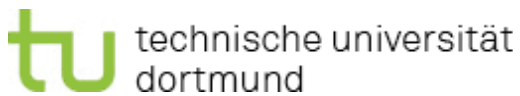




Expertise – Aufbau eines Forschungsdatenzentrums Archäologie und Altertumswissenschaften



Autor:

Prof. Dipl.-Ing. Uwe Schwiegelshohn
TU Dortmund, Institut für Roboterforschung
Otto-Hahn-Str. 8
44221 Dortmund
Uwe.schwiegelshohn@tu-dortmund.de

Version 1.0
30. Mai 2013

Koordination

Förderung



DFG

Autoren: Prof. Dipl.-Ing. Uwe Schwiegelshohn

Titel: Aufbau eines Forschungsdatenzentrums Archäologie und Altertumswissenschaften

Sprache: Deutsch

DOI: 10.13149/000.llwqv2-2

Zitierhinweise:

Uwe Schwiegelshohn (2013) Aufbau eines Forschungsdatenzentrums Archäologie und Altertumswissenschaften. [Version 1.0] Hrsg. IANUS. doi:10.13149/000.llwqv2-2

Kontakt:

ianus-fdz@dainst.de

www.ianus-fdz.de

Lizenz:



Dieses Werk bzw. Inhalt ist lizenziert unter einer Creative Commons – Namensnennung – Weitergabe unter gleichen Bedingungen 4.0 Deutschland Lizenz.

<http://creativecommons.org/licenses/by-sa/4.0/deed.de>

Version Autoren

Datum

Beschreibung

Version	Autoren	Datum	Beschreibung
1.0	Uwe Schwiegelshohn	30.05.2013	Erste finale Fassung

Zusammenfassung

In vielen Bereichen der Wissenschaft gewinnt die Untersuchung und Bearbeitung erheblicher Datenmengen zunehmend an Bedeutung. Diese Daten werden vielfach digitalisiert, was einen einfachen Zugriff von verschiedenen Orten und eine Bearbeitung mit verschiedenen Softwaremethoden ermöglicht. Zu Gunsten einer einfachen späteren (Nach-)Nutzung insbesondere auch für die Forschung sollten die Datensätze auf standardisierte Weise mit Metadaten versehen und archiviert werden. Im Fall der Archäologie und Altertumswissenschaften nimmt die Bewahrung des kulturellen Erbes neben der durch den Nachnutzungsbedarfs und die Erfüllung der Rahmenrichtlinien der guten wissenschaftlichen Praxis bedingten Archivierung von Forschungsdaten einen mindestens gleichwertigen Platz ein. Es ist aus Effizienzgründen geboten, in allen Fällen auf diese Daten mit einem einheitlichen Verfahren zuzugreifen und in ihnen suchen zu können. Neben der Erweiterung der Forschungsmöglichkeiten durch einen vergrößerten Datenbestand ist dies ein weiterer Grund, in das geplante Forschungsdatenzentrum sowohl Daten aus Forschungseinrichtungen und als auch aus Museen und Landesdenkmalämtern einzubinden.

Gegenüber vielen anderen Disziplinen ist die Heterogenität der Daten in den Altertumswissenschaften und der Archäologie sehr groß. Zunächst erschwert dies sowohl die Schaffung eines einheitlichen Zugriffs als auch die Anwendung einfacher Verarbeitungsprozeduren. Allerdings sind dabei keine grundlegenden technischen Herausforderungen zu überwinden, auch wenn einzelne Anpassungen und die Erstellung von Ergänzungskomponenten auf technischer Seite notwendig sein werden. Die Schwierigkeit liegt vielmehr in dem durch das Volumen und die Heterogenität bedingten Arbeitsaufwandes und der organisatorischen Integration von Daten unterschiedlicher Herkunft. Es wird daher zur Reduktion des Aufwandes dringend vorgeschlagen, sich möglichst an existierenden Standards zu orientieren. Dabei sind der Aufbau einer geeigneten Organisationsstruktur und die Durchführung der Datenmigration wesentliche Kernelemente der Umsetzung des geplanten Forschungsdatenzentrums.

Für die Organisationsstruktur wird eine 3-Tier Architektur empfohlen, deren untere (Speicher-)Ebene durch die direkte Datenspeicherung einschließlich der Bereitstellung der notwendigen Hardware gebildet wird. Diese Ebene ist von der inhaltlichen Bedeutung der Daten weitgehend entkoppelt und bietet für den Datenzugriff einheitliche Schnittstellen. Ein Wechsel der Betreiber dieser unteren Ebene ist möglich und nur mit einem relativ geringen Arbeitsaufwand verbunden. Die mittlere (Vermittlungs-) Ebene ist verantwortlich für die inhaltliche Datenpflege und damit insbesondere für die Bewahrung der Einhaltung des Metadatenstandards bei neu eingestellten Daten. Außerdem werden Schnittstellen und einfache Verarbeitungs- und Suchprozeduren für die Anwender Ebene zur Verfügung gestellt. In der oberen (Anwender-) Ebene können die Forscher auf die Daten zugreifen und für die Verarbeitung der Daten eigene Verfahren benutzen, die auf den Prozeduren und Schnittstellen der mittleren Ebene aufbauen. Unter Berücksichtigung der Langfristigkeitsanforderungen nimmt die Vermittlungsebene eine zentrale Rolle ein. Um allen Anforderungen gerecht zu werden, sollte die Vermittlungsebene einen Betreiber haben, der sowohl bezüglich der verschiedenen Datenlieferanten als auch der Datennutzer eine neutrale Position einnimmt. Außerdem muss die Finanzierung des Betreibers langfristig gesichert sein. Aus praktischen Erwägungen bietet sich hier zunächst nur eine primäre Finanzierung durch den Bund an.

Das originale Projekt sieht eine dreijährige Konzeptionsphase und eine dreijährige Umsetzungsphase vor. Dies ist zeitlich nicht realisierbar. Innerhalb der dreijährigen Umsetzungsphase wird vorgeschlagen einen Kern des Forschungsdatenzentrums aufzubauen, der lediglich die Daten einiger

weniger Beteiligten beinhaltet. Neben einer sorgfältigen Auswahl von engagierten Beteiligten ist dazu zunächst eine detaillierte Festlegung des zu verwendenden Metadatenschemas und der Schnittstellen notwendig. Hinzukommen weitere Anforderungen zum Beispiel bezüglich des Rollenkonzeptes, die zur Auswahl einer Basissoftware führen. Nach dem Aufbau eines Grundsystems sollte der Schwerpunkt dann auf der Datenmigration einschließlich einer Datenaufarbeitung liegen. Schließlich sind ausgewählte Nutzer frühzeitig zu beteiligen, um die Anwendungsfreundlichkeit sicherzustellen. Mittelfristig können dann weitere Daten eingebunden werden, bis langfristig ein nationales Forschungsdatenzentrum für die Archäologie und die Altertumswissenschaften entsteht.

Inhaltsverzeichnis

1	Einleitung	6
2	Ausgangspunkt	8
2.1	Potentielle Teilnehmer	8
2.1.1	Außeruniversitäre Forschungseinrichtungen	8
2.1.2	Universitäten	9
2.1.3	Museen	9
2.1.4	Landesdenkmalämter	10
2.1.5	Weitere Beteiligungsaspekte	10
2.2	Funktionale Anforderungen und Vorstellungen	11
3	Anforderungsanalyse	15
3.1	Datenbereitstellung	15
3.1.1	Altdaten	15
3.1.2	Aktuelle Forschungsdaten	17
3.2	Datenverwaltung	17
3.2.1	Forschungsdaten und Metadaten	17
3.2.2	Metadatenextraktion	19
3.2.3	Konvertierung und Validierung	20
3.2.4	Datensicherung	20
3.3	Datenverarbeitung	20
4	Technische Lösungen	22
4.1	Architektur des Forschungsdatenzentrums	22
4.2	Komponenten	23
4.2.1	Lösungskonzepte aus dem kommerziellen Bereich	23
4.2.2	Lösungskonzepte aus dem akademischen Bereich	24
5	Organisatorische Lösungen	28
5.1	Rechte und Pflichten der Beteiligten	28
5.1.1	Speicherebene	28
5.1.2	Verwaltungsebene	28
5.1.3	Nutzer der Daten	29
5.1.4	Besitzer der Daten	29
5.2	Finanzierungskonzept	30
5.3	Rechtliche Fragen	31

6	Umsetzung	32
6.1	Projektablauf.....	32
6.2	Anreizsystem für die Beteiligung von potentiellen Interessenten.....	32
7	Zusammenfassung der Schlussfolgerungen	34

1 Einleitung

Die vorliegende Expertise wurde im Rahmen des von der Deutschen Forschungsgemeinschaft geförderten Projekts „IANUS- Forschungsdatenzentrum Archäologie & Altertumswissenschaften“ für das Deutsche Archäologische Institut (DAI) erstellt. Die Details des Projektes sind auf der Webseite <http://www.dainst.org/de/forschungsdatenzentrum> beschrieben. Daher erfolgt hier nur eine kurze Zusammenfassung. Das Projekt soll aus zwei Phasen bestehen. Die erste Phase dauert drei Jahre und soll Vorgehensweisen für den Aufbau und den Betrieb eines Forschungsdatenzentrums für archäologische und altertumswissenschaftliche Daten erarbeiten. Diese Vorgehensweisen und ihr Reifegrad wurden in der Projektbeschreibung nicht abschließend spezifiziert. Es kann sich lediglich um Konzepte oder um die Beschreibung von Lösungsansätzen oder um konkrete Handlungsanweisungen handeln. Als Methodik wurde die Erarbeitung in speziell eingerichteten Arbeitsgruppen und Workshops gewählt. Dabei sind zunächst die Anforderungen und Bedürfnisse an ein derartiges Forschungsdatenzentrum zu spezifizieren. Neben der offensichtlichen quantitativen Ausweitung sollen durch das Forschungsdatenzentrum auch eine qualitative Verbesserung der Forschungsdaten erreicht werden. In dieser Expertise wird sowohl von einer Verbesserung durch einheitliche Zugriffswerkzeuge als auch von einer auf Metadaten basierenden Verbesserung, die eine bessere Benutzung der Daten ermöglicht, ausgegangen. Eine technische Verbesserung der Forschungsdaten durch geeignete Aufbereitungsmethoden ist dazu orthogonal und wird hier nicht betrachtet. Aufgrund der Ergebnisse der ersten dreijährigen Phase soll in der zweiten dreijährigen Phase der eigentlichen Aufbau und die Inbetriebnahme des Zentrums folgen. Zu diesem Zeitplan wird im Rahmen der Expertise ebenfalls Stellung genommen.

Zum Zeitpunkt der Beauftragung der Expertise lag eine geplante Beschreibung der inhaltlichen, funktionalen, rechtlichen und betriebswirtschaftlichen Ausgestaltung des Forschungsdatenzentrums aufgrund des beschriebenen Zeitplans noch nicht vor. Daher basiert die Expertise im Wesentlichen auf der funktionalen Grobplanung des Zentrums, die sich an Datenzentren anderer Disziplinen orientiert, wobei als funktionaler Schwerpunkt die Langzeitarchivierung von Forschungsdaten festgelegt wurde. Unter dem Begriff *Langzeit* ist dabei ein Zeitrahmen von mindestens 30-50 Jahren gemeint. Die angesprochenen Forschungsdaten sind zunächst statisch und umfassen das gesamte Spektrum von unbearbeiteten Rohdaten bis hin zu wissenschaftlichen Abhandlungen. Auf diese Daten soll der Nutzer bei Vorliegen entsprechender Berechtigungen einen einfachen Zugriff haben. Dieser Zugriff schließt insbesondere auch eine umfassende Suchfunktionalität ein. Technisch soll dieser Zugriff möglicherweise über Webportale oder Webservices realisiert werden. Mittelfristig sollte die Unterstützung bei der Durchführung von laufenden Projekten berücksichtigt werden, was somit auch dynamische Daten einschließt.

Die Initiatoren des Projektes stellen sich weiterhin vor, dass durch das Forschungsdatenzentrum auch weitere Dienstleistungen für die Disziplin angeboten werden. Diese Dienstleistungen stehen allerdings in einem engen Zusammenhang mit den in dem Zentrum gespeicherten Daten und beinhalten Aspekte wie Beratung, Vermittlung von externen Angeboten durch ein Kompetenzzentrum, Unterstützung bei der Einrichtung von Workflows für aufwändige Verarbeitungsroutinen, Definition und Überprüfung der Einhaltung von Standards sowie die Bereitstellung von Informations- und Schulungsmaterial. Die Gesamtarchitektur muss zumindest Erweiterungsmöglichkeiten in diesem Bereich vorsehen.

Die Beteiligung beim Aufbau des Zentrum beschränkt sich nicht auf die Forschungseinrichtungen im engen Sinn, wie Universitäten und wissenschaftliche Einrichtungen des Bundes, sondern soll auch weitere Institutionen in Deutschland berücksichtigen, die ein Interesse an der archäologisch-

altertumswissenschaftlichen Forschung haben. Dazu gehören Landesdenkmalämter, Akademien und Museen. Damit liegt eine sehr heterogene Gemeinschaft vor, deren Interessen nicht überall übereinstimmen.

In diesem Umfeld soll die Expertise Aussagen darüber treffen, ob und in welchem Umfang in Deutschland bereits existierende Infrastruktureinrichtungen für das Forschungsdatenzentrum genutzt und welche Konzepte von vergleichbaren Initiativen verwendbar sind. Hierbei handelt es sich jedoch nicht um eine Umsetzungsanleitung im engeren Sinn sondern um eine vergleichende Darstellung. Aufgrund der noch nicht existierenden genauen Spezifikation des geplanten Forschungsdatenzentrums und des wie in der Expertise dargestellten langen Entwicklungszeitraums sind spezifische Umsetzungsvorschläge zu diesem Zeitpunkt nicht sinnvoll. Vielmehr wird in der Expertise auf grundlegende Aspekte und Rahmenbedingungen eingegangen. Diese Aspekte beschränken sich nicht auf Technologien sondern berücksichtigen auch finanzielle und in einem geringen Umfang rechtliche Rahmenbedingungen. Besonderes Augenmerk liegt auf der organisatorischen Struktur der Lösung.

2 Ausgangspunkt

In diesem Abschnitt wird zunächst die Ausgangslage für ein Forschungsdatenzentrum der Archäologie und der Altertumswissenschaften erläutert und bezüglich dieser Ausgangslage eine erste Bewertung abgegeben.

2.1 *Potentielle Teilnehmer*

Es ist beabsichtigt, in das geplante Forschungsdatenzentrum langfristig alle Forschungsinteressierten aus dem Bereich Archäologie und Altertumswissenschaften einzubeziehen. Zur Bewertung dieser Zielsetzung werden zunächst die potentiellen Beteiligten an einem solchen Forschungszentrum anhand der vier vorgegebenen Kategorien *außeruniversitäre Forschungseinrichtungen, Universitäten, Museen* und *Landesdenkmalämter* betrachtet.

Forschungsdaten werden in vielen Disziplinen vorwiegend oder sogar ausschließlich direkt von Forschern an Universitäten und außeruniversitären Forschungseinrichtungen erhoben und verwaltet, siehe zum Beispiel die Hochenergiephysik oder die Astrophysik. Im Bereich der Archäologie und der Altertumswissenschaften liegt dagegen eine andere Situation vor: Aufgrund des öffentlichen Interesses an der Thematik und ihrer Bedeutung für das kulturelle Erbe bezogen auf Regionen oder Bevölkerungsgruppen werden Museen eingerichtet, die neben der Bewahrung der Forschungsgegenstände und –daten auch die Aufgabe haben, Inhalte der Archäologie und der Altertumswissenschaften der Bevölkerung näher bringen. Weiterhin gibt es Landesdenkmalämter oder ähnliche Landeseinrichtungen, die zumindest teilweise einen expliziten im Gesetz verankerten Forschungsauftrag besitzen und vielfach eng mit Museen zusammenarbeiten. Daraus ergeben sich die oben erwähnten vier Kategorien von Institutionen, die an der Forschung in der Archäologie und den Altertumswissenschaften interessiert sind.

2.1.1 **Außeruniversitäre Forschungseinrichtungen**

Unter außeruniversitären Forschungseinrichtungen werden hier Einrichtungen verstanden, die sich ausschließlich der Forschung in einem speziellen Themenbereich widmen und dabei in der Regel keine oder nur eingeschränkte Lehraufgaben haben. Dabei wird sich auf Einrichtungen beschränkt, die es in einer ähnlichen Strukturierung auch in anderen Disziplinen gibt, wie die Institute der Max-Planck Gesellschaft oder Leibniz-Institute. Als Beispiel für den Bereich der Archäologie sei das Max-Planck Institut für Evolutionäre Anthropologie und das darin enthaltene Max-Planck Weizmann Center für integrative Archäologie und Anthropologie erwähnt. In solchen Forschungseinrichtungen arbeiten vielfach mehr Wissenschaftler in einem Themenbereich als dies in Universitäten der Fall ist, wobei die im Rahmen der Exzellenzinitiative gebildeten Forschungscluster eine Ausnahme bilden. Damit werden in solchen Einrichtungen mehr Forschungsdaten benötigt und es entstehen mehr abgeleitete Forschungsdaten und Publikationen als vielfach an Universitäten. Diese thematische Konzentration ist wesentliches Merkmal von außeruniversitären Forschungseinrichtungen. Um die auch aufgrund dieser Konzentration von Forschern benötigte Menge an Forschungsdaten effizient verwalten zu können, werden in solchen außeruniversitären Forschungseinrichtungen vielfach bereits strukturierte Verfahren für das Datenmanagement eingesetzt. In anderen Disziplinen sind auch aufgrund dieser Erfahrungen außeruniversitäre Forschungseinrichtungen Betreiber von nationalen oder internationalen Forschungsdatenzentren. Als analoges Beispiel aus dem Bereich Klimaforschung kann das World Data Center Climate angeführt werden, das vom Max-Planck Institut für Meteorologie beherbergt wird. Viele außeruniversitäre Forschungseinrichtungen in Deutschland werden vorwiegend vom Bund

finanziert. Auch das Deutsche Archäologische Institut wird hier zu diesen außeruniversitären Forschungseinrichtungen gerechnet, da es als Bundesanstalt zum Geschäftsbereich des Auswärtigen Amtes gehört.

2.1.2 Universitäten

Altertumswissenschaften sind an vielen Universitäten zu finden und decken dabei ein sehr breites Spektrum ab. Allerdings überwiegen an den Universitäten Einzelaktivitäten während lokale Schwerpunkte, wie das Exzellenzcluster der HU und FU Berlin *Topoi* und einige Sonderforschungsbereiche eher die Ausnahme bilden. Hinzu kommt, dass diese Schwerpunkte grundsätzlich Projektcharakter und nur sehr eingeschränkt auf einen nachhaltigen Betrieb ausgerichtet werden. Obwohl von der DFG als dem dominanten Projektträger solcher universitären Schwerpunkte vielfach die Forderung nach Nachhaltigkeit erhoben wird, gestaltet sich eine nachhaltige Umsetzung in der Praxis als sehr schwierig. Sobald die DFG-Förderung beendet ist und keine Aussicht auf eine Anschlussförderung besteht, nimmt die Forschungsaktivität in dem Bereich sehr schnell ab, da die Universitäten aus eigenen Mitteln eine große Forschungsaktivität nicht erhalten können. Ohne eine kritische Größe zu erreichen, besteht in der Regel auch keine Möglichkeit, die notwendige Infrastruktur, sofern sie überhaupt innerhalb der Projektdauer aufgebaut wurde, zu erhalten. Im ungünstigsten Fall führt ein Projektende auch zum Verlust von Forschungsdaten, da gerade junge Forscher nach dem Projektabschluss und ihrer Promotion die Universität verlassen und damit unbeabsichtigt auch Kenntnisse über den Zugriff auf die Daten nicht mehr vorhanden sind. Aus Gründen der Nachhaltigkeit und der im Allgemeinen geringen Forschungskonzentration bietet es sich nicht an, ein Forschungsdatenzentrum an einer Universität anzusiedeln. Weiterhin stehen die Universitäten unter Landeshoheit und sind damit in diesem Bereich automatisch auf das jeweilige Bundesland ausgerichtet. Auch wenn es Fälle gibt, in denen Universitäten nationale Aufgaben übernehmen, wie zum Beispiel im Fall der Universität Stuttgart und der Beteiligung am Gauß-Supercomputing-Center, sind Universitäten für nationale Aufgabe meist nicht die erste Wahl. Allerdings spielen die Forscher an Universitäten aufgrund ihrer Gesamtzahl eine sehr wichtige Rolle als Nutzer des geplanten Forschungsdatenzentrums.

2.1.3 Museen

Museen können Staatseigentum sein, sie können sich in privater Hand befinden oder auch einer Stiftung gehören. Entsprechend heterogen kann die Interessenslage der Museen sein. Eine wichtige Rolle spielt außerdem die Leitung des jeweiligen Museums, der oft ein breiter Gestaltungsspielraum eingeräumt wird. Das Forschungsinteresse von Museen kann Folgen von direkten Aufträgen zur Forschung sein oder es kann sich aus dem Ziel ergeben, Zusammenhänge zu ergründen und dann der Öffentlichkeit besser darstellen zu können. Der Fokus auf die Öffentlichkeit und die Darstellung von Forschungsgegenständen für die Öffentlichkeit wird manchmal dazu führen, dass sich das Forschungsinteresse an Museen von dem Forschungsinteresse von Wissenschaftlern an Universitäten und außeruniversitären Forschungseinrichtungen, das vor allem auf dem Wunsch nach Erkenntnisgewinn basiert, unterscheidet. Weiterhin ist zu berücksichtigen, dass die den Museen zur Verfügung stehenden Ressourcen oft sehr begrenzt sind. Ein eigenständiges Engagement von Museen wird daher selbst bei Interesse schnell an seine Grenzen stoßen. Als Lieferanten von Forschungsdaten sind Museen unverzichtbar, da sie über eine Vielzahl von Artefakten verfügen, die Gegenstand von Forschungsanstrengungen von Wissenschaftlern sein können. Es ist allerdings nicht klar, in welchem Umfang die Forschungsdaten von Museen schon jetzt in digitaler Form vorliegen oder ob zunächst

hier eine umfangreiche Digitalisierung erforderlich ist. Hierzu ist eine separate Untersuchung notwendig.

2.1.4 Landesdenkmalämter

Aufgrund der Kulturhoheit der Länder ist der Denkmalschutz Sache der Länder, die hierzu Denkmalschutzgesetze erlassen haben. Die Aufgaben des Denkmalschutzes werden von den Landesdenkmalämtern wahrgenommen. So definiert die Webseite des Landesdenkmalamtes Berlin die Aufgabe der archäologischen Denkmalpflege als die Untersuchung und Konservierung von „bewegliche(n) und unbewegliche(n) Zeugnisse(n) der Vergangenheit, die sich im Boden oder in Gewässern befinden.“ Der Ausdruck *Untersuchung* definiert damit auch eine Forschungstätigkeit. Landesdenkmalämter können eine obere Landesbehörde mit entsprechenden Kompetenzen oder sie können auch nur eine beigeordnete Behörde sein. Sie haben die Aufgabe gemeinsam mit den Landesmuseen die entsprechend ihres Auftrags erhobenen Daten zu sammeln und sie den Bürgern bei Bedarf zur Verfügung zu stellen. In der Regel liegen ihre Daten in komplett digitalisierter Form vor. Sie besitzen im Allgemeinen auch die Datenhoheit, die schon aus rechtlichen Gründen nicht aufgegeben werden kann. Außerdem können Landesdenkmalämter in manchen Fällen aus diesen Daten auch Erlöse erwirtschaften, was ein weiterer Grund für eine sorgfältige Kontrolle und restriktive Handhabung der Datenweitergabe ist.

2.1.5 Weitere Beteiligungsaspekte

Aufgrund der Heterogenität der Beteiligten, des nationalen Charakters des Forschungsdatenzentrums und des geplanten Archivierungszeitrahmens von mindestens 30 bis 50 Jahren ist der verantwortliche Betrieb des Forschungsdatenzentrums durch eine Bundesinstitution erforderlich. Hierbei ist festzuhalten, dass unter dem Begriff *verantwortlicher Betrieb* keine technischen Aspekte berücksichtigt wurden. So kann aufgrund der heutigen Technik eine Einrichtung den verantwortlichen Betrieb übernehmen ohne einen Datenserver direkt zu betreiben. Hierauf wird unter dem Abschnitt Architektur noch näher eingegangen.

Rechtlich ist der Zusammenhang mit der Kulturhoheit der Länder zu klären. Dies betrifft in der Archäologie und den Altertumswissenschaften nicht nur Forschung und Lehre sondern wie dargestellt auch den Denkmalschutz. Nachteilhaft ist weiterhin die unterschiedliche Verankerung des Themas bei mehreren Bundesinstitutionen. So gehört das DAI in den Geschäftsbereich des Auswärtigen Amtes, die Max-Planck Institute werden vom BMBF gefördert, während das Umweltministerium für den Denkmalschutz auf Bundesebene zuständig ist. Da das Forschungsdatenzentrum Dienstleistungen für alle Forschungseinrichtungen im Bereich der Archäologie und Altertumswissenschaften erbringen soll, ist eine Konsolidierung der Aktivitäten des Bundes in dieser Hinsicht sehr wünschenswert. Andernfalls kann kaum eine gemeinschaftliche Aktivität unter Einbezug der Länder erwartet werden. Die langfristige Rolle der DFG ist dem Verfasser ebenfalls nicht klar. In der Regel ist die DFG für eine zeitlich eng befristete Projektförderung verantwortlich und kommt daher für eine langfristige Finanzierung eher nicht in Frage.

Aufgrund der großen Anzahl von Beteiligten und ihrer starken Heterogenität ist der Aufbau eines Forschungsdatenzentrums in drei Jahren nicht möglich. Es handelt sich vielmehr um einen Prozess, der sich über Dekaden hinziehen wird. Vor der eigentlichen Umsetzung ist zumindest auf der Ebene des Bundes die Klärung der Verantwortlichkeit erforderlich. Allein dieser Prozess wird im ungünstigsten Fall mehrere Jahre in Anspruch nehmen. Dabei ist zu berücksichtigen, dass in der ersten

Phase der Konzepterstellung nach Kenntnis des Verfassers die betroffenen Bundesministerien noch nicht einbezogen sind. Es wird daher vorgeschlagen, in der dreijährigen Umsetzungsphase einen Prototyp für ein Forschungsdatenzentrum aufzusetzen, der eine Abschätzung des Umsetzungsaufwandes ermöglicht, die Funktion des Forschungsdatenzentrums beschreibt und für potentielle Beteiligte Anschauungsmaterial liefert. Dafür ist eine kleine Gruppe von interessierten Einrichtungen zu wählen, die das Projekt tragen. Diese Gruppe muss einerseits die erwähnten Kategorien von Beteiligten möglichst gut vertreten und andererseits so klein sein, dass eine effiziente Zusammenarbeit möglich ist. Ein hohes Maß an Engagement für das Projekt ist selbstverständliche Voraussetzung.

Schlussfolgerung 1: Die Verantwortlichkeit für das Forschungsdatenzentrum soll eine vom Bund finanzierte Einrichtung übernehmen und damit den nationalen Charakter des Forschungsdatenzentrums und die Nachhaltigkeit gewährleisten.

Schlussfolgerung 2: Die Verankerung des Forschungsdatenzentrums beim Bund unter Berücksichtigung der Kulturhoheit der Länder ist in Bezug auf rechtliche Aspekte zu klären.

Schlussfolgerung 3: Die Zuordnung des Forschungsdatenzentrums zu einem Bundesministerium muss festgelegt werden. In diesem Zusammenhang ist eine Konsolidierung der Aktivitäten des Bundes anzustreben.

Schlussfolgerung 4: Innerhalb der vorgesehen Umsetzungszeit von drei Jahren kann nur ein Prototyp eines Forschungsdatenzentrums erstellt werden. Hieran sollte nur eine kleine Gruppe von Einrichtungen beteiligt sein. Die Gruppe sollte die genannten Kategorien der Beteiligten möglichst gut repräsentieren können.

2.2 Funktionale Anforderungen und Vorstellungen

Die Aufgabenstellung der Expertise gibt folgende Untergliederung von Anforderungsaspekten bei gleichzeitiger Angabe einer Priorisierung an.

1. Zentralisierte Datenablage von aktuellen, dynamischen Projektdaten (mittlere Priorität)
2. Online Präsentationsmöglichkeit für Projekte (niedrige Priorität)
3. (Langzeit-)Archivierung von statischen Forschungsdaten (hohe Priorität)
4. Unterstützung von rechenintensiven Analyseverfahren (niedrige Priorität)
5. Online-Bereitstellung von Forschungsdaten via Portalen (hohe Priorität)
6. Online-Bereitstellung von Forschungsdaten via (Web-)Services (hohe Priorität)
7. Spiegelung von Datenbeständen externer Institutionen (mittlere Priorität)

Forschungsdatenzentren (FDZ) verwalten in der Regel primär statische Datensätze. Dabei kann es sich sowohl um Rohdaten als auch um abgeleitete Daten bis hin zu wissenschaftlichen Publikationen zu diesen Rohdaten handeln. Auch wenn die Daten als statisch bezeichnet werden, bedeutet dies nicht, dass sie unveränderlich sind. Notwendig können zum Beispiel nachträgliche Fehlerkorrekturen, Formatänderungen oder Ergänzungen sein. Obwohl Forscher insbesondere bei neuen Aktivitäten vor allem an dem aktuellen Datensatz interessiert sind, müssen Veränderungen des Datensatzes und ihre Gründe später nachvollziehbar sein. Manche Forschungsdatenzentren beherbergen auch Datensätze zu aktuellen Projekten, die häufigen Veränderungen unterliegen und im Allgemeinen nur den Projektteilnehmern zugänglich sind. Diese Datensätze werden als dynamische Daten bezeichnet. In vielen Projekten werden dynamische Daten lokal bei den die Projekte ausführenden Forschern und den zugehörigen Einrichtungen gespeichert. Ausnahmen bilden Projekte mit sehr hohen Datenvolumina oder Projekte, die Daten verwenden, die eine besondere Behandlung erfordern zum Beispiel aus Gründen des Datenschutzes. Im Fall der Archäologie und der Altertumswissenschaften treffen diese Rahmenbedingungen unter Berücksichtigung der aktuellen Speichertechnologie nach Kenntnis des Verfassers nicht zu. Die Speicherung von dynamischen Daten ist daher als eine besondere Dienstleistung zu betrachten, die im Hinblick auf einen späteren Transfer der Daten nach Projektabschluss und frühzeitige Berücksichtigung der für die Aufnahme von statischen Datensätze in das Forschungsdatenzentrum zu beachtenden Regeln sinnvoll sein kann, aber nicht zu den Kernaufgaben eines Forschungsdatenzentrums gehört. Technologisch stellt die Speicherung von dynamischen Datensätzen keine besondere Schwierigkeit dar und kann auch später noch integriert werden. Es wird daher vorgeschlagen, im Rahmen des geplanten Prototyps nicht die Speicherung von dynamischen Daten zu berücksichtigen. Ergänzend sollte hier noch über den Begriff „zentralisiert“ gesprochen werden. Damit sollte kein zentralisierter Speicherort gemeint sein. Vielmehr sollte „zentralisiert“ als eine einzelne Zugriffstelle unabhängig von dem Ort des Zugreifenden aufgefasst werden. Der tatsächliche Speicherort ist für den Nutzer uninteressant, sofern die geforderten Zugriffsrahmenbedingungen gewährleistet sind. Über diesen Speicherort entscheidet daher alleine das Forschungsdatenzentrum.

Schlussfolgerung 5: Die Speicherung von dynamischen Daten wird als potentiell sinnvolle zusätzliche Dienstleistung eines Forschungsdatenzentrums gesehen. Im Rahmen des Prototyps sollte sich allerdings auf den Kernbereich eines Forschungsdatenzentrums beschränkt werden.

Die Online Präsentationsmöglichkeit für Projekte erfordert die Existenz von Projektdaten. Da vorgeschlagen wird, Projektdaten in der ersten Phase nicht zu betrachten, entfällt auch die Grundlage für eine Online Präsentationsmöglichkeit. Auch bei der Online Präsentationsunterstützung handelt es sich um eine Zusatzleistung. Wie bei allen anderen Zusatzleistungen ist vor einer Umsetzung zunächst in Zusammenarbeit mit potentiellen Nutzern zu klären, ob ein ausreichender Bedarf an einer solchen Zusatzleistung besteht.

Schlussfolgerung 6: Vor der Umsetzung von Zusatzleistungen und/oder der Unterstützung von Online Präsentationen ist der voraussichtliche Bedarf zu klären.

Die Langzeitarchivierung von Forschungsdaten ist die Kernaufgabe des Forschungsdatenzentrums und steht daher im Mittelpunkt dieser Expertise.

Die Unterstützung von rechenintensiven Analyseverfahren ist für die Forschungsdatenzentren mancher Communities eine wesentliche Erweiterung in Richtung auf den Aufbau einer virtuellen Forschungsumgebung. Eine solche Erweiterung erlaubt die Durchführung von Workflows ohne

direkte Einbeziehung des Nutzers. Eine solche Erweiterung sollte nur dann in den originären Prototypen integriert werden, wenn ein hoher Bedarf an solchen rechenintensiven Analyseverfahren bereits zum aktuellen Zeitpunkt besteht. Aufgrund von Gespräche mit Wissenschaftlern aus dem Bereich der Archäologie und der Altertumswissenschaften liegt ein solcher Bedarf zurzeit nicht vor. Allerdings sollte eine Erweiterungsmöglichkeit in dieser Richtung vorgesehen werden, da solche Verfahren sich in der Vergangenheit als zunehmend wichtig für viele Disziplinen erwiesen haben.

Schlussfolgerung 7: Die Unterstützung von rechenintensiven Analyseverfahren ist nicht in den Prototypen zu integrieren. Allerdings ist in der Architektur schon jetzt eine Erweiterung in dieser Richtung vorzusehen.

Für den Nutzer besteht ein Bedarf an einem möglichst einfachen Zugriff auf die Daten des Forschungsdatenzentrums. Diese Funktion fällt daher in den Kernbereich des Forschungsdatenzentrums und ist unbedingt im Prototyp zu implementieren. Der Zugriff sollte möglichst unabhängig von dem Betriebssystem des Nutzers und von Zusatzinstallationen sein, damit eine hohe Nutzerfreundlichkeit erreicht wird. Daher bietet sich die Verwendung des Webs und von Portalen an.

Schlussfolgerung 8: Der Zugriff auf die Daten durch Nutzer sollte möglichst unabhängig von dem IT-System des Nutzers erfolgen. Für diesen Zweck bieten sich Webportale wie vorgeschlagen an.

In der Aufgabenstellung wird dem Zugriff über Webservices ebenfalls eine hohe Priorität zugewiesen. Webservices können ein Mechanismus sein, der bei Webportalen zum Einsatz kommt. In diesem Fall ist der Webservice unsichtbar für den Nutzer und sollte als technische Realisierung im Hintergrund nicht separat aufgeführt werden. Des Weiteren können Webservices auch direkt verwendet werden, wenn Anwendungsprogramme des Nutzers auf die Daten ohne zusätzliche Beteiligung des Nutzers zugreifen sollen. Ebenso wie rechenintensive Analyseverfahren erweitert auch diese Eigenschaft das Forschungsdatenzentrum in Richtung auf eine virtuelle Forschungsumgebung. Für den Prototyp bietet sich eine solche Erweiterung nur dann an, wenn sie mit der Unterstützung eines Anwendungsprogramms verbunden ist. Aufgrund der beschränkten Implementierungszeit sollte hierzu nur ein Anwendungsprogramm verwendet werden, das sehr weit verbreitet ist und häufig eingesetzt wird. Ein solches Programm ist dem Verfasser nicht bekannt.

Schlussfolgerung 9: Die explizite Integration von Webservices zum Zugriff auf Daten lohnt sich nur bei gleichzeitiger Implementierung von Anwendungsprogrammen, die diese Webservices verwenden. Ansonsten sollte im Rahmen des Prototyps darauf verzichtet werden.

Die redundante Datenspeicherung ist gerade bei der Langzeitarchivierung unverzichtbar. Sie dient primär der Erhöhung der Datensicherheit und kann sekundär auch die Verbesserung der Zugriffsleistung unterstützen. Gegenwärtig ist nicht zu erkennen, dass die Zugriffsleistung auf die Daten auf Seiten der Speicheranbieter ein Engpass sein wird. Daher sollte zurzeit der ausschließliche Schwerpunkt auf der Datensicherheit liegen. Die Datensicherheit kann zum Beispiel bereits durch eine Spiegelung der gespeicherten Daten gewährleistet werden. Dies kann auf der unteren Ebene der 3-Tier Architektur realisiert und sollte dann durch entsprechende Servicevereinbarungen abgesichert werden. In diesem Fall ist die redundante Datenspeicherung technisch gesehen weder für die Nutzer noch für die Administration des Forschungsdatenzentrums sichtbar. Wenn allerdings einige Datenlieferanten

auf einem Fortbestehen einer eigenen Datenspeicherung bestehen, kann auch auf einer höheren Ebene die notwendige Redundanz geschaffen werden. Allerdings ist dann auch die Kompatibilität der Schnittstelle und die Konsistenz der Daten durch das Forschungsdatenzentrum zu gewährleisten. Deshalb entsteht bei diesem Ansatz potentiell ein höherer Aufwand für das Forschungsdatenzentrum. Aus ökonomischen Gründen ist daher eine Beschränkung der Redundanz auf die untere Speicherebene sinnvoll.

Schlussfolgerung 10: Eine redundante Datensicherung ist unverzichtbar. Sie sollte auf der Ebene der Speicheraanbieter umgesetzt werden.

Die von Forschungsdatenzentren verwalteten Daten sollen in komplett digitalisierter Form vorliegen. Dies ist bei historischen Daten nicht unbedingt gegeben. Langfristig wünschenswert ist in solchen Fällen eine Digitalisierung, die allerdings einen erheblichen Aufwand bedeuten kann. In der Zwischenzeit können auch nur Metadaten mit entsprechenden Verweisen auf die Aufbewahrungsort der Originaldaten verwendet werden. Die Digitalisierung von historischen Daten und der Aufbau von Zwischenstufen sollten von dem Aufbau des Datenzentrums separiert betrachtet werden. Es wird daher vorgeschlagen, im Rahmen des Prototyps sich auf vollständig digitalisierte Daten zu beschränken.

Schlussfolgerung 11: Im Rahmen des Prototyps sollte sich ebenfalls aus Gründen einer Fokussierung und einer schnellen Umsetzung auf die Aufnahme vollständig digitalisierter Daten beschränkt werden.

Eine in der Aufgabenstellung nicht angesprochene Erweiterung des Angebots eines Forschungsdatenzentrums ist die Unterstützung von Methoden zum kollaborativen Arbeiten von mehreren Wissenschaftlern an unterschiedlichen Orten. Dies gilt als einer der Hauptgründe für die Einführung einer virtuellen Forschungsumgebung. Auch diese Zusatzleistung sollte nicht in den Prototypen sofort integriert werden, aber die Voraussetzungen für eine Integration sind zu schaffen, damit später keine Neukonzeption notwendig wird.

Schlussfolgerung 12: Langfristig wird das Forschungsdatenzentrum wahrscheinlich sein Angebot in Richtung auf eine virtuelle Forschungsumgebung ausdehnen müssen. Dies betrifft insbesondere die Zusammenarbeit von Wissenschaftlern an unterschiedlichen Orten. Schon beim Prototyp sollten die Voraussetzung für eine spätere Integration vorgesehen werden. Eine vollständige Umsetzung von entsprechenden Dienstleistungen bereits im Prototyp wird aufgrund des Aufwandes nicht empfohlen.

3 Anforderungsanalyse

In diesem Abschnitt wird auf die technischen Anforderungen des Kernbereichs des Forschungsdatenzentrums eingegangen. Dabei werden vorwiegend die Anforderungen behandelt, die den vorgeschlagenen Prototyp betreffen, während spätere Erweiterungen nur kurz erwähnt werden. Bezüglich der Erweiterungen gibt es auch keinen Anspruch auf Vollständigkeit.

3.1 Datenbereitstellung

Für die in das Forschungsdatenzentrum zu übernehmenden Daten, müssen keine Datenschutzbestimmungen berücksichtigt werden, da nach Auskunft von mit den Daten beschäftigten Personen die Daten nicht personenbezogen sind. Die Daten liegen in einer Vielzahl von Formaten vor. So gehören Textdaten, Bilddaten, Videodaten, Zeichnungen, verschiedene Arten von Messdaten und Simulationsdaten zu den Forschungsdaten. Für viele Daten gibt es unterschiedliche Formate. Es wird nicht empfohlen für einen Datentyp nur ein einziges Format zu akzeptieren, da dies entweder zu einem sehr hohen Konversionsaufwand führt oder die Akzeptanz des Forschungsdatenzentrums bei den potentiellen Nutzern deutlich reduziert. Andererseits kann auch nicht jedes Format als Standard akzeptiert werden, da ansonsten der Aufwand im Forschungsdatenzentrum nicht handhabbar ist. Hier ist ein Kompromiss zu finden, der auf inhaltlichen Argumenten basiert. Die Menge der akzeptierten Formate kann im Bedarfsfall später erweitert werden.

Weiterhin ist für die Integration der Forschungsdaten ein funktionierendes Rechtemanagement notwendig, da einige Einrichtungen, die Daten besitzen, wie zum Beispiel einige Landesdenkmalämter, keinen allgemeinen Zugriff auf diese Daten gestatten und damit eine auf einem Rechtemanagement basierende Zugriffskontrolle Voraussetzung für die Bereitschaft dieser Einrichtungen ist, zu dem Forschungsdatenzentrum beizutragen.

Schlussfolgerung 13: *Die unterstützten Datenformate sind vor dem Beginn der Umsetzung in einer ersten Version festzulegen. Die Menge der unterstützten Formate kann später erweitert werden. Diese spätere Erweiterung kann dann allerdings bereits zu unnötigem Konversionsaufwand zu Beginn der Umsetzung geführt haben. Daher sollte die erste Auswahl der unterstützten Formate bereits sehr sorgfältig erfolgen.*

Schlussfolgerung 14: *Das Forschungsdatenzentrum benötigt ein Zugriffsverfahren, das auf einem Rechtemanagement basiert. Der jeweilige Eigentümer der Daten muss den Zugriff festlegen können. Er ist dabei inhaltlich verantwortlich, dass die entsprechenden rechtlichen Vorgaben bezüglich des Zugriffs gewährleistet werden, während das Forschungsdatenzentrum die Verantwortung für die technische Umsetzung der Zugriffskontrolle trägt. Gleichzeitig muss die Implementierung des Zugriffsverfahrens gewährleisten, dass die Voraussetzung für einen nutzerfreundlichen und effizienten Datenzugriff geschaffen wird.*

3.1.1 Altdaten

In den verschiedenen Einrichtungen gibt es eine erhebliche Menge an Altdaten, die für die Forschung von großer Bedeutung sind. Wenn diese Daten in einem wesentlichen Umfang auf die Dauer unberücksichtigt bleiben, dann wird die Forschung in der Archäologie und in den

Alturwissenschaften potentiell in erheblicher Weise eingeschränkt. Daher sollte es das langfristige Ziel sein, möglichst viele Altdaten in das Forschungsdatenzentrum zu übernehmen. Hierbei müssen unterschiedliche Kategorien von Altdaten betrachtet werden:

1. **Daten, die nicht in digitalisierter Form vorliegen.** Ob und wenn ja wann diese Daten digitalisiert werden sollten, ist separat zu betrachten. Dies ist eine vorwiegend inhaltliche Betrachtung, die Bedarfs- und Aufwandsargumente berücksichtigen soll. Es ist allerdings im Allgemeinen davon auszugehen, dass eine solche Digitalisierung nicht umgehend erfolgt. In der Zwischenzeit kann ein digitaler Katalog eine Zwischenlösung sein. Der entsprechende Katalogeintrag enthält im günstigen Fall die Metadaten zu den Forschungsdaten und einen Verweis auf den Aufbewahrungsort dieser Daten. Er kann in das Forschungsdatenzentrum aufgenommen werden.
2. **Digitalisierte Daten, für die keine Metadaten existieren oder deren Metadatenformat nicht den unterstützten Formaten entspricht.** Im letzteren Fall ist zu entscheiden, ob das betroffene Format in ein unterstütztes Format zu konvertieren ist oder ob die Menge der unterstützten Metadaten erweitert werden soll. Auch diese Entscheidung ist auf inhaltlicher Basis zu fällen. Dieses Thema wird später noch betrachtet. Wenn entschieden wird, eine Metadatenkonversion durchzuführen, dann ist in Abhängigkeit des Quell- und des Zielformats festzustellen, ob die Konversion mit IT-Technik unterstützt werden soll. Diese Entscheidung hängt von den Formaten und von dem Umfang der Daten ab. Im Rahmen des Prototyps sollten solche Daten aus Aufwandsgründen nur in Ausnahmefällen betrachtet werden. Es ist allerdings sinnvoll, eine solche Konversion im Prototyp exemplarisch durchzuführen, um Erfahrungen zu sammeln und den Konversionsweg darzustellen.
3. **Digitalisierte Daten, deren Metadaten im gewünschten Zielformat vorliegen.** Hierbei handelt es sich um Daten, die für den Prototyp von primärem Interesse sind.

Schlussfolgerung 15: Im Rahmen des Prototyps sind die Daten von primärem Interesse, die vollständig digitalisiert sind und deren Metadaten in einem Zielformat vorliegen. Diese Daten sind vorrangig in das Forschungsdatenzentrum aufzunehmen. Daher ist vor dem Beginn der Umsetzung eine sorgfältige Analyse bezüglich des Datenbestandes durchzuführen.

Schlussfolgerung 16: Für Daten, die digitalisiert sind, aber deren Metadaten nicht in einem der Zielformate vorliegen, ist es sinnvoll exemplarisch einen Konversionsprozess im Rahmen des Prototyps durchzuführen. Diese Konversion hat allerdings nicht das Ziel, den Datenbestand des Forschungsdatenzentrums wesentlich zu erweitern sondern soll den Prozess darlegen.

Schlussfolgerung 17: Entscheidungen über die Digitalisierung von Altdaten und die Konversion von digitalisierten Daten, deren Metadaten nicht in einem Zielformat vorliegen, sind auf inhaltlicher Basis und separat vom Aufbau des Forschungsdatenzentrums zu treffen.

3.1.2 Aktuelle Forschungsdaten

Für das Forschungsdatenzentrum ist der Prozess der Übernahme von aktuellen (neuen) Forschungsdaten von größter Bedeutung. Hierbei handelt es sich auch um statische Daten. Dieser Prozess beinhaltet eine Überprüfung der Einhaltung der Standards bezüglich der Daten und der zugehörigen Metadaten. Er ist in einer ersten Form vor der eigentlichen Datenübernahme zu implementieren und später iterativ anzupassen. Die Festlegung der Art der Überprüfung muss inhaltlich begründet sein. Danach muss festgestellt werden, ob diese Überprüfung automatisiert ablaufen kann oder einer endgültigen Freigabe durch einen Verantwortlichen des Forschungsdatenzentrums bedarf. Der Übernahmeprozess ist zukünftigen Datenlieferanten deutlich darzustellen und die besondere Dienstleistung der Kompatibilitätsprüfung ist zu erläutern.

Schlussfolgerung 18: *Für das Einstellen der Daten sollte eine möglichst automatisierte Prüfung der Daten und der zugehörigen Metadaten auf Einhaltung der durch das Forschungsdatenzentrum definierten Standards implementiert werden.*

3.2 Datenverwaltung

In diesem Abschnitt wird detaillierter auf die Anforderungen hinsichtlich der Datenverwaltung in dem Forschungsdatenzentrum eingegangen. Dabei werden die verschiedenen Datentypen nicht näher betrachtet, da die notwendige Auswahl einer inhaltlichen Analyse bedarf. Von Seiten der technischen Implementierung sollte es keine Einschränkungen geben. Insbesondere ist zu berücksichtigen, dass neue Varianten existierender Formate oder neue Formate oder bisher nicht berücksichtigte Formate später als Standards im Forschungsdatenzentrum aufgenommen werden können.

Schlussfolgerung 19: *Für die Aufnahme neuer Formate und neuer Varianten existierender Formate ist ein Prozess zu definieren, um einerseits diese Möglichkeit offen zu halten und andererseits den Aufwand für die Entscheidung transparent zu halten und damit den Verwaltungsaufwand zu beschränken.*

3.2.1 Forschungsdaten und Metadaten

In diesem Abschnitt wird dargestellt, welche funktionalen Anforderungen an das System des Forschungsdatenzentrums hinsichtlich der Ablage von Forschungsdaten bestehen. Dabei sind zunächst jene grundlegenden Operationen wichtig, die in den meisten Systemen für die Handhabung von Daten unterstützt werden, wie

- das Ablegen eines Datensatzes,
- das Lesen eines Datensatzes,
- das Verändern einen Datensatzes und
- das Löschen eines Datensatzes.

Diese Operationen werden häufig unter der Abkürzung CRUD (Create Read Update Delete) zusammengefasst. Änderungen an den Datenbeständen müssen bei Bedarf revidierbar sein, so dass auch alte Zustände wieder hergestellt werden können. Diese Wiederherstellbarkeit kann zeitlich befristet sein. Die endgültige Löschung von Datensätzen ist bei statischen Datensätzen sehr restriktiv zu behandeln. Insbesondere ist zu berücksichtigen, dass im Rahmen der guten wissenschaftlichen Praxis unter anderem alle Forschungsdaten, die zu einem Forschungsergebnis geführt haben,

aufbewahrt werden müssend. Dies hat Konsequenzen bezüglich des Speicheraufwandes und der Nutzungsrechte. In dieser Hinsicht sollte das Forschungsdatenzentrum eine größtmögliche Dokumentation unterstützen.

Schlussfolgerung 20: *Für die Forschungsdaten und ihre Metadaten muss das Speichersystem die Funktionen Speichern, Lesen, Verändern und Löschen unterstützen.*

Schlussfolgerung 21: *Die Metadaten für Forschungsdaten müssen Informationen zu dem Ursprung der Daten und möglicher späterer Veränderung der Daten beinhalten.*

Schlussfolgerung 22: *Das Forschungsdatenzentrum muss vorherige Versionen eines Datensatzes wiederherstellen können.*

Vielfach wird in Datenzentren auch eine Volltextsuche auf dem Datenbestand angeboten. Da es sich bei den Forschungsdaten oft auch um nicht textuelle Daten handelt, ist nicht klar, ob die Einrichtung einer Volltextsuche sinnvoll ist. Vor einer solchen Entscheidung ist eine sorgfältige Bedarfsanalyse erforderlich. Dagegen muss eine Volltextsuche auf den Metadaten möglich sein. Bei der Anzeige der Zugriffsergebnisse müssen die Zugriffsrechte berücksichtigt werden. Auf die Einzelheiten einer solchen Suche wird hier nicht eingegangen, dafür in der Regel Standardwerkzeuge verwendet werden.

Schlussfolgerung 23: *Eine Volltextsuche auf den Daten sollte nur nach einer sorgfältigen Bedarfsanalyse umgesetzt werden.*

Schlussfolgerung 24: *Eine Volltextsuche auf den Metadaten muss implementiert werden. Dabei ist sind die Zugriffsrechte zu beachten.*

Weiterhin erfordert der effiziente Umgang mit großen Mengen von Forschungsdaten eine Strukturierung der Ablage, die über die übliche relativ flache Verzeichnisstruktur einer lokalen Datenhaltung hinausgeht. Nur in einfachen Fällen reicht eine verschachtelte Verzeichnisstruktur aus, die Gruppierung von Datensätzen erlaubt, so dass die Bezeichnung eines Verzeichnisses sowie seine konkrete Position im Verzeichnisbaum zusammen die Semantik einer Kategorie von Dateien bilden, die in diesem Verzeichnis liegen. Diese Verzeichnisstruktur bedingt eine Dominanzordnung der Kategorisierungskriterien. Wenn eine solche Dominanzstruktur nicht vorliegt, dann sollte eine (relationale) Datenbank zum Einsatz kommen. In diesem Fall sollten die Kategorien der einzelnen Datensätze durch ihre jeweiligen Metadaten beschrieben werden. Auch für die einzelne Kategorie können Metadaten eingesetzt werden. Dabei ist zu berücksichtigen, dass eine große Zahl von Kategorisierungsmöglichkeiten den Umgang mit den Daten erschwert. Daher ist ein geeigneter Kompromiss zu finden, der inhaltlich begründet sein muss. Der Umgang mit Kategorien kann durch geeignete (graphische) Nutzerschnittstellen erleichtert werden.

Schlussfolgerung 25: Es muss möglich sein, die Forschungsdaten in Kategorien abzulegen. Wenn es eine Dominanzordnung der Kategorisierungskriterien gibt, dann kann eine geschachtelte Verzeichnisstruktur verwendet werden. Ansonsten ist ein Dokumentenmanagementsystem basierend auf einer relationalen Datenbank mit den entsprechenden Zugriffsverfahren besser geeignet.

Schlussfolgerung 26: Kategorien von Forschungsdaten sollen durch grundlegende Metadaten beschrieben werden können.

Das Forschungsdatenzentrum soll Daten für einen langen Zeitraum beherbergen und ihre Nutzungsmöglichkeit sicherstellen. Die Einschränkung der Nutzungsrechte scheint eine wichtige Voraussetzung für die Integration der Datenbestände mancher Museen und Landesdenkmalämter nach Aussagen von Experten aus diesem Bereich zu sein. Andererseits steht eine Nutzungseinschränkung der uneingeschränkten Verwendung zu Forschungszwecken entgegen. Daher sollten Nutzungseinschränkungen von Forschungsdaten regelmäßig überprüft werden. Ein solches Verfahren kann auf einer grundsätzlich zeitlich beschränkten Nutzungseinschränkung basieren, die vor einer Verlängerung explizit bestätigt werden muss. Informationen zu den Nutzungsrechten sollten in den Metadaten abgelegt sein.

Schlussfolgerung 27: Die Metadaten für Forschungsdaten sollen Informationen über Nutzungsrechte enthalten können.

Schlussfolgerung 28: Das System muss den Ablauf von Fristen automatisch feststellen können und betroffenen Personen Handlungsalternativen vorschlagen.

Weiterhin ist die nachprüfbare Integrität der Daten unabhängig von Nutzungsrechten zu gewährleisten. Dies kann zum Beispiel durch die Aufnahme von Signaturen oder datenbezogenen Prüfsummen in den Metadaten geschehen.

Schlussfolgerung 29: Die Metadaten der Forschungsdaten müssen Möglichkeiten zur Ablage von datenbezogenen Prüfsummen und Signaturen bieten.

3.2.2 Metadatenextraktion

Während die Inhalte von Metadaten bereits Gegenstand des vorherigen Kapitels waren, wird jetzt genauer auf die Erzeugung von Metadaten eingegangen. Dabei liegt der Fokus auf den automatisiert erzeugbaren Metadaten. Grundsätzlich lassen sich Metadaten in technische und fachliche Metadaten unterteilen. Technische Metadaten umfassen zum Beispiel das Erzeugungsdatum eines Datensatzes, sein Datenformat, Informationen zu dem erzeugenden Objekt (z.B. Angabe über das erzeugende Simulationsprogramm) und Daten zur Veränderungsgeschichte, wie bereits im vorherigen Kapitel dargestellt. Technische Metadaten können meist automatisiert erzeugt, ergänzt oder verändert werden. Dabei kann der anstoßende Prozess auf den Daten manuell oder auch automatisiert vorgenommen werden. Die Prozesse zur Erzeugung technischer Metadaten sollten weitgehend unabhängig von der Art der Daten erfolgen.

Schlussfolgerung 30: Das Forschungsdatenzentrum sollte eine möglichst automatisierte Unterstützung für technische Metadaten beinhalten.

Fachliche Metadaten betreffen inhaltliche Aspekte der Daten einschließlich ihrer Entstehungsgeschichte und sind disziplinspezifisch. Der Übergang zwischen fachlichen und technischen Metadaten ist fließend. Fachliche Metadaten werden nur selten automatisiert erzeugt sondern bedürfen in der Regel eine Nutzerinteraktion. Dazu sollte der Nutzer unterstützt werden zum Beispiel durch die Existenz eines Metadateneditors.

Schlussfolgerung 31: Das Forschungsdatenzentrum sollte den Nutzern Werkzeuge zur Bearbeitung der fachlichen Metadaten zur Verfügung stellen.

3.2.3 Konvertierung und Validierung

In der Archäologie und den Altertumswissenschaften wird mit unterschiedlichen Formaten der Daten und Metadaten gearbeitet. Zu Vereinfachung der Handhabung insbesondere mit automatisierten Werkzeugen ist eine Konsolidierung der Formate wünschenswert, soweit dies möglich ist. Wie bereits erwähnt soll im Rahmen des Prototyps eine Konversion exemplarisch dargestellt werden. Aus Aufwandsgründen ist es aber nicht möglich, eine solche Konversion bereits innerhalb dieser ersten Phase im großen Umfang zu betreiben. Bei Konvertierungswerkzeugen sollten existierende Werkzeuge eingebunden werden. Eine eigenständige Entwicklung ist möglichst zu vermeiden.

Schlussfolgerung 32: Das Forschungsdatenzentrum soll die Einbindung von Werkzeugen zur Konvertierung von Daten und Metadaten ermöglichen. Dabei ist der Schwerpunkt auf existierende Werkzeuge zu legen.

Des Weiteren ist die Validierung von Datenformaten und von Metadaten bezüglich Format und Inhalten wünschenswert. Der Aufwand ist von dem jeweiligen Format abhängig. Der Umfang existierender Werkzeuge kann zurzeit nicht beurteilt werden.

Schlussfolgerung 33: Das Forschungsdatenzentrum soll die Einbindung von Werkzeugen zur Validierung von Datenformaten und des Inhaltes von Metadaten erlauben. Zunächst ist dazu eine Analyse über existierende Werkzeuge zu erstellen.

3.2.4 Datensicherung

Da es sich bei Forschungsdaten der Archäologie und den Altertumswissenschaften nicht um personenbezogene Daten handelt, sind keine besonderen Aspekte des Datenschutzes zu berücksichtigen. Wichtig ist dagegen der Schutz der Daten vor Zerstörung durch externe Einflüsse. Dazu sind in der Regel Redundanzverfahren zu benutzen. Diese können allerdings auf der unteren Speicherebene umgesetzt werden, so dass hier nur entsprechende Dienstleistungsvereinbarungen zu treffen sind. Zum Schutz der Daten mit Einschränkung der Zugriffsrechten bei der Übertragung können die üblichen Verschlüsselungsverfahren eingesetzt werden.

3.3 Datenverarbeitung

Die Datenverarbeitung steht, wie bereits dargelegt, nicht im Schwerpunkt dieser Expertise. Allerdings kann langfristig eine Unterstützung von Methoden zur Datenverarbeitung von Interesse sein. Primär

ist hierbei das kollaborative Arbeiten mehrerer Forscher an unterschiedlichen Standorten zu erwähnen. Bisher wurden dazu vor allem der Datenaustausch per E-Mail und das Telefon als Kommunikationsmedium genutzt. Mit Hilfe aktueller Technologien kann die kollaborative Arbeit wesentlich erleichtert werden. Langfristig sollte das Forschungsdatenzentrum daher den Wissenschaftlern verschiedene integrierte Kommunikationsmedien (online oder offline) zur Verfügung stellen. Außerdem sollten die verwendeten Kommunikationsmedien protokollierbar sein und einen direkten Datenaustausch erlauben. Langfristig sollte eine Weiterentwicklung des Forschungsdatenzentrums in Richtung auf eine virtuelle Forschungsumgebung eine gemeinsame und zeitgleiche Bearbeitung von Daten durch mehrere Wissenschaftler erlauben.

4 Technische Lösungen

Die technische Umsetzung des Forschungsdatenzentrums muss vor allem die Langfristigkeit und damit die Nachhaltigkeit berücksichtigt werden. Bei der Auswahl der verwendeten Software ist die Unterstützung bezüglich Pflege und Anpassung an andere Rahmenbedingungen (Updates) zu beachten. Eine Eigenentwicklung führt auf die Dauer zu sehr hohen Kosten und ist zu Gunsten von Standardverfahren abzulehnen. Bezüglich der Softwarekosten sind immer die Gesamtkosten zu betrachten, die auch die langfristige Pflege des Softwareproduktes beinhalten. Viele Anforderungen des Forschungsdatenzentrums lassen sich durch Softwarelösungen, die auf kommerziellen Datenmanagementprodukten basieren, abdecken. Allerdings verlangen viele dieser Produkte Lizenzgebühren für jeden Endnutzer, was zu sehr hohen finanziellen Belastungen aufgrund der hohen Nutzerzahl führen kann. Hier besteht jedoch gegebenenfalls die Möglichkeit Sonderkonditionen zu erhalten.

Schlussfolgerung 34: Zur Sicherstellung der Nachhaltigkeit sind soweit möglich weit verbreitete Standardsoftwareprodukte zu verwenden.

Schlussfolgerung 35: Bei den Kostenbetrachtungen sind die Gesamtkosten einschließlich der Softwarepflege zu berücksichtigen.

4.1 Architektur des Forschungsdatenzentrums

Für die Architektur des Forschungsdatenzentrums wird eine 3-Tier-Architektur vorgeschlagen. In der untersten Schicht werden die Daten in einer Datenbank gespeichert. Für den Betrieb der eigentlichen Speicherressourcen kommen viele Anbieter in Betracht. Darunter fallen kommerzielle Anbieter, wie zum Beispiel T-Systems oder IBM ebenso wie Hochschulrechenzentren. Auf dieser Ebene sollte die Redundanz der Daten zur Gewährleistung der gewünschten Datensicherheit realisiert werden. Mit den Betreibern der Speicherressourcen müssen dann Dienstleistungsverträge abgeschlossen werden. Aufgrund der Verwendung standardisierter Softwareprodukte muss es möglich sein, den Betreiber bei Bedarf zu wechseln. In der mittleren Schicht erfolgt das eigentliche Management des Forschungsdatenzentrums durch den Betreiber des Zentrums. In dieser Schicht werden die Werkzeuge des Forschungsdatenzentrums betrieben und gepflegt sowie das Zentrum administriert. Die mittlere Ebene kapselt die Speicherressourcen von den Nutzern. Die Zugriffsautorisierung kann nur über diese mittlere Ebene erfolgen, auch wenn der eigentliche Datentransfer aus Effizienzgründen direkt zwischen Nutzer und Speicherressource ablaufen kann. Weiterhin stellt die mittlere Ebene Schnittstellen für die Nutzerebene zum Aufbau eigener Werkzeuge zur Verfügung. Die Nutzer befinden sich in der dritten (Nutzer-)Ebene. Aus dieser Ebene werden die Anforderungen auf einen Datenzugriff sowie Organisationsanforderungen an das Forschungsdatenzentrum gestellt. Zu diesen Organisationsanforderungen gehören zum Beispiel Suchanforderungen oder Anforderungen zur Integration neuer Datensätze oder zur Modifikation existierender Datensätze. Unter Verwendung der offen gelegten Schnittstellen zu der mittleren Ebene können in der Nutzerebene auch nutzerspezifische Verarbeitungswerkzeuge entstehen, für deren Pflege der Nutzer selbst verantwortlich ist.

Schlussfolgerung 36: Das Forschungsdatenzentrum soll in Form einer 3-Tier-Architektur aufgebaut werden. Kernstück bildet die mittlere Ebene, die für den eigentlichen Mehrwert des Forschungsdatenzentrums verantwortlich ist. Die Nutzer befinden sich in der oberen Ebene und können selbst gepflegte Werkzeuge über definierte Schnittstellen an das

Forschungsdatenzentrum anbinden. Die untere Ebene deckt die eigentliche Datenspeicherung ab und sollte durch grundsätzlich austauschbare Betreiber von Speicherressourcen ausgefüllt werden.

4.2 Komponenten

In diesem Teil werden einige Alternativen für die Komponenten der Architektur ohne Anspruch auf Vollständigkeit beschrieben. Sobald Einigkeit über den strukturellen Aufbau des Forschungsdatenzentrums besteht, müssen potentielle Kandidaten näher betrachtet werden. Es wird dazu dringend empfohlen, ein Präsentations- und Ausschreibungsverfahren entsprechen dem Umfang des Gesamtprojektes durchzuführen.

Schlussfolgerung 37: Nach Klärung der Rahmenbedingungen und Schnittstellen sollten die einzelnen Komponenten in Form eines Ausschreibungsverfahrens vergeben werden.

4.2.1 Lösungskonzepte aus dem kommerziellen Bereich

Die Handhabung elektronischer Dokumente ist im kommerziellen Bereich ein sehr wichtiges Thema. Zu diesem Zweck werden Dokumentenmanagementsysteme (DMS) eingesetzt. DMS bieten oft eine Komplettlösung an einschließlich Metadatenbehandlung und Archivierung, wobei die Metadaten meistens in XML abgebildet werden. Da in Industrieunternehmen auch eine Vielzahl unterschiedlicher Datenformate auftreten, bestehen meist ähnliche Rahmenbedingungen wie in dem geplanten Forschungsdatenzentrum. Weiterhin werden DMS mit anderen großen Softwaresystemen integriert, wie zum Beispiel ERP-(Enterprise Resource Planning) Systemen. Daher verwenden die meisten DMS offene Lösungen und Baukastensysteme, die eine Anpassung an andere Softwaresysteme unterstützen. Grundsätzlich sind DMS als Grundlage für das Forschungsdatenzentrum geeignet. Hinzu kommen weitere Funktionalitäten, die spezifisch für die Behandlung von Datensätzen in der Archäologie und den Altertumswissenschaften sind, wie zum Beispiel Konversionsverfahren bei Metadaten. Diese Funktionalitäten müssen durch Zusatzentwicklungen oder Konfiguration hinzugefügt werden. Da ein erheblicher Funktionalitätsumfang meist schon in der Basisausstattung (z.B. Suchfunktionen, Versionenverwaltung, Zugriffsverwaltung) enthalten ist und das System meist über eine Entwicklungsumgebung verfügt, wird auf diese Weise der Umsetzungsaufwand reduziert. Viele DMS bieten auch Webschnittstellen für die Nutzer an. Bei Bedarf kann man für diese Produkte auch kommerzielle Umsetzungshilfe allerdings zu einem üblicherweise relativ hohen Preis erhalten. Andererseits beinhalten viele DMS weitere Funktionalitäten, die zwar für das Forschungsdatenzentrum nicht benötigt werden aber trotzdem im Preis enthalten sind. Die Lizenzgestaltung dieser Produkte ist auf Unternehmen mit einer gegenüber dem Forschungsdatenzentrum geringen Nutzeranzahl ausgelegt. Eine direkte Übertragung würde daher zu einem nicht akzeptablen Preis führen. Außerdem ist die langfristige Produktsicherheit aufgrund der Vielzahl der Anbieter bei einem solchen Datenmanagementsystem nicht automatisch gegeben und ein späterer Umstieg führt zu zusätzlichem Aufwand. Trotzdem sollten kommerzielle Lösungen nicht grundsätzlich abgelehnt werden. In dieser Expertise werden aufgrund der großen Anzahl von Anbietern keine einzelnen Produkte vorgestellt.

Die Speicherebene kann als klassisches Hosting einer Softwarelösung mit den dazugehörigen Daten realisiert werden. Sowohl Großunternehmen als auch mittelständige Unternehmen bieten hierzu Lösungen an. Dazu ist ein Anforderungskatalog bezüglich der Zuverlässigkeit und der Effizienz aufzustellen und dann eine Auswahl unter Beachtung des Erfüllungsgrades und des Preises zu treffen.

Eine Migration von einem Anbieter zu einem anderen muss immer mit geringem Aufwand möglich sein.

Für die Nutzerebene wurde bereits in den Anforderungen an die Expertise der Ansatz über Portale und Webservices ausgewählt. Dieser Ansatz bietet den großen Vorteil einer Unabhängigkeit von bestehenden Betriebssystemen auf der Nutzerebene und aufgrund des hohen Verbreitungsgrades nach jetzigen Maßstäben gute Aussichten auf eine langfristig unterstützte Plattform.

Schlussfolgerung 38: Aufgrund der Vorteile bezüglich der Zukunftssicherheit der Softwareprodukte sollten kommerzielle Dokumentenmanagementsysteme zumindest bei der Auswahl berücksichtigt werden.

Schlussfolgerung 39: Die Speicherebene sollte im Outsourcing betrieben werden.

4.2.2 Lösungskonzepte aus dem akademischen Bereich

Im akademischen Bereich gibt es auch mehrere Ansätze und Open-Source Produkte, die bei Forschungsdatenzentren und virtuellen Forschungsumgebungen zum Einsatz kommen. Auf den ersten Blick sind diese Produkte deutlich kostengünstiger als kommerzielle Alternativen. Sie erfordern aber oft einen erheblichen Eigenentwicklungsaufwand und die Produktzuverlässigkeit und eine hohe Marktverbreitung ist nicht immer gegeben. Daher wird das Umsetzungsrisiko dieser Produkte als relativ hoch eingestuft. Einige im akademischen Umfeld eingesetzte Verfahren werden im Folgenden vorgestellt.

dCache/SRM

dCache/SRM¹ (kurz dCache) ist ein im D-Grid eingesetztes System, das vor allem aus dem Bereich der Hochenergiephysik stammt. Das System setzt auf den Middlewares im D-Grid (Globus Toolkit, gLite, UNICORE) auf und optimiert die Datenhaltung. Der Nutzer sieht ein großes virtuelles Dateisystem. Abhängig von der Konfiguration führt dCache/SRM automatisch eine redundante Datenhaltung und eine Lastoptimierung durch. Außerdem kann es Datenverluste selbstständig erkennen und bei Existenz der Redundanzen kompensieren. Allerdings ermöglicht dCache keine Metadatenverwaltung und kommt daher für das Forschungsdatenzentrum nicht in Frage.

iRODS

iRODS (Integrated Rule-Oriented Data System²) ist der Nachfolger des Storage Resource Brokers (SRB³) und kann zum Aufbau von Daten-Grids, digitalen Bibliotheken, persistenten Archiven und Echtzeit-Datensystemen verwendet werden. Es kann sowohl auf PC-Systemen als auch auf Grid-Infrastrukturen mit entsprechenden Middlewares aufsetzen. iRODS ist sehr flexibel und erlaubt vielfache Anpassungsmöglichkeiten. Es basiert auf einem regelbasiertes System, das Datenoperationen über Reihenfolgen von frei wählbaren Schritten ausführt. Grundsätzlich sind die Kernfunktionalitäten von iRODS mit begrenztem Aufwand beliebig erweiter- und veränderbar. Allerdings bietet iRODS nur rudimentäre Möglichkeiten zur Verwaltung von Metadaten, die für die

¹ dCache, URL: <http://www.dcache.org>.

² iRODS, URL: <https://www.irods.org>.

³ SRB - The DICE Storage Resource Broker, URL: http://www.sdsc.edu/srb/index.php/Main_Page.

Anforderungen des Forschungsdatenzentrums nicht ausreichen. Daher wäre bei Auswahl von iRODS zusätzlicher Aufwand notwendig, um für iRODS die erforderliche Metadatenverwaltung zu ergänzen. Da iRODS eine Schnittstelle in der Repository-Software Fedora beinhaltet, sollte allenfalls ein Einsatz zusammen mit Fedora betrachtet werden. I

Stellaris

Stellaris⁴ ist ein Metadatenverwaltungssystem, das im Rahmen des D-Grid Projektes Astro-Grid-D entwickelt wurde. Es erlaubt eine sehr große Flexibilität bei der Struktur und der Ablage von Metadaten. So können gruppenbasierte Zugriffsrechte auf Metadatenkollektionen festgelegt werden. Außerdem existieren umfangreiche Suchfunktionen. Da allerdings die Ablage von reinen Forschungsdaten nicht unterstützt wird und das System eine geringe Verbreitung hat, wird eine Auswahl nicht empfohlen.

OGSA-DAI

OGSA-DAI⁵ ist ein Integrationssystem für verschiedene Datenbanken (oder auch andere Datenquellen), mit dem ein homogener Zugriff auf die Datenquellen ermöglicht wird. Das System selbst kann jedoch ohne eine aufwändige Zusatzimplementierung keine Daten in Dateiform ablegen oder verwalten und ist daher für den angedachten Zweck nicht geeignet.

Fedora

Fedora (Flexible Extensible Digital Object Repository Architecture⁶) ist ein Datenmanagement-Repository, das auf beliebige Dateisysteme aufsetzen kann. Es ist modular aufgebaut und kann an die Bedürfnisse des Forschungsdatenzentrums angepasst werden. Fedora wird zum Beispiel zur Verwaltung digitaler Daten in Bibliotheken in akademischen Einrichtungen eingesetzt. Mehrere internationale Projekte (z.B. ADONIS⁷) arbeiten an der direkten Unterstützung von iRODS durch Fedora. Auf diesen Arbeiten kann grundsätzlich aufgebaut werden, wenn es geplant ist, iRODS als Speicherinfrastruktur für Daten zu verwenden und Metadatenverwaltung über Fedora zu realisieren. Das Projekt WissGrid im D-Grid hat folgende Möglichkeiten zu Implementierung der Schnittstelle zwischen iRODS und Fedora aufgezeigt:

- FUSE⁸
- WebDAV⁹
- Apache VFS¹⁰
- JAGSA¹¹ (Java implementation of the Simple API for Grid Applications)

⁴ Stellaris, URL: <http://stellaris.zib.de>.

⁵ OGSA-DAI, URL: <http://www.ogsadai.org.uk>

⁶ Fedora, URL: <http://fedora-commons.org/>

⁷ ADONIS, URL: <http://www.tge-adonis.fr>

⁸ FUSE, URL: https://www.irods.org/index.php/iRODS_FUSE

⁹ WebDAV, URL: <https://projects.arcs.org.au/trac/davis/wiki/WikiStart>

¹⁰ Apache VFS, URL: <http://www.omii.ac.uk/wiki/CommonsVFSExtensionsForGrids>

¹¹ JAGSA, URL: <http://grid.in2p3.fr/jsaga>

In dem Projekt wurde die FUSE getestet. Dabei ergaben sich Probleme mit größeren Dateien. Die anderen Alternativen haben einen höheren Stabilitätsgrad. Sie erfordern aber zusätzliche Entwicklungen in iRODS oder in Fedora.

Der Nutzer kann bei Fedora mit Hilfe einer Web-Oberfläche direkt auf das System zugreifen. Außerdem existieren Softwareschnittstellen zu anderen Systemen. Beide Zugriffsmöglichkeiten erlauben jederzeit eine Aktualisierung von existierenden Datensätzen. Bezüglich der verwalteten Daten ist Fedora typ- und formatunabhängig. Protokollierung, Versionierung und Katalogisierung kann durch Fedora ebenfalls unterstützt werden. Fedora enthält einfache Suchfunktionen und erlaubt die Integration komplexerer Suchmodule.

Intern arbeitet Fedora auf dem XML-Standard und gewährleistet damit eine relativ hohe Zukunftssicherheit. Für die Rechteverwaltung wird XACML/SAML (EXtensible Access Control Markup Language / Security Assertion Markup Language) verwendet. Schließlich erlaubt Fedora das Erstellen von Prüfsummen und unterstützt eine automatisierte Repository-Spiegelung.

Gegenwärtig befindet sich gerade eine virtuelle Forschungsumgebung für die Sozialwissenschaften im Aufbau, die insbesondere für die sozioökonomische Berichterstattung eingesetzt werden soll¹². Technisch basiert sie auf Fedora zusammen mit iRODS. Des Weiteren wurde die kombinierte Lösung aus iRODS und Fedora im D-Grid Projekt TextGrid¹³ und zur Umsetzung eines Repository für Forschungsdaten im Arbeitspaket Langzeitarchivierung des D-Grid Projektes WissGrid¹⁴ verwendet. Auf in diesem Zusammenhang untersuchte Möglichkeiten zur Implementierung einer Schnittstelle zwischen iRODS und Fedora wurde bereits hingewiesen. Auch die nationale digitale Wissenschaftsbibliothek der USA¹⁵ hat sich für die Kombination aus iRODS und Fedora entschieden. Eine detaillierte Beschreibung der Integration von iRODS und Fedora wird im Rahmen des iRODS Projektes zur Verfügung gestellt¹⁶. Weiterhin gibt es APIs zur Verbindung eines Fedora Commons Repository mit einem IRODS Datengrid im Netz¹⁷. Das Storage Resource Broker Team des San Diego Supercomputer Centers (SDSC), das das iRODS Projekt im Wesentlichen treibt, veröffentlicht immer wieder neue Fedora Module.

Grundsätzlich ist die Fedora/iRODS Kombination als eine realisierbare Implementierungsvariante für das geplante Forschungsdatenzentrum für Archäologie und Altertumswissenschaften zu betrachten. Eine endgültige Entscheidung kann aber erst getroffen werden, wenn sowohl die qualitativen als auch die quantitativen Rahmenbedingungen feststehen. Hierzu wird vorgeschlagen, dass sich das vorgesehene Implementierungsteam des Betreibers des Forschungsdatenzentrums intensiv mit vorhandenen Fedora/iRODS Implementierungen beschäftigt. Falls sich letztendlich für eine Kombination aus Fedora und iRODS entschieden wird und Implementierungsexpertise für

¹² Soziologisches Forschungsinstitut Göttingen an der Georg-August-Universität, URL: <http://www.sofi.uni-goettingen.de/index.php?id=1064>

¹³ TextGrid, URL: <http://www.textgrid.info/>

¹⁴ Aschenbrenner, A; Enke, H; Fritsch, B, et al. (2010): WissGrid-Spezifikation: Grid-Repository, Göttingen, Niedersächsische Staats- und Universitätsbibliothek Göttingen, D3.5.2, 2010.04.30

¹⁵ National Science Digital Library NSDL, URL: <http://nsdl.com>

¹⁶ iRODS, URL: <https://www.irods.org/index.php/Fedora>

¹⁷ irodsfedora, URL: <http://code.google.com/p/irodsfedora/>

Fedora/iRODS bei dem vorgesehenen Betreiber nicht vorhanden ist, so sollten Experten aus den erwähnten deutschen Projekten in dem Prototypprojekt zu beteiligt werden.

Dublin Core

Wie bereits erwähnt benötigt ein Forschungsdatenzentrum eine Möglichkeit zur Abbildung von Metadaten. In diesem Bereich haben sich die so genannten Dublin Core-Standards (Dublin Core Metadata Initiative¹⁸) etabliert, die vor allem im Bibliotheksumfeld eingesetzt werden. Dublin Core unterstützt sowohl technische als auch fachliche Metadaten und setzt Metadaten zur Provenienz von Daten und zur Darstellung von Berechtigungen ein. Auch die Formate Open Document Format (ODF) und HTML bauen auf Dublin Core als Metadatenstandard auf. Damit kann von einer langfristigen Produktsicherheit ausgegangen werden.

Schlussfolgerung 40: Als Lösung aus dem akademischen Umfeld bietet sich eine Kombination von Fedora und iRODS an, die bereits mehrfach auch in Deutschland in vergleichbaren Einrichtungen im Einsatz ist. Für die Implementierung wird vorgeschlagen, die Hilfe einer Einrichtung in Anspruch zu nehmen, die bereits Erfahrung mit der Umsetzung und dem Betrieb eines Fedora-Systems hat.

¹⁸ Dublin Core, URL: <http://dublincore.org>

5 Organisatorische Lösungen

Die Organisationsstruktur orientiert sich an der technischen 3-Tier Architektur. Für jede Ebene gibt es unterschiedliche Verantwortlichkeiten. Daher sollte es zwischen den Beteiligten benachbarter Ebenen Vertragsbeziehungen geben. Dazu sind zunächst die Rechte und Pflichten der Beteiligten aufzuzeigen. Anschließend werden daraus Ansätze für ein Finanzierungskonzept abgeleitet.

5.1 Rechte und Pflichten der Beteiligten

Bei den Beteiligten sollte entsprechend der Architekturebene unterschieden werden.

Schlussfolgerung 41: Über die Rechte und Pflichten der Beteiligten sollte vor dem Beginn des Aufbaus des Forschungsdatenzentrums Einigkeit bestehen.

5.1.1 Speicherebene

Die Betreiber der Speicherebene sind für die zuverlässige Speicherung der Daten und einen effizienten Datenzugriff und Berücksichtigung der Zugriffsrechte verantwortlich. Die Pflichten der Beteiligten müssen in einem Dienstleistungsvertrag spezifiziert werden. In diesem Vertrag werden die verwendeten Schnittstellen spezifiziert und die erwarteten Dienstleistungen sowie Maßnahmen im Fall einer Verletzung des Dienstleistungsvertrages beschrieben. Der Betreiber hat keinerlei eigene Rechte an den Daten. Es wird davon ausgegangen, dass ein im kommerziellen Bereich üblicher Outsourcingvertrag zum Einsatz kommen kann. Bezogen auf den Betreiber handelt es sich nicht um eine Forschungs- sondern um eine reine Dienstleistungsaufgabe. Dies gilt auch, wenn der Betrieb der Speicherebene durch das Rechenzentrum einer Universität oder einer anderen Forschungseinrichtung übernommen wird.

Schlussfolgerung 42: Die Pflichten der Speicherebene werden sich an den üblichen Bedingungen bei Outsourcingverträgen orientieren.

5.1.2 Verwaltungsebene

Der Betreiber des Forschungsdatenzentrums ist alleiniger Verantwortlicher (Alleinunternehmer) gegenüber dem Nutzer. Die Verantwortlichkeiten werden ebenfalls in einem Servicevertrag zwischen dem Nutzer (oder seiner Einrichtung) und dem Forschungsdatenzentrum spezifiziert. Wenn sachlich gesehen nicht der Betreiber des Forschungsdatenzentrums sondern ein Betreiber der Speicherebene für eine nicht erfüllte Dienstleistung verantwortlich ist, so trägt rechtlich gesehen trotzdem der Betreiber des Forschungsdatenzentrums gegenüber dem Nutzer die Verantwortung. Er kann eventuellen Regress seinerseits dann gegenüber dem Betreiber der Speicherebene geltend machen.

Der Betreiber des Forschungsdatenzentrums hat ebenfalls keinerlei eigene Rechte an den Forschungsdaten. Er ist auch nicht inhaltlich für die Forschungsdaten und die zugehörigen Metadaten verantwortlich. Diese Verantwortung liegt beim Besitzer dieser Daten. Einzige Ausnahme bilden technische Metadaten, die automatisch generiert werden. Dazu gehören z.B. Metadaten, die Informationen über den letzten Zugriff enthalten. Diese technischen Metadaten werden dokumentiert und können durch den Besitzer der jeweiligen Forschungsdaten nicht verändert werden. Zusätzlich kann der Betreiber des Forschungsdatenzentrums Unterstützungswerkzeuge für die Bearbeitung der Datensätze und der Metadaten zur Verfügung stellen. Aber auch bei Anwendung dieser Werkzeuge verbleibt die Verantwortung für den Inhalt der Daten bei dem Besitzer der Daten. Dagegen stellt der

Betreiber des Forschungsdatenzentrums sicher, dass nur die durch den Besitzer der Daten genehmigten Zugriffsrechte zur Anwendung kommen. Insbesondere ist er dafür verantwortlich, dass keine Person auf Forschungsdaten zugreifen kann, wenn dies nicht durch die Zugriffsrechte vorgesehen ist. Der Betreiber des Forschungsdatenzentrums stellt den Besitzern von Daten Werkzeuge zur Konfiguration der Zugriffsrechte zur Verfügung. Das Forschungsdatenzentrum kann Bedingungen an die Datenqualität stellen und sich weigern, Forschungsdaten zu speichern, wenn diese Daten die Qualitätsbedingungen nicht erfüllen.

Schlussfolgerung 43: Auch wenn eine wissenschaftliche Einrichtung den Betrieb der Verwaltungsebene übernimmt, so muss trotzdem klar sein, dass es sich bei dem Betrieb der Forschungsdatenzentrums um eine Dienstleistung und nicht um eine wissenschaftliche Aufgabe handelt.

5.1.3 Nutzer der Daten

Den Forschern stehen die Dienstleistungen des Forschungsdatenzentrums auf einer Best-Effort-Basis zur Verfügung. Sie haben das Recht auf Datenzugriff sofern dieser Datenzugriff durch den Besitzer der Daten korrekt gewährt wurde. Das Forschungsdatenzentrum ist verantwortlich dafür, dass dieser Datenzugriff ermöglicht wird. Es besteht aber z.B. kein Rechtsanspruch auf eine bestimmte Datenübertragungsrate, da dieser Wert von unterschiedlichen Rahmenbedingungen beeinflusst wird. Mit dem Zugriff auf diese Daten übernimmt der Nutzer auch die Verantwortlichkeit über diese Daten gegenüber dem Besitzer der Daten. In dieser Hinsicht agiert das Forschungsdatenzentrum lediglich als Makler. Auf Nachfrage wird dem Besitzer von Daten mitgeteilt, wer auf seine Daten zugegriffen hat.

Schlussfolgerung 44: Die Rechte der Datennutzer orientieren sich primär an der Finanzierung. Wenn die Einrichtungen, an denen sich die Datennutzer befinden, sich nicht an der Finanzierung beteiligen, dann kann auch nur ein Best-Effort-Ansatz unterstützt werden.

5.1.4 Besitzer der Daten

Für den Besitzer von Daten wirkt das Forschungsdatenzentrum als ein Outsourcingdienstleister zur Speicherung seiner Daten und als Makler für den Datenzugriff. Der Besitzer der Daten behält alle Rechte an den Daten und ist für die Daten weiterhin verantwortlich. Er muss z.B. durch die Konfiguration von Zugriffsrechten sicherstellen, dass die gesetzlichen Vorgaben erfüllt werden. Er kann für in diesem Zusammenhang auftretende Probleme nur dann den Betreiber des Forschungsdatenzentrums verantwortlich halten, wenn die Probleme eindeutig mit den Werkzeugen des Forschungsdatenzentrums zusammenhängen. Entsprechend sollte ein Dienstleistungsvertrag zwischen dem Besitzer der Daten und dem Forschungsdatenzentrum abgeschlossen werden. Der Besitzer der Daten hat das Recht, seine Daten aus dem Forschungsdatenzentrum entfernen zu lassen.

Schlussfolgerung 45: Das Forschungsdatenzentrum ist nur dann für viele Forscher attraktiv, wenn es einen großen Umfang an Forschungsdaten beinhaltet. Aus diesem Grund müssen viele Besitzer von Daten bereit sein, ihre Daten im Forschungsdatenzentrum zu speichern. Daher ist gerade in Bezug auf die Besitzer von Forschungsdaten eine genaue Definition von Rechten und Pflichten notwendig.

5.2 Finanzierungskonzept

Grundsätzlich sollten alle Beteiligten ein Interesse an dem Aufbau eines Forschungsdatenzentrums für die Archäologie und die Altertumswissenschaften haben, da langfristig aufgrund der in der IT-Technik anerkannten Skalierungseffekte ein einzelnes Forschungsdatenzentrum effizienter betrieben werden kann als eine Vielzahl verteilter Datenzentren. Aus wissenschaftlicher Hinsicht bietet ein umfassendes Datenzentrum deutlich mehr Möglichkeiten als mehrerer voneinander getrennte Zentren, deren Datenbestände alle einzeln betrachtet werden müssen, um die gesuchten Daten zu finden. In anderen Wissenschaftsdisziplinen wie verschiedenen Sparten der Physik und der Klimaforschung wurde der Wert einer zentralen Datenhaltung schon länger erkannt. Es ist wichtig, für den Aufbau ein faires Finanzkonzept aufzustellen.

Im Rahmen des Finanzierungskonzeptes sollte jeder nach den erhaltenen Vorteilen beitragen. Bezüglich der Datennutzung gibt es zwei Möglichkeiten:

1. Es wird ein Verteilungsschlüssel der Bundesländer und des Bundes angewandt unabhängig von der tatsächlichen Nutzung
2. Die Nutzung des Forschungsdatenzentrums wird in Abhängigkeit der Zugriffe oder der heruntergeladenen Daten bestimmt und entsprechend entweder über die Ländern oder die jeweiligen Einrichtungen abgerechnet. Ein dafür notwendiges Abrechnungsverfahren lässt sich einfach installieren.

Die zweite Möglichkeit berücksichtigt die tatsächliche Nutzung. Sie erfordert aber etwas mehr Aufwand und ergibt potentiell weniger Planungssicherheit. Bei einer Umlage auf die einzelnen Institutionen ist zu berücksichtigen, dass dies zumindest in der Anfangsphase zu einer reduzierten Nutzung führen wird. Daher wird dieser Ansatz nicht empfohlen.

Weiterhin ist die Datenspeicherung zu berücksichtigen. Auch hier kann entweder ein Verteilungsschlüssel oder nach benötigtem Speicherplatzbedarf abgerechnet werden. Dabei ist zu berücksichtigen, dass der Speicherplatzbedarf bei digitalisierten Daten auf jeden Fall anfällt. Auch bisher schon durchgeführte Investitionen sprechen aufgrund der kurzen Zykluszeiten in der IT-Technik nicht gegen einen solchen Einsatz.

Weiterhin sind Einmalinvestitionen beim Aufbau des Forschungsdatenzentrums notwendig. Hier sollte der Bund als Initiator des Forschungsdatenzentrums in Vorleistung treten. Wenn die Projektfinanzierung nicht gesichert ist, dann ist zu befürchten, dass andere potentielle Partner aus den Ländern nicht bereit sein werden sich an dem Projekt zu beteiligen.

Schlussfolgerung 46: Das Forschungsdatenzentrum wird nur dann ein Erfolg werden, wenn viele Datenbesitzer ihre Forschungsdaten zur Verfügung stellen. Dies werden die Datenbesitzer nur tun, wenn das Konzept Aussicht auf Erfolg verspricht. Dazu gehört eine ausreichende Projektfinanzierung und ein Bekenntnis des Bunds als Initiators zu dem

Projekt. Deshalb sollte der Bund eine ausreichende Projektfinanzierung über einen entsprechenden Zeitraum zur Verfügung stellen.

Schlussfolgerung 47: *Der Betrieb des Forschungsdatenzentrums sollte entsprechend der Nutzung finanziert werden. Hierbei ist sowohl die Speicherung der eigenen Daten als auch die Nutzung der Daten zu Forschungszwecken zu berücksichtigen. Eine kleinteilige Abrechnung ist möglich kann aber die Nutzung einschränken. Deshalb wird vorgeschlagen, dass die Länder anteilmäßig den Betrieb finanzieren.*

5.3 Rechtliche Fragen

Aus rechtlicher Sicht sind vor allem die Kulturhoheit der Länder und die unterschiedlichen Denkmalschutzgesetze von Relevanz. Die vorgeschlagene Konstruktion führt in dieser Hinsicht zu keinen Einschränkungen. Da jeder Datenbesitzer den Zugriff auf seine Daten selbstständig regeln kann, besitzt er auch weiterhin die Hoheit über diese Daten, auch wenn sie sich physisch nicht auf einem Datenserver vor Ort befinden.

Auch eine wirtschaftliche Tätigkeit mit diesen Daten ist über das Forschungsdatenzentrum möglich. In diesem Fall müsste lediglich für einzelne Nutzer eine Abrechnungssoftware verwendet werden.

Bezüglich der Mehrwertsteuer und der Vergabevorschriften ist zu berücksichtigen, dass es sich bei dem Betrieb des Forschungsdatenzentrums um eine Dienstleistung und nicht um eine wissenschaftliche Tätigkeit handelt.

Schlussfolgerung 48: *Die Organisationsstruktur und die Aufgabenverteilung des Forschungsdatenzentrums werden von den rechtlichen Rahmenbedingungen mitbestimmt. Da eine Änderung der gesetzlichen Rahmenbedingungen üblicherweise eine erhebliche Zeit beansprucht, sollte die Strukturen so gewählt werden, dass sie den existierenden rechtlichen Rahmenbedingungen entsprechen. Insbesondere sollten keine Einrichtung oder kein Staat einen bestehenden Rechtsanspruch zu Gunsten des Forschungsdatenzentrums aufgeben müssen.*

6 Umsetzung

Es wurde bereits am Anfang der Expertise erwähnt, dass sich ein Forschungsdatenzentrum nach Meinung des Verfassers aus organisatorischen Gründen nicht innerhalb der geplanten drei Jahre umsetzen lässt. Daher wird vorgeschlagen, innerhalb der drei Jahre einen Prototyp zu erzeugen, der als überzeugendes technisches Beispiel für das geplante Forschungsdatenzentrum dient. Der Prototyp kann aber noch nicht den geplanten Umfang an Forschungsdaten realisieren. Es wird dringend empfohlen, den Prototyp auch lediglich als einen solchen zu bezeichnen, um nicht falsche Erwartungen zu wecken. Wichtig ist, dass innerhalb der Projektlaufzeit ein System entsteht, das die wesentliche Funktionalität des späteren Forschungsdatenzentrums hat und damit potentielle Interessenten überzeugen kann.

Schlussfolgerung 49: Der Prototyp soll der Überzeugung von Datenlieferanten dienen. Er muss so ausgelegt werden, dass in der Projektlaufzeit ein vollfunktionsfähiges System mit Schnittstellen entsteht, die potentielle Nutzer bezüglich ihrer Handhabung überzeugen.

6.1 Projektablauf

In der Vorbereitungsphase sollte zunächst eine Gruppe von interessierten Projektteilnehmern festgelegt werden. Diese Gruppe sollte den zukünftigen Betreiber des Forschungsdatenzentrums und unterschiedliche Datenlieferanten umfassen. Sie muss einerseits in der Größe eng begrenzt sein, um ein effizientes Arbeiten zu ermöglichen und andererseits sollte sie alle Gruppen von potentiellen Datenlieferanten umfassen. Der Schwerpunkt sollte auf den Datenlieferanten und nicht auf den Datennutzern liegen. Letztere sollten für die Planung und Beurteilung der Nutzerschnittstellen eingebunden werden. Das Forschungsdatenzentrum wird insgesamt aber nur dann erfolgreich sein, wenn es gelingt die Datenlieferanten zu überzeugen.

Eine Auswahl des Softwareproduktes sollte frühzeitig erfolgen. In jedem Fall ist eine Gruppe einzubinden, die mit der Umsetzung des gewählten Produktes bereits Erfahrung hat. Dabei ist nochmals zu betonen, dass diese Umsetzung nicht als wissenschaftliches Projekt aufzufassen ist, dass im Rahmen von Doktorarbeiten durchgeführt werden sollte. Es ist vielmehr ein klassisches Softwareprojekt.

Daher sollten auch die üblichen Strukturen eines solchen Softwareprojektes mit Lenkungskreis sowie Projektleitung mit den entsprechenden Unterstrukturen benutzt werden.

Schlussfolgerung 50: Das Projekt ist als Einführung eines neuen Softwaresystems zu betrachten und entsprechend zu planen.

6.2 Anreizsystem für die Beteiligung von potentiellen Interessenten

Eine notwendige Bedingung für den Erfolg des Projektes ist die Verfügbarkeit einer genügend großen Menge an Forschungsdaten. Wenn diese Menge erst einmal durch das Forschungsdatenzentrum verwaltet wird, dann ist es relativ einfach, andere Datenbesitzer davon zu überzeugen, ihre Daten auch in das Forschungsdatenzentrum zu integrieren, selbst wenn dies mit einigem Aufwand, z.B. für die Erstellung von Metadaten verbunden ist. Daher sollten zunächst die Interessenten angesprochen werden, deren Daten sich einfach in das Forschungsdatenzentrum integrieren lassen.

Schon jetzt müssen die Daten einer Einrichtung verwaltet werden. Eine Einrichtung wird sich vor allem aus zwei Gründen für die Beteiligung an dem Forschungsdatenzentrum entscheiden:

1. Der Aufwand (die Kosten) werden durch das Forschungsdatenzentrum reduziert.
2. In der Zukunft entsteht z.B. durch neue Anforderungen bei der lokalen Datenhaltung ein erheblicher zusätzlicher Arbeitsaufwand, den man gerne vermeiden möchte.

Der zweite Fall tritt vor allem dann auf, wenn sich neue Rahmenbedingungen z.B. durch eine neue Gesetzeslage ergeben. Dies ist im Moment nach Kenntnis des Verfassers nicht abzusehen. Daher ist es wichtig, neben dem Prototyp die Dienstleistungen und die Kosten für die Nutzer genau zu beschreiben. Die Kosten sollten so festgelegt werden, dass ein Anreiz für die Nutzer entsteht.

Für Forscher ist das Forschungsdatenzentrum im Allgemeinen erst interessant, wenn eine genügend große Datenmenge zur Verfügung steht. In diesem Fall ist die Nutzerfreundlichkeit nur noch ein Kriterium unter mehreren und nicht mehr das dominante Kriterium. Daher sollte das Forschungsdatenzentrum primären Wert auf die Erweiterung der Datenmenge legen und des Weiteren sicherstellen, dass Ergonomieanforderungen von modernen Softwaresystemen erfüllt werden.

Schlussfolgerung 51: Primäres Ziel beim Aufbau Forschungsdatenzentrums ist die Vergrößerung der beherbergten Datenmenge. Daher sollten versucht werden zunächst „einfache“ Datensätze zu integrieren.

Schlussfolgerung 52: Auch für Wissenschaftler ist wahrscheinlich der Zugriff auf eine große Menge von Daten das primäre Ziel. Daneben sollte die Benutzerfreundlichkeit berücksichtigt werden.

7 Zusammenfassung der Schlussfolgerungen

In diesem Abschnitt werden die einzelnen Schlussfolgerungen der vorherigen Kapitel nochmals kompakt wiederholt.

1. Die Verantwortlichkeit für das Forschungsdatenzentrum soll eine vom Bund finanzierte Einrichtung übernehmen und damit den nationalen Charakter des Forschungsdatenzentrums und die Nachhaltigkeit gewährleisten.
2. Die Verankerung des Forschungsdatenzentrums beim Bund unter Berücksichtigung der Kulturhoheit der Länder ist in Bezug auf rechtliche Aspekte zu klären.
3. Die Zuordnung des Forschungsdatenzentrums zu einem Bundesministerium muss festgelegt werden. In diesem Zusammenhang ist eine Konsolidierung der Aktivitäten des Bundes anzustreben.
4. Innerhalb der vorgesehenen Umsetzungszeit von drei Jahren kann nur ein Prototyp eines Forschungsdatenzentrums erstellt werden. Hieran sollte nur eine kleine Gruppe von Einrichtungen beteiligt sein. Die Gruppe sollte die genannten Kategorien der Beteiligten möglichst gut repräsentieren können.
5. Die Speicherung von dynamischen Daten wird als potentiell sinnvolle zusätzliche Dienstleistung eines Forschungsdatenzentrums gesehen. Im Rahmen des Prototyps sollte sich allerdings auf den Kernbereich eines Forschungsdatenzentrums beschränkt werden.
6. Vor der Umsetzung von Zusatzleistungen und/oder der Unterstützung von Online-Präsentationen ist der voraussichtliche Bedarf zu klären.
7. Die Unterstützung von rechenintensiven Analyseverfahren ist nicht in den Prototypen zu integrieren. Allerdings ist in der Architektur schon jetzt eine Erweiterung in dieser Richtung vorzusehen.
8. Der Zugriff auf die Daten durch Nutzer sollte möglichst unabhängig von dem IT-System des Nutzers erfolgen. Für diesen Zweck bieten sich Webportale wie vorgeschlagen an.
9. Die explizite Integration von Webservices zum Zugriff auf Daten lohnt sich nur bei gleichzeitiger Implementierung von Anwendungsprogrammen, die diese Webservices verwenden. Ansonsten sollte im Rahmen des Prototyps darauf verzichtet werden.
10. Eine redundante Datensicherung ist unverzichtbar. Sie sollte auf der Ebene der Speicheranbieter umgesetzt werden.

11. Im Rahmen des Prototyps sollte sich ebenfalls aus Gründen einer Fokussierung und einer schnellen Umsetzung auf die Aufnahme vollständig digitalisierter Daten beschränkt werden.
12. Langfristig wird das Forschungsdatenzentrum wahrscheinlich sein Angebot in Richtung auf eine virtuelle Forschungsumgebung ausdehnen müssen. Dies betrifft insbesondere die Zusammenarbeit von Wissenschaftlern an unterschiedlichen Orten. Schon beim Prototyp sollten die Voraussetzung für eine spätere Integration vorgesehen werden. Eine vollständige Umsetzung von entsprechenden Dienstleistungen bereits im Prototyp wird aufgrund des Aufwandes nicht empfohlen.
13. Die unterstützten Datenformate sind vor dem Beginn der Umsetzung in einer ersten Version festzulegen. Die Menge der unterstützten Formate kann später erweitert werden. Diese spätere Erweiterung kann dann allerdings bereits zu unnötigem Konversionsaufwand zu Beginn der Umsetzung geführt haben. Daher sollte die erste Auswahl der unterstützten Formate bereits sehr sorgfältig erfolgen.
14. Das Forschungsdatenzentrum benötigt ein Zugriffsverfahren, das auf einem Rechtemanagement basiert. Der jeweilige Eigentümer der Daten muss den Zugriff festlegen können. Er ist dabei inhaltlich verantwortlich, dass die entsprechenden rechtlichen Vorgaben bezüglich des Zugriffs gewährleistet werden, während das Forschungsdatenzentrum die Verantwortung für die technische Umsetzung der Zugriffskontrolle trägt. Gleichzeitig muss die Implementierung des Zugriffsverfahrens gewährleisten, dass die Voraussetzung für einen nutzerfreundlichen und effizienten Datenzugriff geschaffen wird.
15. Im Rahmen des Prototyps sind die Daten von primärem Interesse, die vollständig digitalisiert sind und deren Metadaten in einem Zielformat vorliegen. Diese Daten sind vorrangig in das Forschungsdatenzentrum aufzunehmen. Daher ist vor dem Beginn der Umsetzung eine sorgfältige Analyse bezüglich des Datenbestandes durchzuführen.
16. Für Daten, die digitalisiert sind, aber deren Metadaten nicht in einem der Zielformate vorliegen, ist es sinnvoll exemplarisch einen Konversionsprozess im Rahmen des Prototyps durchzuführen. Diese Konversion hat allerdings nicht das Ziel, den Datenbestand des Forschungsdatenzentrums wesentlich zu erweitern sondern soll den Prozess darlegen.
17. Entscheidungen über die Digitalisierung von Altdaten und die Konversion von digitalisierten Daten, deren Metadaten nicht in einem Zielformat vorliegen, sind auf inhaltlicher Basis und separat vom Aufbau des Forschungsdatenzentrums zu treffen.
18. Für das Einstellen der Daten sollte eine möglichst automatisierte Prüfung der Daten und der zugehörigen Metadaten auf Einhaltung der durch das Forschungsdatenzentrum definierten Standards implementiert werden.

19. Für die Aufnahme neuer Formate und neuer Varianten existierender Formate ist ein Prozess zu definieren, um einerseits diese Möglichkeit offen zu halten und andererseits den Aufwand für die Entscheidung transparent zu halten und damit den Verwaltungsaufwand zu beschränken.
20. Für die Forschungsdaten und ihre Metadaten muss das Speichersystem die Funktionen Speichern, Lesen, Verändern und Löschen unterstützen.
21. Die Metadaten für Forschungsdaten müssen Informationen zu dem Ursprung der Daten und möglicher späterer Veränderung der Daten beinhalten.
22. Das Forschungsdatenzentrum muss vorherige Versionen eines Datensatzes wiederherstellen können.
23. Eine Volltextsuche auf den Daten sollte nur nach einer sorgfältigen Bedarfsanalyse umgesetzt werden.
24. Eine Volltextsuche auf den Metadaten muss implementiert werden. Dabei ist sind die Zugriffsrechte zu beachten.
25. Es muss möglich sein, die Forschungsdaten in Kategorien abzulegen. Wenn es eine Dominanzordnung der Kategorisierungskriterien gibt, dann kann eine geschachtelte Verzeichnisstruktur verwendet werden. Ansonsten ist ein Dokumentenmanagementsystem basierend auf einer relationalen Datenbank mit den entsprechenden Zugriffsverfahren besser geeignet.
26. Kategorien von Forschungsdaten sollen durch grundlegende Metadaten beschrieben werden können.
27. Die Metadaten für Forschungsdaten sollen Informationen über Nutzungsrechte enthalten können.
28. Das System muss den Ablauf von Fristen automatisch feststellen können und betroffenen Personen Handlungsalternativen vorschlagen.
29. Die Metadaten der Forschungsdaten müssen Möglichkeiten zur Ablage von datenbezogenen Prüfsummen und Signaturen bieten.
30. Das Forschungsdatenzentrum sollte eine möglichst automatisierte Unterstützung für technische Metadaten beinhalten.
31. Das Forschungsdatenzentrum sollte den Nutzern Werkzeuge zur Bearbeitung der fachlichen Metadaten zur Verfügung stellen.

32. Das Forschungsdatenzentrum soll die Einbindung von Werkzeugen zur Konvertierung von Daten und Metadaten ermöglichen. Dabei ist der Schwerpunkt auf existierende Werkzeuge zu legen.
33. Das Forschungsdatenzentrum soll die Einbindung von Werkzeugen zur Validierung von Datenformaten und des Inhaltes von Metadaten erlauben. Zunächst ist dazu eine Analyse über existierende Werkzeuge zu erstellen.
34. Zur Sicherstellung der Nachhaltigkeit sind soweit möglich weit verbreitete Standardsoftwareprodukte zu verwenden.
35. Bei den Kostenbetrachtungen sind die Gesamtkosten einschließlich der Softwarepflege zu berücksichtigen.
36. Das Forschungsdatenzentrum soll in Form einer 3-Tier-Architektur aufgebaut werden. Kernstück bildet die mittlere Ebene, die für den eigentlichen Mehrwert des Forschungsdatenzentrums verantwortlich ist. Die Nutzer befinden sich in der oberen Ebene und können selbst gepflegte Werkzeuge über definierte Schnittstellen an das Forschungsdatenzentrum anbinden. Die untere Ebene deckt die eigentliche Datenspeicherung ab und sollte durch grundsätzlich austauschbare Betreiber von Speicherressourcen ausgefüllt werden.
37. Nach Klärung der Rahmenbedingungen und Schnittstellen sollten die einzelnen Komponenten in Form eines Ausschreibungsverfahrens vergeben werden.
38. Aufgrund der Vorteile bezüglich der Zukunftssicherheit der Softwareprodukte sollten kommerzielle Dokumentenmanagementsysteme zumindest bei der Auswahl berücksichtigt werden.
39. Die Speicherebene sollte im Outsourcing betrieben werden.
40. Als Lösung aus dem akademischen Umfeld bietet sich eine Kombination von Fedora und iRODS an, die bereits mehrfach auch in Deutschland in vergleichbaren Einrichtungen im Einsatz ist. Für die Implementierung wird vorgeschlagen, die Hilfe einer Einrichtung in Anspruch zu nehmen, die bereits Erfahrung mit der Umsetzung und dem Betrieb eines Fedora-Systems hat.
41. Über die Rechte und Pflichten der Beteiligten sollte vor dem Beginn des Aufbaus des Forschungsdatenzentrums Einigkeit bestehen.
42. Die Pflichten der Speicherebene werden sich an den üblichen Bedingungen bei Outsourcingverträgen orientieren.
43. Auch wenn eine wissenschaftliche Einrichtung den Betrieb der Verwaltungsebene übernimmt, so muss trotzdem klar sein, dass es sich bei dem Betrieb der

Forschungsdatenzentrums um eine Dienstleistung und nicht um eine wissenschaftliche Aufgabe handelt.

44. Die Rechte der Datennutzer orientieren sich primär an der Finanzierung. Wenn die Einrichtungen, an denen sich die Datennutzer befinden, sich nicht an der Finanzierung beteiligen, dann kann auch nur ein Best-Effort-Ansatz unterstützt werden.
45. Das Forschungsdatenzentrum ist nur dann für viele Forscher attraktiv, wenn es einen großen Umfang an Forschungsdaten beinhaltet. Aus diesem Grund müssen viele Besitzer von Daten bereit sein, ihre Daten im Forschungsdatenzentrum zu speichern. Daher ist gerade in Bezug auf die Besitzer von Forschungsdaten eine genaue Definition von Rechten und Pflichten notwendig.
46. Das Forschungsdatenzentrum wird nur dann ein Erfolg werden, wenn viele Datenbesitzer ihre Forschungsdaten zur Verfügung stellen. Dies werden die Datenbesitzer nur tun, wenn das Konzept Aussicht auf Erfolg verspricht. Dazu gehört eine ausreichende Projektfinanzierung und ein Bekenntnis des Bunds als Initiators zu dem Projekt. Deshalb sollte der Bund eine ausreichende Projektfinanzierung über einen entsprechenden Zeitraum zur Verfügung stellen.
47. Der Betrieb des Forschungsdatenzentrums sollte entsprechend der Nutzung finanziert werden. Hierbei ist sowohl die Speicherung der eigenen Daten als auch die Nutzung der Daten zu Forschungszwecken zu berücksichtigen. Eine kleinteilige Abrechnung ist möglich kann aber die Nutzung einschränken. Deshalb wird vorgeschlagen, dass die Länder anteilmäßig den Betrieb finanzieren.
48. Die Organisationsstruktur und die Aufgabenverteilung des Forschungsdatenzentrums werden von den rechtlichen Rahmenbedingungen mitbestimmt. Da eine Änderung der gesetzlichen Rahmenbedingungen üblicherweise eine erhebliche Zeit beansprucht, sollte die Strukturen so gewählt werden, dass sie den existierenden rechtlichen Rahmenbedingungen entsprechen. Insbesondere sollten keine Einrichtung oder kein Staat einen bestehenden Rechtsanspruch zu Gunsten des Forschungsdatenzentrums aufgeben müssen.
49. Der Prototyp soll der Überzeugung von Datenlieferanten dienen. Er muss so ausgelegt werden, dass in der Projektlaufzeit ein vollfunktionsfähiges System mit Schnittstellen entsteht, die potentielle Nutzer bezüglich ihrer Handhabung überzeugen.
50. Das Projekt ist als Einführung eines neuen Softwaresystems zu betrachten und entsprechend zu planen.
51. Primäres Ziel beim Aufbau Forschungsdatenzentrums ist die Vergrößerung der beherbergten Datenmenge. Daher sollten versucht werden zunächst „einfache“ Datensätze zu integrieren.

52. Auch für Wissenschaftler ist wahrscheinlich der Zugriff auf eine große Menge von Daten das primäre Ziel. Daneben sollte die Benutzerfreundlichkeit berücksichtigt werden.