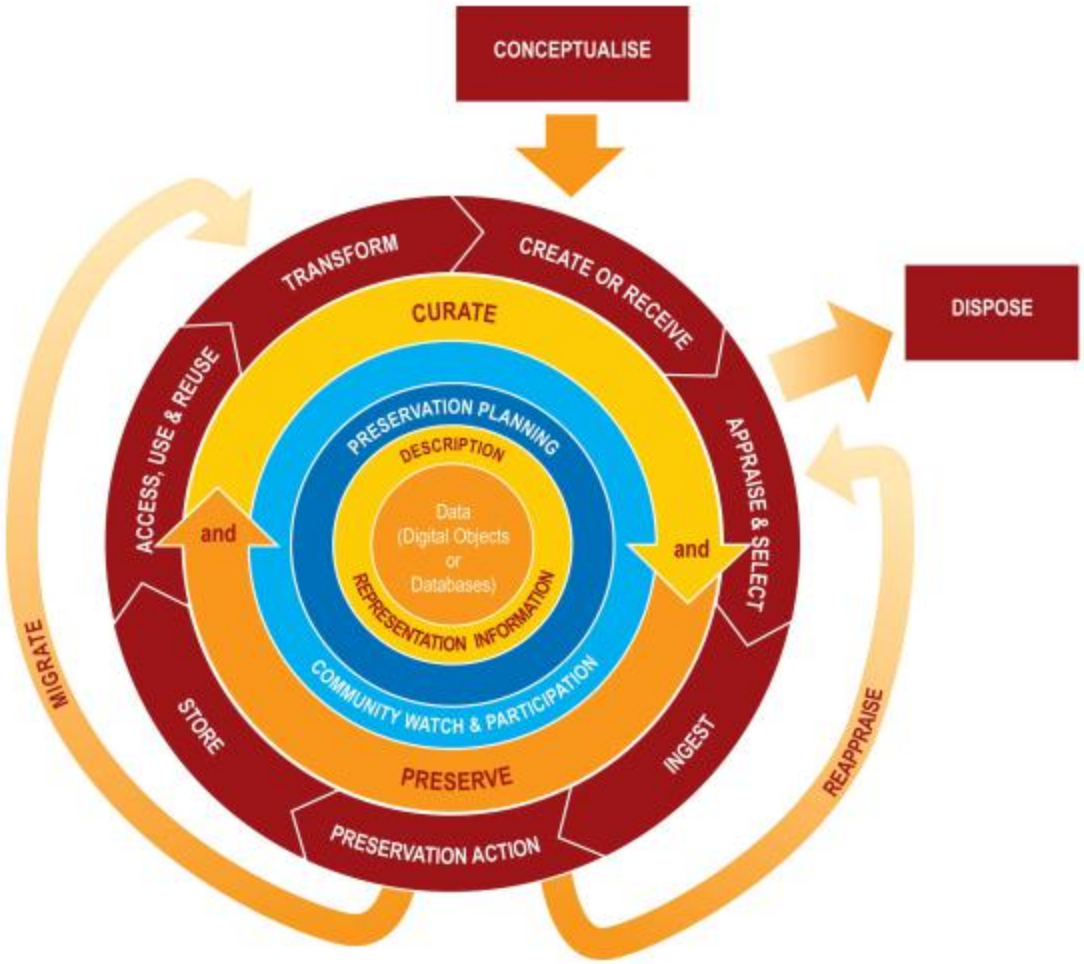


Abbildung 4: The DCC Curation Lifecycle Model (Higgins, 2008:136)

Das Modell enthält des Weiteren sequentielle Etappen mit den dazugehörenden Aktionen (Higgins, 2008:138):

1. Conceptualise (Konzeption): Bevor die Daten überhaupt erstellt werden (bzw. übernommen werden), muss die Datenkreation inklusive des Forschungsinstruments und der Speicherungsoptionen konzipiert und geplant werden.
2. Create or receive (Datenerstellung und -aufnahme): Es folgt die Datenerstellung mit den dazugehörenden Metadaten bzw. die Datenübernahme.
3. Appraise & select (Bewertung & Selektion): Die nächste Etappe beinhaltet die Bewertung und Selektion der Daten, welche langzeitarchiviert werden sollen.
4. Ingest (Datenübernahme): Die ausgewählten Daten werden in eine vorhandene Infrastruktur für die Langzeitarchivierung überführt.

5. Preservation action (Aufbewahrungsaktivitäten): Sobald die Daten in der Infrastruktur sind, müssen nachhaltigkeitssichernde Massnahmen unternommen werden, um den Erhalt von der Authentizität, Vertrauenswürdigkeit, Nutzbarkeit und Integrität der Daten zu gewährleisten.
6. Store (Langzeitarchivierung): Die Daten werden daraufhin in einer sicheren Art und Weise gespeichert.
7. Access, use & reuse (Zugriff, Benutzung & Wiederverwendung): Der Zugang, die Benutzung und die Wiederverwendung der Daten müssen dabei garantiert werden. Dies kann (muss aber nicht) in Form von öffentlich publizierter Information sein.
8. Transform (Transformation): Der letzte Schritt bezieht sich auf die Transformation von den gespeicherten Daten, indem beispielsweise eine Teilmenge erstellt wird, um daraus neue Resultate abzuleiten.

Diese acht Etappen werden unter zwei generellen Begriffen zusammengefasst: Curate (kuratieren, verwalten, organisieren, betreuen) und preserve (erhalten). Die Aktivität des Erhaltens umfasst die eher technischen Aspekte des Lebenszyklus wie die Datenaufnahme in die Infrastruktur, die nachhaltigkeitssichernden Massnahmen und die Speicherung. Dem gegenübergestellt ist die Aktivität der Kuration, welche die Komponenten der Datenerstellung, -bewertung, -auswahl, den Zugang und die Nutzung sowie die Wiederverwendung und die Transformation beinhaltet (Higgins, 2008:138).

Auf der Website des Digital Curation Centers (DCC Curation Lifecycle Model) wird verdeutlicht, dass das Modell als Diskussionsbasis dienen soll. So kann anhand der identifizierten Teilprozesse festgelegt werden, ob alle damit verbundenen Aktivitäten auch stattfinden sollen, wer dafür verantwortlich ist und welche Standards und Technologien benutzt und wann sie eingesetzt werden sollten. Dies führt idealerweise zu dokumentierten Prozessen und Strategien sowie definierten Rollen und Verantwortlichkeitszuteilung (Rümpel et al., 2011:29).

Nachdem nun die Begriffe und Terminologien definiert und erläutert worden sind, soll im folgenden Kapitel auf die Forschung in den Geistes- sowie in den Geschichtswissenschaften näher eingegangen werden, um ihre Besonderheiten gegenüber den Naturwissenschaften herauszuarbeiten. Darauf wird der Begriff der Forschungsdaten in diesem Kontext erklärt und neudefiniert.

3. Forschung

3.1. Forschung in den Geisteswissenschaften

Die Forschung in den Geisteswissenschaften zeichnet sich durch ein vielfältiges methodologisches Spektrum aus. Dies führt zu einer grossen Heterogenität in den Geisteswissenschaften, da stetig neue Methoden dazukommen und dadurch auch neue Arten von Daten generiert werden (Pempe, 2012:137-138).

Es ist sehr schwierig, die Geisteswissenschaften als Ganzes zu beschreiben. Generell kann gesagt werden, dass die darin beinhalteten Disziplinen eher interpretativ als datenorientiert sind, d.h. dass sie eher bereits vorhandene Dokumente bzw. Artefakte kommentieren anstatt neue Daten, bspw. als Konsequenz wissenschaftlicher Experimente erzeugen. Dies bedingt auch, dass die sich aus der Forschungsaktivität ergebenden Daten eine schwer vergleichbare, weniger empirische sondern vielmehr eine eher narrative Struktur haben. Dennoch gibt es Geisteswissenschaftler, welche quantitative bzw. qualitative Studien durchführen, wofür sie Methoden aus den Sozialwissenschaften oder anderen Disziplinen anwenden. Mit dem Gebiet der Digital Humanities werden auch immer mehr technikorientierte Methoden eingesetzt (Borgman, 2007:213).

Obwohl in den Geisteswissenschaften naturwissenschaftliche Methoden durchaus auch ihre Anwendung finden, werden diese Disziplinen grundsätzlich als unterschiedlich zu der Forschung in anderen Gebieten angesehen. Dies bezieht sich sowohl auf die benutzten Methoden, als auch auf die Herangehensweise an Daten (Burrows, 2011:177).

Traditionellerweise wurde die geisteswissenschaftliche Forschung auf der Basis von Primärquellen durchgeführt, welche in einer physikalischen Form in einer Institution wie einem Museum, Archiv oder Bibliothek aufbewahrt werden. In einem weiteren Rahmen können auch Gebäude, archäologische Artefakte oder auch Personen den Untersuchungsgegenstand darstellen. Diesbezüglich kann ein allgemeiner geisteswissenschaftlicher bzw. historischer Forschungsprozess wie folgt aussehen (Blanke & Hedges, 2013:2):

Ein Forschender besucht ein Archiv und durchforscht mit den ihm zur Verfügung stehenden Werkzeugen den Bestand und findet so relevante Dokumente für seine Forschungsarbeit, welche er in einer Sammlung organisiert und annotiert. Eine interessante Stelle in einem Dokument kann den Forschenden zu einem anderen Dokument führen, was in einer Lesekette zwischen sehr vielen Quellen resultiert. Dieser Arbeitsprozess mündet anschliessend in eine Monographie oder einen Artikel.

Für Forschende in den Geisteswissenschaften repräsentieren wissenschaftliche Monographien nach wie vor die wichtigste Publikationsform. Zeitschriften und Konferenzbeiträge sind auch wichtig, haben aber deutlich weniger Bedeutung im Vergleich zu den Natur- und Sozialwissenschaften. Das hat zur Folge, dass in den Geisteswissenschaften im Vergleich zu den übrigen Wissenschaften traditionell der geringere Anteil an Literatur online zur Verfügung steht, da sehr wenige Monographien digitalisiert oder in digitaler Form publiziert werden (Borgman, 2007:214).

3.2. Forschung in den Geschichtswissenschaften

Die Geschichtswissenschaften stellen grundlegend keine einheitliche Disziplin dar, sondern bestehen vielmehr aus einzelnen epochalen, regionalen und sachlichen Teilfächern. Zudem weisen die

Geschichtswissenschaften zahlreiche Referenzen zu geistes- und kulturwissenschaftlichen Fächern auf (Meyer, 2011 (2):41).

Die allen Teilbereichen gemeine Forschungstätigkeit enthält die Sichtung und Bewertung von Material. Anhand von schriftlichen Analysen und Publikationen wird dieses Material erschlossen und in Forschungstendenzen eingeordnet. Diese individuelle Tätigkeit wird im Rahmen eines Forschungsvorhabens durchgeführt und wird zunehmend arbeitsteilend organisiert (Meyer, 2011 (2):41).

Die Arbeit eines Historikers kann grundlegend wie folgt dargestellt werden (Geschichtswissenschaft, Wikipedia, 2012): ein Forscher geht mit einer Fragestellung an einen Untersuchungsgegenstand heran; er sammelt, bewertet und erschliesst die verfügbaren Primär- und Sekundärquellen; er interpretiert die Quellen nach den methodischen Regeln des Fachs; er stellt die Ergebnisse zusammen, um sie zu publizieren.

Monographien sind nach wie vor die häufigste und beliebteste Form, in welcher Ergebnisse der wissenschaftlichen Forschung veröffentlicht werden (Meyer, 2011 (2):41; Jehne, 2009:59). Das führt dazu oder wird dadurch beeinflusst, dass die Anzahl publizierter Monographien und weniger die Anzahl Zitierungen eines der Hauptkriterien ist, mit welchen die Arbeit eines Historikers bewertet wird. Dementsprechend werden bei einer Bewerbung für eine Stelle oder für Forschungsgelder immer zuerst die Liste der veröffentlichten Bücher angeschaut (Jehne, 2009:59).

Doch diese Publikationsart bringt auch einige Nachteile mit sich. Im Vergleich zu der Veröffentlichung eines Artikels in einer Zeitschrift fällt bei einer Monographie die Überprüfung der Qualität durch ein wissenschaftliches Komitee anhand eines Peer Review weg (Meyer, 2011 (2):41). Die Qualitätskontrolle gestaltet sich deshalb in den Geschichtswissenschaften als schwierig. Es gibt Reihen und Verlage, welche für die Qualität der publizierten Werke bekannt sind, doch es gibt auch sehr gute Bücher, die in einem Verlag ohne Qualitätskontrolle erschienen sind. Doch grundsätzlich kann die Qualität historischer Publikationen nur aus deren Lektüre hervorgehen (Jehne, 2009:59). Aus diesem Grunde werden Rezensionen in den Geschichtswissenschaften hoch geschätzt, da sie die Einhaltung von Mindeststandards kontrollieren (Meyer, 2011 (2):41). Die Rezensionen erhalten dadurch eine grosse Wichtigkeit, was eine Spezifität der Geschichtswissenschaften ausmacht.

Eine weitere Besonderheit der Geschichtswissenschaften ist die traditionelle Aufgabe der Nationalgeschichte. Ein italienischer Historiker forscht in Italien über die italienische Geschichte. Die Forschung ist folglich häufig lokal geprägt, was dazu führt, dass die dazugehörenden Publikationen auf Italienisch in Italien veröffentlicht werden (Jehne, 2009:60). Solche Veröffentlichungen interessieren wiederum einen restriktiven Kreis an anderen Forschenden, welche meistens dann auch Deutsche sind. Dies ist ein grosser Unterschied zu den Naturwissenschaften, wo beispielsweise Forschungsergebnisse in der Biologie, welche in den USA erhalten wurden, in Europa gleichviel Bedeutung haben wie in den USA. Obwohl diese Tradition in den Geschichtswissenschaften charakteristisch bleiben wird, findet doch eine Auflockerung statt, und Präsentationen, Austausch und Bewertung von Forschung sind immer mehr von einer Internationalisierung geprägt (Meyer, 2011 (2):42).

Eine weitere Spezifität ist die Eigenschaft, dass in den Geschichtswissenschaften deutlich mehr Wissen kumuliert als aktualisiert wird (Landes, 2008:27). Das hat zur Folge, dass das elektronische Publizieren in den Geschichtswissenschaften weniger verbreitet ist, u.a. weil bezüglich des Veröffentlichens keine Dringlichkeit besteht. Eine Publikation hat meistens eine dermassen persönliche Note, dass ein Historiker selten einem anderen zuvor kommen muss (Jehne, 2009:60).

Online- oder Hybrid-Zeitschriften sind deshalb in den Geschichtswissenschaften eher weniger verbreitet. Zusätzlich dazu ist der rasche Nachweis von Veröffentlichungen in Fachdatenbanken noch nicht zuverlässig (Sahle, 2008:72).

3.3. Forschungsdaten

3.3.1. Forschungsdaten allgemein

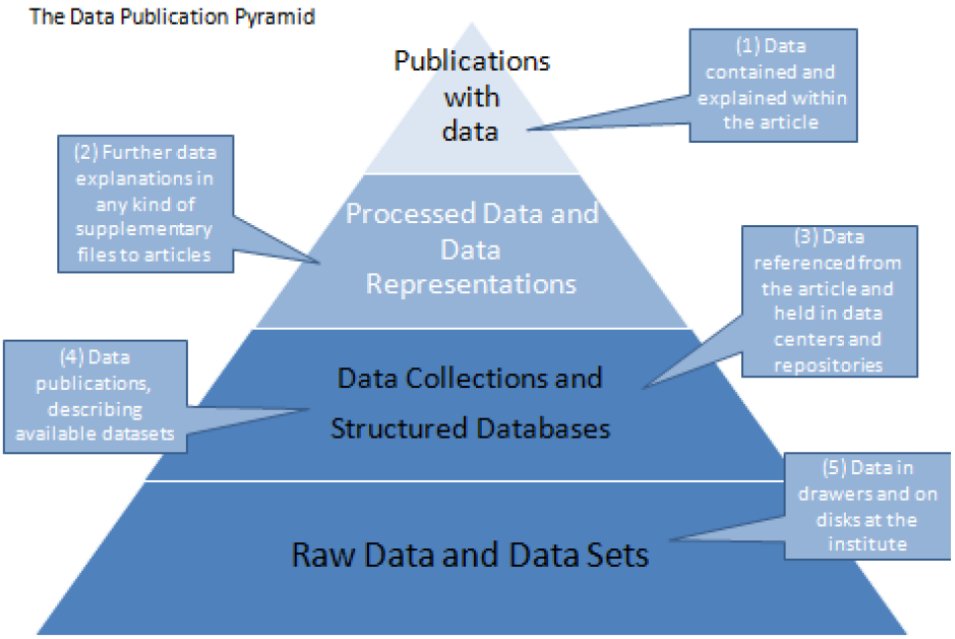
In den bereits installierten Forschungsinfrastrukturen für natur- und auch sozialwissenschaftlich orientierte Disziplinen ist es relativ einfach, die Forschungsdaten zu identifizieren. Diese stammen von Experimenten, Erhebungen, Simulationen und Messungen, welche numerische, d.h. aus Ziffern bestehende, Daten liefern, die in Datenbanken gespeichert werden können.

Die allgemeinste Definition des Begriffs Forschungsdaten stammt von einem Bericht über den legalen Status von Forschungsdaten, welcher ein auf Niederländisch erschienenes Zitat auf Englisch wiedergibt:

> *"A datum is an element that has relevance and semantic value. Data is used to describe features of persons, things, actions, etc. taken from reality."*

(Langerhorst, 1981, zitiert in De Cock Buning et al., 2011:10)

Folglich ist alles als Forschungsdatum zu bezeichnen, was relevant ist und semantischen Wert besitzt sowie Eigenschaften von Personen, Dingen, Handlungen etc. der Realität beschreibt.



Graph 1: The Data Publication Pyramid, developed on the basis of the Jim Gray pyramid, to express the different manifestation forms that research data can have in the publication process. See Chapter 1 for a full explanation.

Abbildung 5: The Data Publication Pyramid (Reilly et al., 2011:6)

In einem Bericht über Daten- und Publikationsintegration vom *Opportunities for Data Exchange* Projekt werden die Forschungsdaten auf vier unterschiedlichen Ebenen eingeteilt, basierend auf der

Arbeit von Jim Gray im Rahmen seiner Arbeit bei Microsoft (siehe Abbildung 5). Diese durch eine Pyramide dargestellten Ebenen enthalten in der untersten Schicht die Rohdaten, welche unbearbeitet und unkorrigiert vorliegen. Auf der zweituntersten Ebene befinden sich Datensammlungen und strukturierte Datenbanken, welche bereinigte Datensätze anbieten. Diese Form von Daten wird meistens in Datenzentren oder -repositorien angeboten. Auf der nächsten Ebene befinden sich die verarbeiteten Daten und Datenvisualisierungen, welche die Ergebnisse der Verarbeitung der Rohdaten darstellen. An der Spitze der Pyramide befinden sich dann die Publikationen, welche die für die vorgestellten Forschungsresultate benötigte Teilmenge an Daten enthalten und erklären.

Es gibt auch andere Typologien, welche zum besseren Verständnis des Begriffs Forschungsdaten angewendet werden. Dabei wird häufig zwischen Primär- und Sekundärdaten unterschieden. In der für das Grossprojekt e-lib.ch entwickelten Konzeptstudie über die Langzeitarchivierung in der Schweiz werden diese beiden Begriffe definiert (Keller-Marxer, 2008:4). Laut dieses Berichts sind Primärdaten Rohdaten, welche anhand von Messungen, Erhebungen oder Simulationen entstanden sind. Diese Rohdaten ermöglichen es, Auswertungen und Resultate zu erzielen. Ebenfalls zu den Primärdaten gehören Metadaten und Dokumentationen, welche eine für dritte verständliche Interpretation der Daten erlaubt. Nicht dazu gehören Unterlagen, welche Auswertungs- und Forschungsprozesse dokumentieren. Diese Unterlagen werden als Forschungsunterlagen bezeichnet. Auswertungsdaten und Daten der Forschungsresultate gehören ebenfalls nicht dazu, weil sie aus den Primärdaten abgeleitet werden können, wenn das Auswertungsverfahren bekannt ist.

Sekundärdaten werden im Vergleich dazu als aus Primärdaten abgeleiteten Publikationen verstanden, welche die Form von elektronischen Zeitschriften und Fachbüchern, Lehr- und Lernunterlagen sowie Hochschulpublikationen annehmen können (Keller-Marxer, 2008:4)

3.3.2. Forschungsdaten in den Geisteswissenschaften

Es gibt in den Geisteswissenschaften quantitative Daten ähnlich denen in den Natur- und Sozialwissenschaften, wie Statistiken oder Datenbanken bzw. wie Interviews, Umfragen und Fragebogen. Die meisten Artefakte in den Geisteswissenschaften unterscheiden sich jedoch von denjenigen in Natur- und Sozialwissenschaften in einigen Aspekten. Die Geisteswissenschaften benutzen das weiteste Spektrum an Informationsquellen und folglich ist die Unterscheidung in diesem Bereich zwischen Dokumenten und Daten am wenigsten klar (Borgman, 2007:214). Des Weiteren unterscheiden sich die Geistes- von den Naturwissenschaften auch darin, dass Datensätze eher gesammelt als generiert werden und heterogener Natur sind (Moulin et al, 2011:5). Viele wissenschaftliche Autoren behaupten, dass die Primärquellen (inkl. Dokumente, Texte und Bilder) die Daten des Forschers sind (Borgman, 2007:215-17; Burrows, 2011:180-181). So haben etwa die von Jim Gray definierten vier Paradigmen der Wissenschaft (das empirische, theoretische, computerorientierte und das daten-explorative Paradigma) in den Geisteswissenschaften gar keine oder nur eine geringe Relevanz (Hey et al., 2009:xviii).

Generell ist es zu einschränkend und zu einfach, den Forschungsdatenbegriff so wie er in den Natur- und Sozialwissenschaften angewendet wird, auf die Geisteswissenschaften zu übertragen. Es stellt sich also die Frage, was denn Forschungsdaten in den Geisteswissenschaften sind.

Im französischen Projekt CORPUS IR (http://www.corpus-ir.fr/) wird eine Forschungsinfrastruktur für die Geistes- und Sozialwissenschaften entwickelt, welche die Archivierung und Bearbeitung von qualitativen Daten übernehmen soll. Auf der Projektwebsite werden qualitative Daten als Informationen interpretiert, die nicht direkt messbar und quantifizierbar sind. Als Informationsträger

dieser qualitativen Daten werden "Carnet de terrain", Manuskripte, Photographien, Skizzen und Zeichnungen, Karten, Audioaufnahmen, etc. aufgeführt. Sie werden in drei Kategorien unterteilt: Texte, Bilder und Tonaufnahmen (siehe Contexte, Corpus - Infrastructure de recherche [online], 2011-2012).

Die Definition des Projekts CORPUS IR wirft die Frage auf, ob die Unterscheidung zwischen quantitativen und qualitativen Daten den Artefakten und "Datensätzen" in den Geisteswissenschaften gerecht werden kann. Während eines Workshop, welcher im Rahmen des JISC Managing Research Programms durchgeführt wurde, fand eine Diskussion rund um den Forschungsdatenbegriff in den "Arts and Humanities" statt (Molloy, 2012). Die Teilnehmer des Workshops waren sich einig, dass der Begriff der Daten unterschiedliche Sachen in unterschiedlichen Kontexten bedeutet. Vieles, was als Datum bezeichnet werden könnte, ist in den Geisteswissenschaften eher sekundär als primär. Dies bedeutet, dass nicht neue Fakten erzeugt, sondern bestehende Fakten gesammelt werden. Deshalb wird in einigen Projekten versucht, einen anderen Begriff als Forschungsdaten zu benutzen. Es wird beispielsweise von der Materialisierung von Forschung gesprochen.

In der Schweiz definiert die Schweizerische Akademie der Geistes- und Sozialwissenschaften (SAGW) primäre Forschungsdaten als digitale Datenobjekte, welche aus einem Forschungsprojekt hervorgegangen sind und meistens in der Form einer Datenbank gespeichert werden (SAGW, 2010). Diese Sichtweise geht von Datenobjekten aus, die auch, aber nicht nur, in numerischer Form vorliegen können. Diese Definition ist eher als vage zu beurteilen, denn alles Digitale wird zum Datenobjekt.

Im Folgenden sollen nun konkrete Beispiele für Forschungsdaten in den Geisteswissenschaften zusammengetragen werden, die in der wissenschaftlichen Literatur gefunden werden konnten. Christine Borgman (2007:216) schlägt vor, dass in den Geisteswissenschaften jedes Dokument, jedes physikalische Artefakt und jede Aufnahme von menschlicher Aktivität als Datenquelle zu bezeichnen ist. Weitere Autoren gehen noch weiter ins Detail und zählen konkrete Dokumente auf, die als Forschungsdaten definiert werden können (Moulin et al., 2011:12; Delaunay, 2012:11; Pempe, 2012:142). Nachkommende Dokumente wurden dabei erwähnt (Auflistung in alphabetischer Reihenfolge):

- Artikel,
- Audio-Aufzeichnungen,
- Berichte,
- Broschüren,
- Bücher,
- Digitalisate,
- Diplomarbeiten,
- Disketten,
- Dokumentation,
- Einladungen,
- Entwürfe,
- Film,
- Flugblätter,
- Fotografien,
- Gemälde,
- graue Literatur,
- Hefte,
- Karteikarten,
- Korrespondenz,
- Laborhefte,
- Lehrbücher,
- Magnetbänder,
- Manuskripte von Artikeln und Werken,
- Monographien ,
- Musik,
- Neuauflagen von Artikeln und Monographien,
- Notizen,
- Papiere,
- Pre-Prints,

- Presseschau,
- Proben,
- Protokolle der Forschung,
- Protokolle von Sitzungen,
- Rohdaten von Texten (OCR),
- Scans,
- Skulpturen,
- Textdaten des Forschers mit inhaltlichem Markup, meistens im XML-Format,
- Übersetzungsdossiers,
- Umfragedaten,
- Video-Aufzeichnungen,
- Volkszählungen,
- Werkzeuge,
- Zeichnungen,
- Zeitschriften,
- Zeitungen.

Diese Aufzählung erschwert aufgrund der Vielfältigkeit eine logisch einwandfreie Formulierung dessen, was ein Forschungsdatum ist, da es sich nicht nur um sehr heterogene Datentypen, sondern teilweise auch um Untertypen bzw. Teile von anderen Typen handelt, bspw. Artikel als Teilmenge von Zeitschriften.

In der eingangs erwähnten Studie der HTW Chur (Zimmermann, Pfister, 2008 (1+2)), auf welcher der Bericht der SAGW an das Staatssekretariat für Bildung und Forschung (SBF) aufbaut (Immenhauser, 2009:15), wurden geisteswissenschaftliche Forschende und Institutionen befragt, was sie als digitale Forschungsdaten erachten. Folgende Dokumente und Dateien wurden dabei erwähnt:

- Durchsuchbarer Text,
- Bilder,
- Bilder mit Texten,
- Audio und Videodateien,
- Datenbanken,
- Statistikdaten,
- digitale Zeichnungsdaten für Pläne,
- Daten für GIS (Geographisches Informationssystem).

In den Geisteswissenschaften kommt noch die Schwierigkeit hinzu, dass sie nicht ausschliesslich mit physikalischen Objekten umgehen, sondern auch mit abstrakten Einheiten wie Texten und Werken, welche sowohl aus konzeptuellen Entitäten als auch aus physikalischen Manifestationen bestehen. Ein Werk von Shakespeare kann unabhängig von einer bestimmten Edition studiert werden oder aber eine bestimmte Edition eines Werks von Shakespeare kann den Untersuchungsgegenstand darstellen (Burrows, 2011:181).

Von daher ist bereits der Umstand, in den Geisteswissenschaften von Forschungsdaten oder Datensätzen zu sprechen, eher verwirrend. Die Forschenden dieser Disziplinen assoziieren mit diesem Begriff meistens ausschliesslich quantitative Daten und fühlen sich dadurch nicht angesprochen. Wie bereits angetönt, sollte in diesem Zusammenhang ein anderer Begriff benutzt werden. Es wäre deshalb zu überlegen, in den Geisteswissenschaften von Forschungsprodukten anstatt von Forschungsdaten zu sprechen. Im Zusammenhang mit einer digitalen Forschungsinfrastruktur für die Geisteswissenschaften scheint es zudem wichtig, das Forschungsdatum bzw. -produkt nicht allzu einschränkend zu definieren. Bisher wurde noch kein vollständiges Verzeichnis aller Daten- und Dokumenttypen erstellt, welche in den Geisteswissenschaften als Forschungsprodukte gelten können. Es kann deshalb auch keine Wertung vorgenommen werden, was in den Begriff miteingeschlossen werden soll und was nicht. Diese Aufgabe kann nur von der jeweiligen Fachgemeinschaft übernommen und beantwortet werden.

In diesem Bericht wird deshalb der Begriff des Forschungsprodukts möglichst weit gefasst und stützt sich auf die allgemeine, zu Beginn dieses Kapitels zitierte Definition: Ein Forschungsprodukt besitzt einen semantischen Wert, liegt in analoger oder digitaler Form vor und ist während einem Forschungsprozess erstellt worden, um Eigenschaften von Personen, Dingen, Handlungen, Konzepten, Werken oder die Forschung selbst zu beschreiben. Forschungsprodukte umfassen somit alle in den vorherigen Listen erwähnten Daten und Dokumente.

Gemäss den oben zitierten Autoren gehören die verschiedensten Dokument- und Datentypen zu den Forschungsdaten der Geisteswissenschaften. Um Ordnung und Struktur in die grosse Anzahl an Informationstypen zu bringen, werden in der wissenschaftlichen Literatur Vorschläge für eine Typologie gemacht. Burrows (2011:182) identifiziert aufgrund diverser Modelle des Forschungszyklus‘ in den Geisteswissenschaften zwei Komponenten für die Forschungsdaten. Die erste Komponente sind Annotationen, Tags, Links, Assoziationen, Bewertungen, Kommentare, welche während des Forschungsprozesses erzeugt werden. Die zweite Komponente sind sogenannte Entitäten, auf welche sich die Annotationen beziehen, wie zum Beispiel Personen, Orte, Ereignisse, Konzepte, aber auch Werke, Texte, Kunstwerke und andere physikalische und digitale Objekte.

Ein weiterer Autor (Delaunay, 2012:11) unterteilt die Forschungsdaten in drei unterschiedliche Archivtypen (gemäss Le Brech, 2011):

- Die Archive, die aus der Forschungsarbeit hervorgegangen sind (im Labor, Feld, in der Bibliothek, in Archiven): Laborhefte, Korrespondenzen, Protokolle von Sitzungen, Notizen, Berichte, Dokumentation, graue Literatur, etc.
- Die Archive, welche die Forschungsresultate festhalten: Berichte, Protokolle der Forschung, Manuskripte von Artikeln und Werken, Proben, Preprints, Artikel, Monographien etc.
- Die Archive bezüglich der Rezeption der Forschungsresultate: Korrespondenz, Presseschau, Übersetzungsdossiers und Neuauflagen von Artikeln und Monographien.

Im Folgenden soll eine weitere in Betracht zu ziehende Typologie erörtert werden, welche aus vier Gruppen besteht: Input, Throughput, Output, Hilfsmittel.

- Input: Zum Input gehören alle (analogen und digitalen) Dokumente, Informationen und Daten, auf welche sich die Forschungsarbeit stützt, beispielsweise Quellen oder Sekundärliteratur.
- Throughput: Zum Throughput gehören alle (analogen und digitalen) Dokumente, Informationen und Daten, welche während des Forschungsprozess‘ produziert werden, aber nicht für die Veröffentlichung vorgesehen sind, beispielsweise Berichte oder Notizen.
- Output: Zum Output gehören alle (analogen und digitalen) Dokumente, Informationen und Daten, welche für die Publikation bestimmt sind, beispielsweise Forschungsergebnisse in Form von Artikeln oder Monographien.
- Hilfsmittel: In diese Kategorie gehören fachspezifische Hilfsmittel, welche beim Forschungsprozess benutzt werden, beispielsweise Quelleneditionen und Verzeichnisse.

Gerade die letzte Typologie macht den teilweise rekursiven bzw. selbstreferentiellen aber auch inkrementellen Prozess der geisteswissenschaftlichen, vor allem aber der historischen Produktion (im Gegensatz zum eher iterativ differenzierenden bzw. abgrenzenden Charakter der naturwissenschaftlichen Produktion) offensichtlich, weil die Ergebnisse der Forschung selbst wieder zum Ausgangsmaterial für weitere Forschung werden können (Sahle, 2008:66). Dies zeigt, dass eine

eindeutige Zuteilung der vorhandenen Materialen in eine der vorgeschlagenen Kategorien selten oder fast nie möglich ist.

4. Digitale Forschungsinfrastrukturen

4.1. Organisatorische Aspekte

4.1.1. Zielsetzungen einer digitalen Forschungsinfrastruktur

Als erstes sollen hier die unterschiedlichen Zwecke aufgelistet werden, welchen eine digitale Forschungsinfrastruktur dienen kann. Die erwähnten Gründe stammen von den Autoren Keller-Marxer (2008:14-15), Neuroth et al. (2007:273) sowie Ball (2012:2) und wurden hier zusammengetragen.

- Integrität der Forschungsprodukte: Der Erhalt der Integrität von Forschungsprodukten erlaubt eine Überprüfung der Forschungsergebnisse durch Dritte. Die Veröffentlichung von Forschungsprodukten und eine offene Kultur verhindern Datenfälschungen.
- Sekundärnutzung: Die Veröffentlichung von Forschungsprodukten und Werkzeugen ermöglicht es, dass sie auch für Dritte von Nutzen sein und Doppelspurigkeiten vermieden werden können.
- Langzeitarchivierung: Die Sicherung der Forschungsprodukte ermöglicht eine zeitlich unbegrenzte Aufbewahrung.
- Zitierbarkeit von Forschungsprodukten: Durch das Zuteilen von einheitlichen Metadaten und persistenten Identifikatoren können auch Forschungsprodukte und nicht nur Publikationen zitiert werden.
- Nachvollziehbarkeit politischer Entscheide: Werden politische Entscheide aufgrund von Forschungsergebnissen und -produkten getroffen, sollen diese Unterlagen für nachkommende Generationen zur Verfügung stehen, damit die Entscheide auch zu einem späteren Zeitpunkt noch nachvollziehbar sind.
- Wissenschaftsgeschichte: Die Dokumentation der Forschungsprozesse bietet eine weitere Grundlage für wissenschaftsgeschichtliche Forschung.
- Schnellere Produktion von Daten: Anhand schon bestehender Daten können Datensätze extrahiert und mit zur Verfügung gestellten Werkzeugen neue Daten generiert werden.
- Interdisziplinäre Zusammenarbeit: Eine bessere Sichtbarkeit der Forschungsprodukte erhöht die interdisziplinäre Zusammenarbeit, welche durch geeignete Infrastrukturen unterstützt und vereinfacht werden kann.
- Supportarbeiten: Die Wartung und Instandhaltung von Forschungsprodukten, aber auch von Werkzeugen kann durch eine Infrastruktur übernommen werden.
- Bessere Forschung: all die oben erwähnten Punkte führen potentiell zu besserer Forschungsqualität. Mit einer Forschungsinfrastruktur können sowohl die Forschungseffizienz als auch der akademische, wirtschaftliche und soziale Einfluss der Forschungsarbeit erhöht werden.

Die aufgelisteten Zwecke sind sehr unterschiedlicher Natur. Um die unterschiedlichen Motivationen, Interessen und Argumente für ein Teilen von Forschungsdaten und -produkten zu beschreiben, hat Borgman (2010:8) zwei charakterisierende Achsen festgelegt (siehe Abbildung 6). Obwohl die von ihr vorgeschlagene Klassifikation spezifisch auf das Teilen von Forschungsdaten ausgerichtet ist, kann

diese Einteilung gut auch auf die oben erwähnten Zwecke übertragen werden. Auf der vertikalen Achse wird festgelegt, ob eine Motivation zum Teilen von Daten eher einen Wert für die Öffentlichkeit oder für die Forschung darstellt. Auf der horizontalen Achse befinden sich die Interessen der Datenerzeuger und der Datenbenutzer, wobei eine Person durchaus beide Perspektiven vertreten kann. Borgman hat daraufhin die von ihr als vier Hauptmotivationen identifizierten Elemente auf diesem Diagramm platziert. Das erste Element betrifft die Einstellung, dass mit öffentlichen Mitteln erstellte Forschungsdaten auch der Öffentlichkeit zur Verfügung gestellt werden sollen (*public goods*). Diese Motivation besteht folglich auf dem Wert der Daten für die Öffentlichkeit und die Bereitstellung findet im Interesse der Datennutzer statt.

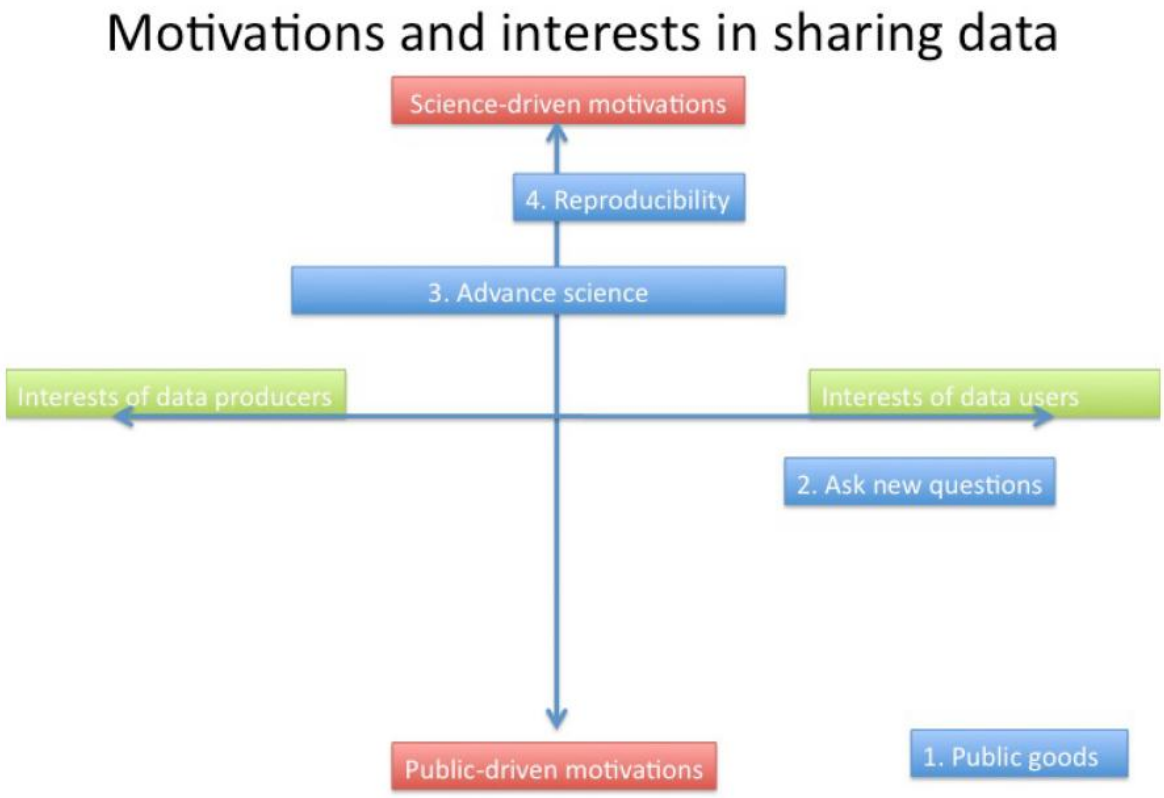


Abbildung 6: Motivations and interests in sharing data (Borgman, 2010:8)

Die zweite Motivation basiert auf der Sekundärnutzung, welche durch die Bereitstellung von Daten ermöglicht wird und welche zu Geldeinsparungen führen kann (*ask new questions*). Auch dieses Element befindet sich im Quadrant der Interessen der Datennutzer und der Öffentlichkeit, wobei es sich aufgrund der forschungsorientierten Nachnutzung schon näher bei den Forschungsmotivationen befindet als das Element des öffentlichen Gutes.

Das dritte Argument bezieht sich auf den möglichen schnelleren Fortschritt, der dank dem Veröffentlichen von Forschungsdaten erreicht werden kann (*advance science*). Dieses Argument bezieht sich jedoch vor allem auf datenintensive Forschungen, wo eine kritische Masse an Daten erreicht werden muss. Dieses Element zeugt von einem Wert der Daten für die Forschung, weshalb es in der oberen Hälfte des Diagramms abgebildet ist, und repräsentiert die Interessen der Datenproduzenten und -nutzer.

Das letzte Argument bezieht sich auf die Wiederproduzierbarkeit der Resultate und somit auf die Integrität der Forschungsdaten (*reproducibility*). Durch die Veröffentlichung von Forschungsdaten soll die Korrektheit von Forschungsergebnissen überprüfbar werden. Dieses Element ist klar von der Forschungsseite motiviert und entspricht eher den Interessen der Datennutzer, damit sie besser die Qualität einer Forschungsarbeit abschätzen bzw. überprüfen können.

Die Abbildung 6 bezieht sich ausschliesslich auf die Argumente für das Teilen von Forschungsdaten. Das grundlegende Diagramm scheint jedoch eine gute Basis darzustellen, um noch weitere Zwecke für die Erstellung einer digitalen Forschungsinfrastruktur einzuordnen. Die in der vorherigen Liste zusammengetragenen Argumente aus der wissenschaftlichen Literatur wurden deshalb von den Verfassern in die Graphik integriert und den beiden Achsen entsprechend platziert. Die aufgeführten Zwecke können danach eingeordnet werden, ob sie den Interessen der Forschung oder der Öffentlichkeit bzw. der Datenerzeuger oder der Datennutzer entsprechen. Diese Erweiterung wird auf der Abbildung 7 dargestellt.

Zwei der vier vorgestellten Argumente entsprechen weiter oben aufgelisteten Zwecken, und zwar die Sekundärnutzung (*ask new questions*) und die Integrität der Forschungsprodukte (*reproducibility*). Das Argument des Fortschritts der Forschung kann allgemeiner als von Borgman vorgeschlagen verstanden werden und würde somit dem Zweck der besseren Forschung gleichkommen. Die anderen aufgeführten Zwecke können aber ebenfalls auf diesem Diagramm eingezeichnet werden.

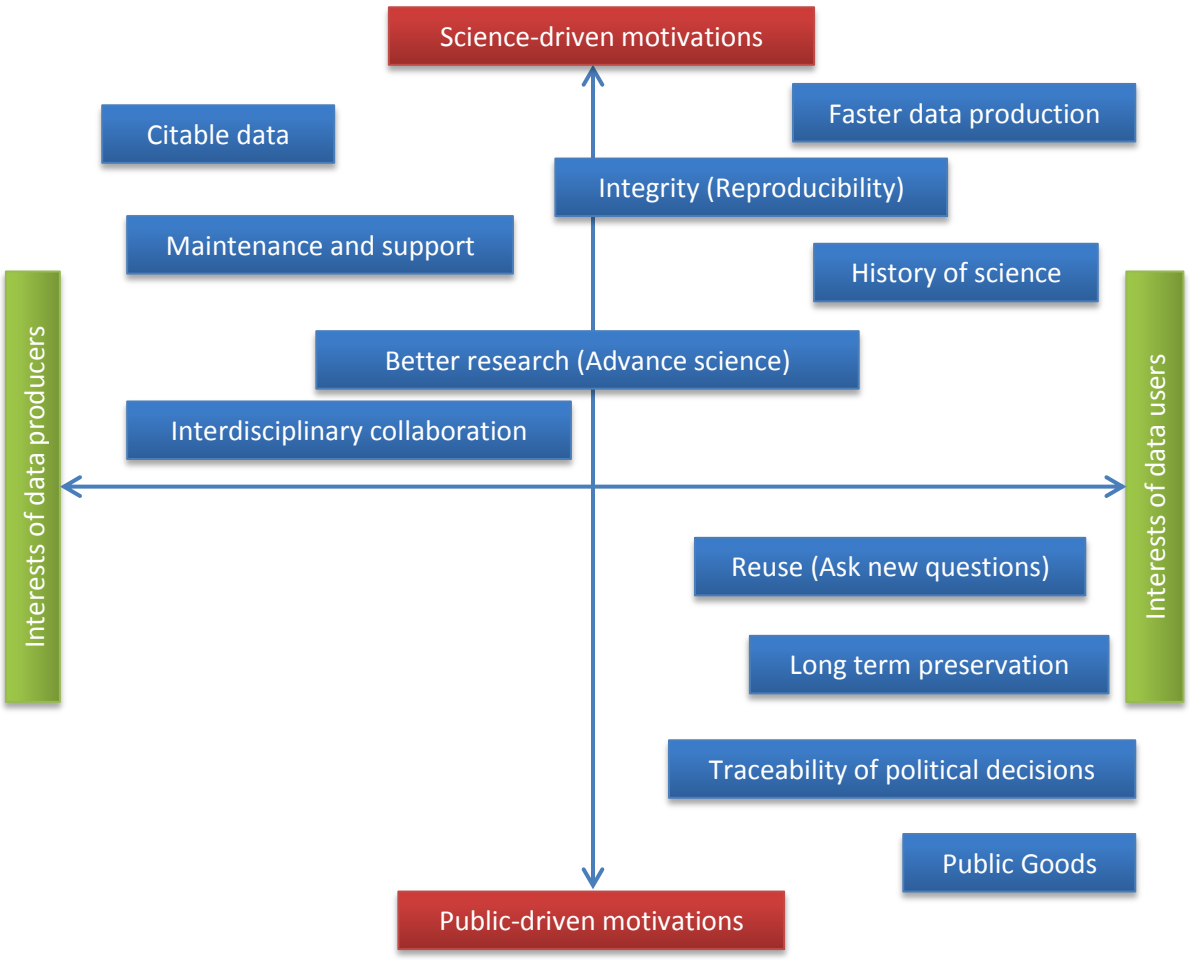


Abbildung 7: Purposes of digital research infrastructures (based on Borgman, 2010:8)

Nach dem Eintragen der restlichen Argumente in das Diagramm fällt als erstes auf, dass es keinen Zweck gibt, der sich im Quadrant Datenproduzenten-Öffentlichkeitsinteresse befindet. Die oben aufgelisteten Argumente sind zwar bestimmt nicht vollständig, greifen aber die wichtigsten Elemente auf. Ist davon auszugehen, dass tatsächlich kein Zweck gefunden werden kann, der sich in den Quadranten Datenproduzenten-Öffentlichkeitsinteresse einordnen lässt, dann ist dies so zu interpretieren, dass Datenerzeuger ihre Forschungsdaten nur aus persönlichem Interesse in eine Infrastruktur übergeben würden. Diese Eigeninteressen sind zwangsläufig forschungsorientiert. Die Schlussfolgerung liegt nahe, dass es schwierig sein wird, Forschende mit altruistischen Argumenten zu motivieren, ihre Forschungsdaten anderen zur Verfügung zu stellen.

Die Zwecke sind in etwa gleichmässig über die restlichen Quadranten verteilt.

Es stellt sich nun die Frage, welche dieser Zwecke auch für die Geisteswissenschaften von Interesse sind. Die einzelnen Funktionen hängen aber wiederum von der Definition von Forschungsdaten bzw. Forschungsprodukten ab. Der Zweck der Erhaltung der Integrität von Forschungsdaten ist beispielsweise bei quantitativen Daten von grösserer Wichtigkeit als bei qualitativen Daten.

In der wissenschaftlichen Literatur werden Zwecke, spezifischer aber noch Funktionen erwähnt, welche von einer geisteswissenschaftlichen Forschungsumgebung unterstützt werden sollten. Eine erste Vision einer solchen Infrastruktur sieht eine Umgebung, welche Instrumente, Daten, Informationen und Werkzeuge virtuell zur Verfügung stellt, so dass Forschende unabhängig von lokalen Ressourcen ihrer Forschung nachgehen können (Neuroth et al., 2007:273). Dies beinhaltet sowohl den virtuellen Zugang zu Programmen, als auch zu Forschungsdaten und Sekundärquellen. Dabei soll die Infrastruktur Aufgaben wie die Datenverwaltung, Langzeitarchivierung und semantische Vernetzung übernehmen.

Auch für die Geschichtswissenschaften wurden Überlegungen bezüglich der Zwecke einer digitalen Infrastruktur bereits unternommen. In dieser Disziplin sollen von einer Forschungsumgebung die Arbeitsprozesse der Datensammlung, -aggregation, -auswertung, Literaturbeschaffung, Kommunikation mit Kollegen und die persönliche Terminplanung unterstützt werden (Meyer, 2011 (2): 38). Weitere Vorteile einer digitalen Forschungsumgebung bestehen für die Historiker auch darin, dass neue methodische Ansätze begünstigt werden können. Dazu gehören beispielsweise Simulationen und virtuelle Rekonstruktionen historischer Ereignisse, Text Mining, semantische Analysen von Quellenkorpora oder bildverarbeitende Verfahren (Sahle, 2008:70).

Sahle (2008:66) unterscheidet für den Übergang zur eScience drei Bereiche in den Geschichtswissenschaften: 1) die Rohstoffe, d.h. historische Überlieferungen im weitesten Sinne und die bisherigen Ergebnisse der Forschung; 2) die Produktion, d.h. die Auswertung der Rohstoffe, welche neue wissenschaftliche Ergebnisse erzeugt; und 3) die Produkte, d.h. Erschliessungsleistungen als Antwort auf Forschungsfragen. Eine digitale Forschungsinfrastruktur sollte dazu dienen, diese durch die Druckkultur stark abgegrenzten drei Bereiche zu integrieren und die Rohstoffe, Produktion und Produkte auf einer Umgebung zusammenzuführen.

4.1.2. Ausrichtung

Eine digitale Forschungsinfrastruktur kann folglich vielen verschiedenen Zwecken gewidmet sein. Dabei stellt sich die Frage, ob all diese Zwecke durch eine einzige Infrastruktur übernommen werden können oder sollen. Bereits existierende digitale Forschungsinfrastrukturen erheben selten den Anspruch, allen Zwecken zu genügen. Zusätzlich dazu werden die einzelnen Argumente von unterschiedlichen Disziplinen auch unterschiedlich gewichtet. Die Gründe für die Erstellung einer

Infrastruktur hängen demnach sowohl von der Disziplin, den in der Disziplin angewendeten Methoden, als auch von der finanzierenden Instanz ab.

Digitale Forschungsinfrastrukturen können verschiedene Ausrichtungen und damit verbunden unterschiedliche Charakteristiken aufweisen. Eine möglichst ausführliche Auflistung haben Blinco und McNeal (2004) mit dem „Glücksrad“ für Repositorien erstellt (siehe Abbildung 8). Dieses Glücksrad besteht aus mehreren Scheiben, welche jeweils einen Aspekt eines Repositoriums repräsentieren. Auf der Scheibe des jeweiligen Aspekts befinden sich die unterschiedlichen Ausprägungen, welche ein Aspekt haben kann. Die Abbildung ist so zu interpretieren, dass jede Scheibe des Rades beweglich ist und beliebig mit den anderen Scheiben kombiniert werden kann.

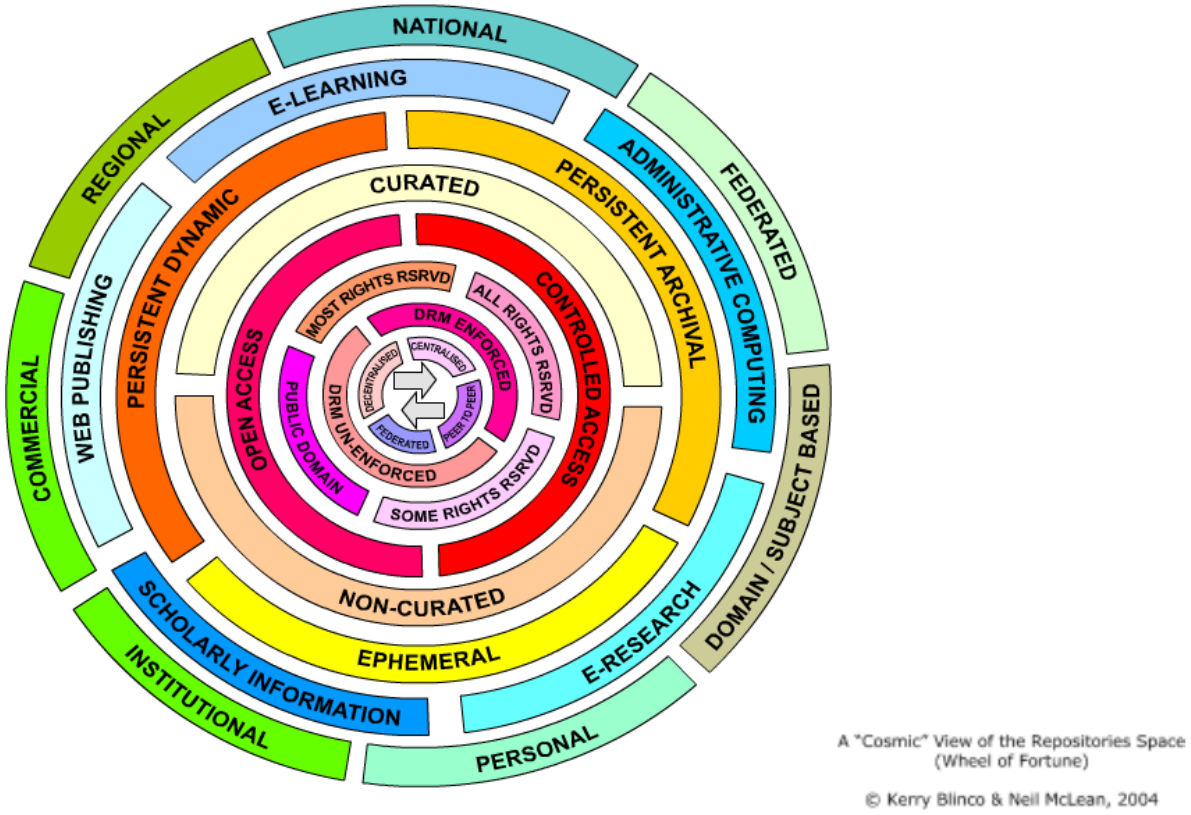


Abbildung 8: A "Cosmic" View of the Repositories Space (Blinco, McLean, 2004)

Von innen nach aussen kann eine Infrastruktur mit folgenden Aspekten beschrieben werden:

- **Datenbankstruktur**: Ein Repositorium kann zentralisiert, dezentralisiert, föderiert bzw. verteilt oder in einem peer-to-peer Netzwerk organisiert sein.
- **Digital Rights Management**: Inhalte können durch DRM-Mechanismen geschützt werden oder nicht.
- **Nutzungsrechte**: Es können alle Rechte, die meisten Rechte oder nur einige Rechte vorbehalten sein bzw. der ganze Inhalt kann in der Public Domain liegen.
- **Zugangskontrolle**: Es kann gar keine Zugangskontrolle stattfinden und die Inhalte liegen als Open Access vor oder es muss eine Erlaubnis für den Zugang eingeholt werden.
- **Datenpflege**: Die Inhalte eines Repositoriums können kuratiert werden oder nicht.

- **Aufbewahrung**: Die Inhalte können von kurzer Lebensdauer sein, dynamisch sein und verändert bzw. ergänzt werden, oder statisch für eine längerfristige Aufbewahrung bestimmt sein.
- **Zweck**: Die Autoren haben fünf Zwecke für ein Repositorium identifiziert. Es kann eine E-Learning-Plattform sein, für die administrative Datenverwaltung bestimmt sein, die E-Research unterstützen, wissenschaftliche Informationen anbieten oder der Online-Publikation dienen.
- **Reichweite**: Ein Repositorium kann regional ausgerichtet sein, beispielsweise für einen bestimmten Kanton, es kann von nationaler Reichweite sein, es kann föderiert sein. Des Weiteren gibt es fachspezifische Repositorien, aber auch persönliche Repositorien. Nicht zuletzt kann eine solche Infrastruktur auch für eine Institution wie beispielsweise eine Universität entwickelt werden oder für die kommerzielle Nutzung gedacht sein.

Diese Abbildung beschreibt sehr ausführlich alle Aspekte, welche eine Forschungsinfrastruktur charakterisieren können. Es wird ersichtlich, dass dies eine grosse Anzahl möglicher Kombinationen darstellt. Wird diese Abbildung im Zusammenhang mit Repositorien in den Geisteswissenschaften betrachtet, fehlt jedoch eine Scheibe, und zwar bezüglich der Natur der zu integrierenden Datenobjekte, welche physikalisch bzw. analog, digitalisiert oder *born digital*, also ‚nativ digital' erstellt sein können.

Soll eine neue Forschungsinfrastruktur entwickelt werden, ist es wichtig, auf jeder dieser Ebenen eine bewusste Entscheidung zu treffen und den Zweck des zu entwickelnden Repositoriums klar zu bestimmen. Soll eine Forschungsinfrastruktur mehrere Aspekte derselben Ebene integrieren, erhöht das automatisch die Komplexität der zu entwickelnden technischen Lösung.

Die Beschreibung der unterschiedlichen Ebenen dieses „Glückrads" ist so allgemein gehalten, dass sie in jedem Fachbereich angewendet werden kann, sowohl in den Geistes- wie auch in den Geschichtswissenschaften. Aus dieser Perspektive wird es eindeutig klar, dass es bereits Repositorien und Infrastrukturen gibt, welche in dieses Schema fallen, u.a. kommerzielle Datenbanken, institutionelle Repositorien und sogar private Archive. Andererseits wird deutlich, dass es wohl kaum zu einer Lösung kommen kann, die alle möglichen Ausprägungen umfassen kann.

4.1.3. Zeitlich begrenzte vs. unbegrenzte Aufbewahrung

Bei der Aufbewahrung von Forschungsprodukten können zwei Arten von Archivierung unterschieden werden, und zwar die zeitlich begrenzte und die zeitlich unbegrenzte Aufbewahrung. Bei der zeitlich begrenzten Aufbewahrung werden die erhaltenen Forschungsprodukte nach einer vorher festgelegten Dauer gelöscht, beispielsweise nach 10 Jahren. Die zeitlich unbegrenzte Aufbewahrung ist gleichzusetzen mit der Langzeitarchivierung, mit welcher die Sicherung der Forschungsdaten über Generationen hinweg angestrebt wird.

In der für das Grossprojekt e-lib.ch entwickelten Konzeptstudie besteht der Autor darauf, dass der Zweck der Aufbewahrung von Forschungsdaten die Ansprüche an die Archivierung bestimmt (Keller-Marxer, 2008:19). Zu den zwei Hauptzwecken werden der Erhalt der Integrität der Forschungsdaten zur Überprüfung von Forschungsergebnissen durch Dritte, sowie die Sekundärnutzung gezählt. Dabei wird jede Nutzung der Primärdaten zu anderen Zwecken, als diejenigen, die ursprünglich vorgesehen waren, als Sekundärnutzung bezeichnet.

Geht es darum, die Integrität der Forschungsdaten zu erhalten, müssen die Daten beispielsweise in ihrem ursprünglichen Format aufbewahrt werden. Des Weiteren werden die Forschenden für diesen

Zweck häufig zu einer Aufbewahrungsdauer von 10 Jahren verpflichtet, welches in der Schweiz der Aufbewahrungsfrist in der Geschäftsbücherverordnung (SR 221.431) des Obligationenrechts entspricht (Keller-Marxer, 2008:5). Steht die Sekundärnutzung im Mittelpunkt im Unterschied zur wissenschaftlichen Integrität, dann müssen die Daten auf Dauer aufbewahrt werden, was höhere Ansprüche an die Dokumentation und die Sicherung stellt. Doch die langfristige, sogar unbegrenzte Aufbewahrung von Primärdaten bedeutet, dass diese – dem OAIS-Referenzmodell folgend - in ein langzeittaugliches Format konvertiert werden müssen (Keller-Marxer, 2008:8). Diese Konversion führt fast zwangsläufig zu Informationsverlusten, da ein langzeittaugliches Format nicht dieselbe Struktur oder Genauigkeit hat, welche ein fachspezifisches Format bietet. Diese Informationsverluste sind zwar für das Archiv akzeptabel und ermöglichen die Sekundärnutzung, doch für den Zweck des Erhalts der Integrität der Forschungsdaten ist dies nicht zufriedenstellend.

Es stellt sich die Frage, wer entscheidet, was zeitlich begrenzt und was unbegrenzt aufbewahrt werden soll. Den Empfehlungen der OECD folgend schlägt die Konzeptstudie vor, dass für die Erhaltung der Integrität die Forschenden verantwortlich sind und dass sie deshalb auch diejenigen sind, welche bestimmen, welche ihrer Daten aufbewahrt werden sollen (Keller-Marxer, 2008:19). Der Entscheid für eine langfristige Archivierung soll jedoch nicht beim Forschenden liegen, sondern von der archivierenden Instanz getroffen werden. Diese Instanz kann je nach Kontext der Forschungsfinanzierer oder der Forschungsträger sein, es handelt sich demnach letztlich um eine institutionelle Aufgabe (Keller-Marxer, 2008:20).

Im Folgenden wird die Aufbewahrungslösung vorgestellt, welche im Rahmen des deutschen Projekts TextGrid gewählt wurde (Pempe, 2012:138; siehe Kapitel 4.4.1). Obwohl das Projekt zuerst auf Textwissenschaften fokussiert war, ist die errichtete virtuelle Forschungsumgebung auch für andere geisteswissenschaftliche Fachrichtungen nutzbar. Diese Forschungsumgebung enthält ein TextGrid Repository, welches in einen dynamischen und in einen statischen Speicherbereich unterteilt ist. Für die Aufbewahrung der darin ablegbaren Forschungsdaten wurden Dienstleistungsvereinbarungen (Service Level Agreements) und Richtlinien entwickelt. Insgesamt werden Service-Levels auf vier unterschiedlichen Ebenen angestrebt Pempe (2012:147):

- Einfache Langzeitarchivierung (*bitstream preservation*) mit Offline Speicherung;
- Einfache Langzeitarchivierung mit online Zugriff;
- Langzeitarchivierung mit Bewahrung der technischen Nachnutzbarkeit (*content preservation*);
- Langzeitarchivierung mit Datenpflege (*data curation*).

Die einfachste archivierende Funktion ist die *Bitstream Preservation*. Dabei wird bei einem digitalen Objekt darauf geachtet, dass es in seiner Integrität erhalten bleibt, ohne aber die dazugehörende verarbeitende Soft- oder Hardware beizubehalten. Dabei kann noch unterschieden werden, ob das Objekt offline gespeichert wird oder ob ein online Zugriff darauf garantiert wird. Auf einer höheren Ebene wird die *Content Preservation* unternommen. Dabei wird neben der *Bitstream Preservation* auch die technische Nachnutzbarkeit sichergestellt. Das bedeutet, dass die digitalen Objekte in offene Formate konvertiert und wenn nötig migriert werden, um die Lesbarkeit des Objekts auf Dauer zu garantieren. Auf dem höchsten Niveau der Langzeitarchivierung befindet sich die *Data Curation*, welche sich zusätzlich zur *Bitstream Preservation* und der *Content Preservation* auch um die kontinuierliche Verlinkung von Daten zu verwandten Daten und deren Verbreitung kümmert.

Für die zeitlich begrenzte Aufbewahrung von Forschungsdaten würden die ersten beiden Stufen der hier vorgeschlagenen Service-Levels ausreichen, d.h. die einfache Langzeitarchivierung mit oder ohne online Zugriff. Bei diesen Niveaus kann auch das Kriterium der Integrität der Forschungsdaten erfüllt

werden. Für eine zeitlich unbegrenzte Aufbewahrung, welche den zwei anderen Service-Levels entspricht, muss die volle Verantwortung über die Forschungsdaten der archivierenden Institution übertragen werden. Denn für Konversionen und Migrationen benötigt das Archiv die entsprechenden Rechte, um Aspekte bezüglich der Form oder der Struktur der Daten verändern zu dürfen (Keller-Marxer 2008:46). Zusätzlich dazu muss die archivierende Institution in der Lage sein, die Daten unabhängig vom Datenproduzenten interpretieren zu können, um eine Langzeitarchivierung garantieren zu können.

Die Konversion von Forschungsdaten in ein einheitliches Standardformat ermöglicht zwar eine zeitlich unbegrenzte Aufbewahrung, ist aber nicht unbedingt erwünscht. Dies liegt einerseits daran, dass es wenig fachübergreifende Standardformate gibt, und andererseits daran, dass die Sekundärnutzung auch im fachspezifischen Format stattfindet (Keller-Marxer, 2008:21). Vor diesem Hintergrund sollten die jeweiligen Bedürfnisse, welche auf eine zeitlich begrenzte bzw. unbegrenzte Aufbewahrung schliessen lassen, genau abgeklärt werden. Des Weiteren bleibt die Frage bestehen, ob diese beiden Aufgaben von einer einzigen, oder aber von zwei unterschiedlichen Infrastrukturen übernommen werden sollte.

Die Studie von iKeep (Keller-Marxer, 2008:45) schlägt eine zweiteilige Lösung vor und zwar in Form von einem Depositorium und einem Repositorium. Das Repositorium ermöglicht es, dass Daten fachspezifisch und häufig genutzt werden, thematisch geordnet und die Dateninhalte durchsucht werden können. Ein Depositorium dagegen dient der Aufbewahrung und langfristigen Sicherung der Verfügbarkeit von Daten, welche eher selten und nicht fachspezifisch genutzt werden, eher homogen und nicht thematisch geordnet sind. Die Dateninhalte können dabei nicht durchsucht werden, sondern nur die Metadaten. Um auf die Inhalte zugreifen zu können, müssen die Datensätze zuerst aus dem Depositorium exportiert werden.

4.1.4. Kosten und Finanzierung

Wie in den vorangehenden Kapiteln verdeutlicht wurde, gibt es nicht eine digitale Forschungsinfrastruktur, welche für eine oder mehrere Disziplinen entwickelt werden kann. Die Kosten und auch die Finanzierungsmodelle hängen demnach stark von den verfolgten Zwecken, der gewählten Ausrichtung sowie dem anzubietenden Dienstleistungsangebot ab. Zusätzlich dazu ist es schwierig abzuschätzen, wie stark eine zu entwickelnde Infrastruktur genutzt werden wird und wie gross die zu verwaltende Datenmenge tatsächlich sein wird.

In der Konzeptstudie für eine zentrale Langzeitarchivierung in der Schweiz hat sich der Verfasser zu keinem seiner Modelle finanziell äussern wollen (Keller-Marxer, 2009:48,58,67). Im Bericht der Schweizerischen Akademie für Geistes- und Sozialwissenschaften (SAGW) werden für die beiden Massnahmenbereiche Data Repository für die Geisteswissenschaften und Vernetzung existierender Infrastrukturen Finanzmittel in der Höhe von 850'000 bis 900'000 CHF jährlich beansprucht (Immenhauser, 2009:4; Details auf Seiten 24-26 und 33-34).

Nicht zu vernachlässigen sind bei der Dienstleistung der Aufbewahrung von Forschungsdaten die Versicherungskosten für Datenverlust, wofür die aufbewahrende Institution je nach Vertrag mit den Forschenden haftet (Pempe, 2012:154).

4.2. Technische Aspekte

4.2.1. Persistente Identifikatoren

Das nachstehende Kapitel resümiert grösstenteils einen Artikel von Emma Tonkin (2008). Persistente Identifikatoren sind aufrechterhaltene Kennzeichen, welche auf ein digitales Objekt verweisen. Das digitale Objekt kann einer oder mehrerer Datei(en) entsprechen und gilt sowohl für Texte, Bilder, Video- und Tonaufnahmen, welche in digitaler Form vorliegen, als auch beispielsweise für ausführbare Dateien. Interessante Identifikatoren sind solche, die auch persistent ausführbar sind, was meint, dass sie wie ein Hyperlink funktionieren und durch Klicken zu dem digitalen Objekt führen. Im Unterschied zum Hyperlink soll ein persistenter Identifikator auch noch dann Zugang zur Ressource bieten, wenn die Datei auf einem anderen Server gespeichert wird oder den Namen ändert. Doch auch ausführbare Identifikatoren sind anfällig dafür, nicht zu funktionieren, vor allem wenn der Identifikator aus dem Namen des Servers besteht. Eine Möglichkeit, diese Gefahr zu umgehen, besteht darin, eine Zwischenschicht zwischen dem Browser und dem Zielobjekt hinzuzufügen. Dabei verweist der Identifikator nicht direkt auf das Objekt, sondern auf eine Beschreibung des Objekts. Indirekte Identifikatoren benötigen einen Resolver (DNS-Auflöser), welcher auf die aktuelle Version des Objekts weiterleitet (Tonkin, 2008).

Allgemein ist festzustellen, dass die Technologie alleine keine persistenten Identifikatoren garantieren kann, denn diese Charakteristik ist vom langfristigen Engagement des jeweiligen Anbieters abhängig. Soll ein persistentes Kennzeichen für ein Objekt via Resolver angeboten werden, muss zusätzlich noch eine langfristige Bindung mit der erhaltenden Organisation des Resolvers eingegangen werden.

Im Folgenden sollen die ausgereiftesten Standards für persistente Identifikatoren vorgestellt werden (Tonkin, 2008).

URN

Ein Standard, welcher den nötigen Reifegrad erreicht hat, um als persistenter Identifikator zu dienen, ist der URN (Uniform Resource Name). Dieser Standard wurde entwickelt, um eine Identität und weniger um eine Lokalität zu beschreiben. So kann ein URN beispielsweise eine ISBN-Nummer enthalten. Um die Zuteilung von URN-Namensräumen kümmert sich IANA, die Internet Assigned Numbers Authority (Tonkin, 2008).

National Bibliography Numbers

Der National Bibliography Numbers (NBN) ist ein Standard, welcher auf dem URN-Standard basiert und nur von Nationalbibliotheken benutzt wird. Er wird verwendet, falls kein anderer Identifikator als eine ISBN-Nummer zur Verfügung steht. Dabei können sowohl digitale sowie analoge Dokumente mit einer NBN versehen werden. Ist keine digitale Version vorhanden, werden bei der NBN die Metadaten hinterlegt (Tonkin, 2008).

The digital object identifier (DOI)

Der Digital Object Identifier (DOI)-Standard, welcher von der International DOI Foundation kontrolliert wird, ist ein indirekter Identifikator, der auf einem Handle Resolver aufbaut. DOIs bestehen aus zwei Sektionen: einer numerischen Identifikation bestehend aus einem Präfix, das den Identifikator als DOI kennzeichnet, und aus einem Suffix, welches den Anbieter (Host) des Dokuments bestimmt. Es folgt eine Zeichenfolge, welche ein bestimmtes

Dokument identifiziert. Die Registrierung eines DOIs generiert Kosten sowohl für die Mitgliedschaft als auch für die Registrierung eines bestimmten Dokuments (Tonkin, 2008).

Bezüglich der DOI-Attribution an Forschungsdaten wurde 2009 das DataCite-Konsortium gebildet (http://datacite.org/). In der Schweiz ist für alle Organisationen und Institutionen bislang die ETH Zürich der Ansprechpartner für die Registrierung der DOIs von Primär- und Sekundärdaten.

Persistent Uniform Resource Locators (PURL)

Ein Persistent Uniform Resource Locator (PURL) besteht aus einer URL und verweist auf einen Resolver, welcher die URL nachschlägt. PURLs sind mit URNs kompatibel, weshalb sie manchmal als Interim-Lösung angesehen wurden, bevor die Benutzung von URNs sich weit verbreitete. PURL ist mit OCLC stark verbunden, welche den Standard entwickelt hat und den ältesten PURL-Resolver zur Verfügung stellt. Bei der Benutzung von PURL entstehen keine Kosten (Tonkin, 2008).

Das Handle System

Das Handle System wird vorwiegend für die Aufschlüsselung von DOIs benutzt, wird aber als Mittel angesehen, welches generell für die Identifikation von digitalen Objekten oder Auflösung von Identifikatoren verwendet werden kann. Das System wird von der Corporation for National Research Initiatives (CNRI) verwaltet und von der Defense Advanced Research Projects Agency (DARPA) unterstützt (Tonkin, 2008).

The Archival Resource Key (ARK)

Der Archival Resource Key ist ein URL Schema, welches von der US National Library of Medicine entwickelt wurde und von der California Digital Library betrieben wird. ARKs können benutzt werden, um sowohl digitale wie auch analoge Objekte zu identifizieren und auf deren Beschreibung zuzugreifen. Der Identifikator nach ARK enthält eine numerisch eindeutige Bezeichnung der namensgebenden Autorität (*Name Assigning Authority Number*) des Objekts und eine Kennzeichnung für die Autorität (*Name Mapping Authority*), welche als Service-Provider (bspw. eine Website) für das Objekt verantwortlich ist. Die Name Mapping Autorität erlaubt es, den Identifikator anklickbar zu machen und ist deshalb auch flexibel: wenn ein Objekt auf einem anderen Server gespeichert wird, ändert sich die *Name Mapping Authority*, während die *Name Assigning Authority Number* und der Name des Objekts immer dieselben bleiben (Tonkin, 2008).

4.2.2. Lizenzen

Das folgende Kapitel basiert hauptsächlich auf einer Anleitung von Alex Ball (2012), welche für das Digital Curation Center in Grossbritannien erstellt wurde. Im Bericht werden unterschiedliche Lizenzen vorgestellt, welche für Forschungsdaten in Frage kommen können. Bei den Forschungsdaten wird davon ausgegangen, dass sie in Form von Datensätzen in einer Datenbank gespeichert sind. Dabei stellt sich generell die Frage der Granularität, da einerseits die Daten, die Datensätze und die Datenbanken jeweils mit einer Lizenz versehen werden können. Wenn Forschungsdaten veröffentlicht werden sollen, sollte klar sein, unter welchen Bedingungen sie benutzt werden können. Gesetzgebungen sind diesbezüglich sehr komplex, vor allem auch weil unterschiedliche Aspekte einer Datenbank unterschiedlich behandelt werden können.

In den USA gibt es eine starke Fokussierung auf den Aspekt der Kreativität. Datenbanken, welche beispielsweise Messdaten beinhalten, sind nicht vom Copyright betroffen. In Australien dagegen ist die Originalität wichtiger als die Kreativität. Doch eine Datenbank, welche für einen originellen Bericht erstellt wurde, wird nicht an sich als originell betrachtet. Unter europäischem Recht ist das Copyright für Datenbanken geltend, sofern der Ersteller eine intellektuelle Arbeit geleistet hat, um die Daten auszuwählen oder zu organisieren. Doch auch wenn eine Datenbank durch Copyright geschützt ist, bleibt die Frage, was gemacht werden darf und was nicht, um das Copyright nicht zu verletzen (Ball, 2012:3).

Die effektivste Methode, um Nutzungsrechte zu kommunizieren, sind Lizenzen. Eine Lizenz wird als legales Instrument verstanden, mit welchem ein Rechteinhaber einer zweiten Partei die Erlaubnis erteilt, Dinge zu tun, die sonst bestehende Rechte verletzen können. Eine Lizenz kann ausschliesslich vom Rechteinhaber gewährt werden. Der Rechteinhaber muss folglich klar identifiziert sein und eventuelle Unklarheiten rechtlich geregelt werden, bevor eine Lizenz erteilt werden kann (Ball, 2012:3).

Creative Commons

Creative Commons Lizenzen bieten den Autoren eine Möglichkeit, einen Zwischenweg zwischen dem Vorbehalten aller Rechte und dem Rechteverzicht einzuschlagen. Aufgrund von vier unterschiedlichen Bedingungen werden insgesamt sechs Hauptlizenzen der Creative Commons abgeleitet. Jede dieser Lizenz enthält die Bedingung der Nennung des Urhebers [Bedingung 1]. Die restlichen Bedingungen können kombiniert werden und enthalten [2] die Untersagung kommerzieller Nutzung des lizenzierten Inhalts, [3] die Veröffentlichung von Derivaten unter derselben Lizenz (auch Copyleft genannt) und [4] die Untersagung der Erstellung von Derivaten des lizenzierten Inhalts. Die Bedingungen [3] und [4] schliessen sich gegenseitig aus und können deshalb nicht in derselben Lizenz angewendet werden (Ball, 2012:6).

Die Creative Commons Lizenzen wurden nicht spezifisch für Forschungsdaten entwickelt, was im Hinblick auf Datenbanken zu Schwierigkeiten führt. Besonders der Unterschied zwischen den individuellen Daten und der ganzen Datenbank, bzw. der Unterschied zwischen der Benutzung der Daten als Teil einer neuen Datenbank und der Benutzung der Daten, um sie zu visualisieren, kann durch die Creative Commons Lizenzen nicht zufriedenstellend beachtet werden (Ball, 2012:6).

Wenn Forschende mehrere Datenbanken miteinander kombinieren wollen, kann es ab einem gewissen Punkt zur administrativen Herausforderung werden, alle Autoren der Daten zu zitieren, wie es die Bedingung der Nennung des Urhebers [1] verlangt.

Auch die Bedingung, dass die Derivate von entsprechend lizenzierten Datensätzen unter derselben Lizenz veröffentlicht werden müssen [3], kann zum Problem werden, weil sie die Kombination dieser Datensätze mit solchen unter einer anderen Lizenz verhindert. Das Derivat könnte nicht beiden Lizenzen entsprechen.

Die Bedingung, dass gar keine Derivate erstellt werden dürfen und dass die lizenzierten Datensätze so benutzt werden sollen, wie sie sind [4], steht zur Debatte. Bezüglich eines Buchs gilt beispielsweise die Übersetzung ein Derivat. Es ist jedoch schwierig zu beurteilen, was bei einer Datenbank als Derivat bezeichnet werden kann und was nicht.

Auch die Bedingung der nicht-kommerziellen Nutzung [2] lässt Raum für Interpretation. Es ist nicht klar, ob Datensätze nicht verwendet werden dürfen, wenn der Autor sie in einem Textbuch oder in einem wissenschaftlichen Artikel verwendet und dafür ein Entgelt erhält (Ball, 2012:7).

Open Data Commons

Das Open Data Commons Projekt, welches im Jahr 2009 in die Open Knowledge Foundation übertragen wurde, hat drei spezifische Lizenzen für Datenbanken entwickelt. Die erste Lizenz ist die Open Data Commons Attribution Licence (ODC-By), welche es erlaubt, eine Datenbank zu kopieren, zu verbreiten, zu benutzen und sie zu modifizieren. Wenn darauf basierend neue Werke oder Datenbanken entstehen, dann soll die benutzte Datenbank benannt werden. Eine weitere Lizenz, die Open Data Commons Open Database (ODC-ODbL), baut auf der ODC-By – Lizenz auf und fügt weitere Bedingungen hinzu. Hinzu kommt die Bedingung des Copylefts für auf der lizenzierten Datenbank aufbauende neue Datenbanken. Des Weiteren dürfen die neu erzeugten Datenbanken nur mit Digital Rights Management (DRM)-Mechanismen versehen werden, wenn eine alternative Kopie ohne Einschränkungen zur Verfügung gestellt wird. Die ODbL-Lizenz kann in Kombination mit der Open Data Commons Database Contents Licence (ODC-DbCL) benutzt werden, um auf die Rechte der Inhalte der Datenbank zu verzichten. Da diese Lizenzen spezifisch für Datenbanken entwickelt wurden, eignen sie sich besser für die Lizenzierung von Forschungsdaten als die Creative Commons (Ball, 2012:8).

Design Science Licence

In den Jahren 1999-2001 entwickelte Michael Stutz die Design Science Licence, welche sich auf Inhalte konzentriert, die einen Unterschied zwischen der Quelle und der Ausgabe des Inhalts machen, wie das beispielsweise bei LATEX-Dokumenten oder Software der Fall ist. Die Lizenz erfordert, dass die Lizenzinformationen mit dem Dokument verteilt werden und dass bei abgeleiteten Werken die Urheberschaft der originalen Teile klar gekennzeichnet ist. Die Bedingung des Copylefts besteht auch hier und es wird zusätzlich dazu verlangt, dass die abgeleiteten Werke einen neuen Titel erhalten. Im Kontext von Forschungsdaten kann die Unterscheidung zwischen Quelle und Ausgabe des Inhalts auf die Quelldaten und die Visualisierungen als Graphen oder Karten übertragen werden. Die Lizenz ist jedoch nicht speziell auf Datenbanken ausgerichtet und macht deshalb auch keinen Unterschied zwischen den Daten und der Datenbank (Ball, 2012:10).

Public Domain - Gemeinfreiheit

Nach deutschem, österreichischem und schweizerischem Recht kann jemand nicht gänzlich auf seine Urheberrechte verzichten, wie es beispielsweise in den USA möglich ist. Es ist jedoch möglich, ein Werk mit einer Lizenz zu versehen, welches eine uneingeschränkte Nutzung erlaubt (Gemeinfreiheit, Wikipedia, 2012). Dafür gibt es die Creative Commons Zero Lizenz, welche einerseits einen Rechteverzicht und eine bedingungslose Lizenz beinhaltet, falls der Rechteverzicht im jeweiligen Land nicht gültig sein sollte. Die Open Data Commons Public Domain Dedication and Licence (PDDL) macht im Grossen und Ganzen dasselbe wie die Creative Commons Zero Lizenz, enthält aber spezifisch den Begriff "Datenbank".

Der Rechteverzicht bzw. die bedingungslose Nutzungslizenzierung ist in Bezug auf Forschungsdaten eher weniger attraktiv, da die Forschenden nicht mehr vor unlauterem Wettbewerb geschützt sind und somit die Verwertung fremder Leistung nicht juristisch verfolgt werden kann. Es gibt jedoch den Vorteil, dass für Nachnutzer keine Unklarheiten mehr bestehen (Ball, 2012:11).

Bezüglich Informationen, welche in Dokumentform vorliegen, ist es weniger wichtig, eine Lizenz zu verwenden, da diese durch das Urheberrecht geschützt sind und es in der wissenschaftlichen Community bereits akzeptierte Normen für das korrekte Zitieren, Paraphrasieren usw. gibt. Falls ein

Autor dennoch wünscht, erweiterte Nutzungsrechte für ein Dokument freizugeben, kann eine entsprechende Lizenz, bspw. von Creative Commons, den Nachnutzer über die Bedingungen aufklären.

In Summe bleibt festzustellen, dass die Problematik der Lizenzierung von Forschungsdaten noch nicht zufriedenstellend gelöst worden ist, weshalb es in diesem Bereich noch viel Bewegung und Dynamik gibt. Die Entwicklung der bestehenden Lizenzen sowie die Erstellung neuer Lizenzen muss bezüglich der Verwaltung von Forschungsdaten weiter verfolgt werden, um in einer spezifischen Infrastruktur die passendste Lizenzierung auswählen und anwenden zu können.

4.2.3. Volumen

Das Datenvolumen, welches in den Geisteswissenschaften jährlich erzeugt wird, ist wegen der Heterogenität der erzeugten Daten und Dokumenten nicht oder nur sehr schwer abschätzbar. Dazu kommt, dass aufgrund der unklaren Definition von Forschungsdaten in diesem Bereich nicht eindeutig bestimmt werden kann, was in eine digitale Forschungsinfrastruktur aufgenommen werden sollte und was nicht. Im nestor Bericht zur Langzeitarchivierung von Forschungsdaten wird davon ausgegangen, dass wenn Digitalisate auch zu den Forschungsdaten gehören, das Volumen sich im Petabyte-Bereich befinden wird (Pempe, 2012:148). Im Vergleich dazu ein Beispiel aus den Naturwissenschaften: Der Large Hadron Collider des CERN produziert um die 15 Petabytes Daten pro Jahr (siehe Worldwide LHC Computing Grid, CERN [online], 2008).

In der Schweiz ergab eine Befragung von 149 Institutionen bzw. Forschenden in den Geisteswissenschaften (Zimmermann, Pfister, 2008 (1):8), dass diese insgesamt 183 Terabytes an digitalen Daten und Dokumenten erstellen bzw. verwenden. Wird dies auf die Totalität der identifizierten Teilnehmenden hochgerechnet (471 an der Zahl), kann mit über 570 Terabytes gerechnet werden. Diese Angabe gilt jedoch für die kumulierten digitalen Daten und entspricht nicht einer jährlichen Produktion.

4.2.4. Metadaten und Standards

Es gibt etliche Standards, welche im Zusammenhang mit digitalen Forschungsinfrastrukturen von Relevanz sein können. Diese Standards betreffen Klassifikationssysteme, kontrollierte Vokabulare, Normen für die Archivierung und die Vertrauenswürdigkeit sowie Metadaten, welche für die Beschreibung von unterschiedlichen Aspekten digitaler Objekte verwendet werden können. Diese Standards können sowohl fachspezifisch wie auch fachübergreifend sein (Jensen et al., 2011:88-91). Welche Standards in einer Forschungsinfrastruktur berücksichtigt werden sollen, hängt davon ab, welchen Zweck die Infrastruktur verfolgt, welche Ausrichtung sie hat und welche Disziplinen damit abgedeckt werden sollen. Für jeden dieser Typen von Standards gibt es meistens etliche spezifische Standards. Bei den Klassifikationssystemen gibt es beispielsweise Normen für geografische Einheiten (ISO 3166, Normen für Sprachen (ISO 639-1:2002), Codes für Berufe (ISCO), fachübergreifende Klassifikationen wie die Dewey-Dezimalklassifikation (DDC) sowie fachspezifische Klassifikationen (Jensen et al., 2011:88). Des Weiteren existieren auch diverse kontrollierte Vokabulare, welche teilweise auch als Ontologie verfügbar sind. Dazu gehören beispielsweise die FOAF Vocabulary Specification, die Bibliographic Ontology Specification oder die GeoNames Ontology. Für eine Auflistung weiterer Vokabulare, siehe Meyer, 2011 (1).

Darüber hinaus gibt es in den Geisteswissenschaften weitere fachspezifische Standards. Dabei wird meistens die Text Encoding Initiative (TEI), die Music Encoding Initiative (MEI) und den Visual Resources Association Core (VRA Core). Diese Standards werden im Folgenden kurz beschrieben:

TEI (http://www.tei-c.org/index.xml)

Die Text Encoding Initiative ist ein auf XML basierendes Dokumentenformat, welches vom TEI-Konsortium reguliert wird. Das Konsortium verfolgt das Ziel der Harmonisierung der digitalen Kodierung von allen möglichen Dokumenten. Dafür werden Richtlinien erarbeitet, welche die Kodierungsmethoden von maschinen-lesbaren Texten vorgibt.

MEI (http://music-encoding.org/)

Aufbauend auf denselben Prinzipien der TEI dient die MEI (Music Encoding Initiative) der digitalen Kodierung von Musiknotationen. Das MEI Metadatenschema beinhaltet Regeln für die Kodierung von intellektuellen und physikalischen Eigenschaften von Musiknotationen, so dass die darin enthaltene Information gesucht, gefunden, angezeigt und ausgetauscht werden kann. Da das Format software-unabhängig aufgebaut ist, kann es auch der Archivierung dienen.

Visual Resources Association Core (http://www.loc.gov/standards/vracore/)

Der Visual Resources Association Core (VRA Core) ist ein auf XML basierender Datenstandard, der für die Beschreibung von Bildern und anderen visuellen Quellen entwickelt wurde.

Doch besonders Metadatenstandards spielen eine wichtige Rolle für digitale Forschungsinfrastrukturen. Deshalb soll im Folgenden näher auf diesen Standardtyp eingegangen werden.

Metadaten sind Daten, welche andere Daten beschreiben und helfen, ein Objekt zu identifizieren. Metadatenstandards existieren, um diese Beschreibung von Objekten zu vereinheitlichen. Dank der Anwendung von Metadatenstandards können Computerprogramme die Metadaten aufrufen und Metadaten aus verschiedenen Quellen kombinieren (siehe Metadata Standards, Research Data Management Toolkit [online], 2012).

Metadatenstandards können Forschungsdaten in folgendem unterstützen:

- Auffindbarkeit von Forschungsdaten
- Identifizierung von Forschungsdaten
- Verbindung der Daten mit Publikationen und verwandten Datensätzen
- Qualitätssicherung und -überprüfung.

Ein Metadatenstandard besteht meistens aus fünf Kernkomponenten (JISC Digital Media, 2013):

- Ein Schema: Das Schema beinhaltet die Kategorien und Felder, mit welchen Objekte beschrieben werden können.
- Ein Vokabular: Ein Metadatenstandard kann ein spezifisches, kontrolliertes Vokabular vorgeben, welches in den jeweiligen Feldern und Kategorien benutzt werden soll.
- Ein konzeptuelles Modell: Jedem Metadatenstandard unterliegt ein Modell, welches die Beziehungen der einzelnen einem Objekt innewohnenden Informationen und Konzepte beschreibt.
- Ein Inhaltsstandard: Es können praktische Standards vorgegeben werden, wie spezifische Informationen in die jeweiligen Felder und Kategorien eingegeben werden sollen.
- Verschlüsselung: Die Verschlüsselung repräsentiert die Metadaten in einer maschinen-lesbaren Art (bspw. XML).

Metadaten bestehen aus einer Vielzahl an Elementen, welche nach ihrer jeweiligen Funktion eingeteilt werden können (Higgins, 2006). Dabei werden häufig die folgenden fünf Kategorien benutzt:

- Beschreibende Metadaten, welche der Identifikation, der Lokalisierung und Auffindbarkeit dienen und oft die Verschlagwortung mit einschliessen.
- Technische Metadaten, welche die technischen Prozesse beschreiben, die für die Erstellung des digitalen Objekts benutzt worden sind, bzw. welche für die Benutzung erforderlich sind.
- Administrative Metadaten, welche der administrativen Verwaltung des digitalen Objekts dienen und Informationen bezüglich der Erstellung, Veränderung, und dem Versioning der Metadaten selber enthalten.
- Nutzungsmetadaten, welche Informationen bezüglich Zugriffen, User Tracking und Versioning des digitalen Objekts integrieren.
- Erhaltungsmetadaten, welche Aktivitäten dokumentieren, die für den Erhalt des digitalen Objekts unternommen wurden, wie beispielsweise eine Migration.

Metadatenstandards sind für unterschiedliche der oben erwähnten Funktionen vorgesehen und können entweder nur eine Funktion, mehrere oder alle Funktionen unterstützen. Zudem werden einige Metadatenstandards in einer fachspezifischen Community entwickelt, um die Ressourcen den Bedürfnissen dieser Community entsprechend so gut wie möglich zu beschreiben. Dabei gibt es meistens ein regulierendes Organ, welches die Aufnahme neuer Elemente in den Standard kontrolliert.

Im Handbuch Forschungsdatenmanagement werden im Kapitel 2.4. (Jensen et al., 2011:93-97) mehr als 50 Standards, Normen und Metadaten Schemata aufgelistet. Dieses Verzeichnis erlaubt es, eine Übersicht der relevantesten Metadatenstandards zu erhalten. Im Folgenden sollen die meistgebrauchten und für Forschungsdaten die relevantesten Metadatenstandards kurz beschrieben werden.

Dublin Core Metadata Element Set (http://dublincore.org/)

Das Dublin Core Metadata Element Set ist ein allgemeiner Standard, welcher einfach verständlich ein implementierbar und deshalb als Standard sehr verbreitet ist. Das Set enthält 15 allgemeine Elemente, welche die beschreibenden, technischen und administrativen Aspekte abdecken, um digitale Objekte zu beschreiben.

PREMIS Data Dictionary for Metadata Preservation (http://www.loc.gov/standards/premis/)

Der PREMIS Data Dictionary for Metadata Preservation wurde von der internationalen Arbeitsgruppe PREMIS (Preservation Metadata: Implementation Strategies) entwickelt, welche 2003 vom Online Computer Library Center (OCLC) und der Research Libraries Group (RLG) ins Leben gerufen wurde. Das Datenmodell des PREMIS-Standards baut auf dem OAIS-Referenzmodell auf und wird von der Library of Congress gewartet.

METS (http://www.loc.gov/standards/mets/)

METS (Metadata Encoding and Transmission Standard) ist ein Standard, welcher die Metadaten von digitalen Objektsammlungen dokumentiert. Er besteht aus sieben Hauptelementen, ist aber flexibel genug, andere Standards wie Dublin Core oder PREMIS zu integrieren (Jensen et al., 2011:92).

NISO MIX (http://www.loc.gov/standards/mix/)

Der NISO MIX (Metadata for Images in XML-Schema) ist ein Metadatenstandard, welcher für die Verwaltung von digitalen Bildsammlungen, spezifisch für die Beschreibung, den Austausch und die Speicherung von digitalen Bilddateien und -sammlungen erstellt wurde. Insbesondere für die Beschreibung von Bilddateien innerhalb von METS-Dateien findet der NISO MIX seine Anwendung.

4.3. Menschliche Aspekte

4.3.1. Anreiz für das Teilen von Forschungsprodukten

Damit Forschende ihre Forschungsdaten und -produkte in einer digitalen Forschungsinfrastruktur ablegen bzw. veröffentlichen können bzw. wollen, braucht es Anreize. Für Forschende muss der persönliche Mehrwert klar ersichtlich sein, da die Überbringung von Forschungsprodukten in eine Infrastruktur wahrscheinlich immer mit einem zusätzlichen Aufwand verbunden sein wird. Denn je nachdem wie eine Forschungsinfrastruktur organisiert ist, müssen die Forschende selber entscheiden, unter welcher Lizenz sie ihre Forschungsprodukte veröffentlichen wollen, die fehlenden Metadaten ergänzen und vielleicht die Dateien in ein akzeptierbares Format konvertieren.

Im Kapitel 4.1.1 wurden die unterschiedlichen Zwecke einer digitalen Forschungsinfrastruktur in Abbildung 7 dargestellt (siehe Seite 23). Die Argumente, welche einen persönlichen Mehrwert für die Forschenden repräsentieren, befinden sich auf der linken Hälfte des Diagramms. Dabei handelt es sich um zitierbare Daten (Citable data), Supportarbeiten (Maintenance and support), interdisziplinäre Zusammenarbeit (Interdisciplinary collaboration) und bessere Forschung (Better research). Die Unterscheidung der beiden Perspektiven Datenproduzent und Datennutzer ist wichtig, weil der Aufwand der Benutzung einer Forschungsinfrastruktur definitiv auf der Seite der Datenproduzenten liegt, während die Datennutzer mit fast keinem Aufwand von der Arbeit anderer profitieren können.

Eine von der HTW Chur durchgeführte Befragung von geisteswissenschaftlichen Forschenden in der Schweiz stellte die Frage nach der Motivation des Teilens von Forschungsdaten (Zimmermann, Pfister, 2008 (1):5). Da Forschende gleichzeitig sowohl Datenproduzenten wie auch Datennutzer sind, wurde diese Frage meistens aus beiden Perspektiven beurteilt. Für die Schnittstelle der Antwortkategorien mit den vier identifizierten Argumenten sind folgende Resultate wichtig (die Prozente beziehen sich auf 149 Befragte):

Bessere Forschung:

- "Durch gemeinsames Teilen erfolgt ein Erkenntnisgewinn, der für die gesamte Disziplin wichtig ist." (56%)

Zitierbare Daten:

- "Andere sollen Querbezüge zu meinen/unseren Daten herstellen können, um so zur Vernetzung der Daten und letztlich des Wissens beizutragen." (51%)
- "Erhöhte nationale Wahrnehmung der eigenen Tätigkeit." (36%)
- "Erhöhte internationale Wahrnehmung der eigenen Tätigkeit." (26%)

Interdisziplinäre Zusammenarbeit:

- "Interdisziplinäre Zusammenarbeit wird dadurch einfacher." (41%)

Keine der von der Umfrage vorgeschlagenen Antwortkategorien entspricht dem Argument der Supportarbeiten. Bei der Angabe, dass das Teilen von Forschungsdaten zu einem wichtigen

Erkenntnisgewinn führt, ist es schwierig abzuschätzen, ob die Befragten dies aus der Perspektive des Datenproduzenten oder Datennutzers beurteilten. Ob dieses Argument als Motivation ausreicht, die eigenen Forschungsprodukte mit anderen zu teilen, ist in Frage zu stellen. Die Antworten gehen tendenziell in Richtung der zitierbaren Daten, da diese Querbezüge erlauben, was wiederum die nationale und internationale Wahrnehmung der Forschungstätigkeit erhöht.

In den Naturwissenschaften entsteht der Anreiz für das Ablegen von Forschungsdaten in einer Infrastruktur dadurch, dass anhand der zur Verfügung stehenden Forschungsdaten die Zitierhäufigkeit des Autors erhöht werden kann (Dallmeier-Tiessen, 2012). Da in diesen Disziplinen Forschungsgelder und Anstellungen oft von der Anzahl Zitierungen bzw. dem H-Index eines Forschers abhängen, kann eine Forschungsdateninfrastruktur die intrinsische Motivation fördern, um eigene Forschungsdaten anderen zur Verfügung zu stellen. In den Geistes-, spezifischer noch in den Geschichtswissenschaften, haben Zitationen einen weniger hohen Stellenwert als in den Naturwissenschaften. Aufgrund der bevorzugten Publikation in Monographien können Zitationen nicht verfolgt werden, da Buchpublikationen häufig in Online-Datenbanken nicht rezensiert werden. Die Publikation an sich wird in dieser Community als wichtiger angesehen als die Anzahl Zitationen (Jehne, 2009:60).

Das Argument der Supportarbeiten kann schon eher das Interesse der Geisteswissenschaftler wecken. Die Übernahme von der Wartung und Instandhaltung von Forschungsprodukten und Werkzeugen durch eine dafür vorgesehene Infrastruktur sollte die Arbeit der Forschenden erleichtern können. Insbesondere wenn es sich bei den Forschungsresultaten um digitale Werkzeuge oder Hilfsmittel handelt, welche beispielsweise in Form einer Datenbank oder einer Website der Öffentlichkeit zur Verfügung gestellt werden, kann die Weiterführung der Werkzeuge durch eine Infrastruktur als hilfreich angesehen werden. Denn aufgrund der überwiegenden Finanzierung von zeitlich begrenzten Projekten ist es schwierig, für die Wartung Forschungsgelder zu erhalten.

Die interdisziplinäre Zusammenarbeit ist ein Argument, bei welchem der effektive Nutzen erst nach einer Bereitstellung der Forschungsprodukte festgestellt werden kann. Es ist relativ schwierig, damit Geisteswissenschaftler zu überzeugen, ihre Forschungsprodukte zu veröffentlichen, da für eine Entstehung interdisziplinärer Zusammenarbeit keine Garantie gegeben werden kann. Es ist hingegen möglich, dass eine Forschungsinfrastruktur, welche Zusammenarbeiten durch technische, administrative und kommunikative Lösungen unterstützt, von Geisteswissenschaftlern durchaus als nützlich beurteilt werden kann.

Das Argument der besseren Forschung ist schwierig zu kommunizieren, da für alle Forschenden andere Voraussetzungen erforderlich sind, damit sie ihre Forschung verbessern können. Während jemand eher digitale Werkzeuge benötigt, braucht ein anderer Forschender einen Bereich für das gemeinsame Bearbeiten von Dokumenten, um seine Forschung effizienter zu gestalten.

Selbstverständlich können auch die Geldgeber und Förderer einen gewissen Zwang auf die Forschenden ausüben, um sie dazu zu bringen, ihre Forschungsprodukte zu veröffentlichen. Dies kann beispielsweise dadurch erreicht werden, dass die Geldgeber die von ihnen geförderten Forschenden verpflichten, ihre Forschungsprodukte zu veröffentlichen oder in einer Infrastruktur zu speichern. Diesbezüglich wurde in der Churer Umfrage (Zimmermann, Pfister, 2008 (1):5) angegeben, dass 39% der 149 Befragten dafür sind, mit öffentlichen Mitteln erstellte Daten auch der Öffentlichkeit zur Verfügung zu stellen. Jedoch wurde die Aussage, dass das Teilen von digitalen Objekten aus einer externen Verpflichtung erwachsen sei, eher ablehnend beantwortet (28%).

Die Gründe und Motivationen für das Teilen von Forschungsprodukten fallen sehr unterschiedlich aus. Digitale Infrastrukturen in den Geistes- bzw. Geschichtswissenschaften müssen deshalb den Bedürfnissen der Forschenden entsprechen, damit diese dann auch genutzt werden. Eine bottom-up Herangehensweise scheint dabei ergebnisreicher zu sein als eine top-down Annäherung.

Wenn der persönliche Mehrwert des Teilens von Forschungsprodukten nicht herausgearbeitet werden kann, muss alternativ ein Paradigmenwechsel bezüglich der Einstellung der Forschenden stattfinden (Molloy, 2011). Denn heutzutage ist die Haltung vorrangig, dass Forschende ihre Arbeit für ein bestimmtes Projekt machen. Doch damit digitale Forschungsinfrastrukturen mit dem Fokus auf das Teilen von Informationen funktionieren, müssen Forschende die Denkweise einnehmen, dass sie ihre Arbeit für ein bestimmtes Projekt und für andere Menschen erledigen. Doch so ein Umdenken hervorzurufen ist kein einfaches Unterfangen und eine digitale Forschungsinfrastruktur sollte deshalb auch nicht darauf aufbauen.

Allgemein kann festgestellt werden, dass es einfacher ist, ein Forschungsdatenmanagement einzuführen, wenn in der jeweiligen Disziplin die Gewohnheit zur Nachnutzung schon vorhanden ist, indem beispielsweise Dateien, Preprints, Daten oder Bilder per E-Mail an andere Forschende verschickt werden. In den Geisteswissenschaften ist im Fachbereich der Archäologie bereits die Nachnutzung stark verbreitet. Folglich würde eine neu entwickelte digitale Forschungsinfrastruktur für die Archäologie einen grösseren Erfolg haben als in einer Disziplin, in welcher die Sekundärnutzung generell noch nicht verbreitet ist (Molloy, 2011).

4.3.2. Rollen

Digitale Forschungsinfrastrukturen betreffen verschiedene Stakeholders/Interessenvertreter und Berufe. Thaesis (2010:10) führt dazu vier grosse Interessengruppen auf: Forschende, Datenmanager, Verleger und Geldgeber (siehe Abbildung 9).

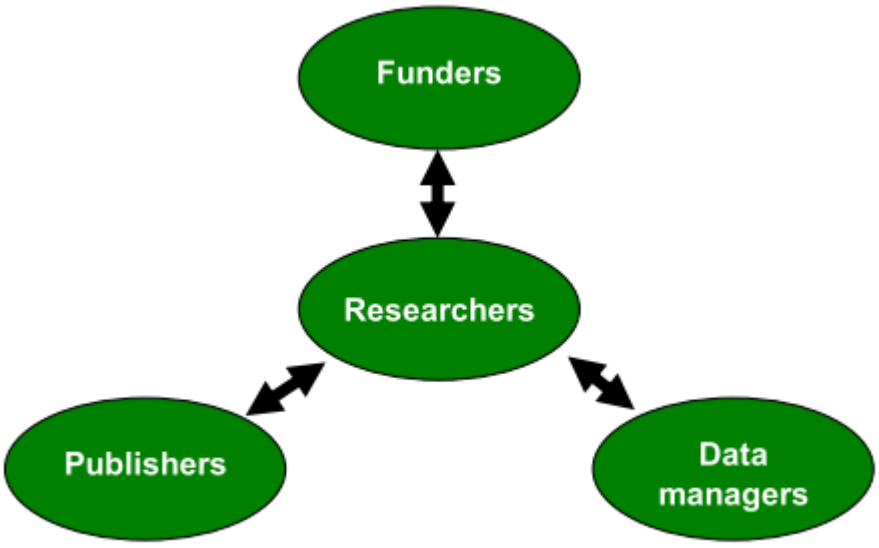


Figure 1: Generalised view on stakeholders in research

Abbildung 9: Stakeholders von digitalen Forschungsinfrastrukturen (Thaesis, 2010:10)

Die Forschenden haben in Beziehung zu Forschungsinfrastrukturen eine Doppelfunktion: sie sind sowohl Erzeuger der Daten als auch deren Nutzer. Die Datenmanager kümmern sich ihrerseits um die

Kuration und die langfristige Aufbewahrung von Daten, während die Verleger für die Verbreitung von Forschungsresultaten verantwortlich sind. Die Geldgeber sind zuständig für die Finanzierung von Projekten, aber auch der Forschungsinfrastrukturen selber. Sie entwickeln Richtlinien, welche sie durchsetzen und kontrollieren.

Wollen Datenerzeuger, also Forschende oder Forschungseinrichtungen, ihre Daten einem internen oder externen Datenzentrum übergeben, müssen die Daten bestimmten Richtlinien (Policies) entsprechen, die meistens vom Datenzentrum vorgegeben werden. Allgemein betreffen die Richtlinien meistens die Datenaufbereitung, die Datenanonymisierung, die Datenübermittlung und die Datenpflege (Pempe, 2012:155). Diese Tätigkeiten können von den Forschenden selbst übernommen werden, oder es kann technisch und wissenschaftlich qualifiziertes Personal angestellt werden, welches allgemein hin als Datenmanager oder "Data Scientist" bezeichnet wird. Dabei stellt sich die Frage, ob ein solcher Datenmanager für eine Forschungseinrichtung angestellt wird und sich mit den dort vertretenen Fachrichtungen auseinandersetzen soll, oder ob es besser fachspezifische Datenmanager gibt, die sich auf eine bestimmte Disziplin spezialisieren (Pempe, 2012:155).

Im Rahmen einer Analyse für Rollen, welche von Informationsspezialisten übernommen werden können, haben Pampel et al. (2009:11) basierend auf einem Blogeintrag (Donnelly, 2008) vier Rollen bezüglich des Forschungsdatenmanagement identifiziert: Data Manager, Data Creator, Data Librarian und Data Scientist (siehe Abbildung 10). Den einzelnen Rollen werden dabei Tätigkeitsbereiche zugeteilt, welche sich auch überschneiden können.

Data Manager (Steuerung)	**Data Creator (Bearbeitung)**
• Juristischer Sachverstand • Nutzungsbedingungen • Notfallplanung / Risk + Desaster Management • Sicherheit und Authentifizierung • Prozess-Monitoring (zus. mit *Data Creator*) • Metadaten (zus. mit *Data Creator*) • Bestandserhaltung (zus. mit *Data Librarian*) • Wert von Daten / Wirtschaftsaspekte (zus. mit *Data Librarian*)	• Dokumentation + Kontext • Prozess-Monitoring (zus. mit *Data Manager*) • Metadaten (zus. mit *Data Manager*) • Datenmodellierung (zus. mit *Data Scientist*)
Data Librarian (Unterstützung)	**Data Scientist (Analyse)**
• Verhandlungsgeschick • Beschwerdemanagement und Kundenerwartungen • Koordination der Praktiken (Verfahrensregelung) • Bewertung und Bestandsaufbau • Promotion / Marketing / Öffentlichkeitsarbeit • Entwicklung von Standards (zus. mit *Data Scientist*) • Bestandserhaltung (zus. mit *Data Manager*) • Wert von Daten / Wirtschaftsaspekte (zus. mit *Data Manager*)	• Informationsmanagement / Wissensmanagement • Datenanalyse / Datenverarbeitung • Merging und Mash-ups / Integration • Informationsextraktion (aus Datenmodellen und Know How von Personen) • Data Modelling (zus. mit *Data Creator*) • Entwicklung von Standards (zus. mit *Data Librarian*)

Abbildung 10: Rollen im Forschungsdatenmanagement (Pampel et al., 2009:11)

Die Forschenden sollten sich dabei im Data Creator wiederfinden. Die anderen Berufstypen bzw. Stellenbeschreibungen erfordern einerseits informationswissenschaftliche Kompetenzen wie unter anderem das Wissensmanagement und die Bestandserhaltung, andererseits Management-Kompetenzen wie beispielsweise das Prozess-Monitoring oder das Risk Management. Diese drei Rollen setzen Querschnittskompetenzen voraus, insbesondere der Data Scientist, welcher in den Bereichen der Informationswissenschaften, dem betroffenen Fachgebiet und in der Informatik Kompetenzen besitzen muss, um die aufgelisteten Tätigkeiten übernehmen zu können.

Am Digital Curation Center gibt es eine Initiative, welche sich mit den Rollen und Berufstypen im Zusammenhang mit digitalen Forschungsinfrastrukturen auseinandersetzt. Die vorläufigen Resultate und Merkblätter bezüglich Berufsgruppen können auf folgender Website eingesehen werden:

http://www.dcc.ac.uk/training/data-management-courses-and-training/career-profiles.

Da es bei einer geisteswissenschaftlichen digitalen Forschungsinfrastruktur um viel mehr als nur "Daten" gehen würde, werden diesbezüglich weitere Kompetenzen notwendig sein. Moulin et al. (2011:9-10) erachten folgende Berufsgruppen als unentbehrlich:

- Bibliotheksspezialisten und Archivare, welche sich um die Kuration und die Bestandserhaltung kümmern;
- Forschende der Digital Humanities, welche sich mit digitalen Werkzeugen auseinandersetzen und die Rolle eines digitalen Mediators übernehmen können;
- IT Spezialisten, welche die entsprechende technische Infrastruktur und Software entwickeln;
- Informationsspezialisten, welche das Suchverhalten der Benutzer untersuchen und Suchprozesse verdeutlichen;
- Geisteswissenschaftler, welche mit den oben erwähnten Kollegen zusammenarbeiten, um über neue Entwicklungen auf dem Laufenden gehalten zu werden und die Bedürfnisse der Community weiterzuleiten.

Da, wie im Kapitel 4.1.1 beschrieben, die Zwecke und Funktionen einer digitalen Forschungsinfrastruktur sehr unterschiedlich ausfallen können, ist es schwierig, allgemeingültige Rollen zu definieren. Die notwendigen Kompetenzen lassen sich zuletzt nur aus dem Pflichtenheft der zu entwickelnden oder der bereits entwickelten Infrastruktur ableiten.

4.4. Infrastrukturen in den Geisteswissenschaften

4.4.1. DARIAH[4]

Auf europäischer Ebene existieren zwei Projekte, welche für digitale Forschungsinfrastrukturen in den Geisteswissenschaften von Bedeutung sind. Das erste Projekt betrifft DARIAH (The Digital Research Infrastructure for the Arts and Humanities). Es hat das Ziel, eine europäische Infrastruktur zu entwickeln, welche computer-basierte Forschungsprojekte und die Erzeugung von Forschungsdaten und Werkzeugen in den Geistes- und Kulturwissenschaften unterstützen soll. Dabei sollen Informationsnutzer, Informationsverwalter und Informationsanbieter vernetzt werden, indem ein technischer Framework zur Verfügung gestellt wird, welcher den Datenaustausch zwischen Forschungsgemeinschaften ermöglicht. Das DARIAH Netzwerk soll dabei so dezentralisiert wie möglich sein, damit sich jeder einzelne Forschende bzw. jede einzelne Institution in das Projekt miteinbringen kann. Damit die jeweiligen Beiträge miteinander vernetzt werden können, werden

[4] http://www.dariah.eu

gemeinsame Standards und Technologien benötigt, welches den Kern von DARIAH ausmacht (ESFRI, 2011:27).

4.4.1. CLARIN[5]

Das zweite europäische Infrastruktur-Projekt ist CLARIN (The Common Language Resources and Technology Infrastructure). Es verfolgt das Ziel einer pan-europäisch organisierten Infrastruktur, welche Sprachquellen und Technologie für Wissenschaftler aller Disziplinen, aber speziell den Geistes- und Sozialwissenschaften zugänglich machen will. Bisherige, fragmentierte Situationen sollen überwunden werden, indem anhand einer GRID-ähnlichen Infrastruktur und Semantic Web Technologien strukturelle und terminologische Unterschiede harmonisiert werden. In zweieinhalb Jahren scheint das Projekt gute Fortschritte gemacht zu haben. Der Schritt von der Exploration hin zur Konvergenz wurde in ersten prototypischen Implementationen vorgenommen (ESFRI, 2011:26).

4.4.2. TextGrid[6]

In Deutschland wird seit 2006 das deutsche Projekt TextGrid entwickelt, welches sich dem Aufbau einer virtuellen Forschungsumgebung für die Geistes- und Kulturwissenschaften widmet. Diese soll unter anderem Werkzeuge für das philologische Editieren und für das kollaborative Arbeiten anbieten. Die virtuelle Forschungsumgebung besteht aus zwei Hauptkomponenten: dem TextGrid Labor (TextGridLab), welches als Einstiegspunkt zur virtuellen Forschungsumgebung dient und Werkzeuge sowie Dienste anbietet, und dem TextGrid Repositorium (TextGridRep), welches für den langfristigen Erhalt der Interoperabilität und des Zugangs von Forschungsdaten zuständig ist (siehe Das Projekt, TextGrid [online], 2012). Bezüglich der Entwicklung eines Datenzentrums für die Geisteswissenschaften in Deutschland könnte TextGrid zentrale Bausteine beitragen bzw. die von TextGrid angebotenen Dienstleistungen auf die ganzen Geisteswissenschaften ausweiten (Pempe, 2012:143).

4.4.3. ADONIS, PROGEDO, CORPUS, BSN[7]

Die Roadmap für Forschungsinfrastrukturen in Frankreich vom Dezember 2008 definierte vier sehr grosse Forschungsinfrastrukturen („très grandes infrastructures de recherche“) für die Geistes- und Sozialwissenschaften, und zwar ADONIS, PROGEDO, CORPUS und BSN. Die Hauptmission von ADONIS ist es, die Nachnutzbarkeit und die Langzeitarchivierung von geistes- und sozialwissenschaftlichen Daten sicherzustellen. Dabei soll ein Raum für die Navigation unter verschiedenen Dokumenten geschaffen werden, was mit der Plattform Isidore erreicht wurde. PROGEDO (Production et gestion de données pour les sciences humaines et sociales) ist ein Dienstleistungszentrum, welches die Produktion von quantitativen Daten unterstützt. Die dritte Infrastruktur ist CORPUS/SHS (Coopération des opérateurs de recherche pour un usage des sources numériques en SHS), welche einerseits eine Plattform für die Zusammenarbeit bezüglich des Zugangs zu Dokumenten der Geisteswissenschaften, andererseits eine Einrichtung für die Finanzierung von anerkannten Konsortien ist. Als letzte Infrastruktur soll die BNS (Bibliothèque scientifique numérique) eine Plattform für die Kooperation bezüglich des Zugangs zu digitaler, wissenschaftlicher Literatur (siehe Les très grandes infrastructures de recherche, 2008) bereitstellen.

[5] http://www.clarin.eu
[6] http://www.textgrid.de/
[7] ADONIS: http://www.tge-adonis.fr/, PROGEDO: http://www.reseau-quetelet.cnrs.fr/spip/article.php3?id_article=189, CORPUS: http://www.corpus-ir.fr/, BSN:http://cleo.cnrs.fr/974

4.4.4. SALSAH[8]

Das Imaging and Media Lab der Universität Basel entwickelt eine virtuelle Forschungsumgebung namens SALSAH (System for Annotation and Linkage of Sources in Arts and Humanities). Diese Umgebung ermöglicht das digitale Arbeiten mit Bildquellen, die Annotation und die Verknüpfung von einzelnen Quellen. Das Angebot ist vollständig webbasiert und ist somit unabhängig von einem bestimmten Computer nutzbar. Die Nutzer dieser Dienstleistung können dabei selbst bestimmen, ob bzw. wer die Annotationen oder Verknüpfungen einsehen kann (siehe Projekt SALSAH, Imaging & media lab, University of Basel [online], 2013).

4.4.5. metagrid.ch[9]

In der Schweiz gibt es das Projekt metagrid.ch, welches vom Projekt Diplomatische Dokumente der Schweiz (DDS) initiiert wurde. Das Projekt dient der Entwicklung eines Webservice, welcher für die Schweizer Geschichtswissenschaften relevante Ressourcen miteinander vernetzt. Dafür wird eine Dienstleistung entwickelt, welche Verbindungen zwischen Entitäten (wie Personen, Organisationen, oder Orte) erstellt, verwaltet und analysiert. Das Ziel ist es, unilaterale Verlinkungen von einer Datenbank zu einer anderen Datenbank zu überwinden und die Vernetzung der Datenbanken als Web Service anzubieten. Den Benutzern steht dank dieses Web Service eine vertrauenswürdige Informationsquelle zur Verfügung, welche auf weiterführende Quellen hinweist (metagrid.ch).

4.4.6. Weitere Projekte, Initiativen und Programme

Die vorhergehende Auflistung einzelner Initiativen ist selbstverständlich nicht vollständig. Doch mittlerweile gibt es eine unüberschaubare Anzahl an Projekten, Initiativen und Programme, die sich auf unterschiedlichen nationalen oder regionalen Ebenen beschränken. Im Folgenden wird auf einzelne Berichte und Arbeiten verwiesen, welche den Versuch einer Inventarisierung dieser Infrastrukturen unternommen haben.

Als erstes sei auf die MERIL-Datenbank (http://portal.meril.eu/converis-esf/publicweb/startpage) verwiesen, welche Forschungsinfrastrukturen in Europa mit mehr als nationaler Reichweite rezensiert. Die Datenbank wird derzeit noch mit Daten gefüllt und Laufe des Jahrs 2013 mit allen Datensätzen zur Verfügung stehen.

In der Konzeptstudie zur Langzeitarchivierung in der Schweiz werden Strategien, Aktionspläne und nationale Infrastrukturen verzeichnet, welche vor allem bezüglich des Themas der Langzeitarchivierung tätig sind (Keller-Marxer, 2009:85ff).

Der Bericht zur digitalen Infrastrukturinitiative für die Geisteswissenschaften listet forschungsgetriebene Infrastrukturen in den Geisteswissenschaften in der Schweiz auf, wobei diese sowohl analog als auch digital ausgerichtet sein können (Immenhauser, 2009:35ff.).

[8] http://www.salsah.org
[9] http://metagrid.ch/

5. Qualitative Fallstudie

5.1. Methodologie

Wie bereits eingangs erläutert, geht es in dieser Studie auch darum, die Einstellungen und Erwartungen, welche innerhalb der Community der Historiker in der Schweiz bezüglich einer elektronischen Infrastruktur für Forschungsdaten bestehen, in Erfahrung zu bringen. Um dies zu tun, wurde in Absprache mit dem Mandaten entschieden, qualitative Interviews durchzuführen. Der Auftraggeber stellte eine Liste mit zu kontaktierenden Personen zur Verfügung, welche für ein Interview eingeladen werden sollten. Die angeschriebenen Personen wurden einerseits gefragt, ob sie an einer Teilnahme an der Studie interessiert seien, andererseits wurden sie gebeten, die Kontakte von anderen möglichen interessierten Historiker anzugeben. Dank dieses Vorgehens konnten zehn Teilnehmende gewonnen werden. Davon waren acht Personen von der Liste der ursprünglich vorgeschlagenen Personen. Zwei dieser acht Teilnehmenden arbeiten für infoclio.ch und repräsentierten den Mandanten. Die Befragung fand mit beiden Personen gleichzeitig statt. Es fanden folglich insgesamt neun Interviews mit zehn Teilnehmenden statt.

Mit den zehn Teilnehmenden wurde jeweils ein semi-strukturiertes Interview durchgeführt, welches durchschnittlich 90 Minuten dauerte. Acht Interviews fanden an einem von den zu interviewenden Personen festgelegten Ort statt, ein Interview wurde per Telefon durchgeführt.

Für die Befragung wurde ein Leitfaden entwickelt (siehe Anhang 9.1) mit Fragen, welche in der vorgegebenen Reihenfolge, aber auch in einer zum Gesprächsverlauf passenden und von der ursprünglich vorgesehenen abweichenden Reihenfolge gestellt wurden.

Den Teilnehmenden wurde die Anonymisierung ihrer Antworten garantiert. Aus diesem Grund werden weder das Geschlecht, die Institution, noch spezielle Forschungsgebiete in der Auswertung erwähnt. Die befragten Personen hatten ebenfalls die Möglichkeit, vor Veröffentlichung der Studie ihre Aussagen noch einmal zu kontrollieren und wenn nötig zu korrigieren.

5.2. Grenzen

An dieser Stelle sei darauf hingewiesen, dass neun von zehn Teilnehmenden vom Mandanten vorgeschlagen worden waren und die Auswahl deshalb nicht als repräsentativ für die Schweizer Geschichtswissenschafts-Community zu sehen ist. Nichtsdestotrotz ermöglichen die Interviews eine Einsicht in die vorhandene Haltung einzelner Historiker gegenüber digitalen Forschungsdateninfrastrukturen.

Eine weitere Beeinflussung entstand dadurch, dass vor allem diejenigen Personen sich für ein Interview bereit erklärt haben, die sich bereits mit dem Thema auseinandergesetzt haben und/oder für ihre Forschung digitale Hilfsmittel einsetzen. Denn einige der angeschriebenen Personen fühlten sich durch die Studie nicht angesprochen und haben deshalb auf andere Historiker verwiesen, die sich ihrerseits wieder nicht angesprochen fühlten. Ein Telefongespräch mit einer dieser Personen ergab, dass sie der Meinung sei, keine Forschungsdaten zu produzieren, deshalb auch gar kein Interesse an einer digitalen Forschungsinfrastruktur habe und nichts zu der Studie beitragen könne.

5.3. Resultate

5.3.1. Forschung in den Geschichtswissenschaften

Die Teilnehmenden der Umfrage wurden während des Interviews gebeten, Auskunft über ihre Forschungstätigkeit zu geben. Acht von zehn Befragte gaben an, dass sie in jetzigen oder früheren Forschungsprojekten eher im Team arbeiteten als alleine. Wobei diese Unterteilung manchmal schwer fiel, da die interviewten Personen teilweise alleine Arbeiten für ein grösseres Forschungsprojekt fertigstellten. Es konnte festgestellt werden, dass keiner der Teilnehmenden einer isolierten Tätigkeit nachging und die Forschungsprojekte mit Studierenden, Assistenten, wissenschaftlichen Mitarbeitern oder anderen Projektpartnern durchgeführt werden.

Auf die Frage, ob Dokumente in Kollaboration erstellt oder alleine redigiert werden, waren die Antworten gemischt. Fünf der Teilnehmenden gaben an, dass sie Dokumente und Unterlagen gemeinsam mit anderen Mitarbeitern erzeugen, während die fünf anderen Befragten ihre Dokumente meistens alleine schreiben.

Die Forschungsprojekte der interviewten Personen finden sowohl in Zusammenarbeit mit anderen Institutionen (sechs von zehn Teilnehmenden) als auch institutionsintern (vier von zehn Teilnehmenden) statt. Die Frage nach der Publikationsform der Forschungsresultate wurde überwiegend mit wissenschaftlichen Artikeln, Monographien und Dissertationen von Studierenden beantwortet. Es ist festzuhalten, dass von den zehn interviewten Personen nur zwei Befragte keine weiteren Typen von Veröffentlichungen angaben. Die anderen nannten beispielsweise Datenbanken, Websites, Blogs bzw. Blogeinträge und Konferenzbeiträge als zusätzliche Publikationsformen.

Bezüglich der Veröffentlichung der Forschungsresultate als Open Access gaben alle Interviewten, welche einer Universität angehören, an, dass in ihrer Institution ein Repositorium für das Speichern einer Publikation als Open Access zur Verfügung steht. Vier der befragten Historiker erläuterten, dass der Reflex zur Open Access Veröffentlichung (noch) nicht existiert, wobei sich ein Teilnehmer die Frage stellte, ob der Nutzen den zusätzlichen Aufwand auch rechtfertige. Lediglich diejenigen Personen, welche eine Datenbank als Dienstleistung anbieten, versuchen Daten und Inhalte als Open Access zur Verfügung zu stellen.

5.3.2. Forschungsdaten in den Geschichtswissenschaften

Die Teilnehmenden der Umfrage wurden gebeten, alle Dokumente, Daten, Inhalte oder Unterlagen aufzuzählen, welche sie als Forschungsdaten der Geschichtswissenschaften erachten. Im Folgenden sind alle erwähnten Begriffe in alphabetischer Reihenfolge aufgelistet:

- Arbeitsplan
- Archivistische Daten
- Beschreibung der Quelle
- Bibliographische Daten
- Bilder
- Biographische Datenbanken
- Biographische Skizzen
- Chronologien / Zeitleisten
- Daten (Wirtschaft, Demographie, etc.)
- Datenbanken
- Datenbanken (Bilder, Ton, Objekte)
- Datenbanken mit Digitalisaten
- Datenbanken mit Metadaten
- Datierung
- Digitalisate
- Eigene Scans
- Exzerpte
- Forschungsnotizen
- Historiographie
- Historisches Lexikon
- Ikonographien
- Inhaltsverzeichnis

- Metadaten zu Quellen
- Publikationen
- Quellen / Archivalien
- Quelleneditionen
- Quellenkritik
- Register
- sonstige Objekte
- Tonaufnahmen
- Transkriptionen
- Übersichten
- Verschlagwortung
- Zusammenfassungen
- Zusammenstellungen von Personen

Im Vergleich zu den in Kapitel 3.3 aufgelisteten Dokumenten, Inhalten und Daten sind auch während der Interviews sehr unterschiedliche Informationstypen als Forschungsdaten der Geschichtswissenschaften qualifiziert worden. Dabei entsprechen nur die in Grau hinterlegten Begriffe denjenigen, die im Kapitel 3.3 erwähnt worden sind. Das hat einerseits mit den spezifisch in den Geschichtswissenschaften benutzten Informationstypen zu tun, wie beispielsweise Quellen und Archivalien, Quelleneditionen sowie Zeitleisten. Andererseits ist dies wohl auch darauf zurückzuführen, dass in den Geschichtswissenschaften quantitative Daten "fehlen", weshalb das Konzept der Forschungsdaten in diesem Bereich nicht klar umrissen ist und verschieden interpretiert werden kann.

Die meisterwähnten Forschungsprodukte waren Bibliographien sowie Metadaten, gefolgt von Digitalisaten. Werden diese Informationstypen anhand der im Kapitel 3.3 vorgeschlagenen Typologie nach Input, Throughput, Output und Hilfsmittel analysiert, ist erkennbar, dass vor allem der Kategorie Input angehörende Elemente wie Quellen, Archivalien, Bilder etc. vorgeschlagen wurden. Forschungsprodukte, welche der Kategorie des Throughputs entsprechen, wie Arbeitspläne, Exzerpte oder Forschungsnotizen, sind eher wenig präsent. Auch Unterlagen, die dem Output zuzuteilen sind, sind mit den Publikationen eher wenig erwähnt worden. Dagegen wurden viele Hilfsmittel aufgelistet, wie beispielsweise Quellendeditionen, Register oder Historiographien. Die Zuteilung der Forschungsprodukte in eine bestimmte Kategorie gestaltet sich aber eher schwierig, da etwa eine Datenbank sowohl zum Input gehören kann, wenn sie den Zugang zu Quellen anbietet, als auch zum Throughput, wenn eine Datenbank in einem Projekt intern für die Organisation der Quellen erstellt wird, oder auch zum Output, wenn es das Ziel eines Forschungsprojekts ist, eine Datenbank der Öffentlichkeit zur Verfügung zu stellen.

Die Antworten der Interviewteilnehmenden zeigen jedoch, dass quasi alles und nichts als Forschungsdatum in den Geschichtswissenschaften bezeichnet werden kann. So hat eine befragte Person angegeben, dass es in den Geschichtswissenschaften keine Forschungsdaten gäbe. Es ist nun die Aufgabe der geschichtswissenschaftlichen Fachgemeinschaft zu definieren, was als Forschungsdaten angesehen werden soll.

5.3.3. Funktionen einer Infrastruktur

Den Teilnehmenden der Studie wurde die Frage gestellt, welches die wichtigste Funktion einer digitalen Forschungsinfrastruktur für die Geschichtswissenschaften ist. Dafür wurden den Befragten sieben mögliche Funktionen zur Auswahl gestellt, welche ergänzt werden konnten:

- Langzeitarchivierung von Forschungsdaten und -unterlagen;
- Zugang zu Quellen;
- Zugang zu Publikationen (Open Access);

- Zugang zu den Anreicherungen im Sinne von Kommentaren, Interpretationen, Verzeichnissen, Listen etc. von anderen Forschern;
- Plattform für die Zusammenarbeit;
- Datenmanagement und Management des Forschungsprozesses;
- Verbindung von Informationen von unterschiedlichen Plattformen (Quellen, Publikationen, Anreicherungen).

Generell gaben die Befragten drei prioritäre Funktionen an, doch zwei interviewte Personen gaben vier Prioritäten an, zwei weitere gaben nur eine Priorität an, während ein Teilnehmer keine Prioritäten setzte, da er die aufgelisteten Funktionen entweder als nicht wichtig oder bereits durch eine Infrastruktur übernommen betrachtete.

Funktion	**Anzahl Erwähnungen**
Langzeitarchivierung von Forschungsdaten und -unterlagen;	5
Zugang zu Quellen;	2
Zugang zu Publikationen (Open Access);	3
Zugang zu den Anreicherungen im Sinne von Kommentaren, Interpretationen, Verzeichnissen, Listen etc. von anderen Forschern;	1
Plattform für die Zusammenarbeit;	3
Datenmanagement und Management des Forschungsprozesses;	3
Verbindung von Informationen von unterschiedlichen Plattformen (Quellen, Publikationen, Anreicherungen).	3

Tabelle 2: Prioritäten bezüglich der Funktionen einer digitalen Forschungsinfrastruktur

Obwohl nicht alle Teilnehmende gleichviele Prioritäten festsetzten, kann doch davon ausgegangen werden, dass die besonders wichtigen Funktionen auch am häufigsten erwähnt wurden. Aus der Tabelle 2 geht hervor, dass die Langzeitarchivierung von Forschungsdaten und -unterlagen am häufigsten als prioritär bezeichnet wurde. Bei einigen Interviews wurde klar, dass die Langzeitarchivierung den Historikern generell sehr am Herzen liegt, da sie für ihre eigene Arbeit Quellen einsetzen, die nur dank der Archivierung noch erhalten sind. Die Langzeitarchivierung ist folglich für die Historiker auch wichtig, damit nachkommende Generationen Quellen der Gegenwart zur Verfügung haben werden.

Die mehr oder weniger gleichmässige Verteilung der Antworten auf die restlichen aufgelisteten Funktionen zeigt, dass das Verständnis der interviewten Historiker bezüglich der Natur einer digitalen Forschungsinfrastruktur sehr heterogen ist. So gab eine befragte Person beispielsweise an, dass für sie die ersten vier Funktionen (Langzeitarchivierung, Zugang zu Quellen und Publikationen sowie zu Anreicherungen) nicht unbedingt mit Forschung zu tun haben und deshalb auch nicht in einer Forschungsinfrastruktur gehören, sondern eher Aufgabe der Archive und Bibliotheken sind (IL3). Diese Sicht wurde von den anderen Teilnehmenden so nicht geteilt.

5.3.4. Modelle

5.3.4.1. Vernetzende Infrastruktur

Den Teilnehmenden der Umfrage wurde während des Interviews ein Modell vorgestellt, welches eine vernetzende Infrastruktur darstellt (siehe Abbildung 11). Im Zentrum dieses Modells stehen die Quellen bzw. deren Digitalisate. Die Digitalisate können mit den Publikationen verlinkt werden, welche sich auf das Dokument beziehen. Dabei ist eine unterschiedliche Granularität vorstellbar, d.h. dass eine Publikation entweder mit der ganzen Quelle verbunden wird, oder aber mit einer spezifischen Textstelle, auf welche sich die Publikation bezieht. Die wissenschaftlichen Publikationen müssen dabei nicht zwangsläufig in derselben Infrastruktur gespeichert sein.

Anhand persistenter Identifikatoren können Links mit anderen Datenbanken hergestellt werden. Solche Vernetzungen könnten entweder durch einen Kurator hinzugefügt werden, durch das System dank der Zitierung von Identifikatoren automatisch generiert werden oder von einem Benutzer der Infrastruktur manuell hinzugefügt werden.

Die Benutzer der Infrastruktur können ganz nach den Prinzipien des Web 2.0 Kommentare oder Interpretationen der Quelle bzw. der Textstelle veröffentlichen. Zudem kann das Digitalisat mit anderen Quellen verlinkt werden, die beispielsweise aus demselben Dossier stammen oder die thematisch eine Einheit bilden. Solche Vernetzungen könnten auch von den Benutzern des Systems hinzugefügt werden.

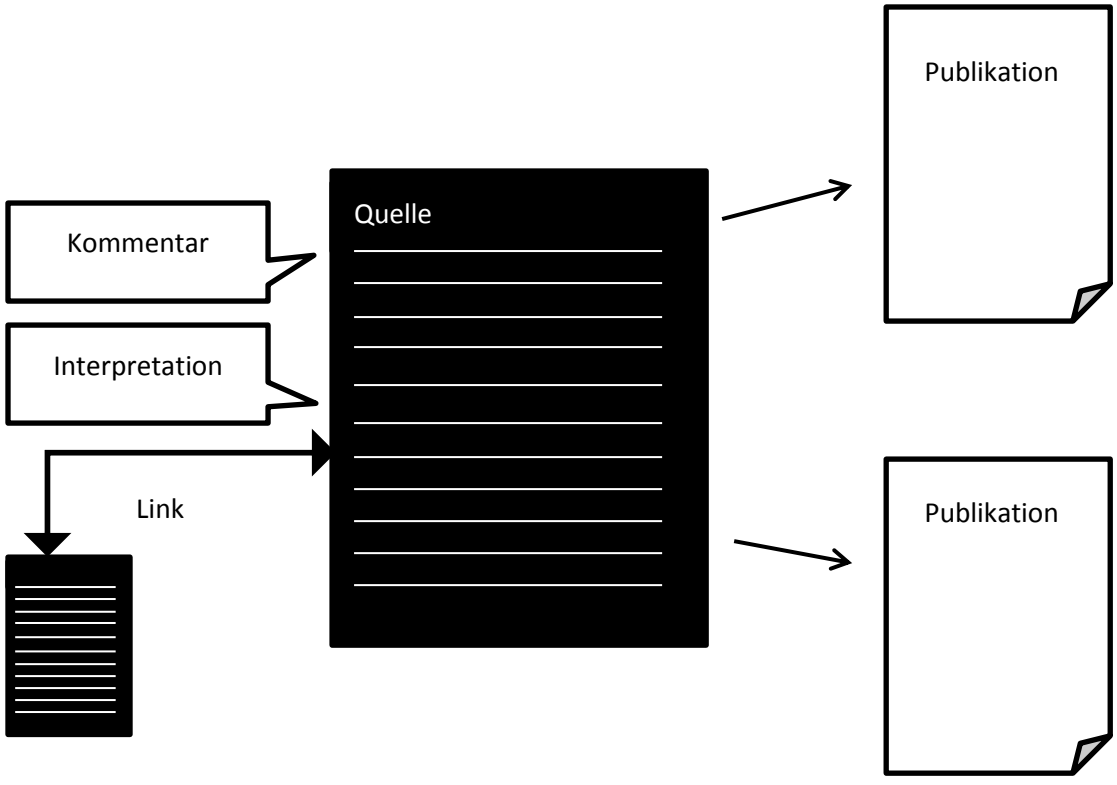


Abbildung 11: Vernetzende Infrastruktur

Die Reaktionen auf dieses Modell von den zehn interviewten Teilnehmenden fielen eher negativ aus. Grundsätzlich wurde es als wünschenswert betrachtet, einen Link zwischen einer Quelle und der darauf basierenden Publikation zu erstellen. Da jedoch die meisten Quellen nicht online zugänglich sind, sei dies schwer realisierbar. Die Teilnehmenden standen besonders den Verlinkungen, welche von einem Benutzer vorgenommen werden können, kritisch gegenüber. Einige stellten sich die Frage

nach der Korrektheit der Vernetzungen. Drei Teilnehmende waren der Meinung, dass eine solche Infrastruktur an Wert gewinnen würde, wenn die Vernetzungen von einer Autorität des jeweiligen Fachgebiets durchgeführt würden.

Bezüglich des Zugangs zu Kommentaren und Interpretationen von Dritten äusserten sich die wenigsten Historiker interessiert. Der Autor eines jeden Kommentars müsste überprüft werden, um herauszufinden, ob es sich um einen Studenten oder um einen Experten handelt. Des Weiteren waren zwei Teilnehmende überzeugt, dass die Möglichkeit Kommentare anderer einzusehen die eigenen Gedanken zu sehr strukturieren und in eine bestimmte Richtung lenken und somit die Forschungsperspektive beeinflussen würde. Dies würde es unter Umständen erschweren, originelle Herangehensweisen und interessante Fragestellungen zu entwickeln.

Eine andere interviewte Person gab an, dass für eine solche Infrastruktur die kritische Masse an Benutzerpartizipation fast unerreichbar ist. Sie glaubt, dass beispielsweise bei ihrer Forschungsthematik in der Schweiz höchstens drei weitere Personen mit denselben Quellen Forschung betreiben wie sie. Zwei weitere Teilnehmende sind auch der Meinung, dass die geschichtswissenschaftliche Forschung sich mit sehr spezifischen Themen auseinandersetzt und für eine bestimmte Quelle nur sehr wenige Personen etwas zu sagen hätten. Eine Moderation wäre aus diesem Grund für die Teilnehmenden zwingend notwendig, doch wegen des hohen Spezialisierungsgrads sehr schwierig und aufwändig umzusetzen.

Zwei Personen sagten, dass solche Infrastrukturen bereits existieren und zitierten annotierte Bibliographien, Zotero, Blogs, TextCreate und GoogleDocs sowie LitLink. Ein weiterer Teilnehmer erwähnte das Forschungsprojekt SALSAH des Imaging and Media Labs der Universität Basel, welches eine virtuelle Forschungsumgebung für die Annotation und Verlinkung zur Verfügung stellen will (Hänggli et al., 2012). Dabei legt ein Befragter viel Wert darauf, dass bestehende Infrastrukturen miteinander verbunden und nicht etliche neue Forschungsumgebungen kreiert werden sollen.

Für einen Befragten würde eine solche Infrastruktur nur Sinn machen, wenn sie innerhalb einer Forschungsgruppe ihre Anwendung findet. Denn einerseits sollen die Kommentare, Interpretationen etc. nicht unbedingt öffentlich zugänglich sein, und andererseits wären solche Informationen für Dritte nur schwer nachvollziehbar, wenn sie nicht gut dokumentiert publiziert werden.

5.3.4.2. Allumfassende Infrastruktur

Den Teilnehmenden der Umfrage wurde während des Interviews ein weiteres Modell vorgestellt, welches den Aufbau einer Infrastruktur aufzeichnet, die alle identifizierten Aufgaben einer Infrastruktur aufnimmt (siehe Abbildung 12). Das Schema verfolgt dabei nicht den Zweck, eine technisch realisierbare Lösung darzustellen, sondern sollte die Diskussion um das Thema einer digitalen Forschungsinfrastruktur während der Befragung anregen und mögliche Funktionen aufzeigen. Das Modell stützt sich dabei auf die Unterteilung, welche im Data Continuum Model beschrieben ist (siehe Kapitel 2.4.2).

Die Infrastruktur ist in der Form einer Pyramide dargestellt. Die Basis der Pyramide bildet die "Private Cloud", in welcher Forscher ihre Dokumente, Daten und Unterlagen erzeugen und für ihre private Nutzung organisieren. Die zweite Schicht wird durch die "Dedicated Cloud" gebildet, in welcher Forscher ihre in der "Private Cloud" erzeugten Dokumente, Daten und Unterlagen ausgewählten anderen Personen zur Verfügung stellen können. In dieser Cloud behalten die Forscher die Kontrolle über die Zugriffsrechte und können über ein Zugriffsberechtigungssystem unter anderem selbst bestimmen, wer auf welche Dateien Zugang hat und ob die Dateien verändert werden können.

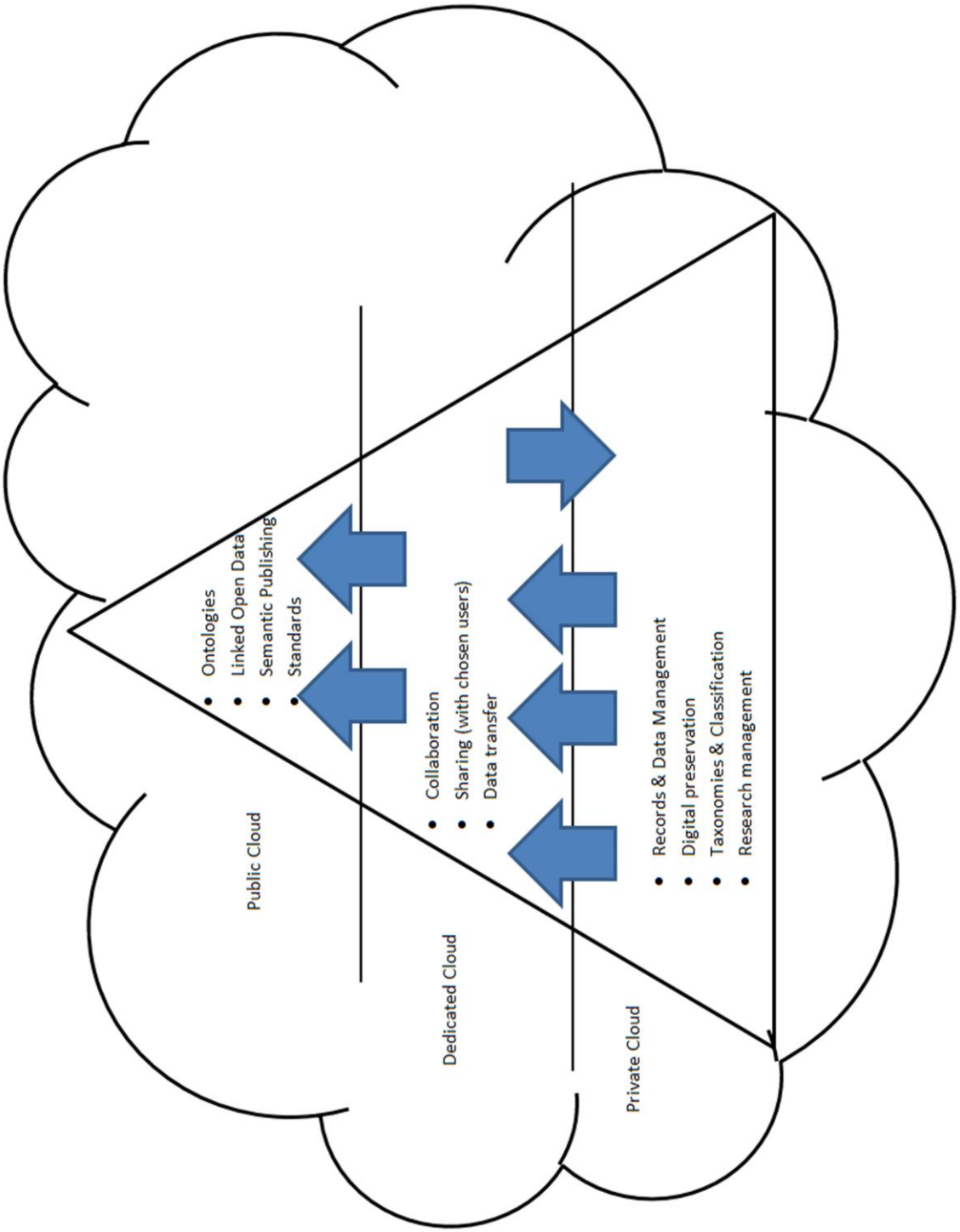


Abbildung 12: Allumfassende Infrastruktur

Die Spitze der Pyramide wird durch die "Public Cloud" gebildet. Diese Cloud dient der Veröffentlichung der vorher in der "Dedicated" bzw. "Private Cloud" produzierten Dokumenten, Daten und Unterlagen, welche im Idealfall mit Ontologien bzw. Linked Open Data unterlegt werden.

Den Teilnehmenden der Befragung wurde während der Präsentation des Modells auch mitgeteilt, dass ein solches System die Forschenden schon während der Dokumenterzeugung dazu auffordern würde, darüber nachzudenken, bis in welche Cloud das Dokument gelangen und ob es längerfristig aufbewahrt werden sollte. Je nach gewählten Parametern wird ein Forschender danach aufgefordert, bestimmte Metadaten anzugeben.

Die Reaktionen der Teilnehmenden auf dieses Model der Umfrage fielen sehr unterschiedlich aus. So wurde beispielsweise die "Private Cloud" einerseits als nicht wichtig empfunden, weil die sich darin befindlichen Aktivitäten ohnehin schon von allen durchgeführt werden, andererseits soll dieser Bereich laut einem anderen Befragten die höchste Priorität erhalten, da dort noch sehr viel Arbeit ansteht. Denn wenn ein Forschender in diesem Bereich gut arbeitet, dann werden direkt gute Bedingungen für die Forschung geschaffen. Alle erwähnten Aktivitäten der "Private Cloud" sollten deshalb in die Unterrichtspraxis integriert werden.

Für eine andere interviewte Person hat jedoch die "Dedicated Cloud" Priorität und stellt nach ihr ein grosses Bedürfnis in den Geisteswissenschaften dar. Dabei erwähnte ein anderer Teilnehmer, dass er schon alle Daten in der Cloud habe und er mit DropBox, GoogleDocs und Litlink gut zurechtkommt, weshalb dieser Bereich nicht speziell gefördert werden sollte.

Die "Public Cloud" dagegen wirkt für einen Befragten zu abstrakt, fast surreal, und würde viele Historiker nicht ansprechen. Andere Teilnehmende waren der Meinung, dass die "Public Cloud" öffentlich gemacht werden muss und dachten darüber nach, ob eine digitale Forschungsinfrastruktur zur Veröffentlichung von Forschungsdaten überhaupt einen Sinn hat, wenn die Publikationen dazu nicht frei zugänglich sind.

Vier der zehn befragten Personen haben sich generell dagegen ausgesprochen, dass alle Funktionalitäten von einer einzigen Infrastruktur übernommen werden. Für jeden Aspekt des Forschungsprozesses soll es den Forschenden überlassen werden zu entscheiden, was ihren Bedürfnissen am besten entspricht und ob sie auf eine Forschungsinfrastruktur zurückgreifen wollen oder nicht. Des Weiteren wünschten sich zwei Teilnehmende keine zu starke Einschränkung und Vereinheitlichung, welche möglicherweise den Forschungsprozess von vornherein genau festlegen.

Ein Befragter äusserte, dass beim Modell klar unterschieden werden müsste, welche Punkte fachspezifisch zu realisieren sind und welche für alle Disziplinen gleich aussehen. Der Interviewte gab als Beispiel den kollaborativen Bereich an, der für alle wie Google Docs aussehen kann. Hierfür werden keine fachspezifischen Lösungen benötigt.

Alle Teilnehmenden der Umfrage haben auch Ergänzungsvorschläge gemacht, um das Modell zu verbessern, bzw. auf Aspekte aufmerksam gemacht, die ihrer Meinung nach fehlten. So wurden unter anderem erwähnt, dass eine Etappe der Qualitätsprüfung fehlt, die auf jeder Ebene stattfinden könnte, und dass politische sowie rechtliche Aspekte nicht berücksichtigt wurden. Zwei Befragte meinten, dass beispielsweise Scans oder Digitalisate nicht so einfach veröffentlicht werden können, ohne vorher die Rechte zu prüfen.

Ein weiteres Element, auf welches drei Mal von interviewten Personen hingewiesen wurde, ist eine Beratungskomponente, welche im Modell nicht dargestellt ist. Ein Teilnehmender äusserte, dass mit einer solchen Infrastruktur auch physische Ansprechpartner nötig sind, welche Kompetenzen in den

Bereichen Projektmanagement, Geschichtswissenschaften und Technologie besitzen. Ein weiterer Befragter gab an, dass solche Berater eine Ahnung von Forschung im Allgemeinen, vom jeweiligen Fachgebiet und von der Informatik haben müssen. Ein solches Beratungszentrum wäre mindestens so aufwendig wie die Realisierung des Modells. Es wurde auch gewünscht, dass die Infrastruktur fähig sein soll, technische Hilfestellungen zu leisten. Beispielsweise könnten bezüglich des Aufbaus und Strukturierung einer Datenbank Best Practices zur Verfügung gestellt werden und Ansprechpartner für eine Beratung zur Verfügung stehen.

5.3.5. Rolle von infoclio.ch

Während der Interviews wurden die Teilnehmenden auch befragt, welche Rolle infoclio.ch bezüglich der Thematik digitaler Forschungsinfrastrukturen in den Geschichtswissenschaften übernehmen könnte bzw. sollte. Die Antworten der Befragten konnten grob in drei grosse Kategorien eingeteilt werden: die Vernetzung von Informationen, die Vernetzung von Historikern sowie das Angebot von Empfehlungen und die Darstellung von Best Practices.

Vernetzung von Informationen

Fünf der zehn interviewten Personen gaben an, dass infoclio.ch verstärkt eine Rolle bezüglich der Vernetzung von Informationen in den Geschichtswissenschaften übernehmen könnte. Dabei wurde erwähnt, dass existierende Datenbanken in diesem Fachgebiet miteinander verbunden und dass generell nicht nur Leute, sondern auch Informationen miteinander verlinkt werden sollten. Infoclio.ch soll weiterhin ein Hilfsmittel sein, welches das Auffinden von Datenbanken, Informationen und Websites unterstützt und als Informationsportal in die verschiedenen Bereiche der Geschichtswissenschaften leitet.

Vernetzung von Historikern

Zwei Teilnehmende erwarten von infoclio.ch, dass es Institutionen, die sich mit denselben Themen auseinandersetzen, koordiniert, damit beispielsweise die Datenbanken im jeweiligen Bereich wenigstens auf denselben Standards und Programmen aufgebaut sind. So könnten für die einzelnen Teilbereiche „Zellen“ errichtet werden, in denen ein Austausch stattfinden kann. Dabei könnte infoclio.ch die Rolle übernehmen, diese Leute zusammenzuführen.

Eine weitere interviewte Person wünscht sich, dass infoclio.ch eine Hilfestellung für die Zusammenarbeit verschiedener Institutionen und eine Struktur für die Vernetzung anbietet.

Empfehlungen und Best Practices

Am häufigsten wurde erwähnt, dass infoclio.ch Empfehlungen und Best Practices als Dienstleistung anbieten sollte. Die meistgenannte Thematik, zu welcher sich drei Teilnehmende geäussert haben, stellt das Daten-, Quellen- bzw. Datenbankmanagement dar.

Eine befragte Person wünschte sich ausserdem, im Forschungsprozess von infoclio.ch unterstützt zu werden und zwei weitere Interviewteilnehmende fanden, dass das Fachportal Empfehlungen zum digitalen Projektmanagement anbieten könnte.

Ein Teilnehmer war der Meinung, dass infoclio.ch nicht diverse Werkzeuge, Plattformen und Infrastrukturen erstellen kann.

Hingegen könnte infoclio.ch auf existierende Hilfsmittel verweisen, sie präsentieren und vielleicht sogar eine Analyse bzw. Bewertung der Werkzeuge anbieten.

Es wurde ebenfalls vorgeschlagen, dass infoclio.ch eine Hilfestellung für das Fundraising, d.h. der Erhalt von Forschungsgeldern, zur Verfügung stellen könnte.

Andere Reaktionen

Darüber hinaus gab es auch eine Reaktion, welche infoclio.ch als überflüssig und als "Web 1.0"-Dienstleistung beschrieb. Die betreffende Person wünscht sich ein Verzeichnis aller Veröffentlichungen, welche einen geschichtswissenschaftlichen Bezug zur Schweiz haben. Allerdings sieht sie eine solche Aufgabe besser bei der Schweizerischen Gesellschaft für Geschichte (SGG) aufgehoben.

Des Weiteren fand ein Befragter, dass infoclio.ch eine digitale Forschungsinfrastruktur für die Geschichtswissenschaften entwickeln sollte, während ein anderer Teilnehmer dagegen war, dass Daten und Inhalte durch infoclio.ch gehostet würden.

Letzterer sah die Zukunft von infoclio.ch in einer stärker interdisziplinären Ausrichtung mit den Digital Humanities, um auch andere, historisch organisierte Geisteswissenschaften wie beispielsweise die historischen Musikwissenschaften zu integrieren.

Eine weitere Aufgabe für infoclio.ch könnte auch das Lobbying für Open Access innerhalb der Schweizer Geschichtswissenschafts-Community darstellen, damit mehr Publikationen frei zugänglich zur Verfügung stehen.

5.3.6. User stories

Basierend auf den Interviews, welche für diese Studie durchgeführt wurden, wurden User Stories extrahiert. User Stories sind Software-Anforderungen, welche in Alltagssprache formuliert sind. Sie werden häufig in der agilen Software-Entwicklung eingesetzt und dienen der Steuerung eines agilen Projekts. Die Anwendung dieser Methode für die Auswertung der Umfrage wurde gewählt, weil User Stories sich auch dazu eignen, Bedürfnisse und Erwartungen knapp und prägnant darzustellen. Die Anwendererzählungen, wie User Stories auch genannt werden, werden immer nach derselben Struktur formuliert (Nazarro, Suscheck, 2010):

"Als <Rolle> möchte ich <Ziel/Wunsch>, um <Nutzen>."

Es ist dabei nicht zwingend nötig, den Nutzen immer in die User Story zu integrieren. Die Tabelle 3 beschreibt die Anforderungen, welche die interviewten Historiker bezüglich digitaler Forschungsinfrastrukturen geäussert haben.

Rolle	Ziel/Wunsch	Nutzen
Als Historiker	möchte ich eine einzige Schnittstelle für alle Repositorien in der Schweiz,	damit ich schneller relevante Dokumente finden kann und nur eine Schnittstelle erlernen muss.
Als Historiker	möchte ich Hilfsmittel für die Organisation von Dokumenten, Quellen und Wissen,	damit ich meine Forschung effizienter gestalten kann.
Als Historiker	möchte ich einen zentralen Zugang zu den Online-Angeboten in den Geschichtswissenschaften haben,	damit ich schneller relevante Dokumente finden kann und nur eine Schnittstelle erlernen muss.

Rolle	Ziel/Wunsch	Nutzen
Als Historiker	möchte ich Zugang zu einer Liste für die Forschung relevanter Werkzeuge (am besten mit einer Bewertung) haben,	damit ich effizient ein passendes Tool auswählen kann.
Als Historiker	möchte ich die freie Wahl haben, welches Werkzeug ich für meine Forschung benutzen will,	damit ich der Fragestellung angepasste Methoden wählen kann.
Als Historiker	möchte ich Zugang zu mehr wissenschaftlicher Literatur und Quellen haben,	damit ich die Qualität meiner Forschung verbessern kann.
Als Historiker	möchte ich Datenbanken an eine Infrastruktur übergeben,	damit die Datenbank nachhaltig gesichert werden kann.
Als Historiker	möchte ich eine Plattform für die Zusammenarbeit zur Verfügung gestellt bekommen,	damit ich in einem sicheren Bereich mit Forschern von anderen Institutionen zusammenarbeiten kann.
Als Historiker	möchte ich Software über eine Infrastruktur beziehen oder in einer Infrastruktur nutzen,	damit meine Projektpartner dieselben technischen Voraussetzungen haben und die Daten einfacher auszutauschen sind.
Als Historiker	möchte ich eine Hilfskraft erhalten,	damit ich meine Inhalte in eine digitale Forschungsinfrastruktur integrieren kann.
Als Historiker	möchte ich keine Einschränkung meines Forschungsprozesses durch eine Infrastruktur,	damit ich der Fragestellung angepasste Methoden wählen kann.
Als Historiker	möchte ich genau wissen, welche Qualität von anderen zur Verfügung gestellte Daten haben,	damit ich eine Entscheidung zur Nachnutzung treffen kann.
Als Historiker	möchte ich Hilfestellung für die Zusammenarbeit und eine Struktur für die Vernetzung,	damit mehr kollaborativ ausgerichtete Forschungsprojekte entstehen können.

Tabelle 3: User Stories

6. Zum Problem einer geisteswissenschaftlichen Forschungsinfrastruktur

6.1. Gründe für einen fehlenden Infrastruktur-Framework in den Geisteswissenschaften

Aufgrund einer Analyse der digitalen Landschaft konnte Burrows (2011:181-182) einige Gründe ableiten, weswegen die Geisteswissenschaften noch keinen e-Research Framework haben:

- Es ist schwierig, die Forschungsdaten in einer maschinen-lesbaren Art zu definieren.
- Es ist schwierig, einen generischen Forschungsprozess zu modellieren.
- Es gibt eine Tendenz zu projektspezifischen digitalen Lösungen, welche isoliert dastehen.
- Es ist schwierig, die Analyse und Forschungsergebnisse von Quellenmaterialien zu unterscheiden – die Publikationen eines Forschers können schnell die Quelle eines anderen Forschers werden.
- Es gibt eine Kluft zwischen den Forschungsprozessen von Wissenschaftlern und den Kuratierungsprozessen von kulturellen Institutionen.
- Die Digitalisierung von Quellenmaterial wird als Ersatz oder Äquivalent zur e-Research gesehen, da Quellenmaterial als Daten angesehen werden.

In den Geisteswissenschaften ist es überhaupt schwierig, Forschungsdaten zu definieren, unabhängig davon, ob sie in einer maschinen-lesbaren Art zur Verfügung stehen oder nicht. Im Kapitel 3.3 wurde ein Versuch unternommen, eine für die Geisteswissenschaften akzeptable Definition festzulegen. Dabei scheint es insbesondere wichtig, Forschungsdaten nicht auf ein digitales Datenobjekt zu beschränken, sondern auch auf analog vorhandene Informationen zu erweitern. In diesem Bereich muss das Konzept der Forschungsdaten bzw. -produkte von den betroffenen Disziplinen noch erarbeitet werden.

Aufgrund der grossen Heterogenität und der Methodenvielfalt, welche in den Geisteswissenschaften herrscht, ist es logischerweise schwierig, einen allgemeingültigen Forschungsprozess zu modellieren. Einige Wissenschaftler haben dies dennoch versucht und mit dem Begriff *Scholarly Primitives* bedacht. Mit *Scholarly Primitives* sind die grundlegenden Funktionen gemeint, welche Disziplinen in den Geisteswissenschaften in ihren Forschungsaktivitäten fachübergreifend gemeinsam haben. Fachübergreifend wurden beispielsweise die Aktivitäten Suchen, Sammeln, Lesen, Schreiben, Zusammenarbeiten sowie Querschnittsaktivitäten identifiziert (Palmer et al., 2009).

Spezifisch für die Geisteswissenschaften wurden auch Versuche unternommen, *Scholarly Primitives* festzulegen. Diese können wie folgt lauten: Entdecken, Annotieren, Vergleichen, Verweisen, Proben und Darstellen (Unsworth, 2000). Bezüglich dieser über mehrere Disziplinen hinweg gültigen Aktivitäten gibt es etliche solche Definitionen, welche pro Funktion noch detailliertere, weitere Aktivitäten aufschlüsseln. Bei einem Infrastrukturprojekt wird dann versucht, diese Aktivitäten zu integrieren. Dabei gibt es nach wie vor keine allgemein anerkannte *Scholarly Primitives*.

Die Tendenz zu projektspezifischen, isolierten, digitalen Lösungen ist definitiv auch in der Schweiz beobachtbar. Dies liegt teilweise am Finanzierungssystem (Fördermechanismus), welches eben projektbasiert ist. Dies führt zu einer weiteren Heterogenität, diesmal bezüglich der unterschiedlichen Schnittstellen, welche Forschende immer wieder aufs Neue erlernen müssen. Der Vorschlag der Schweizerischen Akademie für Geistes- und Sozialwissenschaften geht mit dem Massnahmenbereich der Vernetzung auf diesen Punkt ein. Mit dem Projekt metagrid.ch besteht in

den Geschichtswissenschaften bereits der Versuch, fachspezifische Online-Ressourcen miteinander zu verbinden.

Auf die mangelnde Unterscheidung zwischen Forschungsergebnissen und Quellenmaterialien wurde auch im Kapitel 3.3 kurz eingegangen. Dieses Phänomen liegt an dem teilweise rekursiven Natur der Forschungsprozesse in den Geistes-, besonders aber in den Geschichtswissenschaften. Diese Besonderheit, welche in den Naturwissenschaften weniger zum Tragen kommt, muss bei der Entwicklung von fachspezifischen Infrastrukturen berücksichtigt werden.

Ein weiteres Hindernis stellt die Diskrepanz zwischen den Forschungsprozesse und den Kuratierungsprozessen von kulturellen Institutionen dar. Die Art und Weise der Datenpflege dieser Institutionen entspricht nicht unbedingt den Bedürfnissen der Forschenden. Dies ist wahrscheinlich in den Naturwissenschaften genau gleich, doch diese Tatsache ist in den Geisteswissenschaften schwerwiegender, weil dort der Grossteil der Quellenmaterialien von Bibliotheken oder Archiven verwaltet wird. Dies wiederum führt zur Frage der Aufgabenverteilung; wer soll für eine fachgerechte Kuratierung zuständig sein?

Eine weitverbreitete, aber stark vereinfache Sichtweise ist diejenige, dass Digitalisate die Forschungsdaten der Geisteswissenschaften darstellen. Wird diese Perspektive beibehalten, so kann unter Umständen ein grosser Reichtum an anderen Forschungsprodukten verloren gehen, die für eine Sekundärnutzung wertvoll sein könnten.

Die hier vorgestellten Gründe, weshalb ein Infrastruktur-Framework in den Geisteswissenschaften noch nicht entstehen konnte, zeigt eindrücklich, dass die in den Naturwissenschaften gemachten Fortschritte nicht einfach auf die Geisteswissenschaften übertragen werden können (Moulin et al., 2011:5).

6.2. Verantwortlichkeiten und Stakeholder

Die Auflistungen im Kapitel 3.3 bezüglich der Dokumente, welche als Forschungsdaten in den Geisteswissenschaften gezählt werden können, zeigt auf, dass sie ein unübersichtliches Ganzes ergeben, welches gleichzeitig zu den Archiven, zum Kulturerbe, zur Forschung, zur Dokumentation und zur Bibliothek gehört (Delaunay, 2012:11). Gerade darin liegt das grosse Problem, die Zuständigkeiten für Forschungsdaten festzulegen. Denn die meisten der in Frage kommenden Institutionen für die Forschungsdatenverwaltung interessieren sich bestenfalls für einen Teil dessen, was als Forschungsprodukt bezeichnet werden könnte. Bibliotheken interessieren sich grundsätzlich nur für Publikationen, während Archive sich auf Records konzentrieren. Digitale Infrastrukturen können nur digitale Objekte aufnehmen, während sich jemand anderes um die Papierdokumente kümmern soll.

Bei den Forschungsdaten ist es nicht einmal überall klar, ob diese nun der Forschungsinstitution gehören, da sie während von ihr finanzierten Arbeit erstellt wurden, oder ob sie den Besitz von Forschenden darstellen. Sogar das forschungsfördernde Organ kann Anspruch auf diese Daten erheben, da es die Forschung finanziert hat.

Als einziges scheint dabei festzustehen, dass die Archive der kantonalen Hochschulen über zu wenig finanzielle Ressourcen verfügen, um die digitale Archivierung von Forschungsdaten zu garantieren (Keller-Marxer, 2008:30). Im Bezug zu digitalen Forschungsinfrastrukturen für die Geisteswissenschaften sollte also nicht nur das Konzept der Forschungsdaten genauer definiert, sondern auch die Zuständigkeit der unterschiedlichen Stakeholder geklärt werden.

Im Bericht "Digitale Infrastrukturinitiative für die Geisteswissenschaften" der Schweizerischen Akademie für Geistes- und Sozialwissenschaften werden sehr ausführlich die verschiedenen Institutionen beschrieben, welche im Zusammenhang mit dem Thema der Forschungsdaten respektive der digitalen Forschungsinfrastruktur eine Verantwortlichkeit haben (können). Die einzelnen Organisationen sind im Bericht detailliert beschrieben, weshalb hier darauf verzichtet wird und nur die erwähnten Institutionen aufgelistet werden (Immenhauser, 2009:9-13):

- Nationalbibliothek;
- Bundesarchiv;
- Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST);
- Schweizer Stiftung für die Forschung in den Sozialwissenschaften (FORS);
- Memoriav;
- Schweizerische Nationalphonothek;
- SWITCH;
- Institutionelle Repositorien;
- Imaging & Media Lab;
- E-lib.ch – Elektronische Bibliothek Schweiz;
- Universitätsbibliotheken und Bibliothek der ETH.

Eine Auflistung der forschungsgetriebenen Infrastrukturen in den Geisteswissenschaften (Immenhauser, 2009:35-38) weist auf die finanzielle Zuständigkeit bestehender Infrastrukturen. Dabei werden folgende Organe erwähnt:

- Schweizer Nationalfonds (SNF);
- Schweizerische Akademie für die Geistes- und Sozialwissenschaften (SAGW);
- Der Bund;
- Eine Universität;
- Ein Kanton;
- Die EU;
- Ein Förderverein.

Auch hier wird wieder aufgezeigt, wie unterschiedlich digitale Infrastrukturen ausfallen können und wie die Finanzierung zwangsläufig aus sehr diversen Quellen kommt. Es stellt sich die Frage, ob die Zuständigkeiten für Forschungsprodukte in der Schweiz ein für alle Mal geregelt werden können, oder ob diese von Fall zu Fall, d.h. von Infrastruktur zu Infrastruktur, neu definiert werden müssen. Wahrscheinlich wird eher der zweite Fall eintreten.

6.3. Organisatorischer Aufbau von Infrastrukturen

In Europa gibt es eine kaum noch überschaubare Anzahl an Forschungsprojekten, die sich auf die Verfügbarkeit und langfristige Aufbewahrung von digitalen Informationsressourcen beziehen. Diese Aktivitäten lassen sich in fünf Typen unterteilen (Keller-Marxer, 2008:76):

- Internationale bzw. europäische Programme;
- Fachspezifische, länderübergreifende Projekte;
- Fachübergreifende, nationale Initiativen;
- Fachspezifische, nationale Projekte;
- Fachspezifische, institutsbezogene Projekte.

Diese Typen von Programmen werden nach zwei Charakteristiken beschrieben: der Reichweite (International, länderübergreifend, national, institutionell) und dem Spezialisierungsgrad (fachübergreifend, fachspezifisch).

Bezüglich der Reichweite hat Moulin et al. (2011:5) unterschiedliche Verantwortungsbereiche herausgearbeitet. Auf europäischer bzw. pan-europäischer Ebene sollen Programme gemeinsame Standards, Metadatenschemata, Protokolle und Formate entwickeln, fördern und implementieren. Zusätzliche Dienstleistungen sollen dann fachgemeinschaftsgetrieben und eher nationaler Natur sein. Dies kann Werkzeuge, Software oder Aufbewahrungsdienstleistungen beinhalten. Auf einer lokalen bzw. institutionellen Ebene können zusätzlich Dienstleistungen für die Unterstützung vom Forschungsworkflow angeboten werden, damit eine leichtere Überführung der Daten in eine fachspezifische Infrastruktur stattfinden kann.

Eine weitere Aufteilung, welche für alle wissenschaftlichen Disziplinen bezüglich virtueller Forschungsumgebungen gleich aussieht, besteht aus folgenden drei Schichten: Forschung, Forschungsinfrastruktur und Basisinfrastruktur (Neuroth et al., 2007:273). Die Basisinfrastruktur kann beispielsweise auf einer GRID-Technologie basieren und Speichermenge sowie Rechenkapazität anbieten. Darüber hinaus kann sie auch für die Authentifizierung, Autorisierung und für die Kostenabrechnung zuständig sein.

Die Forschungsinfrastruktur dagegen wirkt als Mediator zwischen der Basisinfrastruktur und den Forschenden. Sie kann den Wissenschaftler fachspezifische Werkzeuge und Dienstleistungen anbieten.

Die letzte Schicht repräsentiert die Forschung, welche durch die Forschungsinfrastruktur unterstützt stattfinden kann. Dabei sollen Inhalte, Daten, Werkzeuge und Menschen miteinander verbunden werden können.

Wenn nun diese drei Schichten auf die vorhin definierten Eigenschaften der Reichweite und des Spezialisierungsgrads übertragen werden soll, scheint die Basisinfrastruktur am besten auf nationaler, fachübergreifender Ebene platziert zu sein. Die Schweiz hat dafür beispielsweise SWITCH (http://www.switch.ch) als Kompetenzzentrum für Dienstleistungen bezüglich Informations- und Kommunikationstechnologien im Dienst von Lehre und Forschung, welches mit dem SWITCHaai bereits eine nationale Authentifizierung für Forschende, Lehrpersonen und Studierende anbietet.

Die Schicht der Forschungsinfrastruktur sollte am besten fachspezifisch und national ausgerichtet sein und Werkzeuge sowie Dienstleistung spezifisch auf die Bedürfnisse der Forschenden des Landes anpassen. Länderübergreifende bzw. europäische Programme können solche fachspezifische nationale Infrastrukturen unterstützen, indem sie Standards, Formate und Protokolle vorgeben.

Die Frage stellt sich, auf welcher Ebene die zeitlich begrenzte bzw. unbegrenzte Aufbewahrung von Forschungsdaten angeboten werden soll und ob beide Archivierungsarten auf demselben Niveau durchgeführt werden sollen. Wie bereits im Kapitel 4.1.3 erwähnt, kann eine zweiteilige Lösung mit einem Repositorium und einem Depositorium angestrebt werden (Keller-Marxer, 2008:45). Das Repositorium ermöglicht es, dass Daten fachspezifisch und häufig genutzt werden, thematisch geordnet und die Dateninhalte durchsucht werden können. Ein Depositorium dagegen dient der Aufbewahrung und langfristigen Sicherung der Verfügbarkeit von Daten, welche eher selten und nicht fachspezifisch genutzt werden, eher homogen und nicht thematisch geordnet sind. Die Dateninhalte können dabei nicht durchsucht werden, sondern nur die Metadaten. Um auf die Inhalte zugreifen zu können, müssen die Datensätze zuerst aus dem Depositorium exportiert werden.

Dabei sollte generell festgehalten werden, dass eine fachübergreifende Infrastruktur wegen der unterschiedlichen Fachterminologien und -konventionen sowie der grossen Heterogenität der Quellen eine einheitliche Erschliessung nur auf einer sehr allgemeinen Ebene erlaubt. Jede detailliertere Erschliessung müsste durch den Datenproduzenten geschehen (Keller-Marxer, 2008:21).

Die einzige eindeutige Tendenz ist diejenige, dass ein zentrales System, welches alle Aufgaben übernimmt, nicht realistisch und auch nicht unbedingt erwünscht ist.

6.4. Risikoanalyse

Abschliessend sollen hier noch einige Risiken beschrieben werden, welche zwar nicht zum Scheitern führen können, jedoch die besondere Problematik und die damit verbundenen Schwierigkeiten beim Aufbau einer Forschungsdateninfrastruktur für die Geisteswissenschaft darstellen.

Risiko 1: Die Technologie wird nicht als notwendiges Werkzeug für die Forschung angesehen.

Je nachdem, wie ein Forschungsprojekt ausgerichtet ist, sind die Forschungsprozesse nicht unbedingt auf Technologie, welche über Textverarbeitungsprogramme hinausgeht, angewiesen. Des Weiteren gibt es Forschende, welche der Technologie generell eher abneigend gegenüber stehen. Diese beiden Gründe können dazu führen, dass Technologie als nicht notwendiges Werkzeug für ihre Forschung angesehen wird. Daraus können eine mangelnde Unterstützung von Infrastrukturprojekten sowie die Nichtbenutzung existierender Infrastrukturen erfolgen.

Risiko 2: Die Schnittstelle der digitalen Forschungsinfrastruktur ist zu kompliziert zu benutzen und erzeugt dadurch einen Mehraufwand.

Benutzerunfreundliche Interfaces können eine Hürde bei der Benutzung von digitalen Infrastrukturen darstellen. Dies kann beispielsweise zur Folge haben, dass vorhandene Datensätze nicht gefunden werden können, die Datenübergabe in die Infrastruktur zu umständlich und zeitaufwendig gestaltet ist oder allgemein die Vertrauenswürdigkeit der dahinter steckenden Infrastruktur in Frage gestellt wird. Im schlimmsten Fall kann eine schlechte Schnittstelle dazu führen, dass Forschende die Infrastruktur nicht mehr benutzen.

Risiko 3: Die Forschenden werden für das Verwalten von Daten nicht bezahlt.

Die Benutzung und Bereitstellung von Forschungsdaten bzw. -produkten führt zwangsläufig zu einem Mehraufwand. Da die Forschenden für diesen Mehraufwand nicht bezahlt werden, können sie sich weigern, diese Arbeitszeit auf sich zu nehmen. Alternativ dazu müsste dieser Aufwand bei der Beantragung von Forschungsprojekten und im Dauerbetrieb einer jeden Forschungseinrichtung berücksichtigt werden. Damit Forschende bereit sind, diesen Mehraufwand notfalls auf eigene Kosten zu betreiben, müssen sie darin einen klaren persönlichen Mehrwert erkennen können.

Risiko 4: Der Mehraufwand wird im Vergleich zum Nutzen als nicht verhältnismässig angesehen.

Besteht der Anreiz für die Übergabe eigener Forschungsprodukte in eine Forschungsinfrastruktur darin, die Sichtbarkeit der Forschungstätigkeiten der Forschenden zu erhöhen, kann die Nichterfüllung dieses Anreizes die Abwendung von der Infrastruktur bedeutet. Es ist durchaus möglich, dass bereitgestellte Forschungsprodukte nicht von anderen Forschenden benutzt oder zitiert werden. Sollte dies vermehrt der Fall sein, kann das zur Einstellung führen, dass sich der Mehraufwand der Aufbereitung der Forschungsprodukte für die Veröffentlichung nicht lohnt.

Infolgedessen können Forschende, welche ihre Forschungsprodukte einmal zur Verfügung gestellt haben, dies danach nicht mehr tun.

Risiko 5: Der persönliche Mehrwert für die Bereitstellung eigener Forschungsprodukte wird nicht wahrgenommen.

Im Kapitel 4.1.1 wurden in der Abbildung 7 die unterschiedlichen Gründe für den Einsatz einer Forschungsinfrastruktur dargestellt. Dabei befinden sich auf der Seite des Forscherinteresses nur vier Argumente (Zitierbare Forschungsdaten, Supportarbeiten, bessere Forschung, interdisziplinäre Zusammenarbeit). Obwohl die Auflistung der Gründe in dieser Abbildung sicher nicht vollständig ist, ist dennoch klar ersichtlich, dass es nicht gerade viele Argumente gibt, welche Forschende in den Geisteswissenschaften für die Benutzung einer Infrastruktur motivieren können. Die intrinsische Motivation, Forschungsprodukte anderen zur Verfügung zu stellen, muss durch einen konkreten Anreiz geschaffen werden. Fehlt dieser klar formulierte Anreiz, können einige Forschende nicht bereit sein, den nötigen Mehraufwand für die Bereitstellung eigener Forschungsprodukte zu leisten.

Risiko 6: Die benutzte Terminologie (bspw. Forschungsdaten) wird von der Community missverständlich aufgenommen und die Forschenden fühlen sich deswegen davon nicht angesprochen.

Die qualitative Studie (Kapitel 5) hat bereits gezeigt, dass der Begriff der Forschungsdaten von Geschichtswissenschaftlern sehr unterschiedlich aufgenommen wird. Das kann so weit gehen, dass ein Historiker findet, es gäbe keine Forschungsdaten in den Geschichtswissenschaften. Werden in der Kommunikation rund um eine Forschungsinfrastruktur für die Geisteswissenschaften solche missverständlichen Begriffe verwendet, kann es soweit kommen, dass sich die Forschenden in diesen Disziplinen nicht angesprochen fühlen und deshalb die Infrastruktur nicht benutzen. Eine klare Definition, aber auch eine Sensibilisierung der Fachgemeinschaft mit allen Implikationen eines konsequenten Innovations- und Changemanagement sind diesbezüglich nötig

Risiko 7: Mangelnde Sensibilisierung gegenüber der Volatilität von digitalen Dateien.

Digitale Dateien haben eine kürzere Lebensdauer als Papierdokumente, wenn sie der Verantwortung von Forschenden überlassen werden. Diesbezüglich kann aber immer noch mangelndes Wissen unter Forschenden bestehen, welche die Problematik als nicht wichtig einschätzen. Die Verwaltung von digitalen Dateien durch die Forschenden kann zu einem ungewollten Verlust von wertvollen Forschungsprodukten führen. Ob dies selbstverschuldet ist oder nicht, spielt dabei keine Rolle. Der Volatilität von digitalen Daten sind sich aber einige Forschende nicht vollständig bewusst. Diesbezüglich wäre eine Sensibilisierung der Forschungsgemeinschaft hilfreich.

Risiko 8: Mangelnde Kontrolle über die Sekundärnutzung.

Wenn Forschungsprodukte für die Nachnutzung zur Verfügung gestellt werden sollen, können Forschende dabei das Gefühl haben, die Kontrolle über ihre Forschungsprodukte zu verlieren. Einige Forschende würden gerne kontrollieren, welche Personen auf ihre Forschungsprodukte zugreifen dürfen. Je nach Infrastruktur kann dies entweder gar nicht möglich sein, oder die Kontrolle wird durch von der Infrastruktur angestellten Personen durchgeführt, was wiederum grosses Vertrauen in diese Personen verlangt. Sollte eine solche Kontrolle nicht möglich sein, ist es möglich, dass Forschende eine zur Verfügung gestellte Forschungsinfrastruktur deshalb nicht benutzen.

Risiko 9: Eine langfristige Finanzierung einer Infrastruktur kann nicht garantiert werden.

Damit Forschungsprodukte zeitlich unbegrenzt aufbewahrt werden können, benötigt es eine Infrastruktur mit einer langfristigen Finanzierung. Gerade wenn dafür eine neue Infrastruktur

entwickelt wird, können Forschende die langfristige Finanzierung anzweifeln. Je nach Finanzierung ist es möglich, dass diese beispielsweise jährlich schwankt und folglich nicht jedes Jahr dasselbe Serviceniveau angeboten werden kann. Oder eine wirtschaftlich angespannte Lage kann zur kompletten Streichung der Mittel führen, wie beispielsweise der Arts and Humanities Data Service, welcher nach 12 Jahren Betrieb im Jahr 2008 keine Gelder für die Weiterführung mehr erhielt (Tiedau, 2008). Je vertrauenswürdiger die langfristige Finanzierung einer Forschungsinfrastruktur ist, umso wahrscheinlicher werden Forschende ihre Forschungsprodukte in diese Infrastruktur übergeben.

7. Schlussfolgerung

Diese Studie verfolgte das Ziel, die Thematik der digitalen Forschungsinfrastrukturen und -daten aus der Perspektive der Geisteswissenschaften ebenso wie der Geschichtswissenschaften zu analysieren. Dafür wurden nach der Festlegung des Kontexts in Kapitel 2 die geläufigen Begriffe definiert, unter anderem die Begriffe eScience oder cyberinfrastructure. Daraufhin wurde im dritten Kapitel die Forschung sowie die Fachkommunikation in den Geistes- und Geschichtswissenschaften genauer betrachtet, um ihre Spezifität gegenüber den Naturwissenschaften herauszuarbeiten. Dies führte zu einer Neudefinition des Begriffs der Forschungsdaten, respektive zu einer neuen Wortbildung, dem Forschungsprodukt. Im darauffolgenden Kapitel wurden einzelne Aspekte, welche beim Aufbau einer Infrastruktur beachtet werden müssen, genauer dargestellt und in einen Bezug zu den Geisteswissenschaften gesetzt. Das Kapitel 5 stellte eine auf die vorhergehenden Kapitel basierende qualitative Studie vor, welche die Einstellung von Historiker/innen zum Thema Forschungsdaten erfragte. Dafür wurden neun semi-strukturierte Interviews mit Geschichtswissenschaftler durchgeführt, welche sich zum Thema der Forschung, der Forschungsdaten und Forschungsinfrastrukturen äussern konnten. Die wichtigste Erkenntnis der qualitativen Studie ist es, dass keine einheitliche Meinung oder Erwartung gegenüber Forschungsinfrastrukturen bestehen. Dies gilt sowohl für die Definition von Forschungsdaten in den Geschichtswissenschaften als auch für die Ausrichtung von einer Forschungsinfrastruktur. Schliesslich wurden im Kapitel 6 Gründe für die verlangsamte Entwicklung von Cyberinfrastrukturen in den Geisteswissenschaften erörtert, ein möglicher organisatorischer Aufbau skizziert und auf mögliche Risiken hingewiesen.

Aus dieser Studie kann geschlossen werden, dass die Entwicklung von digitalen Forschungsinfrastrukturen in den Natur- und auch in den Sozialwissenschaften weiter fortgeschritten als in den Geisteswissenschaften ist. Dies lässt sich einfach an der Anzahl an existierenden Infrastrukturen, Programmen oder Initiativen erkennen. Die vorliegende Studie zeigt auf, dass die Geisteswissenschaften nicht ohne Weiteres mit den Naturwissenschaften zu vergleichen sind. Dies liegt grösstenteils an der viel grösseren Heterogenität innerhalb der Geisteswissenschaften bezüglich der angewendeten Methoden, die teils von Forscher zu Forscher variieren. Ein weiterer Grund dafür kann auch in der Natur der Artefakte liegen, auf welchen die geisteswissenschaftliche Forschung beruht, denn jedes Objekt kann zur Informationsquelle werden. Folglich ist es in diesen Disziplinen sehr schwierig, einen schlüssigen Forschungsdatenbegriff allgemeingültig zu definieren, bzw. einen neuen Begriff dafür zu entwickeln.

Diese Uneindeutigkeit führt zu schwerwiegenden Konsequenzen bezüglich fast aller Aspekte, welche eine digitale Forschungsinfrastruktur betreffen. Dabei werden Fragen aufgeworfen wie diejenige des Zwecks, welchem eine Infrastruktur dienen soll, oder diejenige, welche Dateien und Objekte denn nun aufbewahrt werden sollen. Wegen der grossen Heterogenität an digitalen Objekten ist davon auch eine grosse Anzahl an Metadatenstandards betroffen, welche für eine Forschungsinfrastruktur in Frage kommen. Des Weiteren führen die Heterogenität und auch die unterschiedliche Komplexität der Objekte zu einer ständigen Beurteilung der erwünschten Granularität, nach welcher die Objekte beschrieben, identifiziert oder lizenziert werden sollen.

Es ist wichtig festzustellen, dass aufgrund der hohen Diskrepanzen zwischen den Natur- und den Geisteswissenschaften die in den Naturwissenschaften erreichten Ergebnisse und Erfolge bezüglich Forschungsinfrastrukturen nicht auf einfache Art und Weise auf die Geisteswissenschaften übertragen werden können.

Die Geschichtswissenschaften reflektieren die Eigenschaften der Geisteswissenschaften auf einer niedrigeren Ebene, denn auch diese Disziplin ist durch eine hohe Heterogenität und Methodenvielfalt geprägt. Die qualitative Studie, für welche Interviews mit Historikern durchgeführt wurden, ergab ihrerseits, dass auch die Erwartungen bezüglich einer Forschungsinfrastruktur sehr unterschiedlich ausfallen. Daraus lässt sich schliessen, dass noch eine enorme Arbeit auf die geschichtswissenschaftliche Fachgemeinschaft zukommt, um einerseits zu definieren, was Forschungsdaten bzw. -produkte sind, welche davon es wert sind, anderen zur Verfügung zu stellen, und welches die grundsätzlichen Bedürfnisse der Community bezüglich einer digitalen Infrastruktur sind. Sind diese Definitionen einmal festgelegt, muss eine Sensibilisierung der Forschenden stattfinden, damit die Begriffe allgemein anerkannt werden.

Ein weiterer zu leistender Aufwand ist die Abklärung und Abgrenzung der Verantwortlichkeiten der einzelnen Stakeholder in der Schweiz. Solange die Rollen und Zuständigkeiten nicht eindeutig zugeteilt worden sind, wird jede Institution oder Organisation eine digitale Forschungsinfrastruktur als ausserhalb ihres Aufgabenbereichs festsetzen, was zwangsläufig auch zu Finanzierungsproblemen führen wird.

8. Bibliographie

About the DCC. In : *Digital Curation Center* [online]. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://www.dcc.ac.uk/about-us.

ACLS Commission on Cyberinfrastructure. In : *ACLS American Council of Learned Societies* [online]. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://www.acls.org/programs/Default.aspx?id=644.

BALL, Alexander, 2012. *How to License Research Data* [online]. Edinburgh. Digital Curation Center. [Konsultiert am 8. Januar 2013]. DCC How-to Guides. Verfügbar unter: http://www.dcc.ac.uk/webfm_send/332.

BBI 2012 3099 , 2012. *Botschaft über die Förderung von Bildung, Forschung und Innovation in den Jahren 2013-2016* [online]. Bern. [Konsultiert am 18. Januar 2013]. Verfügbar unter: http://www.admin.ch/ch/d/ff/2012/3099.pdf.

BERNERS-LEE, Tim, 2009. Linked Data - Design Issues. In : *W3C* [online]. 18 juin 2009. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://www.w3.org/DesignIssues/LinkedData.html.

Big Data. In: *Wikipedia* [online]. Letze Änderung vom 09.01.2013 um 23:55 Uhr. [Konsultiert am 11 Januar 2013]. Verfügbar unter: http://en.wikipedia.org/wiki/Big_data.

BLANKE, Tobias und HEDGES, Mark, 2013. Scholarly primitives: Building institutional infrastructure for humanities e-Science. In : *Future Generation Computer Systems*. Februar 2013. Vol. 29, n° 2, pp. 654‑661. DOI 10.1016/j.future.2011.06.006.

BLINCO, K. und MCLEAN, N., 2004. *The Wheel of Fortune: A « Cosmic » View of the Repositories Space* [online]. 2004. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://www.rubric.edu.au/extrafiles/wheel/main.swf.

BORGMAN, Christine L., 2007. *Scholarship in the digital age: infomation, infrastructure, and the Internet*. Cambridge, Mass : MIT Press. 336 S.

BORGMAN, Christine L., 2010. *Research Data: Who will share what, with whom, when, and why? In : China-North America Library Conference* [online]. Peking : China-North America Library Conference, 2010. [Konsultiert am 8. Januar 2013]. Verfügbar unter: http://works.bepress.com/borgman/238.

BÜCHLER, Georg et al., 2012. *Referenzmodell für ein Offenes Archiv-Informations-System: Deutsche Übersetzung* [online]. nestor-Arbeitsgruppe OAIS-Übersetzung/Terminologie. nestor-materialien 16. Frankfurt am Main : nestor. [Konsultiert am 16. Januar 2013]. Verfügbar unter: http://files.d-nb.de/nestor/materialien/nestor_mat_16.pdf.

BURROWS, Toby, 2011. Sharing humanities data for e-research: conceptual and technical issues. In : *Sustainable data from digital research: Humanities perspectives on digital scholarship* [online]. Proceedings of the conference held at the University of Melbourne, 12-14th December 2011. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://hdl.handle.net/2123/7938.

CCSDS, 2005. *Modèle de référence pour un Système ouvert d'archivage d'information (OAIS) : Recommandation pour les normes sur les systèmes de données spatiales* [online]. CCSDS 650.0B-1 (F). (Livre Bleu). Washington, DC, USA. CCSDS. [Konsultiert am 16. Januar 2013]. Verfügbar unter: http://pin.association-aristote.fr/lib/exe/fetch.php/public/documents/norme_oais_version_francaise.pdf.

CCSDS, 2012. *Reference Model for an Open Archival Information System (OAIS) : CCSDS 650.0-M-2* [online]. Recommended Practice, Issue 2 (Magenta Book). Washington, DC, USA: CCSDS. [Konsultiert am 15. Januar 2013]. Verfügbar unter: http://public.ccsds.org/publications/archive/650x0m2.pdf.

CHIGNARD, Simon, 2012. *L'open data: comprendre l'ouverture des données publiques*. Limoges : Fyp. Entreprendre. 191 S.

Contexte, 2011-2012. In : *Corpus - Infrastructure de recherche* [online]. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://www.corpus-ir.fr/index.php?page=presentation2.

COYLE, Karen, 2012. *Semantic web and linked data*. In : Library Technology Reports. Juni 2012. Vol. 48, n° 4, pp. 10‑14.

DALLMEIER-TIESSEN, Sünje, 2012. *Forschungsdaten in der digitalen Wissenschaft - Umgang und Herausforderung*. Präsentation im Rahmen der 5. Herbstschule "New Services in Library and Information Science" in Bern, 21.11.2012.

Das Projekt, 2012. In : *TextGrid* [online]. 2012. [Konsultiert am 18. Januar 2013]. Verfügbar unter: http://www.textgrid.de/ueber-textgrid/projekt/.

DCC Curation Lifecycle Model. In : *Digital Curation Center* [online]. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://www.dcc.ac.uk/resources/curation-lifecycle-model.

DE COCK BUNING, Madeleine, VAN DINTHER, Barbara, JEPPERSEN DE BOER, Christina G. und RINGNALDA, Allard, 2011. *Legal Status of research data in the four partner countries* [online]. Utrecht. Centre for Intellectual Property Law (CIER), Molengraaff Institute for Private Law, Utrecht University. [Konsultiert am 8. Januar 2013]. Knowledge Exchange Report. Verfügbar unter: http://www.knowledge-exchange.info/default.aspx?id=461.

Definitionen - Open Access. In : *Website der Schweizerischen Akademie der Geistes- und Sozialwissenschaften* [online]. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://www.sagw.ch/de/sagw/laufende-projekte/open-access/oa-definitonen.html.

DELAUNAY, Guillaume, 2012. *Les archives scientifiques en sciences humaines et sociales : état de l'art* [online]. Mémoire d'études. Lyon : Université de Lyon, Enssib. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://memsic.ccsd.cnrs.fr/mem_00686499.

DONNELLY, Martin, 2008. RDMF2: Core Skills Diagram. In : *Research Data Management Forum* [en ligne]. 17. Dezember 2008. [Consulté le 23 janvier 2013]. Verfügbar unter: http://data-forum.blogspot.ch/2008/12/rdmf2-core-skills-diagram.html.

ESFRI, 2011. *Strategy Report on Research Infrastructures: Roadmap 2010* [online]. Luxembourg : Publications Office of the European Union. European Strategy Forum on Research Infrastructures (ESFRI). [Konsultiert am 9. Januar 2013]. Verfügbar unter: http://ec.europa.eu/research/infrastructures/pdf/esfri-strategy_report_and_roadmap.pdf#view=fit&pagemode=none.

FRASER, Michael, 2005. Virtual Research Environments: Overview and Activity. In : *Ariadne* [online]. Juli 2005. n° 44. [Konsultiert am 2. Januar 2013]. Verfügbar unter: http://www.ariadne.ac.uk/issue44/fraser.

Gemeinfreiheit. In: *Wikipedia* [online]. Letze Änderung vom 09.01.2013 um 21:41 Uhr. [Konsultiert am 11 Januar 2013]. Verfügbar unter: http://de.wikipedia.org/wiki/Gemeinfreiheit.

Gene patents in the United States. In : *Wikipedia* [online]. Letze Änderung vom 18.12.2012 um 08:49 Uhr. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://en.wikipedia.org/wiki/Gene_patent.

Geschichte, 2012. In : *Informationsplattform Open Access* [online]. 29. Juni 2012. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://open-access.net/ch_de/allgemeines/was_bedeutet_open_access/geschichte/.

Geschichtswissenschaft. In: *Wikipedia* [online]. Letze Änderung vom 20.12.2012 um 02:13 Uhr. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://de.wikipedia.org/wiki/Geschichtswissenschaft.

HEY, Tony, TANSLEY, Stewart et TOLLE, Kristin, 2009. Jim Gray on eScience: A Transformed Scientific Method. In : *The Fourth Paradigm: Data-Intensive Scientific Discovery* [Online]. Microsoft Research. [Konsultiert am 11. Februar 2013]. Verfügbar unter: http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_jim_gray_transcript.pdf.

HIGGINS, Sarah, 2006. What are Metadata Standards. In : *Digital Curation Center* [en ligne]. August 2006. [Konsultiert am 11. Februar 2013]. Verfügbar unter: http://www.dcc.ac.uk/resources/briefing-papers/standards-watch-papers/what-are-metadata-standards.

HIGGINS, Sarah, 2008. The DCC Curation Lifecycle Model. In : *International Journal of Digital Curation*. 2. Dezember 2008. Vol. 3, n° 1, pp. 134‑140. DOI 10.2218/ijdc.v3i1.48. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://www.ijdc.net/index.php/ijdc/article/view/69.

Human Genome Project. In : *Wikipedia* [online]. Letze Änderung vom 12.09.2012 um 10:04 Uhr. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://en.wikipedia.org/wiki/Human_Genome_Project.

IMMENHAUSER, Beat, 2009. *Digitale Infrastrukturinitiative für die Geisteswissenschaften: Bericht zuhanden des Staatssekretariats für Bildung und Forschung* [online]. Bern. Schweizerische Akademie der Geistes- und Sozialwissenschaften. [Konsultiert am 9. Januar 2013]. Verfügbar unter: http://www.sagw.ch/dms/sagw/laufende_projekte/infrastrukturinitiative/Bericht_def/Bericht-def.pdf.

JAMIESON, Brian, 2000. *Good Scientific Practice in Research and Scholarship* [online]. Science Policy Briefing, 10. Strasbourg : European Science Foundation. [Konsultiert am 7. Januar 2013]. Verfügbar unter: http://www.esf.org/index.php?eID=tx_ccdamdl_file&p[file]=15189&p[dl]=1&p[pid]=4052&p[site]=European%20Science%20Foundation&p[t]=1358159496&hash=7cacae1d31b5f9ee36c61dab948856b0&l=en.

JEHNE, Martin, 2009. Publikationsverhalten in den Geschichtswissenschaften. In : *Publikationsverhalten in unterschiedlichen wissenschaftlichen Disziplinen Beitraege zur Beurteilung von Forschungsleistungen* [online]. Zweite erweiterte Auflage. Bonn : Alexander von Humboldt-Stiftung. Diskussionspapiere der Alexander von Humboldt-Stiftung, 12. pp. 59‑63. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://www.humboldt-foundation.de/pls/web/wt_show.text_page?p_text_id=1073898.

JENSEN, Uwe, KATSANIDOU, Alexia und ZENK-MÖLTGEN, Wolfgang, 2011. Metadaten und Standards, Meta data and standards. In : *Handbuch Forschungsdatenmanagement* [online]. Bad Honnef : Bock + Herchen. S. 83-100. [Konsultiert am 17. Januar 2013]. Verfügbar unter: http://opus4.kobv.de/opus4-fhpotsdam/frontdoor/index/index/docId/198.

JISC DIGITAL MEDIA, 2013. Putting Things in Order: a Directory of Metadata Schemas and Related Standards. In : *JISC Digial Media* [online]. 2013. [Konsultiert am 17. Januar 2013]. Verfügbar unter: http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/putting-things-in-order-links-to-metadata-schemas-and-related-standards.

KELLER-MARXER, Peter, 2008. *Konzeptstudie zur Entwicklung eines Modells für eine zentrale Langzeitarchivierung von digitalen Primär- und Sekundärdaten der Forschung für die Schweiz* [online]. Bern: Ikeep. [Konsultiert am 7. Januar 2013]. Verfügbar unter: http://e-collection.library.ethz.ch/view/eth:1286.

LANDES, Lilian, 2008. Open Access und Geschichtswissenschaften. In : *LIBREAS. Library Ideas*. 2008. n° 14, S. 26–30. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://edoc.hu-berlin.de/libreas/14/landes-lilian-26/PDF/landes.pdf.

LANGERHORST, R. P., 1981. *Gegevensanalyse*. Den Haag: Academic Service.

LE BRECH, Goulven, 2011. Archives des SHS, mémoire et science. In : *ArchiSHS* [online]. 2. März 2011. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://archishs.hypotheses.org/411.

Les très grandes infrastructures de recherche: feuille de route française, 2008 [online]. Ministère de l'enseignement supérieur et de la recherche, France. [Konsultiert am 21. Januar 2013]. Verfügbar unter: http://www.corpus-ir.fr/uploads/Feuille_de_Route_TGRI_E.pdf.

LOCHER, Hansueli, 2011. *Webarchiv Schweiz : Merkblatt Archivieren* [online]. Bern. Schweizerische Nationalbibliothek NB. [Konsultiert am 15. Januar 2013]. Verfügbar unter: http://www.nb.admin.ch/nb_professionnel/01693/01695/01705/index.html?lang=de&download=NHzLpZeg7t,lnp6I0NTU042l2Z6ln1acy4Zn4Z2qZpnO2Yuq2Z6gpJCDeH93gmym162epYbg2c_JjKbNoKSn6A--.

Metadata Standards, 2012. In : *Research Data Management Toolkit* [online]. University of Western Australia. 13. Dezember 2012. [Konsultiert am 14. Januar 2013]. Verfügbar unter: http://guides.is.uwa.edu.au/content.php?pid=319161&sid=2616083.

metagrid.ch [online]. DODIS. [Konsultiert am 16. Januar 2013]. Verfügbar unter: http://metagrid.ch/.

MEYER, Thomas, 2011 (1). Historisches Forschungsnetz. Eine virtuelle Forschungsumgebung für die Geschichtswissenschaften. In : *Metadaten & Vokabularien* [online]. Graz. 25. November 2011. [Konsultiert am 17. Januar 2013]. Verfügbar unter: http://conference.ait.co.at/digbib/index.php/digbib2011/metavok/paper/view/17.

MEYER, Thomas, 2011 (2). Virtuelle Forschungsumgebungen in der Geschichtswissenschaft – Lösungsansätze und Perspektiven. In : *LIBREAS. Library Ideas*. 2011, n° 18, S. 38‑54. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://libreas.eu/ausgabe18/texte/05meyer.htm.

MOLLOY, Laura, 2011. Oh, the humanities! A discussion about research data management for the Arts and Humanities disciplines. In : *JISC MRD: Evidence Gathering* [online]. 16. Dezember 2011. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://mrdevidence.jiscinvolve.org/wp/2011/12/16/oh-the-humanities-a-discussion-about-research-data-management-for-the-arts-and-humanities-disciplines/.

MOULIN, Claudine, CIULA, Arianna und NYHAN, Julianne, 2011. *Research Infrastructures in the Digital Humanities* [online]. Science Policy Briefing, 42. Strasbourg : European Science Foundation. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://www.esf.org/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/be_user/research_areas/HUM/Strategic_activities/RIs_in_the_Humanities/SPB42_44p-5oct_FINAL.pdf&t=1322938661&hash=aaa4161943a3d131ccddadd0bd3b09b66920a8ff.

NAZZARO, William F. et SUSCHECK, Charles, 2010. Scrum Alliance - New to User Stories? In : *ScrumAlliance* [en ligne]. 19. April 2010. [Konsultiert am 25. Januar 2013]. Verfügbar unter: http://www.scrumalliance.org/articles/169-new-to-user-stories.

NEUROTH, Heike, ASCHENBRENNER, Andreas und LOHMEIER, Felix, 2007. e-Humanities – eine virtuelle Forschungsumgebung für die Geistes-, Kultur- und Sozialwissenschaften. In : *Bibliothek Forschung und Praxis*. 2007, Vol. 31, n° 3, S. 272–279. DOI 10.1515/BFUP.2007.272.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD), 2007. *OECD Principles and Guidelines for Access to Research Data from Public Funding* [online]. OECD Publishing.

[Konsultiert am 7. Januar 2013]. Verfügbar unter: http://www.oecd.org/science/scienceandtechnologypolicy/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm.

PALMER, Carole L., TEFFEAU, Lauren C. et PRIMANN, Carrie M., 2009. *Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development* [Online]. Dublin, OH : OCLC Research. [Konsultiert am 25. Januar 2013]. Verfügbar unter: http://www.oclc.org/resources/research/publications/library/2009/2009-02.pdf.

PAMPEL, Heinz, HOBOHM, Hans-Christoph und BERTELMANN, Roland, 2009. „Data Librarianship" - Rollen, Aufgaben, Kompetenzen. In : *Tagungsband des 98. Deutschen Bibliothekartag* [online]. Erfurt : Deutscher Bibliothekartag. 2009. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://eprints.rclis.org/14896/1/RatSWD_WP_144.pdf.

PEMPE, Wolfgang, 2012. Geisteswissenschaften. In : *Langzeitarchivierung von Forschungsdaten* [online]. Boizenburg : vwh. S. 137‑159. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://nestor.sub.uni-goettingen.de/bestandsaufnahme/index.php.

Projekt SALSAH, 2013. In : *Imaging & media lab, University of Basel* [online]. 2013. [Konsultiert am 21. Januar 2013]. Verfügbar unter: http://www.iml.unibas.ch/index.php/de/forschung/salsah.

REILLY, Susan, SCHALLIER, Wouter, SCHRIMPF, Sabine, SMIT, Eefke und WILKINSON, Max, 2011. *Report on Integration of Data and Publications* [online]. Opportunities for Data Exchange (ODE). [Konsultiert am 15. Januar 2013]. Verfügbar unter: http://www.alliancepermanentaccess.org/wp-content/plugins/download-monitor/download.php?id=ODE+Report+on+Integration+of+Data+and+Publications.

ROCKWELL, Geoffrey, 2010. As Transparent as Infrastructure: On the Research of Cyberinfrastructure in the Humanities. In : *Online Humanities Scholarship: The Shape of Things to Come* [online]. Houston, Texas : Rice University Press. März 2010. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://cnx.org/content/m34315/latest/.

RÜMPEL, Stefanie, 2011. Der Lebenszyklus von Forschungsdaten, The life cycle of research data. In : BÜTTNER, Stephan, HOBOHM, Hans-Christoph und MÜLLER, Lars (Hrsg.), *Handbuch Forschungsdatenmanagement* [online]. Bad Honnef : Bock + Herchen. S. 25‑34. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://opus4.kobv.de/opus4-fhpotsdam/frontdoor/index/index/docId/193.

SAGW, 2010. Sicherung der digitalen Informationsversorgung für die Geisteswissenschaften. In : *Bulletin SAGW*. 2010. Vol. 4, S. 15‑18. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://www.sagw.ch/dms/sagw/bulletins_sagw/bulletins_2010/bulletin4-10/bulletin4-10.pdf.

SAHLE, Patrick, 2008. eScience History (?). In : HECKMANN, Marie-Luise und RÖHRKASTEN, Jens (Hrsg.), *Von Nowgorod nach London. Studien zu Handel, Wirtschaft und Gesellschaft im mittelalterlichen Europa. Festschrift für Stuart Jenks zum 60. Geburtstag* [online]. Göttingen : Vandenhoeck & Ruprecht unipress. S. 63–74. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://kups.ub.uni-koeln.de/2512/2/eScienceHistory_print.pdf.

SBF, 2011. *Schweizer Roadmap für Forschungsinfrastrukturen: Schlussbericht* [online]. Bern, Staatssekretariat für Bildung und Forschung, Ressort Nationale Forschung. [Konsultiert am 18. Januar 2013]. Verfügbar unter: http://www2.unil.ch/fors/IMG/pdf/11_03_30_NFO_Roadmap Forschungsinfrastrukturen_d_2_-2.pdf.

science, n. In: *Oxford English Dictionary Online* [online]. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://www.oed.com/view/Entry/172672?redirectedFrom=science.

SNF, 2013. Open Access. In : *Schweizerischer Nationalsfonds zur Förderung der wissenschaftlichen Forschung* [online]. 2013. [Konsultiert am 25. Januar 2013]. Verfügbar unter: http://www.snf.ch/D/Aktuell/Dossiers/Seiten/open-access.aspx.

THAESIS, 2010. *Insight into digital preservation of research output in Europe* [online]. PARSE Insight. [Konsultiert am 8. Januar 2013]. Insight Report. Verfügbar unter: http://www.parse-insight.eu/downloads/PARSE-Insight_D3-6_InsightReport.pdf.

TIEDAU, Katrin, 2008. Enabling Digital Resources for the Arts and Humanities. In : *The Arts and Humanities Data Service* [online]. 28. März 2008. [Konsultiert am 11. Febraur 2013]. Verfügbar unter: http://www.ahds.ac.uk/.

TONKIN, Emma, 2008. Persistent Identifiers: Considering the Options. In : *Ariadne* [online]. Juli 2008, n° 56. [Konsultiert am 11. Februar 2013]. Verfügbar unter: http://www.ariadne.ac.uk/issue56/tonkin.

TRELOAR, Andrew, 2011. *Private Research, Shared Research, Publication, and the Boundary Transitions* [online]. [Konsultiert am 30. Dezember 2012]. DOI 10.1045/september2007-treloar. Verfügbar unter: http://andrew.treloar.net/research/diagrams/data_curation_continuum.pdf.

TRELOAR, Andrew, GROENEWEGEN, David und HARBOE-REE, Cathrine, 2007. The Data Curation Continuum. In : *D-Lib Magazine* [online]. Septembre 2007. Vol. 13, n° 9/10. [Konsultiert am 30. Dezember 2012]. DOI 10.1045/september2007-treloar. Verfügbar unter: http://www.dlib.org/dlib/september07/treloar/09treloar.html.

TRELOAR, Andrew und HARBOE-REE, Cathrine, 2008. Data management and the curation continuum : how the Monash experience is informing repository relationships. In : *VALA 2008 : Libraries, changing spaces, virtual places : conference proceedings*, 14th Biennial Conference & Exhibition, 5-7 February 2008, Melbourne Convention Centre, Australia [online]. [Konsultiert am 30. Dezember 2012]. Verfügbar unter: http://www.vala.org.au/docman/vala2008-proceedings/vala2008-session-6-treloar-paper/download.

UNSWORTH, John, 2000. *Scholarly Primitives: what methods do humanities have in common, and how might our tools reflect this?* [online]. Symposium on Humanities Computing: formal mehtods, experimental practice, King's College, London,13. Mai 2000. [Konsultiert am 25. Januar 2013]. Verfügbar unter: http://people.lis.illinois.edu/~unsworth//Kings.5-00/primitives.html.

Worldwide LHC Computing Grid, 2008. In : *CERN - European Organization for Nuclear research* [online]. 2008. [Konsultiert am 15. Januar 2013]. Verfügbar unter: http://public.web.cern.ch/public/en/lhc/Computing-en.html.

ZIMMERMANN, Hans-Dieter und PFISTER, Joachim, 2008 (1). *Auswertung der Umfrage Bedarfsanalyse für ein Angebot « Digitale Langzeitarchivierung » in den Geisteswissenschaften (data repository) [online*]. Chur. Hochschule für Technik und Wirtschaft Chur. Schweizerisches Institut für Informationswissenschaft. [Konsultiert am 9. Januar 2013]. Verfügbar unter: http://www.sagw.ch/dms/sagw/laufende_projekte/infrastrukturinitiative/Auswertung/Auswertung.pdf.

ZIMMERMANN, Hans-Dieter und PFISTER, Joachim, 2008 (2). *Auswertung der Umfrage Bedarfsanalyse für ein Angebot « Digitale Langzeitarchivierung » in den Geisteswissenschaften (data repository): Zusammenfassung [online*]. Chur. Hochschule für Technik und Wirtschaft Chur. Schweizerisches Institut für Informationswissenschaft. [Konsultiert am 9. Januar 2013]. Verfügbar unter: http://www.sagw.ch/dms/sagw/laufende_projekte/infrastrukturinitiative/Zusammenfassung/Zusammenfassung.pdf.

9. Anhang

9.1. Leitfaden für die qualitative Studie - Deutsch

Interviewfragen für die Geschichtswissenschaften

- Einverständniserklärung unterschreiben lassen.
- Aufnahmegerät einschalten.

DIE FORSCHUNG IN DER GESCHICHTSWISSENSCHAFT

1) Ein Forschungsprojekt wird in der Regel eher alleine oder in einem Team durchgeführt (Doktoranden, Studentenarbeiten, …)?
2) Wird ein Forschungsprojekt eher in Zusammenarbeit mit anderen Institutionen oder in einer einzigen Institution durchgeführt?
3) In welcher Form können die Resultate veröffentlicht werden (Artikel, Dissertationen, Thesen, Bücher, Websites, Datenbanken)?
4) Werden die Resultate auch in Open Access Repositories veröffentlicht?
5) Welche Dokumenttypen werden während eines Forschungsprojekts produziert, welche nicht zur Publikation bestimmt sind (Berichte, Notizen, Zusammenfassungen)?
6) Wie werden diese Dokumente aufbewahrt?
7) Werden diese Dokumente mit anderen Personen geteilt oder werden sie in Zusammenarbeit mit anderen Personen erstellt?
8) Welche dieser Dokumente könnten für Dritte nützlich oder von Interesse sein?
9) Existieren in Ihrer Institution definierte Prozesse und Regeln für das Management eines Forschungsprojekts?

FORSCHUNGSDATEN IN DER GESCHICHTSWISSENSCHAFT

10) Was bedeutet für Sie der Begriff "Forschungsdaten in den Geschichtswissenschaften"?
11) Schauen Sie sich die Tabelle unten an. Sie sehen drei Niveaus, mit welchen die Forschung grob beschrieben wird. Auf dem ersten Niveau befindet sich der Untersuchungsgegenstand, welcher ein physisches Objekt, ein Phänomen oder ähnliches sein kann. In der Astronomie beispielsweise sind Sterne der Untersuchungsgegenstand. Auf dem zweiten Niveau befinden sich die Messungen des Untersuchungsgegenstands, bzw. die Beobachtung des Phänomens. Auf dem dritten Niveau befinden sich die Publikationen, die sich auf den erhaltenen Messungen oder Beobachtungen basieren. Können Sie mir Ihre Vorstellung einer Entsprechung dieser drei Niveaus für die Geschichtswissenschaften geben?

1. Niveau	Sterne	Quellen
2. Niveau	Messungen der Sterne (Distanz, Bewegung, Zusammensetzung, etc.)	
3. Niveau	Publikationen	Publikationen

12) Ein Autor hat diese Etappen für die Geschichtswissenschaft wie folgt definiert:

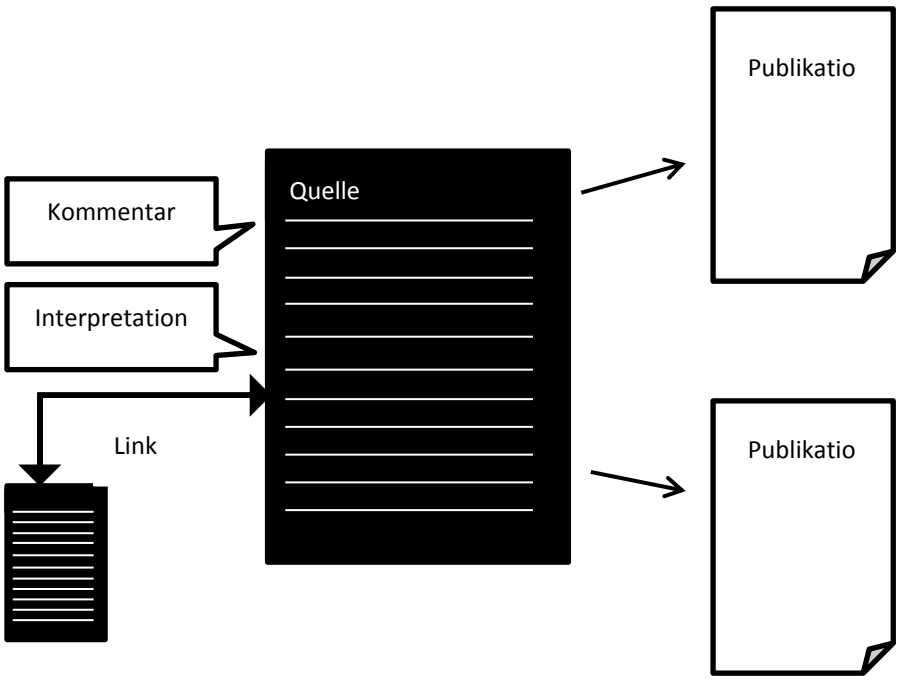
1. Niveau	Die Quelle
2. Niveau	Die Anreicherungen wie Annotationen, Kommentare, Links, Assoziationen, etc., welche vom Forscher gemacht wurden.
3. Niveau	Die Publikationen

Erscheint es Ihnen nützlich Zugang zu den im zweiten Niveau beschriebenen Informationen zu haben?

13) Wären Sie bereit, diesen Typ von Information während einem bestimmten Zeitpunkt im Prozess des Forschungsprojekts zu veröffentlichen?

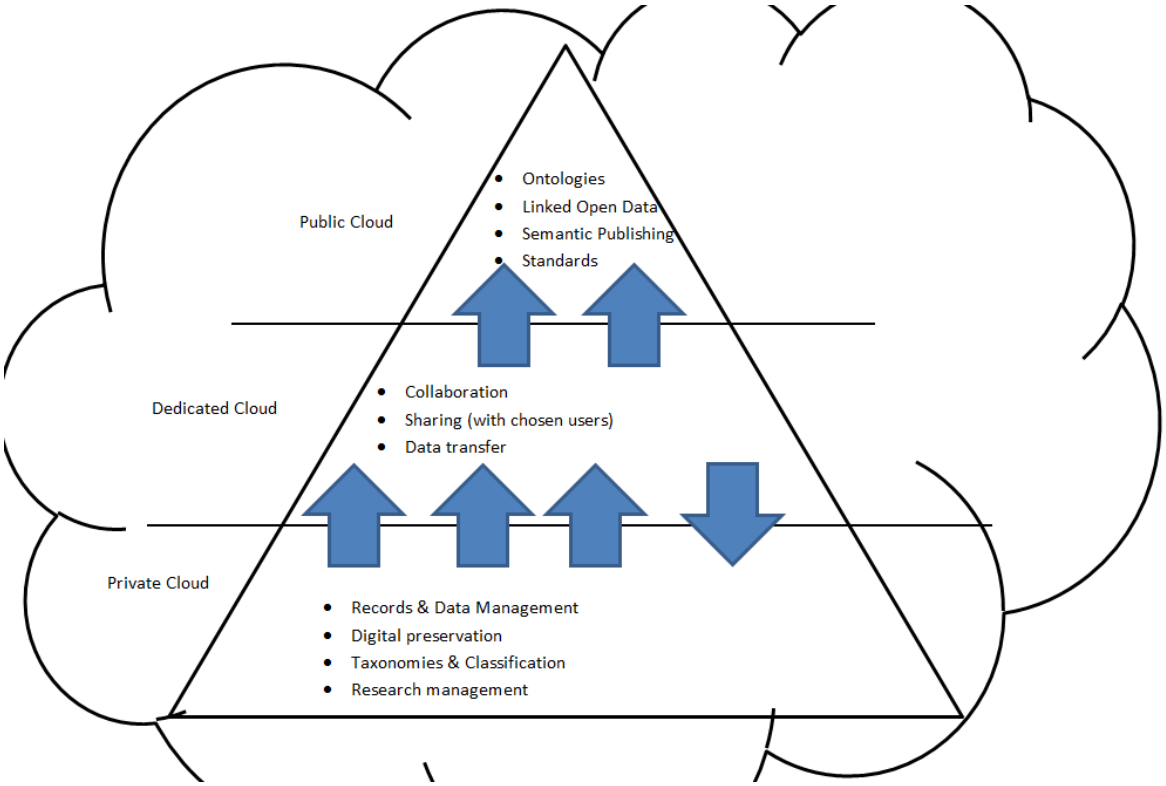
FORSCHUNGSINFRASTRUKTUR

14) Was stellen Sie sich vor, wenn ich Ihnen von einer Forschungsinfrastruktur für die Geschichtswissenschaften spreche?
15) Welche ist Ihrer Meinung nach die wichtigste Funktion einer Forschungsinfrastruktur für die Geschichtswissenschaften?
    - Langzeitarchivierung
    - Zugang zu Quellen
    - Zugang zu Publikationen (Open Access)
    - Zugang zu den Anreicherungen wie oben erwähnt
    - Plattform für die Zusammenarbeit
    - Datenmanagement und Management des Forschungsprozesses
    - Verbindung von Informationen von unterschiedlichen Plattformen (Quellen, Publikationen, Anreicherungen)
16) Werden in Ihrer Institution oder von einer anderen Organisation schon eine oder mehrere der oben zitierten Funktionen übernommen? Existieren Kollaborationen, um die erwähnten Funktionen zu gewährleisten?
17) Könnte infoclio.ch diesbezüglich eine bestimmte Rolle übernehmen?
18) Ich werde Ihnen zwei Visionen einer möglichen Infrastruktur für die Geschichtswissenschaften darlegen:
    Erscheint Ihnen die Infrastruktur des ersten Beispiels nützlich für Ihre Forschung?



Beispiel 1

19) Erscheint Ihnen die Infrastruktur des zweiten Beispiels nützlich für Ihre Forschung?



Beispiel 2

20) Nachdem Sie nun mehrere Möglichkeiten gesehen haben, die eine Infrastruktur bieten kann, welche Entwicklungen sind für Sie die wichtigsten, die in den nächsten Jahren realisiert werden sollten?

9.2. Leitfaden für die qualitative Studie - Français

Questions à poser en sciences historiques

- Faire signer le formulaire de consentement
- Allumer le dictaphone

LA RECHERCHE EN SCIENCES HISTORIQUES

Les questions concernent l'expérience que le participant a eu par le passé ou a pu observer parmi ses collaborateurs.

1) Un projet de recherche se fait plutôt en équipe ou seul (doctorant, travaux d'étudiants)?
2) Un projet de recherche se fait plutôt en collaboration avec plusieurs institutions ou dans une seule institution?
3) Sous quelle forme les résultats sont-ils publiés? (Articles, dissertations, thèses, site web, base de données etc.)?
4) Est-ce que ces publications sont publiées dans des archives ouvertes? Est-ce qu'il y a un service dans votre institution qui s'en occupe?
5) Quels types de documents sont produits lors de la recherche qui ne sont finalement pas publiés (rapports, notes, résumés, littérature grise)?
6) Où est-ce que ces documents sont stockés?
7) Est-ce que ces documents sont partagés respectivement élaborés avec des autres personnes?
8) Lesquels de ces documents pourraient être utiles à des tiers?
9) Est-ce qu'il existe dans votre institution des processus ou règles définis pour la gestion d'un projet de recherche?

DONNÉES DE RECHERCHE EN SCIENCES HISTORIQUES

10) Qu'est-ce qu'évoque le terme "données de recherche en sciences historiques" pour vous?
11) Regardez le schéma en-dessous. Vous voyez trois niveaux pour décrire les étapes du processus de recherche. Au premier niveau, il y a l'objet de recherche qui peut être un objet physique, mais aussi un phénomène par exemple. Dans l'astronomie, ce sont par exemple des étoiles. Au deuxième niveau, il y a la mesure de cet objet respectivement l'observation de ce phénomène. Au troisième niveau, il y a des publications qui se basent sur les mesures ou observations obtenus. Est-ce que vous pouvez me donner votre idée de l'équivalent de ce schéma pour les sciences historiques?

1er niveau	Etoiles	
2ème niveau	Mesures des étoiles (distance, mouvement, composition, etc.)	
3ème niveau	Publications	

12) Un auteur a défini ces étapes par:

1er niveau	La source
2ème niveau	Les enrichissements tels qu'annotations, commentaires, liens, associations fait par le

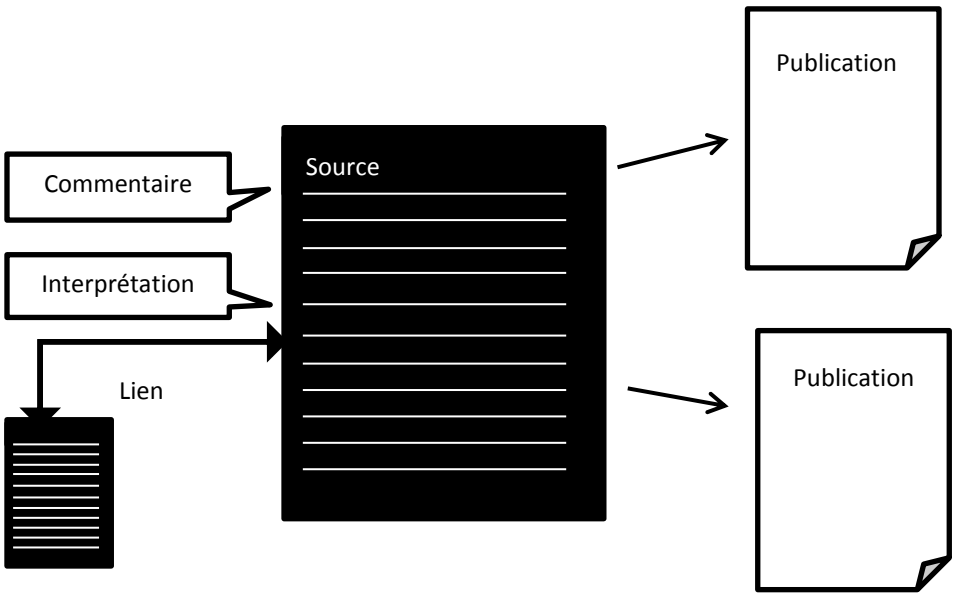
	chercheur
3ème niveau	Les publications

Est-ce que ça vous parait utile d'avoir accès à tout ce qui est contenu dans le deuxième niveau?

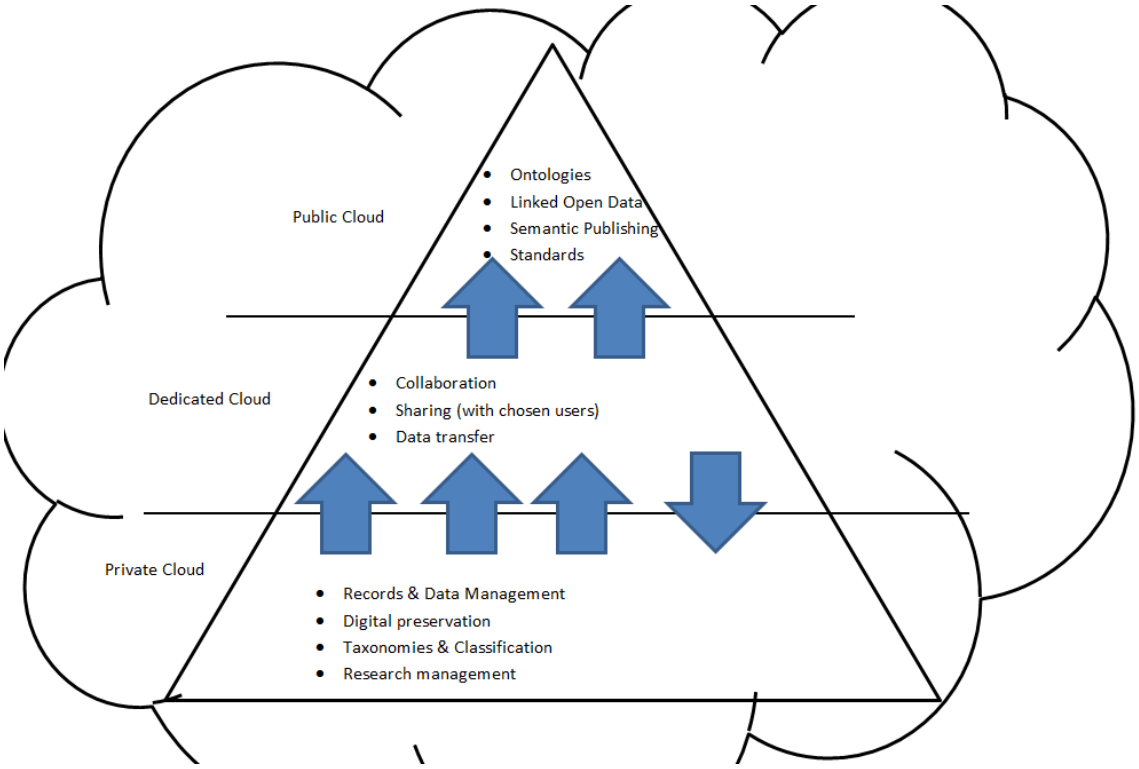
13) Seriez-vous prêt(e)s à divulguer ce genre d'information à un certain moment du processus de la gestion du projet de recherche?

INFRASTRUCTURE DE RECHERCHE

14) Qu'est-ce que vous imaginez si je vous parle d'infrastructure de recherche pour les historiens?
15) Quelle est selon vous la fonction principale d'une infrastructure de recherche pour les sciences historiques?
    - Archivage à long terme, pérennité
    - Accès aux sources
    - Accès aux publications (Open Access / archives ouvertes)
    - Accès aux enrichissements cités au 2ème niveau
    - Plateforme de collaboration
    - Gestion des données et gestion du processus de recherche
    - Liaison des informations entre plusieurs plateformes (des sources, des publications, des enrichissements)
16) Est-ce que dans votre institution, il y a déjà une ou plusieurs fonctions citées en haut qui sont **prises en charge** par votre institution ou par un autre organisme? Est-ce qu'il existe des collaborations afin d'assurer des fonctions citées en haut?
17) Est-ce que infoclio.ch pourrait jouer un rôle pour une ou plusieurs de ces fonctions dans le domaine des sciences historiques?
18) Je vous explique deux visions d'une infrastructure idéale pour les sciences historiques:
    Est-ce que l'infrastructure dans le premier exemple vous parait utile pour vos recherches?



19) Est-ce que l'infrastructure dans le deuxième exemple vous parait utile pour vos recherches?



20) Après avoir vu les possibles développements grâce à une infrastructure, quel est selon vous le développement le plus important à réaliser dans les prochaines années?

9.3. Einverständniserklärung

Einverständniserklärung

Studie über Forschungsdaten und -infrastrukturen in den Geschichtswissenschaften

Forschungsinstitution:	Haute école de gestion de Genève, Filière Information documentaire
Verantwortlicher Professor:	Prof. Dr. René Schneider
Forschungsassistentin:	Jasmin Hügi
Mandant:	Infoclio.ch, Enrico Natale

Beschreibung der Studie

Die Studie hat als Ziel

- Forschungsdaten in den Geschichtswissenschaften zu definieren,
- die Möglichkeiten einer Forschungsinfrastrukturen in den Geschichtswissenschaften zu erklären,
- die Bedürfnisse von Historikern bezüglich einer Forschungsinfrastruktur zu beschreiben.

Ablauf des Interviews

- Leitfadeninterview von einer Dauer zwischen einer bis zwei Stunden (wurde vorher festgelegt)
- Aufnahme des Interviews für die Analyse und Auswertung (fakultativ)

Vertraulichkeit

- Die Informationen werden vertraulich behandelt.
- Es werden Pseudonyme eingesetzt.
- Ausschliesslich der verantwortliche Professor und die Forschungsassistentin kennen die Identität der Interviewteilnehmer.
- Infoclio.ch wird keinen Zugang zur Identität der Interviewteilnehmer haben.
- Keine Information, die den Interviewteilnehmer identifizieren könnte, wird im Rahmen dieser Forschung an Dritte weitergegeben.

Freiwilligkeit der Teilnahme

- Sie haben das Recht, das Interview zu jedem beliebigen Zeitpunkt abzubrechen.
- Sie haben das Recht, einige Fragen nicht zu beantworten.
- Sie haben das Recht, die Aufnahme des Interviews zu verweigern.
- Sie haben das Recht, die Aufnahme des Interviews zu jedem beliebigen Zeitpunkt zu unterbrechen.

Unterschriften

Als Interviewteilnehmer bestätige ich

- Den Inhalt dieser Einverständniserklärung gelesen und verstanden zu haben.
- Dass ich weiss, dass es mir frei steht, an diesem Interview teilzunehmen und dass ich das Interview zu jedem beliebigen Zeitpunkt abbrechen kann.

Sind Sie damit einverstanden, dass das Interview aufgenommen wird?

☐ Ich bin damit einverstanden.

☐ Ich möchte nicht, dass das Interview aufgenommen wird.

Ich, der Unterzeichnende, akzeptiere an diesem Interview teilzunehmen:

Name: _____

Unterschrift : _____

Datum : _____

Als Interviewer bestätige ich,

- Die Elemente der Einverständniserklärung dem Unterzeichnenden erklärt zu habe,
- Alle Fragen des Unterzeichnenden beantwortet zu haben,
- Dem Unterzeichnenden verdeutlicht zu haben, dass er die Freiheit hat, zu jedem beliebigen Zeitpunkt seine Teilnahme am Interview zu beenden.

Name des Interviewers : _____

Unterschrift : _____

Datum : _____

9.4. Formulaire de consentement

Formulaire de consentement

Etude sur les données et infrastructures de recherche en sciences historiques

Institution de recherche: Haute école de gestion de Genève,
Filière Information documentaire

Professeur responsable: Prof. Dr. René Schneider

Assistante de recherche: Jasmin Hügi

Mandant: Infoclio.ch, Enrico Natale

Description de l'étude

Cette étude a pour but de

- définir ce que sont les données de recherche en sciences historiques,
- expliquer les possibilités d'une infrastructure de recherche en sciences historiques,
- décrire les besoins des historiens par rapport à une infrastructure de recherche.

Déroulement de l'étude

- Entretien semi-directif d'une durée entre une ou deux heures (choisie au préalable)
- Enregistrement de l'entretien pour l'analyse ultérieure (facultatif)

Confidentialité

- Garantie du traitement confidentiel des informations.
- Utilisation de pseudonyme.
- Uniquement le professeur responsable et l'assistante de recherche connaîtront l'identité du répondant.
- Infoclio.ch n'aura pas accès à l'identité des répondants.
- Aucune information permettant de vous identifier ne sera transmise à un tiers dans le cadre de cette recherche.

Liberté de participation et de retrait

- Vous avez le droit d'arrêter l'entretien à n'importe quel moment.
- Vous avez le droit de ne pas répondre à certaines questions.
- Vous avez le droit de refuser l'enregistrement de l'entretien.
- Vous avez le droit d'arrêter l'enregistrement à n'importe quel moment.

Signatures

En tant que participant, je certifie

- Avoir lu et compris le contenu du présent formulaire.
- Que je sais que je suis libre d'y participer et que je peux quitter à n'importe quel moment si je le désire.

Êtes-vous d'accord que l'entretien soit enregistré?

☐ Je suis d'accord. ☐ Je souhaite que l'entretien ne soit pas enregistré.

Je, soussigné, accepte de participer à cette étude :

Nom: _____

Signature : _____

Date : _____

En tant que chercheur, je certifie

- Avoir expliqué au signataire les termes du présent formulaire de consentement;
- Avoir répondu aux questions qu'il m'a posées à cet égard;
- Lui avoir clairement indiqué qu'il demeure libre, à tout moment, de mettre un terme à sa participation au présent projet de recherche;

Nom du chercheur : _____

Signature : _____

Date : _____