

# Konzeptstudie zur Entwicklung eines Modells für eine zentrale Langzeitarchivierung von digitalen Primär- und Sekundärdaten der Forschung für die Schweiz

Situation, Bedarf, Nutzenziele, Modelle,  
Anspruchsgruppen, Rahmenbedingungen und  
europäisches Umfeld

**Report**

**Author(s):**

Keller-Marxer, Peter

**Publication date:**

2008

**Permanent link:**

<https://doi.org/10.3929/ethz-a-006070903>

**Rights / license:**

[In Copyright - Non-Commercial Use Permitted](#)

Konzeptstudie zur

# Entwicklung eines Modells für eine zentrale Langzeitarchivierung von digitalen Primär- und Sekundärdaten der Forschung für die Schweiz

Situation, Bedarf, Nutzenziele, Modelle, Anspruchsgruppen,  
Rahmenbedingungen und europäisches Umfeld

Bericht erstellt für



**ETH-Bibliothek**  
Eidgenössische Technische Hochschule Zürich  
Swiss Federal Institute of Technology Zurich

ETH-Bibliothek, Eidgenössische  
Technische Hochschule Zürich

**E-lib.ch**

Elektronische Bibliothek Schweiz  
Bibliothèque électronique suisse  
Swiss electronic library

Projekt Elektronische Bibliothek Schweiz,  
Schweizerischen Universitätskonferenz

**INTERNES DOKUMENT**

080401-01– Version 1.4 – 2008-12-22 – (Schlussversion)

## Autoren

Peter Keller-Marxer, Dr. phil. nat.  
Niklaus Bütikofer, lic. phil.

ikeep AG  
Morgenstrasse 129  
CH-3018 Bern  
+41 31 998 42 80  
peter.keller@ikeep.com

## Dokument-Historie:

Version	Datum	Status	Autoren	Bemerkungen
1.0	2008-10-01	interner Entwurf	PK,NB	interner Review
1.1	2008-10-25	interner Entwurf	PK,NB	Erweiterungen gemäss internem Review
1.2	2008-11-06	Entwurf	PK,NB	Abgabe zum Review durch den Kunden
1.3	2008-12-09	Entwurf Schlussversion	PK,NB	Änderungen gemäss Review mit Kunde am 2.12.08
1.4	2008-12-22	Schlussversion	PK,NB	Kleine Korrekturen gem. Wunsch Kunde 22.12.08

# Inhalt

<b>1</b>	<b>Management-Zusammenfassung</b> .....	<b>1</b>
<b>2</b>	<b>Einleitung und Übersicht</b> .....	<b>3</b>
2.1	Begriffe.....	4
2.2	Ausgangslage .....	5
2.2.1	Sekundärnutzung.....	5
2.2.2	Integrität der Forschung .....	6
2.2.3	Archivierungsgesetze und Anbietepflicht.....	7
2.3	Schwerpunktsetzung .....	8
2.4	Modelle .....	10
2.5	Struktur des Berichts .....	12
<b>3</b>	<b>Situation, Bedarf und Umfang</b> .....	<b>13</b>
3.1	Primärdaten aus der Forschung .....	13
3.1.1	Kurzcharakterisierung der Unterlagen .....	13
3.1.2	Möglicher Nutzen der Aufbewahrung und Archivierung .....	14
3.1.3	Bedarf und Nutzungshäufigkeit.....	15
3.1.3.1	Integritätsrichtlinien der ESF und OECD .....	16
3.1.3.2	Aufbewahrungsregeln in der Schweiz und Deutschland.....	16
3.1.3.3	Zitierbarkeit von Forschungsdaten .....	18
3.1.3.4	Aufbewahrung für die Sekundärnutzung.....	19
3.1.4	Archivierungssituation .....	20
3.2	Digitale Publikationen .....	22
3.2.1	Kurzcharakterisierung der Unterlagen .....	22
3.2.2	Möglicher Nutzen der Archivierung .....	23
3.2.3	Bedarf und Nutzungshäufigkeit.....	24
3.2.4	Archivierungssituation.....	24
3.3	Digitale Lehr- und Lernunterlagen .....	25
3.3.1	Kurzcharakterisierung der Unterlagen .....	25
3.3.2	Möglicher Nutzen der Archivierung .....	25
3.3.3	Bedarf und Nutzungshäufigkeit.....	25
3.3.4	Archivierungssituation.....	25
3.4	Digitalisate .....	26
3.4.1	Kurzcharakterisierung der Unterlagen .....	26
3.4.2	Möglicher Nutzen der Archivierung .....	26
3.4.3	Nutzungshäufigkeit und Bedarf.....	26
3.4.4	Archivierungssituation.....	27
<b>4</b>	<b>Anspruchsgruppen</b> .....	<b>28</b>
4.1	Rechtlicher Rahmen .....	32

4.1.1	Bundesgesetz über die Archivierung.....	33
4.1.2	Kantonales Recht .....	35
4.1.3	Bibliotheksrecht .....	35
<b>5</b>	<b>Modelle für Primärdaten .....</b>	<b>36</b>
5.1	Modell «Aufbewahrung mit Selbstregulierung» .....	36
5.1.1	Strategieziele.....	37
5.1.2	Begründung.....	38
5.1.3	Operative Ziele .....	39
5.1.3.1	Kosten und Finanzierung .....	43
5.1.4	Erläuterungen .....	45
5.1.5	OAIS-Erweiterung der operativen Ziele.....	47
5.1.5.1	Kosten und Finanzierung .....	48
5.1.6	Übersicht.....	48
5.2	Modell « Archivierung mit Anbietepflicht» .....	52
5.2.1	Strategieziele.....	54
5.2.2	Begründung.....	55
5.2.3	Operative Ziele .....	55
5.2.3.1	Kosten und Finanzierung .....	58
<b>6</b>	<b>Modell für Sekundärdaten .....</b>	<b>60</b>
6.1	Strategieziele .....	60
6.2	Begründung.....	61
6.3	Operative Ziele für eJournals und eBooks .....	64
6.3.1	Kosten und Finanzierung.....	67
6.4	Operative Ziele für weitere Datentypen .....	68
6.4.1	Wissenschaftliche Datenbanken .....	68
6.4.2	eLearning-Objekte.....	69
6.4.3	Scanning-Masterkopien.....	70
<b>7</b>	<b>Alternativmodelle .....</b>	<b>73</b>
7.1	Strategieträger / nationale Task Force.....	73
7.2	Nationales Kompetenz-Zentrum .....	74
7.3	Nationales Kompetenz-Netzwerk .....	74
7.4	Register, Inventar und Portal .....	75
<b>8</b>	<b>Europäisches Umfeld .....</b>	<b>76</b>
8.1	EU-Projekte im Rahmen von FP6 und FP7 .....	77
8.2	Institutionelle Repositorien .....	78
8.3	Nationale Depot-Bibliotheken.....	79
8.4	Alliance for Permanent Access to the Records of Science.....	80
8.4.1	PARSE.insight.....	82
8.5	Deutsche Allianz-Initiative Digitale Information.....	82
8.6	Weitere ähnliche Aktivitäten .....	84
8.6.1	Strategien, Aktionspläne und nationale Infrastrukturen.....	85
8.6.1.1	Internationale Aktivitäten.....	85

8.6.1.2	Nationale Aktivitäten .....	85
8.6.2	Fachspezifische Aufbewahrung von Primärdaten der Forschung.....	89
8.6.2.1	Sozialwissenschaften .....	89
8.6.2.2	Naturwissenschaften.....	91
8.6.3	Archivierung elektronischer Verlagszeitschriften .....	92
8.6.3.1	Niederlande: e-Depot Koninklijke Bibliotheek.....	92
8.6.3.2	Deutschland: Kopal.....	92
8.6.3.3	Schweiz: e-Helvetica .....	93
8.6.3.1	Frankreich: Hyper Article en Ligne (HAL) .....	93
8.6.3.2	International: Portico .....	93
<b>9</b>	<b>Bibliographie.....</b>	<b>95</b>



# 1 Management-Zusammenfassung

Der überwiegende Teil moderner wissenschaftlicher Prozesse basiert auf elektronischen Daten: Digital erfasste, dokumentierte und gespeicherte Mess-, Simulations- und Erhebungsdaten (Primärdaten) bilden die Grundlage für Auswertungen und Forschungsergebnisse, welche zunehmend ebenfalls in elektronischer Form formuliert und publiziert (Sekundärdaten) werden, und fließen durch Sekundärnutzung wiederum in neue Forschungsprozesse ein.

Eine wachsende Zahl von wissenschaftlichen Publikationen und Lehrmaterialien wird zudem nur noch in elektronischer Form genutzt und publiziert (eJournals, eBooks, Datenbanken). Dadurch werden Strategien, Methoden und technische Lösungen nötig, mit welchen die Verfügbarkeit dieser Sekundärdaten für die Forschung gesichert werden kann, nachdem Nutzungslizenzen auslaufen oder Verlage den Online-Zugriff oder ihre Geschäftstätigkeit einstellen.

Beim Umgang mit Primärdaten hat ein Umdenken stattgefunden: Nach Abschluss des Forschungsprojekts stellen sie nicht obsolekte Artefakte dar, sondern sind Grundlage der Überprüfbarkeit der Integrität von Forschungsergebnissen und bleiben Ressource für die Sekundärnutzung.

In der Schweiz haben zahlreiche Forschungsinstitutionen sowie alle Organe der Forschungsförderung (aufgrund des seit März 2008 gültigen neuen Artikel 11a des Forschungsgesetzes) Vorschriften erlassen, welche unter anderem die sichere und dokumentierte Aufbewahrung von Primärdaten auch nach Abschluss des Forschungsprojektes verlangen. In der Schweiz muss der Aufbewahrungszeitraum «für das Fachgebiet adäquat» sein, in Deutschland und einigen anderen Ländern ist eine explizite Frist von mindestens zehn Jahren vorgeschrieben.

Die Verantwortung für diese sichere Aufbewahrung wird dabei heute einzig an die Forschenden delegiert, denen aber in der Regel, nicht zuletzt wegen der hohen Volatilität und Mobilität im Arbeitsmarkt des Wissenschaftsbetriebs, die nötigen organisatorischen und technischen Mittel fehlen, um eine solche langfristige Verantwortung wahrzunehmen.

Die gesamteuropäische Initiative «Alliance for Permanent Access to the Records of Science in Europe» und die «Allianz-Initiative ‚Digitale Information‘» von neun deutschen Forschungsgesellschaften erkennen deshalb bei der Bewahrung von Primärdaten zur Sicherung der Integrität der Forschung und der Sekundärnutzung einen dringenden Handlungsbedarf.

Die Schweizerische Universitätskonferenz, die Rektorenkonferenz der Schweizer Universitäten, der ETH-Rat, das Bundesamt für Berufsbildung und Technologie haben das Thema der Langzeitarchivierung von Primär- und Sekundärdaten als Teilprojekt in das nationale Vorhaben «E-lib.ch» (Elektronische Bibliothek Schweiz) aufgenommen. Die Konferenz der Universitätsbibliotheken hat die ETH-Bibliothek beauftragt, Modelle für eine nationale, zentrale Dienstleistung in diesem Bereich zu entwickeln.



Im Auftrag der ETH-Bibliothek haben wir den Rahmen, den Bedarf und die Nutzenziele untersucht und unter Berücksichtigung der Vorhaben und Strategien im europäischen Umfeld zwei komplementäre Modelle für Primärdaten und ein Modell für Sekundärdaten formuliert.

Das vorgeschlagene Modell für Sekundärdaten entspricht im Wesentlichen der in Deutschland verfolgten Strategie, berücksichtigt jedoch zusätzlich auch Synergien mit möglichen Partnern im Bereich der Sicherung von eLearning-Objekten und «Master»-Kopien von Digitalisaten.

Das erste Modell für Primärdaten verfolgt einen «Self Deposit»-Ansatz und fokussiert pragmatisch auf Bedürfnisse des Forschungsbetriebs hinsichtlich der neuen Richtlinien zur Integrität in der Forschung. Es sieht keine Regulierung vor und geht von der Selbstverantwortung und Eigeninitiative der Forschenden und Forschungsinstitutionen aus. Sie selber sollen die Anforderungen an Aufbewahrungsdauer, Umfang, Dokumentation und Formate der von ihnen in der Dienstleistung deponierten Primärdaten im Hinblick auf die für sie jeweils geltenden Integritätsrichtlinien und den Wert für eine längerfristige Sekundärnutzung bestimmen und umzusetzen. Zur Unterstützung sieht das Modell ein Beratungsangebot durch ein Kompetenznetzwerk vor.

Dabei bietet das Modell den Forschenden und Forschungsinstitutionen neben der sicheren und nachweislich integren Aufbewahrung der Daten auch Zusatznutzen, etwa die Möglichkeit, die Zugänglichkeit deponierten Daten selber zu bestimmen und freiwillig öffentlich zugänglich zu machen («Open Access»), die Bildung von thematischen oder institutionsbezogenen Inventaren von Forschungsdaten sowie die Zitierbarkeit von Primärdaten. Letzteres stellt ein wesentliches Merkmal dar, da Primärdaten in der Regel nicht publiziert werden, wodurch Forschende insbesondere bei der Sekundärnutzung von Daten heute keine Möglichkeit haben, in Publikationen auf eine einfache, eindeutige und allgemein nachvollziehbare Weise auf einen verwendeten Datenbestand zu verweisen.

Das zweite, zum ersten komplementäre Modell verfolgt einen regulierenden Ansatz und berücksichtigt, dass auch die Archivierung von dauerhaft wertvollen Primärdaten im Sinne der staatlichen Archivgesetze (Bund und Kantone) bisher ungelöst ist. Es sieht ein systematisches und doch pragmatisches Vorgehen zur archivischen Bewertung von Primärdaten vor und schafft so für Forschungsinstitutionen die Voraussetzungen zur Wahrnehmung ihrer Pflichten im Rahmen der Archivgesetze. Dieses Modell entspricht insbesondere auch den Erwartungen der Öffentlichkeit, dass Forschung in ethisch und politisch heiklen Gebieten und dort, wo sie grosse Auswirkungen auf Mensch und Umwelt haben oder den politischen und gesetzgeberischen Prozess beeinflussen kann, längerfristig oder dauerhaft nachvollziehbar bleibt.

Zusätzlich skizzieren wir vier prinzipiell anders gewichtete Alternativmodelle, bei denen die nationale Dienstleistung selber keine Daten archiviert, sondern strategisch-koordinierende und beratende Aufgaben sowie die Organisation eines nationalen Kompetenznetzwerks „Langzeitarchivierung“ nach den Vorbildern in Deutschland und Frankreich übernimmt.

Das Engagement der «E-lib.ch»-Partner erfolgt aufgrund der Erkenntnis, dass die langfristige Sicherung und Verfügbarkeit wissenschaftlicher digitaler Information Voraussetzung für eine gute wissenschaftliche Informationsversorgung an den Hochschulen über Generationen ist.

Es entspricht damit auch den Empfehlungen der OECD von 2007 zum Umgang mit Daten aus öffentlich finanzierter Forschung: Strategien, Projekte und Dienstleistungen zur Aufbewahrung und Sekundärnutzung von Primärdaten sollten als integrale Bestandteile der nationalen Forschungsinfrastruktur eines Landes verstanden, geplant und implementiert werden.

## 2 Einleitung und Übersicht

Das Projekt, in dessen Rahmen die vorliegende Analyse- und Konzeptstudie erstellt wurde, ist ein Einzelvorhaben des nationalen Projekts «E-lib.ch – Elektronische Bibliothek Schweiz»<sup>1</sup>. «E-lib.ch» wird von der Schweizerischen Universitätskonferenz (SUK) für die Beitragsperiode 2008 – 2011 getragen, wobei die Teilnahme der Institutionen des ETH-Bereichs durch den ETH-Rat, jene der Fachhochschulen durch Unterstützung des Bundesamtes für Berufsbildung und Technologie (BBT) ermöglicht wird. Innerhalb von E-lib.ch werden von einer Vielzahl von Partnern Teilprojekte realisiert, die einen einfachen und schnellen Zugang zu Informationsressourcen unterschiedlichster Art gewährleisten. Die Projektleitung und Projektkoordinierung für das Gesamtprojekt E-lib.ch ist an der ETH-Bibliothek angesiedelt, und die Konferenz der Universitätsbibliotheken der Schweiz (KUB) begleitet das Gesamtprojekt.

Auftraggeberin des vorliegenden Einzelvorhabens mit dem Projekttitel «Konzeptentwicklung eines Modells für eine zentrale Langzeitarchivierung von digitalen Primär- und Sekundärdaten für die Schweiz», ist die KUB. Ausführende Organisation ist die ETH-Bibliothek. Der Projektantrag der Direktion der ETH-Bibliothek, der Grundlage der Projektfinanzierung durch die KUB ist, nennt folgende Projektziele [1] (gekürzt):

- *«Das Ziel des Projektes ist die Entwicklung eines Konzeptes mit einem oder mehreren Modellen bzw. Modellvarianten für eine zentrale vertrauenswürdige Langzeitarchivierung von digitalen Primär- und Sekundärdaten für die Schweizer Hochschulen, Forschungszentren und Hochschulbibliotheken.»*
- *«Das Ergebnis des Projektes soll ein Konzept mit einem oder mehreren Modellen bzw. Modellvarianten für eine zentrale Langzeitarchivierung digitaler Primär- und Sekundärdaten für die Schweizer Hochschulen, Forschungszentren und Hochschulbibliotheken sein, das zum Beispiel Vorschläge vorsieht für die Entwicklung nationaler Standards und Guidelines im Bereich der digitalen Langzeitarchivierung, für eine schweizweite Service-Einrichtung zur digitalen Langzeitarchivierung sowie für ein digitales nationales vertrauenswürdige Langzeitarchiv.»*

In Rahmen dieser Projektziele wurden wir beauftragt, die rechtlichen und regulativen Randbedingungen zu analysieren und darzustellen, die potenziellen Anspruchsgruppen zu identifizieren, die Bedarfs- und Nutzenaspekte einer Langzeitaufbewahrung nach Unteragentypen zu beurteilen, die Situation des Vorhabens im europäischen und internationalen Umfeld darzustellen und schliesslich Modellvarianten vorzuschlagen, welche dem vom Projekt angestrebten zentralen und nationalen Dienstleistungsanspruch gerecht werden und aus unserer Sicht unter realen Bedingungen auch realisierbar sind. Die Konzeption von Modellen für experimentelle Prototypen („Testbeds“) ist dagegen nicht Bestandteil dieser Studie.

---

<sup>1</sup> <http://www.e-lib.ch>

## 2.1 Begriffe

Unter «**Primärdaten**» werden die ursprünglichen, von offensichtlichen Fehlern bereinigten Rohdaten (Messdaten, Erhebungsdaten und Simulationsdaten) verstanden, welche als Grundlage für weitere Auswertungen und schlussendlich für die Ableitung von wissenschaftlichen Resultaten dienen. Dagegen werden Auswertungsdaten und Daten der Forschungsergebnisse nicht als Primärdaten bezeichnet, da sie grundsätzlich aus den Primärdaten reproduziert werden können, wenn das Auswertungsverfahren bekannt ist.

Mitgemeint sind bei «Primärdaten» jedoch *immer* auch diejenigen Zusatzinformationen (Metadaten und Dokumentation), durch welche die Daten mindestens längerfristig für Dritte verständlich und interpretierbar werden (z.B. Mess- bzw. Erhebungsgrößen, Masseinheiten, Datentypen, Kalibrierungen von Messinstrumenten, Erhebungs- und Messkontext, Laborjournale etc.). *Nicht* mitgemeint sind dagegen alle weiteren Unterlagen, welche den *Auswertungs- und Forschungsprozess* dokumentieren und dadurch nachvollziehbar machen. Sind diese ebenfalls mitgemeint, so sprechen wir stattdessen von «**Forschungsunterlagen**».

Unter «**Sekundärdaten**» verstehen wir elektronischen Ausgaben der von Bibliotheken lizenzierten wissenschaftlichen Zeitschriften (eJournals) und Fachbücher (eBooks), die von den Universitäten hergestellten oder gekauften elektronischen Lehr- und Lernunterlagen (eLearning) sowie elektronische Hochschulpublikationen (Dissertationen, Masterarbeiten, Preprints etc.)

Wir unterscheiden «Aufbewahrung» und «Archivierung»:

- **Aufbewahrung** bedeutet die *weitgehend formfreie* Vorhaltung von Informationen, bei der die Anforderungen an Datenformate, Dokumentation und Ordnung im Einzelfall bestimmt werden, sei es durch den Datenproduzenten oder durch unverbindliche Empfehlungen eines Fachgebiets. Aufbewahrung kann freiwillig erfolgen oder durch Gesetze<sup>2</sup> oder supranationale Regulierungen vorgeschrieben sein.
- **Archivierung**<sup>3</sup> bedeutet die *formgebundene* Vorhaltung von Daten, bei der unabhängig vom Einzelfall gültige und verbindliche Anforderungen an die Selektion der zu archivierenden Unterlagen sowie an deren archivtaugliche Dokumentation, Ordnung und technische Aufbereitung definiert und vorgeschrieben sind. Archivierung kann freiwillig erfolgen oder durch Gesetze<sup>4</sup> oder supranationale Regulierungen vorgeschrieben sein.
- **Archivierung nach staatlichen Archivgesetzen** erfolgt bezüglich Form und Umfang nach den Bestimmungen dieser Gesetze und den Vorgaben der staatlichen Archive. Sie beschränkt sich auf Unterlagen, die durch die Archive in Zusammenarbeit mit den Unterlagenproduzenten als archivwürdig bestimmt und dauernd aufbewahrt werden.

Sowohl Aufbewahrung wie Archivierung erfolgen in der Regel mit definierten Zweckbestimmungen und mit Blick auf heute bereits bekannte Nutzergruppen.

<sup>2</sup> Zahlreiche Gesetze schreiben die Aufbewahrung bestimmter Unterlagen über Zeiträume von 1 – 20 Jahren vor. Dazu gehört zum Beispiel das Obligationenrecht.

<sup>3</sup> In der digitalen Informationstechnologie ist „Archivierung“ im Sinne von „Auslagern und sicher Speichern“ seit über 50 Jahren fester Bestandteil der Fachterminologie. Diese Bedeutung berücksichtigen wir hier nicht.

<sup>4</sup> Zahlreiche Gesetze und Regulierungen schreiben die Archivierung von bestimmten Unterlagen über bestimmte Zeiträume von 1 – 30 Jahren *in einer definierten Ordnung und Form* vor. Dazu gehören zum Beispiel das Mehrwertsteuergesetz oder kantonale Archivverordnungen zum Grundbuch.

Aufbewahrung und Archivierung können zeitlich beschränkt oder unbeschränkt sein. Ist der Zeitraum bei digitalen Daten länger als zehn Jahre, spricht man von Langzeitarchivierung, da diese Zeiträume bereits länger sind als die Lebensdauer heutiger Informatik- und Speichersysteme. Gesetzliche Archivierung erfolgt in der Schweiz dauerhaft.

Weiter unterscheiden wir in dieser Studie nur vier Gründe für die Aufbewahrung bzw. Archivierung von wissenschaftlichen Primär- und Sekundärdaten:

- **«Integrität der Forschung»:** Beschränkte Aufbewahrung (5 – 10 Jahre) von Primärdaten zur Erfüllung der Richtlinien zur „Integrität der Forschung“ und „Guten wissenschaftlichen Praxis“.
- **Sekundärnutzung:** Aufbewahrung von wertvollen Primärdaten für die Sekundärnutzung, häufig zeitlich beschränkt.
- **Sicherung:** Dauerhafte Aufbewahrung oder Archivierung von wissenschaftlichen Verlags-Publikationen zur Gewährleistung der Verfügbarkeit wissenschaftlicher digitaler Publikationen auch nach Ende einer Lizenz und der Existenz des Verlags.
- **Archivgesetze:** Archivierung von Primärdaten und Sekundärdaten durch staatliche Archivstellen aufgrund staatlicher Archivgesetze.

In den Zielen dieses Teilprojekts von E-lib.ch werden allerdings nur der zweite und dritte Grund explizit angesprochen. Grundsätzlich besteht die Aufgabe der Archivierung aus der Gesamtheit aller nötigen Arbeitsschritte, um Unterlagen auf Grund eines bleibenden Wertes für ein bezeichnetes Publikum

- zu identifizieren, auszuwählen und in «das Archiv» zu überführen,
- nach einheitlichen und systematischen Kriterien zu ordnen und zu verzeichnen,
- in authentischer Form technisch zugreifbar, lesbar und benutzbar
- sowie im Entstehungszusammenhang verständlich zu erhalten.

Die Ziele der Aufbewahrung sind in der Regel weniger umfassend.

## 2.2 Ausgangslage

### 2.2.1 Sekundärnutzung

Dieses Teilprojekt von E-lib.ch geht in seinen Zielsetzungen von einer *prinzipiellen* Relevanz der langfristigen Aufbewahrung von digitalen Primärdaten und dauerhaften Archivierung von Sekundärdaten für die Forschung aus [1]:

- *«Der langfristige Erhalt und die dauerhafte Verfügbarkeit von wissenschaftlicher digitaler Information (Primär- und Sekundärdaten) ist Voraussetzung für eine gute wissenschaftliche Informationsversorgung an den Schweizer Hochschulen über Generationen hinaus.»*
- *«Der überwiegende Teil moderner wissenschaftlicher Prozesse wird über elektronische Daten abgewickelt: Auf der Basis von digital erfassten Messdaten oder aufgrund elektronisch dokumentierter Erhebungen (Primärdaten) erfolgen Auswertungen, die ohne mehr oder minder intensive technische Unterstützung nicht mehr zu bewältigen wären. Die Ergebnisse dieses Prozesses werden wiederum in elektronischer Form formuliert und publiziert (Sekundärdaten).»*

Nachdem Primärdaten (Messdaten, Erhebungsdaten und Simulationsdaten) bereits seit längerem fast ausschliesslich nur noch in elektronisch digitaler Form anfallen, werden heute zunehmend auch die Forschungsergebnisse entweder zusätzlich oder sogar ausschliesslich in elektronischen Ausgaben von wissenschaftlichen Zeitschriften (eJournals) publiziert. Hinzu kommen bei den Sekundärdaten auch vermehrt elektronische wissenschaftliche Fachbücher (eBooks) sowie Lehr- und Lernunterlagen (eLearning), für die es keine papiergebundenen Exemplare mehr gibt.

Dabei stellen die Auswertung von Primärdaten deren Primärnutzung dar, ihre spätere Nutzung durch den Datenproduzenten die Nachnutzung und ihre spätere Nutzung durch Dritte die Sekundärnutzung dar. Bei Sekundärdaten stellt die Publikation der Forschungsergebnisse die Primärnutzung (zur Bekanntmachung der Resultate im Wissenschaftsbereich) und die Nutzung der Publikation als Quelle oder Zitat die Sekundärnutzung dar. Sowohl die Primärdaten wie Sekundärdaten können also wiederum in neue Forschungsprozesse einfließen.

Bei Primärdaten bedingt dies jedoch, dass die betreffenden Daten nicht nur bis zum Schluss des Forschungsprojektes, sondern auch darüber hinaus längerfristig verfügbar sind. Es bedingt auch, dass die Daten in einer Form aufbewahrt werden, in der sie durch Sekundärnutzer später noch verstanden und interpretiert werden können.

Bei Sekundärdaten bedingt es, dass die digitalen Publikationen auch nach Ablauf von Lizenzen oder Verlagsauflösungen noch zugreifbar bleiben. Es bedingt auch, dass die Sekundärdaten in einer Form archiviert werden, die eine dauerhafte Benutzbarkeit und Verstehbarkeit ermöglicht. Dies kann die periodische Konversion in neue Formen bedeuten, wenn die bisher verwendeten technischen Grundlagen zu stark veraltet sind.

### 2.2.2 Integrität der Forschung

Bei der Publikation von Forschungsergebnissen in Zeitschriften übernehmen die Verlage, deren «Editorial Boards» und «Peer Review»-Gruppen grosse Anstrengungen, die Zuverlässigkeit, Güte, Zitierbarkeit und effiziente Verbreitung der publizierten Inhalte zu garantieren. Immer häufiger werden jedoch Betrugsversuche, Manipulationen und Fälschungen von Forschungsergebnissen bekannt, die nicht zuletzt auch auf einen stark zunehmenden Erfolgs- und Publikationsdruck auf die Forschenden zurückzuführen sind.

In den letzten drei Jahren haben deshalb weltweit Forschungsorganisationen und -verbände Empfehlungen und Richtlinien zur „Integrität in der Forschung“ und der „Guten wissenschaftlichen Praxis“ veröffentlicht, darunter auch die OECD und die European Science Foundation, welche fordern, dass die Primärdaten, auf welchen die publizierten Resultate beruhen, längerfristig aufbewahrt werden sollen, damit die Korrektheit und Verlässlichkeit von Resultaten nach deren Publikation nachgeprüft werden können, sei es durch andere Forschende oder (bei Betrugsverdacht) auch durch universitäre Fachgremien.

In der Schweiz haben letztes und dieses Jahr die Organe der Forschungsförderung – der Nationalfonds und die vier wissenschaftlichen Akademien – neue Richtlinien zur „Guten wissenschaftlichen Praxis“ erlassen, zu deren Einhaltung sie ihre Beitragsempfänger/innen auf Grund des seit März 2008 gültigen neuen Artikels 11a des Forschungsgesetzes verpflichten müssen.

Diese Richtlinien der Organe der Forschungsförderung beinhalten auch eine sichere und unveränderte Aufbewahrung der originalen Primärdaten über einen längeren, adäquaten Zeitraum. Dies nicht zu tun, bedeutet einen Verstoss gegen die „Gute wissenschaftliche Praxis“ und gilt in Untersuchungsverfahren bei Betrugsverdacht prinzipiell bereits als ein Indiz für einen Betrug. Zahlreiche Universitäten und Hochschulen sind den Beispielen der Organe der Forschungsförderung bereits gefolgt und haben selber auch entsprechende Reglemente erlassen.

Die *generelle*, wenn auch in der Schweiz zeitlich nicht klar definierte Aufbewahrungspflicht –zum Beispiel in Deutschland gilt eine 10-Jahre-Regel –, stellt für die Forschenden in der Schweiz eine grundsätzlich neue Situation dar, denn die Richtlinien delegieren die Verantwortung für die Aufbewahrung der Primärdaten direkt an sie und nicht an die Institution. Angesichts der Volatilität von Forschungsprojekten und der hohen Stellenmobilität der Forschenden ist eine langfristige Verantwortung für die sichere Aufbewahrung von Daten aus zahlreichen Projekten für die einzelne Forscherin, den einzelnen Forscher eine kaum handhabbare Aufgabe.

Die OECD sowie zum Beispiel die neun grossen Forschungsgesellschaften in Deutschland sehen bezüglich der Aufbewahrung oder Archivierung von Primärdaten dringenden Handlungsbedarf.

### 2.2.3 Archivierungsgesetze und Anbietepflicht

Die meisten staatlichen Archivgesetze enthalten einen umfassenden Unterlagenbegriff, der unabhängig vom Informationsträger alle aufgezeichneten Informationen einschliesst, welcher der Bund bzw. die Kantone in Erfüllung ihrer öffentlichen Aufgaben empfangen oder erstellen.

Archivwürdig sind bspw. gemäss dem Bundesgesetz über die Archivierung (BGA) [2] jene Unterlagen, *«die von juristischer oder administrativer Bedeutung sind oder einen grossen Informationswert haben»*. Archiviert werden Unterlagen gemäss BGA ausserdem dann, wenn sie *«rechtlich, politisch, wirtschaftlich, historisch, sozial oder kulturell wertvolle Unterlagen des Bundes»* sind. Analoges liesse sich aus den meisten Archivierungsgesetzen der Kantone zitieren.

Ob bestimmte Primärdaten (und Forschungsunterlagen) im Einzelfall im Sinne der genannten Bewertungskriterien archivwürdig sind, liegt in der Bewertungskompetenz der zuständigen Archivstelle. Diese Archivstellen sind das Bundesarchiv für die Bundesversammlung, den Bundesrat, die Parlamentsdienste sowie die zentrale und dezentrale Bundesverwaltung. Die autonomen Anstalten des Bundes und ähnliche bundeseigene Institutionen führen dagegen eigene Archive und archivieren ihre Unterlagen nach den Grundsätzen des BGA und in Absprache mit dem Bundesarchiv selbstständig. Zu diesen Stellen gehören auch die beiden ETHs, der ETH-Rat sowie EMPA, EAWAG, WSL, PSI und Swissmedic.

Für die kantonalen Universitäten und Fachhochschulen sind die zuständigen Archivstellen die kantonalen Staatsarchive, wobei die meisten Universitäten eine eigene Archivstelle entweder mit eigenem Archiv im Auftrag des Staatsarchivs oder als Vermittlerstelle zwischen Universität und Staatsarchiv führen.

Die den Archivierungsgesetzen unterstehenden Stellen haben die Anbietepflicht zu erfüllen: Sie müssen alle Unterlagen spätestens wenn sie sie nicht mehr ständig benötigen dem zuständigen Archiv zur Archivierung anbieten und auf ihre Archivwürdigkeit beurteilen lassen.

Die im Sinne der Archivgesetze als archivwürdig bestimmte Unterlagen werden schliesslich archiviert. Die Aufbereitung für die Langzeitarchivierung muss in einzelnen Rechtskreisen (z.B. beim Bund) von den anbietepflichtigen Stellen nach den Vorgaben des Archivs durchgeführt werden, in einzelnen Kantonen ist diese Arbeit dem zuständigen Archiv aufgetragen.

Die staatlichen Archivstellen haben sich, wenn überhaupt, bisher nur am Rande mit Forschungsunterlagen beschäftigt. Die (in Kapitel 3.1.3 diskutierten) Aufbewahrungspflichten von wissenschaftlichen Primärdaten aus Gründen der „Integrität der Forschung“ gelten allerdings unabhängig von der Anbietepflicht bzw. die Archivwürdigkeit im Sinne der staatlichen Archivgesetze. Beide Aufgaben schliessen sich nicht gegenseitig aus, sie ergänzen sich aber auch nicht unbedingt: Einerseits müssen die im Sinne der Gesetze *nicht* archivwürdigen Primärdaten trotzdem wegen den Integritätsrichtlinien aufbewahrt werden (allerdings nur zeitlich begrenzt).

Umgekehrt ist eine Archivierung von archivwürdigen Primärdaten nicht zwingend gleichzeitig auch der Erfüllung der Integritätsrichtlinien durch die Forschenden dienlich, da die Primärdaten für die dauerhafte Archivierung in der Regel aufbereitet und zum Beispiel in langzeittaugliche Formate konvertiert werden müssen. Es ist vor allem bei komplexen Binärformaten wahrscheinlich, dass eine solche Konversion zu Verlusten an Originalität und Information (z.B. Präzision, Datentypen, Struktur etc.) führt, die aus der Sicht der Archivierung zwar akzeptabel sind, jedoch das exakte Überprüfen der Forschungsergebnisse, die aus den Primärdaten in ihrer *ursprünglichen* Form gewonnen wurden, verunmöglichen. Eine Konversion mit dieser Konsequenz wäre im Sinne der bestehenden Vorschriften zur Forschungsintegrität unzulässig, und die Primärdaten müssten zusätzlich zur Archivierung auch noch in der Originalform aufbewahrt werden.

### 2.3 Schwerpunktsetzung

Die Ziele dieses Teilprojekts von E-lib.ch fokussieren bezüglich der Bewertung und Selektion von Primärdaten für die Aufbewahrung auf die Selbstregulierung der Wissenschaft und den direkten Nutzen für die Forschenden und Forschungsprojekte [1]:

- *«Bewertung und Selektion von digitalen Daten und Objekten: Eine inhaltliche Bewertung wissenschaftlicher Messdaten soll und kann nicht allein durch eine Archiveinrichtung vorgenommen werden. Sie ist im Wesentlichen Sache der Produzentinnen und Produzenten. Diese müssen vor der Datenablieferung eine Vorstellung davon entwickeln, welche Daten und Objekte sie in welcher Form zukünftig wieder benötigen könnten oder welche sie anderen Wissenschaftlerinnen und Wissenschaftlern zur Verfügung stellen möchten.»*

Den Anspruch der gesetzlichen Archivierung sehen die Projektziele dagegen als unabhängig von diesem Projekt und durch staatliche Archivstellen zu erfüllen:

- *«Ein anderer Teil der zu archivierenden digitalen Daten und Objekte wird bereits von vorgelegerten Stellen gesammelt und aufbereitet. Dieser Prozess beinhaltet in der Regel bereits eine Auswahl und/oder Bewertung sowie eine Entscheidung über die Erhaltungswürdigkeit.»*

Wir sind jedoch der Meinung, dass eine Fokussierung auf den Sekundärnutzen und die selbstbestimmte Bewertung (*«Sie ist im Wesentlichen Sache der Produzentinnen und Produzenten.»*) nicht den Gesamtnutzen erfasst, den eine nationale Archivilösung erbringen könnte, wenn sie neben den Bedürfnissen der Forschenden auch die Ansprüche der staatlichen Archivgesetze

berücksichtigen würde. Dies lässt sich am Beispiel des im BGA genannten Kriteriums „grosser Informationswert“ für die Archivwürdigkeit verdeutlichen: Primärdaten, die aus Sicht der wissenschaftlichen Einzeldisziplin (in deren Rahmen sie erzeugt wurden) einen „grossen Wert für eine spätere Sekundärnutzung“ durch eben diese Fachdisziplin besitzen, haben dadurch möglicherweise auch einen „grossen Informationswert“ im Sinne des BGA.

Wir halten uns mit dem vorgeschlagenen Basismodell für Primärdaten an den vom Projektauftrag gesetzten Fokus, versuchen aber trotzdem, eine zweite Modell-Komponente zu skizzieren, welche in den Grundzügen den Ansprüchen der gesetzlichen Archivierung dient.

Im Sinne des Projektauftrags beschränken wir uns dabei aber ebenfalls auf Primärdaten, also die eigentlichen Daten zusammen mit einer Dokumentation, durch welche die Daten für Dritte verständlich und interpretierbar sowie die Forschungsergebnisse nachprüfbar und reproduzierbar bleiben. Dabei gehen wir davon aus, dass auch diese Dokumentation in elektronischer Form vorliegt (entweder original oder in retrospektiv digitalisierter Form).

Nicht inbegriffen sind dagegen die *Forschungsergebnisse* selber sowie alle weiteren Unterlagen eines Forschungsprojekts, welche den *Projektlauf* sowie den *Auswertungs- und Forschungsprozess* nachvollziehbar machen. Wir gehen davon aus, dass diese *Forschungsunterlagen* im weiteren Sinne – sie liegen zum Beispiel in Form von E-Mails sowie mehr oder weniger geordneten Papierdokumenten und elektronischen Dokumenten vor – von den Forschungsinstitutionen in ihren normalen Ablagen (z.B. Papierdossiers) wenn nötig aufbewahrt bzw. von staatlichen Archiven im Rahmen der Anbietepflicht wie normale Verwaltungsakten bewertet und übernommen werden können.

Der Anspruch der KUB-Projektziele, «*ein nationales digitales Langzeitarchiv der Schweizer Hochschulen, Forschungszentren und Hochschulbibliotheken*» für Primärdaten und Sekundärdaten zu schaffen, beschränkt sich aus unserer Sicht somit auf die eigentlichen Rohdaten (inkl. Dokumentation) sowie die Verlagspublikationen und befasst sich insbesondere nicht mit der Aufbewahrung von Papierunterlagen und auch nicht mit der Aktenführung in Forschungsprojekten.

In Bezug auf die gesetzliche Archivierung müssen wir einschränken, dass ein Anspruch im Bereich der *staatlichen Archivgesetze* komplexe rechtliche und organisatorische Fragen aufwirft. Diese könnten umfassend nur durch die zahlreichen von einem solchen Vorhaben betroffenen Archivstellen des Bundes und der Hochschulkantone gemeinsam beantwortet werden.

Die Vorstellung, dass Tausende von Forschungsprojekten aus Hunderten von hochspezialisierten Forschungsdisziplinen ihre Primärdaten systematisch einem staatlichen Archiv anbieten, das diese dann inhaltlich bewerten soll, scheint uns unrealistisch. Darüber hinaus müssten sich die zahlreichen betroffenen Kantonsarchive und diversen Archive der Forschungsanstalten auf gemeinsame, einheitliche rechtliche, organisatorische und technische Bedingungen für eine zentrale, nationale Archivdienstleistung einigen können.



## 2.4 Modelle

Wir schlagen, entsprechend der in Abschnitt 2.3 diskutierten Schwerpunktsetzung, ein Modell für die zentrale Archivierung von digitalen Verlagspublikationen sowie zwei Modellvarianten für Primärdaten vor. Zusätzlich skizzieren wir bezüglich Primärdaten vier, prinzipiell andere Alternativmodelle.

Das **Modell für Verlagspublikationen** (Kapitel 6) entspricht weitgehend der Strategie «Allianz-Initiative ‚Digitale Information‘» der Forschungsgesellschaften in Deutschland (vgl. Kapitel 8.5). Es soll die dauerhafte Verfügbarkeit von digitalen Verlags-Zeitschriften (eJournals) und -Büchern (eBooks) über den Ablauf von Lizenzen und die Verfügbarkeit von Online-Diensten der Verlage hinaus sicherstellen.

Die **erste Modellvariante für Primärdaten** (Kapitel 5.1, Modell «Aufbewahrung mit Selbstregulierung») funktioniert nach dem «Self Deposit»-Ansatz und fokussiert pragmatisch auf die Bedürfnisse der Forschenden hinsichtlich der neuen in der Schweiz geltenden Richtlinien zur „Integrität der Forschung“. Die Dienstleistung sieht keine Regulierung vor und geht konsequent von der Selbstverantwortung und Eigeninitiative der der Forschenden und Forschungsinstitutionen aus, die Anforderungen an Dauer, Umfang und Formate der aufzubewahrenden Primärdaten bezüglich Integritätsrichtlinien und eines Wertes für eine längerfristige Sekundärnutzung selber zu bestimmen und umzusetzen.

Die Dienstleistung soll für Forschende und Forschungsprojekte grundsätzlich kostenlos sein, geht aber davon aus, dass die Forschenden auf eigene Kosten die Primärdaten und die dazugehörige Dokumentation soweit aufbereiten, dass sie den Anforderungen der Reglemente zur wissenschaftlichen Integrität genügen.

Die Dienstleistung bietet dazu auf Nachfrage Beratung und Unterstützung, garantiert ansonsten aber nur «Bitstream Preservation» sowie beweiskräftige Nachweisbarkeit der Integrität der Daten für den Zeitraum der Deponierung, erleichterte Zitierbarkeit der deponierten Daten und optional deren öffentliche Katalogisierung und Zugänglichkeit.

Sie bietet den *Forschungsinstitutionen* zudem die Möglichkeit, jederzeit, spätestens aber bei Ablauf der von den Forschenden definierten Aufbewahrungsfrist, die Daten *freiwillig* einer fachlich-inhaltlichen Bewertung zuzuführen, die Aufbewahrungsfrist entsprechend zu verlängern oder die Daten bei *dauerndem* Wert einer entsprechenden dauerhaften Archivierung durch die Dienstleistung zu übergeben, was auch unter dem Titel der staatlichen Archivgesetze erfolgen kann. Diese Modellvariante berücksichtigt damit die Empfehlungen der OECD (siehe Kapitel 3.1.3.4), bei der dauerhaften Aufbewahrung von Primärdaten für die Sekundärnutzung eine vorschnelle Selektion durch periodische Überprüfungen zu ersetzen.

Diese Modellvariante orientiert sich funktional an RFC 4810 «Long-Term Archive Service Requirements» der «Long-Term Archive and Notary Services Charter» (LTANS), jedoch *nicht* an ISO 14721:2003 «Open Archival Information System Reference Model» (OAIS).

Die **zweite Modellvariante für Primärdaten** (Kapitel 5.2, Modell « Archivierung mit Anbieterpflicht») verfolgt einen regulierenden Ansatz, der von einer generellen Anbieterpflicht für Forschungsunterlagen auf Grund der staatlichen Archivierungsgesetze ausgeht. Die Umsetzung der Anbieterpflicht und der archivischen Bewertung setzt dort an, wo Forschungsprojekte

ohnehin von Fachgremien beurteilt werden: bei der Bewilligung bzw. Finanzierung, der Begutachtung bzw. Evaluation der Ergebnisse sowie beim Reporting bzw. beim Abschlussbericht.

Die jeweiligen Gutachter werden angehalten, die (zeitlich beschränkte oder dauernde) Archivwürdigkeit aus der Sicht der Fachdisziplin zu beurteilen. Diese Vorschläge werden systematisch der jeweils zuständigen Archivstelle zugeleitet, welche darüber entscheidet. Die Gutachter und Forschenden beurteilen hauptsächlich den langfristigen Nutzen der Primärdaten für die Sekundärnutzung im wissenschaftlichen Fachbereich, während die zuständige Archivstelle zusätzlich auch den dauerhaften Wert aus Sicht des Bundesgesetzes über die Archivierung bzw. der jeweiligen Archivgesetzgebung im betreffenden Kanton berücksichtigt.

Sind die Primärdaten dauerhaft archivwürdig, werden sie in diesem Modell von der nationalen Archivdienstleistung in Zusammenarbeit mit den Forschenden, der Forschungsinstitution sowie der zuständigen staatlichen Archivstelle für die dauerhafte Archivierung aufbereitet und von der nationalen Archivdienstleistung im Auftrag der staatlichen Archivstelle übernommen und dauerhaft betreut. Richtlinien und Standards für die Dokumentation der Daten und für ihre Aufbereitung und Archivierung werden von fachbereichsspezifischen Archivkommissionen erarbeitet.

Dieses Modell orientiert sich bei dauerhaft zu archivierenden Primärdaten an ISO 14721:2003 «Open Archival Information System Reference Model» (OAIS). Das bedeutet *insbesondere*, dass die nationale Archivdienstleistung die fachlich-technische Gesamtverantwortung für die erfolgreiche Langzeit-Archivierung dieser Primärdaten übernimmt und deshalb über alle dafür nötigen Entscheidungskompetenzen bezüglich künftiger Konversionen bzw. Migrationen verfügen müssen.

Dieses zweite Modell deckt allerdings nur die Archivierung von jenen Primärdaten ab, welche durch die staatlichen Archivstellen auf der Grundlage der Archivgesetzte als archivwürdig bezeichnet werden. Eine *generelle* Aufbewahrung von Primärdaten, mit welcher die Forschenden ihre Pflichten bezüglich der Integrität der Forschung erfüllen können, wird dadurch nicht geboten, weder für *nicht* archivwürdige Primärdaten noch für alle archivwürdige Primärdaten, die durch die Archivierung so aufbereitet werden müssen, dass sie nicht mehr genügend den Originaldaten im Sinne der Integritätsrichtlinien genügen.

Die vier **Alternativmodelle** gehen alle davon aus, dass die nationale Dienstleistung selber keine Daten archiviert, sondern „nur“ strategisch-koordinierende und beratende Aufgaben sowie die Organisation eines nationalen Kompetenznetzwerks „Langzeitarchivierung“ nach den Vorbildern in Deutschland und Frankreich übernimmt.

## 2.5 Struktur des Berichts

Der Bericht ist wie folgt aufgebaut:

- **Kapitel 3 – Situation, Bedarf und Umfang:** Kurzcharakterisierung der einzelnen Unterlagentypen, Identifikation des aktuellen Bedarfs und Nutzens ihrer Aufbewahrung, Prognose der zu erwartenden Nutzungshäufigkeit während der Aufbewahrungszeit sowie Beschreibung der aktuellen Situation bezüglich Aufbewahrung und Archivierung.
- **Kapitel 4 – Anspruchsgruppen:** Charakterisierung der mögliche Interessen, Rollen und gesetzlichen Rahmenbedingungen der potenziellen Stakeholders an bzw. bei einer zentralen, nationalen Archivdienstleistung.
- **Kapitel 5 – Modelle für Primärdaten:** Empfohlenen Ziele und Spezifikation der von uns vorgeschlagenen Basismodelle für eine zentrale, nationale Archivdienstleistung für wissenschaftliche Primärdaten in der Schweiz.
- **Kapitel 6 – Modell für Sekundärdaten:** Empfohlenen Ziele und Spezifikation des von uns vorgeschlagenen Modells für eine zentrale, nationale Archivdienstleistung für wissenschaftliche digitale Publikationen in der Schweiz.
- **Kapitel 7 – Alternativmodelle:** Charakterisierung von vom Primärmodell prinzipiell abweichenden Alternativmodellen.
- **Kapitel 8 – Europäisches Umfeld:** Übersicht, Erläuterung und Typisierung der Projekte, Initiativen und Strategien zur digitalen Langzeitarchivierung im europäischen Umfeld, mit Abgrenzung zu Primärdaten und Sekundärdaten. Liste und Kurzcharakterisierung von Aktivitäten mit ähnlichen Zielen in der Schweiz und im Ausland.
- **Kapitel 9 – Bibliographie**

## 3 Situation, Bedarf und Umfang

Archivierungssituation, -bedarf und -umfang der zu archivierenden digitalen Unterlagen unterscheiden sich je nach Unterlagentyp stark. Wir differenzieren deshalb im Folgenden nach den Unterlagen-Typen, die im Rahmen der Auftragsvergabe erwähnt worden sind:

- Forschungsunterlagen (Primärdaten und Forschungsdokumentation)
- Elektronische wissenschaftliche Publikationen
- Elektronische Lehr- und Lernunterlagen („eLearning Objects“)
- Digitalisate aus retrospektiven Digitalisierungen

Die nachfolgenden Unterkapitel gliedern sich in die Abschnitte:

- Kurzcharakterisierung der Unterlagen
- Möglicher Nutzen der Archivierung
- Konkreter Bedarf und Nutzungshäufigkeit
- Archivierungssituation

Die *Geschäftsunterlagen im engeren Sinne*<sup>5</sup>, d.h. solche der Administration, Planung und Führung der Forschungsinstitution, sind *nicht* Gegenstand dieses Konzepts.

### 3.1 Primärdaten aus der Forschung

#### 3.1.1 Kurzcharakterisierung der Unterlagen

Primäre Forschungsdaten entstehen zum Beispiel in Rahmen von Experimenten, Messreihen, Computer-Simulationen oder statistischen Erhebungen. Charakteristisch für Forschungsdaten ist die grosse Vielfalt an Formaten der Daten und der sie beschreibenden Metadaten in den verschiedenen Bearbeitungsstadien, die bei der Forschung an Hochschulen und Forschungsanstalten anfallen. In einer Umfrage an der ETH Zürich [3] gaben nur 51% der Antwortenden an, dass die zu archivierenden Daten in einem offenen Standardformat vorlägen.<sup>6</sup>

Ebenso vielfältig sind die Begleitunterlagen, die nötig sind, um die Datengewinnung zu dokumentieren und den Forschungsprozess nachvollziehbar zu machen. Auch die Datenmengen sind sehr unterschiedlich. Eine Untersuchung zur Situation in Deutschland aus dem Jahre 2004

---

<sup>5</sup> Für die Bewertung und Archivierung der Geschäftsunterlagen der Hochschulen auf Bundesebene sind die selbstständig archivierenden Stellen gemäss Bundesgesetz über die Archivierung, bei den kantonalen Universitäten und Hochschulen die jeweiligen kantonalen Staatsarchive zuständig.

<sup>6</sup> Die Rücklaufquote bei der an alle Departemente versandten Umfrage betrug allerdings nur 17%.

[19] nennt für sprachwissenschaftliche, geisteswissenschaftliche und soziologische Institutionen insgesamt 100 GB pro Jahr und für die Gesamtmenge der für die Aufbewahrung relevanten Primärdaten in Deutschland 1000 – 2000 TB pro Jahr, mit stark steigender Tendenz.

Die längerfristige Aufbewahrung von wissenschaftlichen Primärdaten bzw. Forschungsdaten im Rahmen der öffentlich finanzierten Forschung hat in den vergangenen Jahren unabhängig von staatlichen Archivgesetzen stark an Bedeutung gewonnen, verursacht durch spektakuläre Fälle von Fälschungen und Manipulationen bei wissenschaftlichen Publikationen. Die Aufbewahrung von Forschungsdaten für die Überprüfbarkeit von publizierten Forschungsergebnissen ist deshalb heute Bestandteil der „Guten wissenschaftlichen Praxis“ bzw. „Integrität der Forschung“. Zunehmend sprechen weltweit Organisationen der Forschungsförderung Gelder nur noch mit der Auflage, dass Forschungsunterlagen nach Abschluss eines Forschungsprojektes während eines adäquaten Zeitraums sicher und lesbar aufbewahrt werden.

### 3.1.2 Möglicher Nutzen der Aufbewahrung und Archivierung

Eine längerfristige oder dauerhafte Aufbewahrung von Forschungsunterlagen kann im Wesentlichen folgenden Zwecken dienen:

- **Nachvollzug und Überprüfung der Forschungsergebnisse** sind zentrale Forderungen in den Grundsätzen zur „Integrität der wissenschaftlichen Forschung“ und „Guten wissenschaftlichen Praxis“. Eine Nutzung aus diesen Gründen dürfte für den Grossteil der Forschungen allerdings nach rund 10 Jahren nicht mehr notwendig sein.
- **Erhalt nicht-reproduzierbarer Daten:** Zahlreiche Typen von Messdaten sind an den absoluten Zeitpunkt ihrer Erhebung gebunden und deshalb nicht reproduzierbar. Einige wenige Beispiele: Wetterdaten, Messdaten zu Erdbebenereignissen, Bevölkerungsstatistiken, Meinungsumfragen, Erhebungen zur Artenvielfalt in Fauna und Flora sowie gewisse Messdaten aus Klimaforschung und Astronomie. Sie haben deshalb grundsätzlich einen langfristigeren und höheren historischen und wissenschaftlichen Wert als reproduzierbare Daten.
- **Sekundärnutzung:** Daten können auch nach Jahren unter neuen Gesichtspunkten, mit anderen Methoden oder in Bezug auf neue Theorien nochmals auf dieselbe Fragestellung hin ausgewertet werden. Oder Forschende nutzen vorhandene Daten aus anderen Forschungsprojekten für eigene Arbeiten mit eigenen Fragestellungen. Ein dritter Fall der Sekundärnutzung ist die Bildung von historischen Zeitreihen, zum Beispiel in der Klimaforschung oder Meteorologie, wo Daten aus Einzelprojekten in projektübergreifende Datensammlungen akkumuliert werden, aus denen Zeitreihen für neue Forschungsprojekte erzeugt werden können. Die Sekundärnutzung bedingt in besonderem Masse eine klare Dokumentation der Daten und ihrer Gewinnung.
- **Zitierbarkeit von Primärdaten:** Mit der Sekundärnutzung verbunden ist die Notwendigkeit, in Publikationen auf unzweideutige und nachvollziehbare Weise die verwendeten Primärdaten Dritter zu bezeichnen. Dies ist jedoch nicht möglich, da Primärdaten in der Regel nicht publiziert werden. Die Aufbewahrung oder Archivierung der Daten bei einer zentralen und unabhängigen Stelle, welche für die Datensätze zitier-

fähige Identifikatoren vergibt, kann hier eine wesentliche Vereinfachung und bessere Transparenz ermöglichen.

- **Nachvollziehbarkeit politischer Entscheide:** Forschungsergebnisse können politische Entscheidungen (z.B. in der Gesetzgebung) oder das Verwaltungshandeln des Staates beeinflussen. (Beispiele sind Daten aus der Risikoforschung zur Anwendung von gentechnologischen Verfahren oder zu Gesundheitsrisiken der Mobiltelefonie). Staatspolitische Grundprinzipien verlangen, dass das Handeln der Verwaltung, der politischen Behörden und schliesslich auch des Parlaments langfristig nachvollziehbar und überprüfbar bleiben. Dies kann bedeuten, dass in heiklen Fällen auch die Primärdaten, auf denen die Forschungsergebnisse beruhen, langfristig oder dauerhaft aufbewahrt werden.
- **Wissenschaftsgeschichte:** Sie untersucht die Entwicklung der Wissenschaften anhand ihrer Interessen, Fragestellungen, Methoden und Ergebnisse. Sie ist weniger an den Forschungsdaten selber, als an den Begleitunterlagen interessiert.

### 3.1.3 Bedarf und Nutzungshäufigkeit

Die Häufigkeit der Nutzung aufbewahrter Forschungsdaten hängt stark davon ab, wie gut die Forschungsunterlagen zugänglich sind. Bezüglich der Sekundärnutzung (siehe 3.1.2 oben) ergab eine Untersuchung in Deutschland aus dem Jahre 2004 [4], dass 60% der befragten Institutionen regelmässig Primärdaten mit anderen Institutionen austauschen, die Reaktivierung von Datensätzen mit der Zeit aber stark abnimmt und nach 10 Jahren kaum eine Rolle mehr spielt.

Ausnahmen mit intensiverer Nutzung wurden dort erkannt, wo es um nicht-reproduzierbare Primärdaten (3.1.2) handelt und die Datenhaltung und -verbreitung fachbereichsspezifisch und zentral organisiert werden, wie bspw. in den über 50 Weltdatenzentren des International Council for Science<sup>7</sup> (ICSU). Auch die Nutzungszahlen anderer Datenarchive in diesem Bereich deuten teilweise auf hohe Benutzungsfrequenzen hin. So meldet zum Beispiel der Economic and Social Data Service (ESDS) in Grossbritannien für das Geschäftsjahr 2003-2004 17'800 Zugriffe auf den Bestand von 3'500 Umfragedaten-Kollektionen.[5] Wesentlich für einen hohen Nutzungsgrad ist allerdings in der Regel die Möglichkeit eines Online-Zugriffes.

In einer Umfrage an der ETH Zürich [3] haben 80% der Antwortenden Interesse an einer zentralen digitalen Langzeitaufbewahrung für ihre Forschungsdaten bekundet, und über 55% haben erklärt, dass in ihrem Bereich Datensätze vorhanden sind, die dauerhaft bzw. länger als 10 Jahre archiviert werden sollten. Diese Antworten sagen zwar nichts über die Nutzungshäufigkeit aus, einen Nutzungsbedarf darf man aber trotzdem daraus ableiten.

Aus den folgend beschriebenen Regeln zur Sicherung der Integrität in der Forschung ergibt sich zudem per se ein Aufbewahrungsbedarf für *sämtliche* Primärdaten aus öffentlich finanzierter Forschung für Zeiträume in der Grössenordnung von rund zehn Jahren.

<sup>7</sup> <http://www.icsu.org>

### 3.1.3.1 Integritätsrichtlinien der ESF und OECD

Die European Science Foundation (**ESF**) empfiehlt seit 2000 in ihren Richtlinien «Good scientific practice in research and scholarship» [6] für Forschungsdaten eine Aufbewahrungszeit von mindestens zehn Jahren (Ziff. 37): *«Institutions must pay particular attention to documenting and archiving original research and scholarship data. Several codes of good practice recommend a minimum period of 10 years, longer in the case of especially significant or sensitive data.»*

Die **OECD** empfiehlt seit 2007 in ihren «Principles and Guidelines for Access to Research Data from Public Funding» [7], die für die OECD-Mitgliedstaaten allerdings nicht rechtlich bindend sind, dass Forschungsdaten aus öffentlich finanzierter Forschung *grundsätzlich offen und nachhaltig zugänglich* sein sollen, was eine Aufbewahrung nach Ende des Forschungsprojektes voraussetzt und die Überprüfbarkeit von publizierten Forschungsergebnissen einschliesst: *«Promote a culture of openness and sharing of research data among the public research communities within member countries and beyond. ... Open access to research data from public funding should be easy, timely, user-friendly and preferably Internet-based.»* Insbesondere:

- *«Research institutions and government organisations should take formal responsibility for ensuring that research data are effectively preserved, managed and made accessible in order that they can be put to efficient and appropriate use over the long term.» (lit. F)*
- *«Due consideration should be given to the sustainability of access to publicly funded research data as a key element of the research infrastructure.» (lit. M)*

### 3.1.3.2 Aufbewahrungsregeln in der Schweiz und Deutschland

In Deutschland ist eine 10-jährige Aufbewahrungspflicht für Forschungsdaten bereits seit 1998 Bestandteil der Verträge der **Deutschen Forschungsgemeinschaft** (DFG) mit den Empfängern von DFG-Forschungsbeiträgen [8]. Diese Vorgabe haben seither fast alle deutschen Hochschulen und öffentlichen Forschungseinrichtungen in ihre internen Regelwerke übernommen.

In der Schweiz sind die Institutionen der Forschungsförderung (die Akademien der Wissenschaften sowie der Nationalfonds) seit dem 25.2.08 durch den neuen Artikel 11a des Forschungsgesetzes (Bundesgesetzes über die Forschung [9] (SR 420.1) verpflichtet, darauf zu achten, *«dass die von ihnen geförderte Forschung nach den Regeln der guten wissenschaftlichen Praxis durchgeführt wird»*.

Die **Akademien der Wissenschaften Schweiz** bzw. die vier schweizerischen Akademien der Wissenschaften<sup>8</sup> haben am 28. Juni 2007 gemeinsame «Grundsätze und Verfahrensregeln zur wissenschaftlichen Integrität» [10] verabschiedet, welche auch Vorgaben für die Aufbewahrung von Forschungsdaten und -unterlagen enthält:

- *«Damit Forschung überprüft und Versuche reproduziert und Daten auch nach anderen Gesichtspunkten analysiert werden können, sind alle Daten (inkl. Rohdaten) vollständig, klar und genau zu dokumentieren. Daten und Materialien müssen in einer Weise aufbewahrt werden, welche Beschädigung, Verlust oder Manipulation ausschliesst. Dies gilt sowohl für handschriftliche wie auch für elektronische Daten. Besondere Vorfälle, z.B. Datenverlust und Abweichungen vom*

<sup>8</sup> Akademie der Naturwissenschaften Schweiz SCNAT, Schweizerische Akademie der Medizinischen Wissenschaften SAMW, Schweizerische Akademie der Geistes- und Sozialwissenschaften SAGW, Schweizerische Akademie der Technischen Wissenschaften SATW.

*ursprünglichen Forschungsplan, müssen dokumentiert werden. Die Projektleitung ist dafür verantwortlich, dass Daten und Materialien nach Abschluss des Projektes während einer für das Fachgebiet adäquaten Dauer aufbewahrt bleiben. Sie hat für ihre Haltbarkeit und Sicherung zu sorgen.» (Abs. 3.1)*

- *«Nach Abschluss des Projektes und nach Vorliegen der Ergebnisse sollen die für eine Überprüfung notwendigen Daten offen gelegt und Materialien, die für eine Wiederholung des Projektes notwendig sind, nach Möglichkeit zur Verfügung gestellt werden.» (Abs. 3.2)*

Im Gegensatz zur Regelung in Deutschland ist hier keine explizite Dauer, sondern „eine einer für das Fachgebiet adäquate Dauer“ vorgesehen.

Der **Schweizerische Nationalfonds** (SNF) verpflichtet seine Betragsempfänger im neuen (am 1.3.2008 in Kraft getretenen) Beitragsreglement [11] *«die Forschungsarbeiten ... nach den Regeln der guten wissenschaftlichen Praxis ... durchzuführen»* (Art. 23). Welches diese Regeln sind, wird nicht ausdrücklich erwähnt (wie auch nicht eine Aufbewahrungspflicht für Forschungsdaten), doch dürfte ein impliziter Bezug auf die Regeln der Akademien Schweiz und der OECD und ESF (siehe Abschnitt 3.1.3.1) naheliegen. Eine weitgehende Aufbewahrungspflicht ergibt sich im SNF-Reglement jedoch bereits implizit aus dem Artikel 44 in [11]:

- *«Die mit Beiträgen des SNF erhobenen Daten sind auch anderen Forschenden für die Sekundärforschung zur Verfügung zu stellen und gemäss den Vorschriften des SNF in anerkannte wissenschaftliche Datensammlungen einzubringen.»<sup>9</sup>*

Die Regeln der Akademien Schweiz und des SNF gelten nur für die von diesen Institutionen mit Beiträgen geförderten Forschungsvorhaben. Allerdings haben 2007 bereits auch mehrere Organe der Hochschulforschung eine Aufbewahrungspflicht für Forschungsdaten in ihre eigenen Regelwerke aufgenommen. Zum Beispiel gilt im «Reglement über die wissenschaftliche Integrität» der Universität Bern vom 1.5.2007 [13] als Verstoß *«insbesondere ... die Beseitigung aufbewahrter Primärdaten vor Ablauf der vorgeschriebenen Aufbewahrungsfrist gemäss den rechtlichen Grundlagen ...»*, ohne dass jedoch die rechtlichen Grundlagen genannt werden.

Die «Richtlinien für Integrität in der Forschung und gute wissenschaftliche Praxis an der ETH Zürich» [14] verlangen: *«Primärdaten müssen so abgelegt und geschützt werden, dass auf sie bei späterer Verwendung oder Überprüfung sicher zugegriffen werden kann.»*

- *«Sämtliche Verfahrensschritte im Umgang mit den Primärdaten (statistische Analysen, Umformungen etc.) sind in einer der jeweiligen Disziplin angepassten Form (z.B. Laborjournal, andere Datenträger) derart zu dokumentieren, dass die aus den Primärdaten gewonnenen Ergebnisse vollständig reproduziert werden können.» (Art. 11 Abs. 2)*
- *«Die Projektleitung [...] sorgt insbesondere dafür, dass Daten und Materialien nach Abschluss des Projektes während der für das Fachgebiet massgebenden Frist aufbewahrt bleiben und gegebenenfalls innerhalb der gesetzlich vorgegebenen Frist ordnungsgemäss vernichtet werden.» (Art. 11 Abs.4)*

Als Primärdaten gelten dabei die *«ursprünglichen experimentellen oder durch eine andere Methode erhobenen, nicht weiter bearbeiteten Originaldaten»*, und Materialien sind *«jegliche*

<sup>9</sup> Weitere Einzelheiten dazu im Reglement über die Information, die Valorisierung und die Rechte an Forschungsergebnissen vom 17.6.08 [12]



*Art von vor oder während der Forschungsarbeit gewonnenen Proben und erzeugten Produkten in materieller (z.B. Prototypen) oder nichtmaterieller (z.B. Programmiercodes) Form».*

Die ETH Zürich weist die Verantwortung für die Aufbewahrung denjenigen Personen zu, die auch für die Durchführung des wissenschaftlichen Projekts verantwortlich sind, «*namentlich Professoren, Senior Scientists sowie Oberassistenten oder Postdoktoranden*» (Art. 3).

Obwohl die hier zitierten Reglemente von einer „für das Fachgebiet adäquaten“ bzw. „massgebenden“ Aufbewahrungsfrist sprechen, dürfte sich bei der praktischen Auslegung auch in der Schweiz die im europäischen Ausland gebräuchliche „10-Jahre“-Regel durchsetzen.

#### 3.1.3.3 Zitierbarkeit von Forschungsdaten

Mit der Aufbewahrung von Forschungsdaten zum Zwecke der Überprüfbarkeit ist auch die Frage nach deren Zitierbarkeit verknüpft. Die Autorenschaft von Publikationen, d.h. die Urheberschaft von Forschungsdaten und insbesondere auch Sekundärnutzer/innen dieser Daten, benötigen eine Methode, sich in ihren Publikationen unzweideutig auf Datensätze beziehen zu können, so wie dies bei publizierter Sekundärliteratur möglich ist. Damit soll einerseits der Bezug auf den Umfang und Zustand von Datensätzen zu einem bestimmten Zeitpunkt hergestellt werden, andererseits aber auch auf deren Urheberschaft.

Nun werden Forschungsdaten selber in der Regel nicht publiziert. Eine verlässliche Zitierbarkeit wird aber auch dadurch erreicht, dass die Primär- bzw. Rohdaten nach ihrer Auswertung bei einer unabhängigen, von der Forschungsgemeinde anerkannten Drittstelle deponiert werden, welche die Datensammlung mit einem einheitlichen Identifikator versieht, katalogisiert sowie sicher und unveränderbar aufbewahrt. Hier trifft sich der Wunsch nach einer Zitierbarkeit von nicht publizierten Roh- oder Primärdaten mit den Reglementen zur Forschungsintegrität (siehe 3.1.3.2), welche eine Aufbewahrung verlangen, bei der die Daten vor Beschädigung, Verlust oder Manipulation geschützt sind.

Erste Pilotprojekte für die Zitierbarkeit von Primärdaten werden in Deutschland von der Deutschen Forschungsgemeinschaft (DFG) gefördert, zum Beispiel «CODATA – Datenpublikation und die Registrierungsagentur für wissenschaftliche Daten an der Technischen Informationsbibliothek Hannover»<sup>10</sup> im Bereich der Geowissenschaften.

Die im September 2007 von OECD und ESF sowie dem U.S. Office of Research Integrity (ORI) durchgeführte «First World Conference on Research Integrity» empfiehlt für Forschungsdaten sogar die Umkehrung des Zitierprinzips ([15], S.28):

- *«An important development will be the establishment of public digital repositories for primary research data with links to the published articles. In the USA an Inter Agency Working Group on Digital Data has been set up to propose such a repository system. In Europe an Alliance for Permanent Access to the Digital Records of Science has been created by major stakeholders in science and science information to help establish a European Digital Information Infrastructure.»*

Analog zu den öffentlichen Forschungsfinanzierern und -organen haben in der letzten Zeit auch einige wissenschaftliche Zeitschriften eine „Data Availability Policy“ erlassen und nehmen nur

<sup>10</sup> <http://www.tib-hannover.de/de/die-tib/projekte/codata/>

noch Manuskripte entgegen, wenn die zugrundeliegenden Daten gut dokumentiert und in nicht proprietärer Form mitgeliefert und für die Überprüfung der Forschungsergebnisse durch andere Forscher freigegeben werden.<sup>11</sup>

#### 3.1.3.4 Aufbewahrung für die Sekundärnutzung

Werden Primärdaten auch zu anderen Zwecken genutzt als denjenigen, zu denen sie ursprünglich erhoben wurden, so spricht man von Sekundärnutzung.

Alle in den Abschnitten 3.1.3.1 und 3.1.3.2 zitierten Reglemente und Empfehlungen, insbesondere jene der OECD, der Akademien Schweiz und des Nationalfonds, erwähnen ausdrücklich die Sekundärnutzung als Grund für eine längerfristige oder gar dauerhafte Aufbewahrung von Primärdaten. Als einzige Institution verlangt der Nationalfonds nicht nur, dass die Daten *«auch anderen Forschenden für die Sekundärforschung zur Verfügung»* gestellt werden, sondern auch *«gemäss den Vorschriften des SNF in anerkannte wissenschaftliche Datensammlungen einzubringen»* sind ([11], Art. 44), was eine dauerhafte Aufbewahrung impliziert.

Zwischen der Aufbewahrung von Primärdaten wegen der „wissenschaftlichen Integrität“ bzw. wegen der „Sekundärnutzung“ gibt es drei wesentliche Unterschiede: Letztere erfolgt in der Regel zeitlich unbegrenzt (da der Zeitpunkt der Sekundärnutzung nicht bestimmbar ist), stellt höhere Ansprüche an die langfristige Verstehbarkeit bzw. gute Dokumentation der Daten und ist schliesslich auch abhängig vom angenommenen wissenschaftlichen Wert der Daten. Vor allem dann, wenn eine permanente Aufbewahrung mit erheblichen Kosten verbunden ist, wird man davon absehen, wenn die Primärdaten als „nicht relevant“ oder als „von geringem Nutzen“ für eine Sekundärnutzung erscheinen. (Die Beurteilung dieser Qualität ist allerdings im Einzelfall schwierig und sollte auch nicht ausschliesslich durch Datenerzeuger selber erfolgen.) Die Richtlinien der Akademien Schweiz und der OECD legen zudem bei der Aufbewahrung auch grosses Gewicht auf die vollständige Dokumentation der Daten.

Die Richtlinien der OECD [7] empfehlen, bei der dauerhaften Aufbewahrung von Daten für die Sekundärnutzung die Problematik einer vorschnellen Selektion oder Bewertung durch periodische Überprüfungen und Bewertungen („Assessments“) zu ersetzen:

- *«While publicly funded research data are subject to the default rule of openness ..., this does not mean that all such data should be preserved permanently. The data archiving community should carry out cost-benefit assessments periodically and constantly develop and refine retention protocols to ensure that those data sets with the greatest potential utility are preserved and made accessible. Use of accepted retention protocols and thorough documentation of data should help to reduce unnecessary duplication of effort as well as to establish the necessary selectivity in preservation.»* ([7], Kap. K)
- *«[Long-term retention] ... can be a difficult task, given that most research projects, and the public funding provided, have a limited duration, whereas ensuring access to the data produced is a long-term undertaking. Research funding agencies and research institutions, therefore, should*

---

<sup>11</sup> Zum Beispiel: *«It is the policy of the American Economic Review to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Authors of accepted papers that contain empirical work, simulations, or experimental work must provide to the Review, prior to publication, the data, programs, and other details of the computations sufficient to permit replication.»* [5]

*consider the long-term preservation of data at the outset of each new project, and in particular, determine the most appropriate archival facilities for the data.» ([7], Kap. M)*

Bemerkenswert an diesen OECD-Empfehlungen ist, dass sie (anders als z.B. die Richtlinien der ETH Zürich) die Überprüfung der Aufbewahrungswürdigkeit und die Verantwortung für die dauerhafte Aufbewahrung nicht einfach an die Datenerzeuger bzw. Projektleitungen delegieren, sondern als Aufgaben einer eigentlichen „archiving community“ bzw. der Forschungsfinanzierer und Forschungsträger definieren, also als institutionelle Aufgaben.

#### 3.1.4 Archivierungssituation

Die Archivierungssituation bei Forschungsdaten ist sehr heterogen. Die staatlichen Archive haben sich, wenn überhaupt, bisher nur am Rande mit Forschungsunterlagen beschäftigt. Lediglich dort, wo der Staat selber grosse Erhebungs- und Messaufgaben übernommen hat, begannen einzelne Nationalarchive in den 1970er und 1980er-Jahren, die entsprechenden (bereinigten) Primärdaten zu archivieren.

Seit den 1960er Jahren entstand eine Reihe von Daten-Archiven, die sich zunächst auf sozialwissenschaftliche Daten konzentrierten. Später entstanden auch Daten-Archive im naturwissenschaftlichen Bereich. Diese fachbereichsspezifischen Datenarchive wurden alle früher oder später an nationale Wissenschaftsorganisationen und Akademien angegliedert und in nationalem Rahmen finanziert, obwohl sich diese Datenarchive teilweise zu internationalen Clustern zusammenschlossen [16].

Dieses heterogene Bild gilt auch für die Schweiz. Einige Fachbereiche, die international stark vernetzt und auf den Austausch und die mittel- bis langfristige Sekundärnutzung von Daten angewiesen sind, haben sich selber organisiert und erhalten dafür zum Teil auch öffentliche Gelder. So liefern einige naturwissenschaftlichen Disziplinen regelmässig Daten an internationale, fachspezifische Datenzentren<sup>12</sup>. Und die Sozialwissenschaften in der Schweiz betreiben mit SIDOS<sup>13</sup> seit 1992 ein eigenes Datenarchiv.

Die Klima- und Geowissenschaften sowie die Sozialwissenschaften sind bezüglich der Aufbewahrung und Zugänglichmachung von Primärdaten wohl die am besten organisierten Fachdisziplinen<sup>14</sup>. Dies darf allerdings nicht darüber hinwegtäuschen, dass es hierbei erstens um

<sup>12</sup> Auch die Schweiz finanziert einige solche Datenzentren. Pressemitteilung des Eidg. Departements des Innern vom 06.06.2008: «Ein 2007 von MeteoSchweiz zusammen mit Partnerinstitutionen erstelltes Inventar hat jedoch gezeigt, dass die Finanzierung einiger bedeutender Messreihen nicht gesichert ist. Betroffen sind die Beobachtungen von Kohlendioxid, Eisbedeckung von Seen, Schnee-Wasseräquivalent, Gletscher, Permafrost und Phänologie sowie die beiden Weltzentren, 'World Glacier Monitoring Service' an der Universität Zürich und 'Euro-Climhist' an der Universität Bern. Mit dem heutigen Bundesratsbeschluss ist die Weiterführung dieser Messreihen und Datenzentren nun nachhaltig gesichert. Im Jahr 2010 mit 0.9 Mio Franken und ab 2011 mit jährlich 1.6 Mio Franken.»

<sup>13</sup> Der Schweizerische Informations- und Datenarchivdienst für die Sozialwissenschaften (SIDOS) ist eine Stiftung der Schweizerischen Akademie für Geistes- und Sozialwissenschaften (SAGW). Gegründet 1992 bestehen die Hauptaufgaben von SIDOS in der Archivierung von Daten sowie der Erstellung eines Forschungsinventars. SIDOS ist seit dem 1. Januar 2008 Teil von FORS (Schweizer Stiftung für die Forschung in den Sozialwissenschaften) und an der Université de Lausanne lokalisiert.

<sup>14</sup> Fast in jedem europäischen Land existiert ein Archivdienst für Primärdaten aus den Sozialwissenschaften. Bisher haben sich 20 solche nationale Dienste zum Dachverband Council of European Social Science Data Archives (CESSDA, <http://www.cessda.org>) zusammengeschlossen und katalogisieren ihre Datensätze gemeinsam. In den Klima- und Geowissenschaften spielen die World Data Centers des des International Council for Science (ICSU, <http://www.ngdc.noaa.gov/wdc/>) eine zentrale Rolle.

Einzelinitiativen, zweitens fast ausschliesslich um nicht reproduzierbare Primärdaten für die Sekundärnutzung und drittens auch nur um ausgewählte Datensätze geht. Von einer flächendeckenden mittel- oder langfristigen Aufbewahrung von Forschungsdaten kann weder international noch auf die Schweiz bezogen die Rede sein.

Das grundsätzliche Problem eines *fachübergreifenden*, nationalen Dienstleistungsangebots zur Aufbewahrung von Primärdaten stellt die enorme Vielfalt an Datenformaten und -strukturen dar, die bereits innerhalb desselben Fachgebiets gross ist. Diese Vielfalt und die unterschiedlichen Fachterminologien und -konventionen erlauben eine einheitliche Erschliessung (Ordnung und Verzeichnung) nur auf einer sehr rudimentären Ebene.

Bei den Datenformaten ist die Situation noch schwieriger: Eine Erhebung in Deutschland ergab [4], dass umgerechnet auf die produzierten Datenvolumina 97,8% der Daten in binären, fachspezifischen Formaten gespeichert werden, nur 0,3% in XML und 1,9% in reinen Textformaten. Die fachspezifischen binären Formate sind zwar in der Regel offen dokumentiert. Trotzdem ist bei den strukturierten Forschungsdaten in binären Formaten eine Konversion dieser Formate in einheitliche Standardformate<sup>15</sup>, wie es zum Beispiel bei der Langzeitarchivierung von dokumentenzentrierten Unterlagen üblich ist, in der Regel illusorisch und sogar unerwünscht. Erstens existieren überhaupt nur sehr wenige fachübergreifende Standardformate. Zweitens erfolgt auch eine Sekundär- oder Nachnutzung nach mehreren Jahren wiederum in diesen fachspezifischen Formaten.

Und drittens beziehen die in den Abschnitten 3.1.3.1 und 3.1.3.2 zitierten Richtlinien zur Integrität in der Forschung auf die originalen Forschungsdaten, auf denen die Forschungsergebnisse basieren. (Die ETH-Richtlinie [14] bezieht sich auf die «*ursprünglichen experimentellen oder durch eine andere Methode erhobenen, nicht weiter bearbeiteten Originaldaten*», die Richtlinie der Akademien Schweiz [10] bezieht sich auf «Rohdaten».) Eine Konversion der spezifischen binären Datenformate in Standardformate müsste somit ohne jeglichen Verlust an Information (Daten-Werte, -Typen und -Struktur) möglich sein, was jedoch kaum vorstellbar ist.

Zusammenfassend lässt sich sagen, dass *fachübergreifende* Archivlösungen, die Daten aus vielen unterschiedlichen und heterogenen Quellen aufnehmen sollen, nur auf sehr allgemeiner Ebene sinnvoll und möglich sind. Dabei handelt es sich in der Regel als reine Depots für „Datenpakete“, die Dateien in binären, fachspezifischen Datenformaten enthalten und durch das Archiv auch nur als „Paket“ beschrieben bzw. katalogisiert werden. Eine tiefer gehende Erschliessung kann nur durch die von Datenproduzenten erstellte und im Paket inbegriffene Dokumentation erfolgen. Ebenso ist keine eigentliche Archivierungsstrategie möglich, welche sich auf die langfristige Migration von im Paket enthaltenen und möglicherweise obsolet werdenden, fachspezifischen binären Datenformate bezieht. Diese Aufgabe kann nicht das Archiv (also der Depositär) übernehmen, sondern nur der einzelne Datenproduzent.

Trotzdem wird die mittel- und langfristige Aufbewahrung von Forschungsdaten in Zukunft unter den Titeln „Integrität der Forschung“ und „Gute wissenschaftliche Praxis“ national wie international stark an Bedeutung gewinnen (siehe dazu auch Kapitel 8: Europäisches Umfeld), um Primärdaten aus öffentlich finanzierter Forschung auch nach Projektschluss in gut dokumen-

---

<sup>15</sup> Zum Beispiel in das „Portable Document Format for Archiving“ (PDF/A), ISO 19005-1:2005.

tiertes Form längerfristig oder gar dauerhaft aufzubewahren und Dritten für die Sekundärnutzung zugänglich zu machen. Es wächst das Bewusstsein dafür, dass sich so die teilweise erheblichen Investitionen zur Gewinnung der Forschungsdaten besser zu nutzen lassen.<sup>16</sup>

Dabei dürfte sich in der schweizerischen Praxis ebenfalls die im europäischen Ausland gebräuchliche „10-Jahre“-Regel durchsetzen, zumal dies auch der Aufbewahrungsfrist in der schweizerischen Geschäftsbücherverordnung (SR 221.431) des Obligationenrechts entspricht.

Es wird allerdings rasch deutlich werden, dass die in den schweizerischen Richtlinien vorgesehene Delegation der Verantwortung für die Aufbewahrung alleine an die Forschenden bzw. Projektverantwortlichen in der Praxis nicht umsetzbar ist: Einerseits macht die hohe Anstellungsmobilität im Forschungsbereich eine langfristige Verantwortung einer Person für zahlreiche, an unterschiedlichen Forschungsstandorten lokalisierte Datenbestände unrealistisch. Andererseits bedingt eine langfristige und sichere Aufbewahrung technische und finanzielle Ressourcen, die nicht unter der Kontrolle der Forschenden sind bzw. längerfristig bleiben können.

Einen realistischeren Weg weisen die Empfehlungen der OECD (siehe Abschnitt 3.1.3.1), nach denen die längerfristige oder gar dauerhafte Aufbewahrung und Zugänglichmachung von Forschungsdaten nicht primär in der Verantwortung der einzelnen Forschenden liegen kann, sondern eine institutionelle Aufgabe darstellt, welche durch die Organe der öffentlichen Forschung und Forschungsfinanzierung massgeblich mitgetragen werden muss.

Für Primärdaten aus öffentlich finanzierter Forschung fordert die OECD zudem den Grundsatz des „Open Access“ mit entsprechenden institutionellen Repositorien und Selektionsverfahren zur dauerhaften Aufbewahrung relevanter Bestände für die Sekundärnutzung.

Bei öffentlich finanzierter Forschung sprechen weder die staatlichen Archivgesetze noch die Regeln der Forschungsintegrität dagegen, dass Forschende ihre Daten nach Projektschluss einem Dritten zur Aufbewahrung übergeben, der in ihrem Auftrag die langfristige Integrität (Unverändertheit), Authentizität und sichere Speicherung ausführt. Dies gilt insbesondere dann, wenn es sich bei diesem Dritten um eine staatliche und von den Organen der Forschung und Forschungsfinanzierung selber kontrollierten und finanzierten Institution handelt.

## 3.2 Digitale Publikationen

### 3.2.1 Kurzcharakterisierung der Unterlagen

Bei den digitalen Publikationen handelt es sich vorwiegend um elektronische wissenschaftliche Zeitschriften (eJournals), Bücher (eBooks), Datenbanken sowie Hochschulpublikationen (z.B. Dissertationen, Masterarbeiten) sowie graue Literatur in elektronischer Form (z.B. Preprints).

Grundsätzlich sind drei verschiedene Beschaffungsarten von digitalen Publikationen zu unterscheiden:

- a) Physische Beschaffung der digitalen Objekte und lokale Nutzung bzw. Aufbewahrung über den Dokumentenserver *der Bibliothek(en)*.

---

<sup>16</sup> Siehe dazu auch zum Beispiel [16]

- b) Beschaffung nur der Zugriffslizenz und Nutzung (über Internet) der auf dem Server *des Verlags* gespeicherten Dokumente oder Datenbanken.
- c) Wie b), aber *zusätzlich* die physische Beschaffung der digitalen Objekte nur zum Zweck der lokalen Sicherungsaufbewahrung durch die Bibliothek (d.h. ohne Zugriff durch normale Benutzer).

Der Zugriff auf digitale Publikationen über den Server des Verlages bzw. der Bibliothek ist durch die lizenzrechtlichen Bestimmungen zum Teil auf einzelne Hochschulen beschränkt. Auch der Verwendungszweck der physisch bezogenen Sicherungsexemplare (Variante c) ist in der Regel lizenzrechtlich stark eingeschränkt.

Bei der Bedarfsanalyse für eine Archivierung von digitalen Dokumenten (durch die Bibliothek), aber auch bei der Risikoanalyse einer dafür tauglichen Archivlösung, ist in jedem Fall zu berücksichtigen, ob die Bibliothek zusätzlich auch Papierexemplare der Publikationen besitzt oder nur die digitalen Versionen (was bei Zeitschriften immer häufiger vorkommt).

Zusätzlich kommt hinzu, dass in geringerem Umfang auch ältere Jahrgänge von Zeitschriften, die bisher nur in Papiausgaben vorhanden waren, retrospektiv digitalisiert werden.

#### 3.2.2 Möglicher Nutzen der Archivierung

Primäres Ziel einer Archivierung im Bereich der Publikationen ist die Sicherstellung der langfristigen Verfügbarkeit der erworbenen digitalen Exemplare, vor allem wenn die Bibliothek(en) keine Papierexemplare besitzen. Diese Verfügbarkeit ist auf verschiedene Weise bedroht:

- Mit einer Kündigung der Lizenz geht in der Regel auch der Zugriff auf die bezahlten Jahrgänge einer Zeitschrift oder eines eBooks auf dem Verlagsserver verloren.
- Ältere Zeitschriftennummern sind auf den Verlagsservern ab und zu nicht mehr auffind- bzw. zugreifbar.
- Die Auflösung oder der Verkauf einer Zeitschrift oder eines ganzen Verlags bedroht die Verfügbarkeit der erworbenen Zeitschriften auf den Verlagsservern.
- Der Umstieg auf neue Technologien auf Seiten des Verlags kann dazu führen, dass ältere, weniger gefragte Nummern aus Kostengründen nicht mehr in die neuen Systeme migriert werden und das entsprechende Angebot eingestellt wird.

In diesen Fällen ist nicht mehr sichergestellt, dass die betroffene Bibliothek ihrem Benutzerkreis die wissenschaftlichen Zeitschriften weiterhin zur Verfügung stellen und somit ihren Versorgungsauftrag für die Wissenschaft noch erfüllen kann.

Diese grundsätzliche Problematik wird im Rahmen des Modells für Sekundärdaten in Kapitel 6 nochmals etwas vertieft diskutiert.

Als Investitionsschutz für bezahlte Lizenzen und zur Garantie der langfristigen Erfüllung des Versorgungsauftrages bleibt somit nur die in Abschnitt 3.2.1 aufgeführte Variante c: Die Bibliothek muss die digitalen Exemplare von Verlag physisch erwerben und sie selber archivieren, so dass diese Archivexemplare zu Nutzexemplaren werden können, wenn der Zugriff über die Verlagsserver nicht mehr möglich ist.

Bei der Archivierung zeichnet sich jedoch ein grosses Synergiepotenzial ab:

- Es macht keinen Sinn, wenn jede Bibliothek eine eigene Archivlösung für ihre digitalen Zeitschriften und Publikationen sowie Digitalisate von Papierexemplaren aufbaut. Vielmehr lässt sich diese Archivierung in einer einzigen, nationalen Lösung relativ einfach zentralisieren. Ein einziges, fachgerecht archiviertes und gut gesichertes Exemplar reicht für eine beliebig grosse Anzahl von Institutionen und Zeitschriften. Lizenzrechtliche Probleme ergeben sich daraus nicht, solange es sich ausschliesslich um ein Archiv ohne direkten Zugriff der Bibliotheksbenutzer handelt.
- Auf eine *zusätzliche* Anschaffung von Papierversionen der Zeitschriften und Bücher kann verzichtet werden, womit auch der Aufwand für deren bibliothekarische Bearbeitung und die Aufbewahrung eingespart werden kann.

Die Einsparungen durch Realisierung der Synergiepotenziale können durchaus höher sein als der Aufwand für eine fachgerechte und sichere digitale Archivierung.

#### 3.2.3 Bedarf und Nutzungshäufigkeit

Die Nutzungshäufigkeit dürfte sehr gering sein, da die Langzeitarchivierung nur der langfristigen Sicherung dient, während die laufende Nutzung über Dokumentenserver entweder der Verlage oder der Bibliotheken erfolgt. Der Bedarf nach Lösungen für die Langzeitarchivierung ist dagegen auf Grund der in Abschnitt 3.2.2 skizzierten Risikolage unbestritten. Sie stellt eine Voraussetzung für die Umsetzung der Strategien und Lösungen für eine langfristige Benutzbarkeit und Zugänglichmachung dar, wie sie 2005 – 2008 im Rahmen des Projektes «E-Archiving»<sup>17</sup> des Konsortiums der Schweizer Hochschulbibliotheken erarbeitet wurden. Die Frage ist einzig, ob Lösungen zur Langzeitarchivierung zentralisiert oder dezentral operieren.

#### 3.2.4 Archivierungssituation

Bei den publizierten eJournals, eJournals und Datenbanken ist zu unterscheiden zwischen den in der Schweiz verlegten und den durch Hochschulen im Ausland erworbenen bzw. lizenzierten Publikationen. Während für erstere die Sammelaufträge der Nationalbibliothek (Helvetica) und der Kantonsbibliotheken (Kantonalia) Anwendung finden, obliegt die Aufbewahrung der Publikationen ausländischer Verlage den Hochschul- und Universitätsbibliotheken.

In diesem Konzept geht es auftragsgemäss um den zweiten Bereich, also Publikationen, welche die Hochschulbibliotheken *im Ausland* erwerben und ihren Nutzerkreisen zur Verfügung stellen.

Die Aufbewahrung und die Nutzbarmachung von digitalen Publikationen, die zu den Kategorien der Helvetica, Kantonalia und der Hochschulschriften gehören, sind zwar noch nicht vollständig gelöst, aber die Zuständigkeiten und der organisatorische Rahmen sind geklärt. Im Bereich der Helvetica erarbeitet die Nationalbibliothek in Zusammenarbeit mit den Kantons- und Hochschulbibliotheken entsprechende Lösungen im Projekt «e-Helvetica»<sup>18</sup>.

---

<sup>17</sup> <http://lib.consortium.ch>. Der Schlussbericht ist noch nicht öffentlich zugänglich, weshalb wir auf die Resultate einer Vorstudie [17] aus dem Jahr 2005 verweisen.

<sup>18</sup> [www.e-helvetica.admin.ch](http://www.e-helvetica.admin.ch)

Eine fachgerechte Archivierung von im Ausland erworbenen digitalen Publikationen ist in der Schweiz heute noch praktisch inexistent. Meist lagern die physisch erworbenen digitalen Exemplare mehr oder weniger kontrolliert auf losen Datenträgern (CD oder Magnetband) in den einzelnen Institutionen. In der Regel wird das in den Lizenzen zugestandene Recht, zusätzlich zum Zugriff auf die Verlagsserver auch physische Kopien zu erwerben, mangels Archivlösung gar nicht ausgeübt.

## 3.3 Digitale Lehr- und Lernunterlagen

### 3.3.1 Kurzcharakterisierung der Unterlagen

Digitale Lehr- und Lernunterlagen entstehen heute in vielfältigen, meist multimedialen Formen. Es kann sich dabei um selbstständige virtuelle Lerneinheiten, nur um unterstützende Materialien wie bspw. Präsentationen oder um audiovisuelle Aufzeichnungen von ganzen Unterrichtseinheiten handeln.

### 3.3.2 Möglicher Nutzen der Archivierung

Da die Erstellung digitaler Unterrichtsmaterialien zum Teil mit einem hohen Aufwand verbunden ist, macht es Sinn, diese Investitionen entsprechend zu schützen und über den Primärzweck hinaus nutzbar zu machen. Zwei Nutzungen stehen im Vordergrund:

- Wiederverwendung in unveränderter Form oder als Bausteine für neue Lehrmaterialien
- Wissenschafts- und Bildungsgeschichte

### 3.3.3 Bedarf und Nutzungshäufigkeit

Die Nutzung für die Wiederverwendung dürfte bei leichter Zugänglichkeit beträchtlich sein. Da sich aber die Wissenschaften und die Unterrichtsformen relativ rasch weiterentwickeln, dürfte die Sekundärnutzung nach wenigen Jahren stark abflachen und die Unterlagen dürften nur noch aus historischen Gründen von Interesse sein.

### 3.3.4 Archivierungssituation

Eine eigentliche organisierte Archivierung von digitalen Lehr- und Lernunterlagen existiert zumindest in der Schweiz nicht. Nach eigenem Bekunden sind in der Schweiz sowohl die ETH Zürich als auch die Hochschulstiftung SWITCH daran, ein Verzeichnis und eine gemeinsame Ablage von digitalen Lehr- und Lernmaterialien zu errichten, um den Austausch und die Wiederverwendung zu fördern.

Eine Langzeitarchivierung ist aufgrund der Vielfalt der Unterlagen und aufgrund des multimedialen Charakters und der Komplexität der Unterlagen sehr aufwändig, obwohl mit SCORM<sup>19</sup>

---

<sup>19</sup> Das SCORM (Sharable Content Object Reference Model) ist ein Referenzmodell für austauschbare elektronische Lerninhalte der Advanced Distributed Learning Initiative des us-amerikanischen Verteidigungsministeriums.



grundsätzlich ein Standard existiert. Da langfristig nur ein historischer Nutzungsbedarf bleiben wird, ist es sinnvoll, nur eine kleine Auswahl langfristig zu archivieren.

## 3.4 Digitalisate

### 3.4.1 Kurzcharakterisierung der Unterlagen

Auch in der Schweiz werden in zahlreichen Projekten Archivalien auf Papier oder analogen audiovisuellen Medien retrospektiv digitalisiert, um ...

- die Archivalien leichter zugänglich und nutzbar zu machen;
- die Originale vor Benutzungsschäden zu bewahren;
- die Originale durch digitale Kopien zu ersetzen, wenn die weitere Erhaltung der Originale nicht mehr möglich oder zu kostspielig geworden ist.

Digitalisierte Unterlagen werden in der Regel wie digitale Publikationen (siehe Abschnitt 3.2) über spezialisierte Nutzungssysteme und in spezieller Aufbereitung zur Verfügung gestellt.

Nach Erstellung der verschiedenen Benutzungsversionen bleiben die nicht oder nur wenig bearbeiteten Roh- bzw. „Master“-Digitalisate in höchster Auflösung zur Archivierung zurück. Man kann davon ausgehen, dass heute Master-Digitalisate durchgehend in archivtauglichen Standardformaten erstellt werden.

Da heute gewöhnlich mit einer hohen optischen Auflösungen digitalisiert wird, sind auch die entstandenen und laufend neu entstehenden Datenmengen sehr gross und bei jedem Projekt schnell in der Grössenordnung von mehreren Terabytes.

### 3.4.2 Möglicher Nutzen der Archivierung

Master-Digitalisate werden zu folgenden Zwecken aufbewahrt und benutzt:

- Sie bilden das Ausgangsmaterial für neue, künftige und spezielle Benutzungsversionen, da sie dem „Originalobjekt“ am nächsten sind.
- Digitalisate werden dort zu eigentlichen Archiv-Mastern, wo das ursprüngliche Material rasch altert und nicht mehr oder nur mit sehr grossem Aufwand erhalten werden kann (Ersatzdigitalisierung).

Die aktuelle Entwicklung deutet allerdings darauf hin, dass überall dort, wo die Vorlagen für die Digitalisierung weiterhin adäquat aufbewahrt werden (können), an Stelle einer langfristigen Archivierung von Master-Digitalisaten eher eine spätere Neudigitalisierung ins Auge gefasst wird, da die Fortschritte in der Digitalisierungstechnik laufend markante Effizienz- und Qualitätsgewinne ermöglichen.

### 3.4.3 Nutzungshäufigkeit und Bedarf

Die Nutzungshäufigkeit dürfte sehr gering sein, da die eigentliche Benutzung der Digitalisate über spezifische Benutzungssysteme erfolgt.

Der Bedarf, die Master-Digitalisate fachgerecht und sicher aufzubewahren, besteht jedoch und dürfte angesichts der grossen Zahl von Digitalisierungsprojekten beträchtlich sein. Ob es dazu allerdings einer nationalen Lösung bedarf, darf bezweifelt werden.

#### 3.4.4 Archivierungssituation

Ein genauer Überblick über die Archivierungssituation besteht nicht. Je nach Institution wird die Situation unterschiedlich sein, in den meisten Fällen dürften aber die Master-Digitalisate nicht fachgerecht und genügend sicher archiviert sein. In der Regel mangelt es in erster Linie an genügenden Speicherkapazitäten auf professionellen Speichersystemen.

## 4 Anspruchsgruppen

Wie bereits in der Einleitung (Kapitel 2) und Kapitel 3 angetönt, tangiert eine zentrale Lösung für die Langzeitarchivierung von digitalen Primär- und Sekundärdaten viele Institutionen und Gruppen, die diesbezüglich unterschiedliche Interessen und Ansprüche haben. Im Hinblick auf die Identifizierung möglicher Träger und der Klärung der Finanzierung ist deshalb die Frage nach den möglichen Anspruchsgruppen von hoher Wichtigkeit für die Realisierbarkeit einer Lösung. Welche der *möglichen* Anspruchsgruppen schlussendlich tatsächlich relevante Anspruchsgruppen sowie Träger und Nutzer einer Lösung sind, hängt von den gewählten strategischen Zielen des Modells ab.

Umgekehrt hängt die Realisierbarkeit davon ab, wie sich Anspruchsgruppen bezüglich dieser Vorschläge positionieren werden. Im Rahmen der Ressourcen des vorliegenden Projektes war es leider nicht möglich, eine umfangreiche Konsultation der Anspruchsgruppen durchzuführen.

Mögliche Anspruchsgruppen und ihre Bedürfnisse und Positionierungen haben wir deshalb auf Grund unserer Fachkenntnisse, der Auswertung der gesetzlichen Grundlagen und von Fachdokumenten sowie den Resultaten eines Workshops<sup>20</sup> [18] mit Vertreterinnen und Vertretern möglicher Anspruchsgruppen zu charakterisieren versucht. Die nachstehende Tabelle nennt die wichtigsten Anspruchsgruppen, ihre wichtigsten Interessen und Erwartungen sowie allfällige kritischen Punkte, die zu beachten sind.

Um die Komplexität zu reduzieren, wird hier (im Gegensatz zu Kapitel 3) keine Differenzierung nach Unterlagentypen vorgenommen. Es ist aber zu berücksichtigen, dass die Erwartungen der Anspruchsgruppen unterschiedlich sind, je nachdem, ob es sich um Forschungsdaten, Publikationen oder Digitalisate handelt (siehe dazu Kapitel 3).

Wie die nachstehende Tabelle zeigt, können einzelne Anspruchsgruppen bei der Einrichtung und dem Betrieb eines zentralen Langzeitarchivs für Primär- und Sekundärdaten mehrere Rollen spielen, da verschiedene Akteure sowohl als Kunden (als Auftraggeber und/oder Nutzer) wie auch als Träger oder Leistungserbringer in Frage kommen.

---

<sup>20</sup> Workshop II, 26.08.2008, ETH-Bibliothek in Zürich. Teilnehmer/innen: Kay Bieri (Physikalisches Institut Uni Bern, Climate and Environmental Physics / Climat unibe), Martin Borer (ETH-Informatikdienste), Pascalia Boutsouci (ETH-Bibliothek), Urs Brander (Paul Scherrer Institut / PSI), Rolf Brugger (SWITCH), Niklaus Bütikofer (ikeep AG), Jean-Marc Comment (Bundesarchiv / BAR), Michael Ehrismann (Konsortium der Schweizer Hochschulbibliotheken), Michael Gasser (ETH-Archiv, ETH-Bibliothek), Angela Gastl (Projekte Digitale Bibliothek, ETH-Bibliothek), Susanne Grulich Zier (ETH-Archiv, ETH-Bibliothek), Patrick Hinni (Nationalbibliothek / SNB), Martin Kaiser (Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen / KOST), Peter Keller-Marxer (ikeep AG), Rudolf Mumenthaler (ETH-Bibliothek, Projekt e-rara), Christian Rohrer (SWITCH), Marco Schmidt (Functional Genomics Center Zurich, ETHZ und Uni ZH), Nicole Simonet (Konsortium der Schweizer Hochschulbibliotheken), Matthias Töwe (Konferenz der Universitätsbibliotheken der Schweiz / KUB). Bericht zum Workshop: [18].

4 Anspruchsgruppen

Stakeholder	Kurzbeschreibung	Interessen / Erwartungen	Kritische Punkte
<b>BIBLIOTHEKEN</b>			
<b>Schweizerische Nationalbibliothek</b>	Zentrale Stelle des Bundes für die Sammlung und dauerhafte Aufbewahrung der publizierten Helvetica.	Interesse an optimaler schweizerischer Kooperation bei der Archivierung <i>schweizerischer</i> Publikationen  Bietet selber Angebot zur Speicherung von Datenpaketen (Schliessfachprinzip) zum Selbstkostenpreis	
<b>KUB (Konferenz der Universitätsbibliotheken der Schweiz)</b>  <b>Universitätsbibliotheken</b>  <b>Fachhochschul-Bibliotheken</b>	Die KUB betreut eine Reihe von Koordinationsaufgaben der Hochschulbibliotheken. Sie betreibt u.a. das «Konsortium» sowie die Projekte e-Archiving und E-Lib.ch.	Die KUB ist Projektauftraggeberin des vorliegenden Projekts zur Langzeitarchivierung von Primär- und Sekundärdaten, das ein Teilvorhaben im Rahmen des nationalen Projekts E-lib.ch ist.  Zentrale Lösung zur Langzeitarchivierung von elektronischen Zeitschriften (eJournals) der einzelnen Universitäten und Hochschulen.	Abgrenzung / Aufgabenteilung bzw. Kooperation im Bereich der Primärdaten zwischen Archiven und Bibliotheken
<b>ARCHIVE</b>			
<b>Schweizerisches Bundesarchiv (BAR)</b>	Zentrale Archivierungsstelle des Bundes für Bundesrat, Parlament und Bundesverwaltung gemäss Bundesgesetz über die Archivierung (BGA).	Aufsicht über die Umsetzung des BGA im Bereich der ETHs und der Forschungsanstalten des Bundes (siehe unten). Bietet Dritten die Archivierung von Unterlagen zum Selbstkostenpreis an. Diese müssen jedoch bez. Formaten und Strukturen den Anforderungen des BAR entsprechen.	
<b>Selbstständig archivierende Stellen gemäss BGA</b>	Hier relevant sind die ETH, EMPA, EAWAG, WSL, PSI. Sie archivieren ihre Geschäftsunterlagen selbstständig nach den Grundsätzen des BGA und einer Vereinbarung mit dem Bundesarchiv.	Möglichkeit der Aufbewahrung von Primärdaten für rund 10 Jahre zur Erfüllung der Vorschriften zur Integrität in der Forschung. Erfüllung der Archivierung gemäss BGA. Das PSI bewahrt Primärdaten aus der Forschung am PSI mehr oder weniger systematisch selber für 10 Jahre auf.	

#### 4 Anspruchsgruppen

<b>Kantonsarchive</b>	Zentrale Archivierungsstellen der Regierungen, Parlamente und Verwaltungen der jeweiligen Kantone	Aufsicht über und teilweise Durchführung der Archivierung von Geschäftsunterlagen der kantonalen Universitäten und Fachhochschulen. Die Hoheit über die Beurteilung der Archivwürdigkeit liegt bei den Archiven.	Primärdaten der Universitäten und Fachhochschulen werden den Archiven heute weder angeboten noch von diesen bewertet und archiviert.
<b>KOST (Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen)</b>	Die Stelle wird getragen von 18 Kantonen, dem Fürstentum Liechtenstein und dem Bund. Sie leitet und begleitet digitale Archivprojekte, erarbeitet Anforderungen und verfolgt die Entwicklungen im Bereiche der digitalen Archivierung.	Interesse, den Kantonsarchiven geeignete Archivierungsdienstleistungen vorwiegend im Bereich der Speicherung und „Bitstream Preservation“ zu vermitteln, z.B. für die Master- oder Rohdigitalisate der von Archiven für die Bestandserhaltung und Vermittlung digitalisierten Unterlagen (Bild, Ton, Film).	Ein langfristiges Engagement aller an der KOST beteiligten Stellen dürfte schwierig zu erreichen sein. Die KOST selber hat keine Entscheidungskompetenz.
<b>Archive der kantonalen Universitäten und Fachhochschulen</b>	Führen Archivierungsaufgaben an den jeweiligen Institutionen durch, sind aber den jeweiligen kantonalen Archivierungsgesetzen unterstellt und fachlich den jeweiligen kantonalen Archiven zugeordnet.	Fehlende Verfahren und Infrastrukturen für die digitale Archivierung.  Sicherungs- und Bewertungsverfahren (Diese Archive sind i.d.R. die primären Ansprechpartner der Forschungsgruppen)	Diese Archive wie auch die Kantonsarchive, denen sie zugeordnet sind, haben kaum finanzielle Ressourcen für die digitale Archivierung.  Die Archivierung von Forschungsdaten benötigt viel Spezialwissen über die jeweiligen Forschungsgebiete.
<b>Spezialarchive</b> (einzelner Wissenschaftsbereiche)	Von einzelnen Wissenschaftsgebieten selber initiierte und betriebene Archive für Forschungsdaten (bspw. SIDOS für die Sozialwissenschaften)	Wahrscheinlich kein Interesse an einem zentralen Angebot.	Sehr heterogenes Feld mit vielen bereichsspezifischen Lösungen. Die Daten sind in der Regel fachspezifisch erschlossen.
<b>UNIVERSITÄTEN, HOCHSCHULEN UND FORSCHUNGSANSTALTEN</b>			
<b>Leitung der Hochschulen und Forschungsanstalten</b>		Erfüllung der Regeln zur Integrität bzw. Nachvollziehbarkeit der Forschungen an ihrer Institution.  Investitionsschutz: Sekundärnutzung von Forschungsdaten und Erhaltung der lizenzierten digitalen Publikationen (und der retro-digitalisierten Unterlagen).	

#### 4 Anspruchsgruppen

<b>Forschende und Forschungsgruppen</b>	<p>Wichtigste Zielgruppe. Alle mit öffentlichen Geldern finanzierte oder geförderte Forschende und Forschungsprojekte aus allen Wissenschaftsbereiche an allen Forschungsinstitutionen (Universitäten, Fachhochschulen, ETHs, Forschungsanstalten des Bundes und weitere.)</p>	<p>Die Vorschriften in den Integritätsrichtlinien der Hochschulen und den Beitragsreglemente von Forschungsförderungsstellen delegieren die Verantwortung zur adäquaten Aufbewahrung von Primärdaten alleine an die Forschenden bzw. Projektleiter. Diesen fehlen aber i.d.R. das Wissen und die nötigen Infrastrukturen und finanziellen Mittel dazu. Sie sind also an einfachen Lösungen interessiert, die sie bei der fachgerechten Aufbewahrung effizient unterstützen.</p> <p>Interesse der Forschenden an der Sekundär- oder Nachnutzung eigener oder fremder Primärdaten.</p> <p>Interesse an der Zitierbarkeit eigener und fremder Primärdaten.</p>	<p>Damit eine zentrale Archivierungsdienstleistung Erfolg haben kann, muss sie im Bereich der Integritätsregeln für die Forschenden einfach nutzbar sein, darf keinen Zwang darstellen und muss grundsätzlich für die Forschenden kostenfrei sein bzw. darf das Forschungsbudget der Projekte nicht belasten.</p>
<b>SWITCH</b>	<p>Non-profit Organisation der Schweizerischen Wissenschaftsinstitutionen, welche hauptsächlich Netzwerkinfrastrukturen zur Verfügung stellt.</p>	<p>Unterstützt die Universitäten bei der Erstellung, Verwaltung und Zugänglichmachung von digitalen Lern- und Lehrmaterialien (eLearning).</p> <p>Möglicher Infrastrukturpartner für die Archivierung.</p>	
<b>FORSCHUNGSFÖRDERUNG und FORSCHUNGSPOLITIK</b>			
<b>Nationalfonds (SNF)</b> <b>Akademien Schweiz</b> <b>Staatssekretariat für Bildung und Forschung (SBF)</b> <b>Bundesamt für Berufsbildung und Technologie (BBT)</b>	<p>Der Nationalfonds und die Akademien müssen ihre Beitragsempfänger seit 2008 zur Einhaltung der „Guten wissenschaftlichen Praxis“ verpflichten (gem. Art. 11a des Forschungsgesetzes). Zahlreiche Universitäten und Hochschulen haben die Regeln der Akademien ebenfalls übernommen.</p>	<p>Diese Stellen sind daran interessiert, dass die von ihnen selber aufgestellten Richtlinien zur Aufbewahrung von Primärdaten im Rahmen der „Integrität der Forschung“ von den Forschenden möglichst vollständig und effizient erfüllt werden können.</p> <p>Investitionsschutz: Sekundärnutzung von Forschungsdaten.</p>	<p>Gleiche Vorbehalte wie bei den Forschenden.</p>

Mögliche **Kunden** einer zentralen Langzeitarchivierung sind bei Primärdaten in erster Linie die Forschenden, Forschungsprojekte und Forschungsgruppen sowie bei digitalen Publikationen die Bibliotheken. Als Kunden sind beide zudem gleichzeitig Auftraggeber und Benutzer, da sie Unterlagen zur zentralen Langzeitarchivierung übergeben und archivierte Unterlagen selber nutzen.

Mögliche **Träger** sind, soweit es sich bei den Archivierungsaufgaben um gesetzlich vorgeschriebene Archivierung handelt, die vom Gesetz bezeichneten Stellen (siehe Abschnitt 4.1 unten), also bei elektronischen Zeitschriften in jedem Fall die Bibliotheken. Einer Zentralisierung der Archivierung bei einer Universitäts- oder Hochschulbibliothek stehen höchstens allfällige Restriktionen in den Lizenzbedingungen der Verlage im Wege.

Die Organe Forschungsförderung sowie die Universitäten und Hochschulen, welche kürzlich die neuen Richtlinien zur Aufbewahrung von Primärdaten im Rahmen der „Integrität der Forschung“ (siehe Kapitel 3.1.3) erlassen haben, werden daran interessiert sein, dass ihre Richtlinien von den Forschenden möglichst vollständig und effizient erfüllt werden können.

Als möglichen **Leistungserbringer** gibt es unter den Anspruchsgruppen jedoch niemanden, der bereits eine Langzeitarchivierung für Primär- und Sekundärdaten anbietet. Verschiedene Institutionen unter ihnen bieten aber Elemente an, die beim Aufbau einer zentralen Langzeitarchivierung für Primär- und Sekundärdaten aus der Forschung genutzt werden können.

## 4.1 Rechtlicher Rahmen

In den einleitenden Abschnitten 2.2.3 und 2.3 wurde die grundsätzlich unterschiedlichen Ansprüche der Forschenden und Forschungsinstitutionen (Aufbewahrungspflichten im Rahmen der Vorschriften zur Integrität in der Forschung) und der staatlichen Archivstellen (Archivierungsgesetze) diskutiert. Diese Ansprüche und die zeitlich begrenzte Aufbewahrung bzw. dauerhafte Archivierung können sich bei Primärdaten nur beschränkt ergänzen (Kapitel 2.3). Aus diesem Grund werden wir in Kapitel 5 auch zwei grundsätzlich unterschiedliche Modellvarianten für Primärdaten präsentieren.

Alle staatlichen Hochschulen und Forschungsinstitute haben eine gesetzliche Grundlage und unterstehen staatlichem Archivierungsrecht. Hochschulen und Forschungsanstalten des Bundes unterstehen dem Bundesgesetz über die Archivierung vom 26.6.1998 (BGA) [2], die kantonalen Universitäten und Fachhochschulen dem jeweiligen kantonalen Archivierungsrecht.

Allerdings können die Datenproduzenten die Aufbewahrung von Primärdaten, die als nicht archivwürdig im Sinne der staatlichen Archivgesetze (Bund und Kantone) bewertet sind, nach eigenen Bedürfnissen regeln und durchführen oder eine beliebige, geeignete Institution damit beauftragen, sofern dabei die Benutzungsvorschriften der betreffenden Archivierungsgesetze, Datenschutzgesetze und Öffentlichkeitsgesetze nicht verletzt werden. Sinnvollerweise handelt es sich dabei jedoch um eine vertrauenswürdige, mit Vorteil staatlich getragene Stelle.

Primärdaten, die im Sinne der staatlichen Archivgesetze als archivwürdig bewertet sind, müssen nach den Vorgaben der zuständigen Archivstellen aufbereitet und unter deren Kontrolle archiviert werden.

### 4.1.1 Bundesgesetz über die Archivierung

Das Bundesgesetz über die Archivierung [2] (BGA) unterscheidet drei Kreise:

- Die **anbietepflichtigen Stellen**, welche alle nicht mehr ständig benötigten Unterlagen der zuständigen Archivstelle anbieten müssen. Diese bewertet die Archivwürdigkeit der angebotenen Unterlagen, übernimmt die archivwürdigen, verwahrt sie dauerhaft und vermittelt sie auf eigene Kosten den Benutzerinnen und Benutzern des Archivs. Die für eine dauerhafte Archivierung nötige Aufbereitung der abzuliefernden Unterlagen obliegt den anbietepflichtigen Stellen und muss die Vorgaben des Bundesarchivs erfüllen.
- Die **selbstständig archivierenden Stellen** gemäss BGA sind die in der Verordnung VBGA [19] zum BGA genannten Stellen, die ihre Unterlagen nicht dem Schweizerischen Bundesarchiv zur Archivierung anbieten, sondern die Archivierungsaufgaben im Rahmen des BGA und VBGA vollumfänglich selber und auf ihre Kosten durchführen. Einzelheiten regelt das Bundesarchiv mit diesen Stellen in Vereinbarungen.
- Stellen oder Personen, welche Bundesaufgaben in einem **privatrechtlichen Auftragsverhältnis** ausführen (VBGA Art. 2 Abs. 3 und Art. 9): Mit diesen Stellen bzw. Personen muss die Auftrag erteilende Stelle in Absprache mit dem Bundesarchiv die Archivierung vertraglich regeln.

Der vorliegende Bericht ist auftragsgemäss auf Hochschulen und Forschungsanstalten ausgerichtet. Die beiden Hochschulen des Bundes (ETH Zürich und EPF Lausanne) und die Forschungsanstalten des Bundes gehören zu den selbstständig archivierenden Stellen gemäss BGA, die in VBGA, Anhang 2 Bst. a, abschliessend aufgelistet sind:<sup>21</sup>

- Eidgenössische Anstalt für Wasserversorgung, Abwasserreinigung und Gewässerschutz
- Eidgenössische Forschungsanstalt für Wald, Schnee und Landschaft
- Eidgenössische Materialprüfungs- und Forschungsanstalt
- Eidgenössische Technische Hochschulen (Lausanne und Zürich)
- Paul Scherrer Institut
- Rat der Eidgenössischen Technischen Hochschulen

Die rechtlich vorgesehenen Vereinbarungen mit dem Bundesarchiv hat die ETH Zürich bereits vor einiger Zeit abgeschlossen. Man darf davon ausgehen, dass die anderen, teilweise noch ausstehenden Vereinbarungen grundsätzlich dieselben Bestimmungen enthalten. Im Rahmen dieser Bestimmungen könnten die Archive der Hochschulen und Forschungsanstalten des Bundes bei der Archivierung zusammenarbeiten und ein gemeinsames digitales Langzeitarchiv unterhalten. Solange die archivierten Unterlagen im Eigentum und unter Kontrolle dieser Stellen bleiben, sprechen keine gesetzlichen Bestimmungen gegen dieses Vorgehen.

Im Hochschul- und Forschungsbereich des Bundes dürfte auch eine grössere Zahl privatrechtlicher Auftragsverhältnisse bestehen (z.B. Drittmittelforschung). Im Rahmen dieses Konzeptes

<sup>21</sup> Hier werden nur die Hochschulen und Forschungsanstalten wiedergegeben. Die Liste in der VBGA umfasst noch weitere Stellen, die aber hier nicht relevant sind.



konnte keine Erhebung gemacht werden, in welchem Umfange in diesen Auftragsverhältnissen die Archivierung explizit geregelt wird. Man muss davon ausgehen, dass dies, wenn überhaupt, nur bei einer kleinen Zahl von Forschungsaufträgen der Fall ist.

Das Bundesrecht gilt für sämtliche *Unterlagen* der Stellen im Zuständigkeitsbereich. Die Legaldefinition von ‚Unterlagen‘ ist dabei sehr umfassend:

- *«Unterlagen im Sinne dieses Gesetzes sind alle aufgezeichneten Informationen, unabhängig vom Informationsträger, welche bei der Erfüllung öffentlicher Aufgaben des Bundes empfangen oder erstellt worden sind, sowie alle Hilfsmittel und ergänzenden Daten, die für das Verständnis dieser Informationen und deren Nutzung notwendig sind.» (BGA Art. 3 Abs. 1) [2]*

Diese Definition wird in den Rechtstexten selber nicht eingegrenzt, wohl aber in der Botschaft des Bundesrates über das Bundesgesetz über die Archivierung vom 26.2.1997 [20]:

- *«Bei Bundesstellen, deren Aufgabe im Sammeln und Archivieren von bestimmten Dokumenten besteht (z.B. die zahlreichen Archive im ETH-Bereich, die SLB usw.), fallen selbstverständlich nur die Geschäftsunterlagen, nicht aber deren Sammlungsgut in den Geltungsbereich.»*

Der Begriff ‚Unterlagen‘ meint also nicht Sammlungen von Literatur, Daten und Archivalien, die von Dritten erworben worden sind. Die erworbenen bzw. gesammelten Literatur- und Datenbestände der ETH-Bibliothek und anderer Instituts- bzw. Anstaltsbibliotheken fallen deshalb nicht unter den Geltungsbereich der Archivierungsgesetzgebung.

Forschungs- und Entwicklungstätigkeiten gehören dagegen zu den Kerngeschäften der Hochschulen und Forschungsanstalten. Da nicht nur bei Forschungsarbeiten häufig Unterlagen Dritter beigezogen werden müssen, besteht eine Grauzone, in der ein klares Kriterium benötigt wird, ob Unterlagen von Dritten Teil der Geschäftsunterlagen oder Teil der Sammlungen sind. In der Fachdiskussion wird hier das Kriterium der Geschäftsrelevanz beigezogen: Unterlagen (inkl. Daten) von Dritten gehören dann zu den Geschäftsunterlagen, wenn sie für den Fortgang und die Nachvollziehbarkeit eines Geschäftes wesentlich sind und sie nicht ohne weiteres jederzeit über andere Kanäle bezogen werden können (wie bspw. Publikationen).

Das Archivierungsgesetz des Bundes (BGA) umreißt grob, nach welchen Kriterien die Archivwürdigkeit zu bestimmen ist:

- *Rechtlich, politisch, wirtschaftlich, historisch, sozial oder kulturell wertvolle Unterlagen des Bundes werden archiviert. (BGA Art. 2 Abs. 1) [2]*
- *Archivwürdig sind Unterlagen, die von juristischer oder administrativer Bedeutung sind oder einen grossen Informationswert haben. (BGA Art. 3 Abs. 3)*

Im Rahmen der erwähnten Vereinbarung zwischen selbstständig archivierenden Stellen und Bundesarchiv können erstere ein Bewertungskonzept erstellen und dieses selbstständig umsetzen. Die ETH Zürich hat ein solches Konzept in Form der *Richtlinien für die Bewertung von Unterlagen am Archiv der ETH Zürich* vom 29.1./23.4.02 [21] erstellt. Die Richtlinien gehen allerdings nicht explizit auf Forschungsunterlagen ein, so dass für die in diesem Bericht fokussierten Primärdaten aus der Forschung keine Vorentscheide vorliegen und zu berücksichtigen sind.

Unterlagen, die von den jeweils zuständigen Organen (selbstständig archivierende Stellen bzw. Bundesarchiv) als nicht archivwürdig beurteilt worden sind, sind kein Archivgut und unterstehen

nicht mehr dem BGA.<sup>22</sup> Die Eigentümer sind deshalb frei, diese solange aufzubewahren wie sie das für nötig und sinnvoll erachten bzw. solange es administrative Fristen oder andere Gesetze als das BGA verlangen.

### 4.1.2 Kantonales Recht

Die kantonalen Universitäten unterstehen dem Archivierungsrecht der jeweiligen Kantone. Die Universitäten unterhalten aber in der Regel einen eigenen Archivdienst im Rahmen der gesetzlichen Vorschriften, welche jedoch vorwiegend nur als Zwischenstationen der Universitätsunterlagen auf dem Weg ins zuständige Kantonsarchiv dienen.<sup>23</sup>

Die Universitäten unterstehen einer je nach Kanton mehr oder weniger explizit formulierten Anbietepflicht, die eine mehr oder weniger strikte Bewertungshoheit des zuständigen Kantonsarchivs impliziert.

Die Fachhochschulen sind kantonale oder interkantonale öffentlich-rechtliche Anstalten mit eigener Rechtspersönlichkeit mit Selbstverwaltung im Rahmen der jeweiligen Gesetze und Verträge, auf denen sie basieren. Bei denjenigen Fachhochschulen, die auf einem Konkordat mehrerer Kantone beruhen, ist nach aussen hin nicht klar, welche kantonalen Archivierungsgesetze wo gelten. Die entsprechenden Konkordatsverträge gehen, soweit es im Rahmen dieses Berichts abgeklärt wurde, nicht auf die Archivierungsfrage ein.

Die neueren kantonalen Archivierungsgesetze benutzen alle den Begriff ‚Unterlagen‘ aus dem Bundesrecht mit sehr ähnlicher Definition und ebenfalls, ohne diesen weiter zu definieren.

Die Möglichkeit, für die Archivierung von digitalen Unterlagen Dienstleistungen Dritter in Anspruch zu nehmen, wird, soweit es im Rahmen dieses Berichts abgeklärt wurde, nirgendwo explizit ausgeschlossen.<sup>24</sup> Der Entwurf für ein neues bernisches Archivierungsgesetz [22] sieht in Art. 7 Abs. 2 für die Archivierung von digitalem Archivgut sogar ausdrücklich die Möglichkeit vor, Dienstleistungen von geeigneten Unternehmen in Anspruch zu nehmen.

### 4.1.3 Bibliotheksrecht

Bibliotheken haben in der Regel in ihren gesetzlichen Grundlagen einen Sammlungs- und Versorgungsauftrag. Dieser ist bei der Nationalbibliothek, bei den Kantonsbibliotheken und bei den wissenschaftlichen Bibliotheken zeitlich nicht begrenzt. Sie sind deshalb eigentliche ‚Archivbibliotheken‘. Die Möglichkeit von Ausscheidungen aus dem Bibliotheksgut wird in ihren Rechtsgrundlagen nicht erwähnt. Ausscheidungen kommen aber in der Praxis vor und liegen im Ermessen der einzelnen Institutionen. Als dauernd zu archivierender Kern gelten jedoch die Helvetica und die Kantonalia.

---

<sup>22</sup> Die Bestimmungen des BGA zur Zugänglichkeit (3. Abschnitt), zur gewerblichen Nutzung (Art. 19) und zur Unveräusserlichkeit (Art. 20) beziehen sich nur auf Archivgut.

<sup>23</sup> So bspw. in den Universitäten Basel, Bern und Zürich

<sup>24</sup> Eine Aufbewahrung im Ausland müsste allerdings rechtlich genauer abgeklärt werden.

## 5 Modelle für Primärdaten

Die Konferenz der Universitätsbibliotheken der Schweiz (KUB) verfolgt mit dem vorliegenden Projekt die «Entwicklung eines Modells für eine zentrale Langzeitarchivierung von digitalen Primär- und Sekundärdaten aus der Forschung für die Schweiz» (Projekttitle<sup>25</sup>).

Wir schlagen in den folgenden Kapiteln 5.1 und 5.2 zwei Modellvarianten für Primärdaten vor, ausgehend von den Projektzielen und -vorgaben, unserer Analyse in den Kapiteln 3 (Bedarf und Nutzen), 4 (Anspruchsgruppen) und 8 (Europäisches Umfeld) sowie der im einleitenden Abschnitt 2.3 diskutierten Schwerpunktsetzung.

### 5.1 Modell «Aufbewahrung mit Selbstregulierung»

Dieses Modell für Primärdaten funktioniert nach einem «Self Deposit»-Ansatz und fokussiert pragmatisch auf die Bedürfnisse der Forschenden hinsichtlich der neuen in der Schweiz geltenden Richtlinien zur „Integrität der Forschung“. Die folgend als «die Archivdienstleistung für Primärdaten» bezeichnete nationale Dienstleistung sieht keine Regulierung vor und geht konsequent von der Selbstverantwortung und Eigeninitiative der Forschenden, Forschungsbereiche und Forschungsinstitutionen aus, die Anforderungen an Dauer, Umfang und Formate der aufzubewahrenden Primärdaten bezüglich Integritätsrichtlinien und eines Wertes für eine längerfristige Sekundärnutzung selber zu bestimmen und umzusetzen.

Diese Modellvariante orientiert sich funktional an RFC 4810 «Long-Term Archive Service Requirements»<sup>26</sup> [23] der «Long-Term Archive and Notary Services Charter» (LTANS), jedoch *nicht* an ISO 14721:2003 «Open Archival Information System Reference Model» (OAIS).

Die Archivdienstleistung erleichtert Forschenden bzw. Projektleitungen, ihren Aufbewahrungspflichten bezüglich Primärdaten im Sinne der „Integrität der Forschung“ nachkommen zu können und bietet zusätzlichen Nutzen durch Beratung, Zitierbarkeit und eine freiwillige öffentliche Katalogisierung der deponierten Daten. Sie soll für die Forschenden und Forschungsprojekte grundsätzlich kostenlos sein.

Die Archivdienstleistung führt folgende Basisfunktionen aus:

- Zur Aufbereitung der Primärdaten wird den Forschenden und Forschungsinstitutionen auf Nachfrage Beratung und Unterstützung durch ausgewählte Fachpersonen eines Kompetenznetzwerks geboten.

---

<sup>25</sup> Auf die Bedeutungsunterschiede zwischen *Archivierung* und *Aufbewahrung* bzw. *Langzeitarchivierung* und *Langzeitaufbewahrung*, die sich aus staatlichen Archivgesetzen und archivfachlichen Gepflogenheiten herleiten, wurde in den Kapiteln 2.1 eingegangen.

<sup>26</sup> <http://www.ietf.org/rfc/rfc4810.txt>

- Die deponierten Daten werden katastrophensicher gespeichert und vor unberechtigten Zugriffen geschützt.
- «Bitstream Preservation» für die deponierten Datenobjekte.
- Beweiskräftige Nachweisbarkeit der Integrität (Unverändertheit) der Daten durch qualifizierte elektronische Zeitstempel gemäss Signaturgesetz (ZertES).
- Zitierbarkeit der deponierten Daten in Publikationen mit Hilfe eines von der Archivdienstleistung erzeugten und verwalteten Digital Object Identifiers (DOI).
- Forschende können einem Dritten den einmaligen lesenden Zugriff auf von ihnen deponierten Daten ermöglichen (je nach Datenmenge online oder als schriftlicher Auftrag an die Archivdienstleistung, die Daten an eine bestimmte Person zu liefern).
- Forschende können von ihnen deponierte Daten freiwillig in einem öffentlich zugänglichen Katalog verzeichnen lassen.

Ergänzend bietet die Archivdienstleistung den *Forschungsinstitutionen* zudem die Möglichkeit, jederzeit, spätestens aber bei Ablauf der von den Forschenden definierten Aufbewahrungsfrist, die Daten *freiwillig* einer fachlich-inhaltlichen Bewertung zuzuführen, die Aufbewahrungsfrist entsprechend zu verlängern oder die Daten bei *dauerndem* Wert einer entsprechenden dauerhaften Archivierung durch die Dienstleistung zu übergeben. Diese Modellvariante berücksichtigt damit die Empfehlungen der OECD (siehe Kapitel 3.1.3.4), bei der dauerhaften Aufbewahrung von Primärdaten für die Sekundärnutzung die Problematik einer vorschnellen Selektion durch periodische Überprüfungen zu lösen.

### 5.1.1 Strategieziele

Die folgenden sechs strategischen Ziele sollen mit dem Modell erreicht werden:

- G1. **National:** Die Archivdienstleistung für Primärdaten soll Forschenden an allen öffentlichen Forschungsinstitutionen in der Schweiz dienen können: neben dem ETH-Bereich insbesondere auch den kantonalen Universitäten und Fachhochschulen und sonstigen öffentlich finanzierten oder geförderten Forschungsinstitutionen.
- G2. **Fachübergreifend:** Die Archivdienstleistung für Primärdaten soll Forschenden in allen Wissenschaftsbereichen dienen, sich also nicht nur an einzelne Fachbereiche oder Institutionen mit besonderen Voraussetzungen richten.
- G3. **Unterstützend:** Die Archivdienstleistung für Primärdaten soll die praktische Umsetzung der Aufbewahrungsrichtlinien „Integrität der Forschung“ der Organe der Forschungsförderung, der Universitäten sowie Art. 11a des Forschungsgesetzes pragmatisch und effizient unterstützen, die Zitierbarkeit von Primärdaten in Publikationen erlauben, die Sekundärnutzung von Primärdaten erleichtern und die Verfügbarkeit von dauerhaft wertvollen Primärdaten ermöglichen.
- G4. **Forschungspolitisch:** Die Archivdienstleistung für Primärdaten soll ermöglichen, dass die Verantwortung für eine sachgerechte Aufbewahrung von wissenschaftlichen Daten nicht alleine an die einzelnen Forschenden (Primärdaten) delegiert wird, sondern im Sinne der Empfehlungen der OECD [7] als Aufgabe und Bestandteil der nationalen Forschungsinfrastruktur und -kultur betrachtet werden kann.

- G5. Integral:** Die Archivdienstleistung für Primärdaten soll aus Gründen der Transparenz und wegen ihrer organisatorischen und technischen Komplexität sowie Kostenintensität von einer einzelnen, geeigneten Institution angeboten werden, die von den Organen der Forschung und Forschungsförderung als vertrauenswürdig anerkannt und entsprechend mandatiert und mitgetragen wird.
- G6. Nicht-regulierend:** Die Archivdienstleistung soll *seitens der Trägerschaft* mit keiner Pflicht zur Nutzung verbunden werden, d.h. die Forschenden und Forschungsinstitutionen sollen sie auf freiwilliger Basis nutzen können.

### 5.1.2 Begründung

Die Problematik der Aufbewahrung von Primärdaten der Forschung wurde in Kapitel 2.2.3 und 3.1 diskutiert (siehe dazu auch das europäische Umfeld im den Kapiteln 8.4 und 8.5).

Die Entwicklung von Strategien, Projekten und Dienstleistungen zur Primärdaten-Archivierung als integrale Bestandteile der nationalen Forschungsinfrastruktur eines Landes wird von der OECD [7] seit kurzem gefordert bzw. empfohlen (siehe Abschnitte 3.1.3.1 und 3.1.3.4). Hinzu kommt, dass in der Schweiz erst seit kurzem die systematische Aufbewahrung von Primärdaten über einen für den Forschungsbereich „adäquaten“ Zeitraum (im Ausland i.d.R. 10 Jahre) als Teil der „Integrität der Forschung“ bzw. der „Guten wissenschaftlichen Praxis“ sowie für die Sekundärnutzung der Daten gefordert wird. In der Schweiz haben die wissenschaftlichen Akademien und der Nationalfonds (auf der Basis des neuen Art. 11a des Forschungsgesetzes) sowie mehrere Universitäten und Hochschulen entsprechende Richtlinien verabschiedet (siehe Kapitel 3.1.3.2).

Diese Richtlinien delegieren die Verantwortung für die sachgerechte Aufbewahrung alleine an die Forschenden bzw. die Projektleitung. Angesichts der Volatilität von Forschungsprojekten und der hohen Stellenmobilität der Forschenden ist eine langfristige Verantwortung für die sichere Aufbewahrung von zahlreichen, an unterschiedlichen Forschungsstandorten lokalisierte Datenbestände für die einzelne Forscherin, den einzelnen Forscher eine kaum handhabbare Aufgabe.

Hinzu kommt, dass der Aufwand für eine langfristige oder gar dauerhafte Aufbewahrung / Archivierung eines Primärdaten-Satzes nicht durch eine einmalige Investition gedeckt werden kann, sondern laufende Infrastruktur-Betriebskosten verursacht.

Beide Aspekte – Verantwortung und Kosten – bedeuten, dass die Aufbewahrung in der Praxis nicht durch die einzelne Forscherin, den einzelnen Forscher erfolgen kann, sondern bestenfalls durch die Forschungsinstitution, an der die Primärdaten erzeugt wurden. Diese müsste somit im Auftrag der einzelnen Forschenden sicherstellen, dass die relevanten Daten auch nach dem Weggang der Forschenden von der Institution und nach der Löschung von deren Computer-Benutzerkonti noch sicher und auffindbar aufbewahrt bleiben. Darauf sind Forschungsinstitutionen heute in der Regel weder organisatorisch noch technisch vorbereitet, abgesehen von der Bereitstellung reiner Speicherkapazität.

Die Notwendigkeit einer integrierteren Aufbewahrung von Primärdaten und die damit für die Forschenden verbundenen, unbefriedigenden Situation werden erst seit kurzem von der OECD [7] und European Science Foundation [6] thematisiert und auf nationaler sowie inter-

institutioneller Ebene ernsthaft wahrgenommen. Leadership übernehmen dabei in Europa unter anderem die 2006 gegründete und prominent zusammengesetzte «Alliance for Permanent Access to the Records of Science in Europe» (Alliance PARSE) [24] (siehe Kapitel 8.4) sowie in Deutschland die «Allianz-Initiative ‚Digitale Information‘» [25] von neun deutschen Forschungsgesellschaften (Kapitel 8.5), die einen „einen dringenden Handlungsbedarf“ in allen Wissenschaftseinrichtung feststellen. Die Schweiz ist in der PARSE-Allianz nicht vertreten. Auch war bzw. ist der Forschungsplatz Schweiz bei den mit insgesamt über 70 Mio. EUR dotierten 18 Projekten der Europäischen Kommission zu «*IST-2005-2.5.10: Access to and preservation of cultural and scientific resources*» (siehe Kapitel 8) im Forschungsrahmenprogramm FP6 deutlich schwächer beteiligt<sup>27</sup> als zum Beispiel Österreich.

Damit gelten in der Schweiz zwar ähnliche Aufbewahrungsrichtlinien und –ziele für Primärdaten wie im europäischen Ausland, und es bestehen diesbezüglich auch dieselben organisatorischen, technischen und finanziellen Problemstellungen, ohne dass die Schweiz bisher am Prozess der internationalen Diskussion und Lösungsfindung zur Aufbewahrung von Primärdaten substantiell beteiligt ist bzw. von diesem profitieren kann.

Das vorliegende Projekt der KUB will deshalb einerseits die Diskussion auf nationaler Ebene in Gang bringen und gleichzeitig einen ersten, pragmatischen Lösungsschritt initiieren, von dem alle Forschenden profitieren können, insbesondere auch kleinere Forschungsbereiche, die bezüglich der Aufbewahrung von Primärdaten schlecht organisiert sind.

Dass dabei die Hochschul- und Universitätsbibliotheken die Initiative ergreifen, muss nicht erstaunen, da sie ihre „klassischen“ Aufgaben nahe bei den Forschenden erfüllen und die Initiative auch im europäischen Ausland primär von den Bibliotheken und nicht von den staatlichen Archiven ausgeht.

### 5.1.3 Operative Ziele

Wir schlagen für Primärdaten die folgenden operativen Ziele im Rahmen der eingangs aufgeführten, übergeordneten Strategieziele vor.

- P1. **Praxisnähe:** Die Archivdienstleistung für Primärdaten soll auf einfache Weise und direkt durch einzelne Forschende (insbesondere auch Doktorierende) und Forschungsprojekte nutzbar sein, unabhängig von der Projektgrösse<sup>28</sup>.
- P2. **Kostenlosigkeit:** Die Archivdienstleistung zur Erfüllung der Aufbewahrungspflicht für Primärdaten im Rahmen der „Integrität der Forschung“ soll für einzelne Forschende prinzipiell und für Forschungsprojekte grundsätzlich<sup>29</sup> kostenlos nutzbar sein.

<sup>27</sup> Beteiligung des Schweiz. Bundesarchivs an PLANETS (Preservation and long-term access to our cultural and scientific heritage), der SRG/SSR an MEMORIES (Design of an audio semantic indexation system allowing information retrieval for the access to archive content), der Archives Internationales de Musique Populaire an ETHNOARC (Linked European archives for Ethno-musicological research) und der Universität Genf an MULTIMATCH (Multilingual/Multimedia access to cultural heritage).

<sup>28</sup> Gemeint ist hier primär die finanzielle und organisatorische Grösse. Natürlich müssen bezüglich der reinen Datenmenge sinnvolle Grenzen *nach oben* gesetzt werden. So kann und soll es nicht Ziel sein, singuläre „Data Hot Spots“ zu bedienen, die pro Jahr Dutzende von Terabytes erzeugen. Diese Forschungsprojekte verfügen jedoch ohnehin über eigene Infrastrukturen zur Speicherung und ggf. auch Archivierung ihrer Daten (Beispiel: PSI).

- P3. **Standards:** Die Archivdienstleistung für Primärdaten orientiert sich an RFC 4810 «Long-Term Archive Service Requirements»<sup>30</sup> [23] der «Long-Term Archive and Notary Services Charter»<sup>31</sup> (LTANS).
- P4. **Anmeldung:** Forschende, welche die Archivdienstleistung nutzen möchten, beantragen ein Depot über ihre Forschungsinstitution, an der das Forschungsprojekt durchgeführt wird. (Diese wird in der Regel auch die Eigentümerin der Daten sein.) Die für den Zugang relevanten Benutzerdaten sind die in der von SWITCH betriebenen «Authentifizierungs- und Autorisierungs-Infrastruktur (AAI)»<sup>32</sup> aktuell hinterlegten, soweit die Forschungsinstitution daran angeschlossen ist. Beim Weggang des/der Forschenden aus der Schweiz ist der Zugang im Einzelfall zu regeln.
- P5. **Self-Deposit:** Die Archivdienstleistung für Primärdaten bietet Forschenden *mindestens* die Möglichkeit, ihre Daten in der Originalform selbstständig und effizient, über eine sichere Internetverbindung oder per Datenträger, der Archivdienstleistung zu übergeben, ohne dass Änderungen an den Daten nötig oder besondere Anforderungen an Formate erfüllt werden müssen. Eine sach- und fachgerechte Dokumentation der Daten sowie die Wahl von möglichst langzeit-tauglichen Datenformaten z.B. gemäss Best Practices des jeweiligen Fachgebiets, liegt in der Selbstverantwortung der Forschenden.
- P6. **Bewertungsmöglichkeit:** Die Archivdienstleistung für Primärdaten ermöglicht es der betreffenden Forschungsinstitution, vor der Deponierung der Daten und *beim Ablauf der Aufbewahrungszeit* die deponierten Daten durch wissenschaftliche Fachpersonen hinsichtlich ihres langfristigen oder sogar dauerhaften Wertes bewerten zu lassen.
- P7. **Beratungsmöglichkeiten:** Gemäss P5 basiert das Self-Deposit auf von den Forschenden selber bestimmten Formaten und Dokumentationen der Originaldaten. Beides kann möglicherweise für eine langfristige Aufbewahrung und Verständlichkeit ungenügend oder untauglich sein. Die Archivdienstleistung für Primärdaten bietet den Forschenden deshalb die Möglichkeit, sich vor der Deponierung der Daten bezüglich langfristig tauglicher Formate und Datendokumentationsweisen von Fachpersonen mit Erfahrung zur Datenaufbewahrung im gleichen Forschungsgebiet beraten zu lassen.
- P8. **«Bitstream Preservation»:** Die Archivdienstleistung für Primärdaten übernimmt im Auftrag der Forschenden während des vereinbarten Zeitraums *mindestens* die Verantwortung für die Verfügbarkeit und sichere Erhaltung des Bitstroms der ihr übergebenen Originaldaten. Dies schliesst auch den Schutz gegen unberechtigte Zugriffe und Katastrophen sowie die fachgerechte langfristige Speicherung mit ein.
- P9. **Integrität:** Die Archivdienstleistung für Primärdaten bietet Forschenden die Möglichkeit, die Integrität (d.h. die Unverändertheit) der von ihnen deponierten Daten nach dem Depot-Zeitpunkt jederzeit und rechtlich beweiskräftig nachweisen zu können. Dazu versieht die Archivdienstleistung die Daten bei der Deponierung automatisch mit

---

<sup>29</sup> Es ist denkbar, dass gewisse Regeln aufgestellt werden, aus denen sich bei bestimmten Projekttypen oder -größen für die Nutzung der Archivdienstleistung eine Verpflichtung ergibt, die Kosten einer langfristige Aufbewahrung der Primärdaten bereits bei der Beantragung des Forschungskredits zu berücksichtigen.

<sup>30</sup> <http://www.ietf.org/rfc/rfc4810.txt>

<sup>31</sup> <http://www.ietf.org/html.charters/ltans-charter.html>

<sup>32</sup> <http://www.switch.ch/aai/>

einem qualifizierten Zeitstempel gemäss dem schweizerischen Bundesgesetz über die elektronische Signatur ZertES<sup>33</sup> [26]. Sie sorgt zudem für eine rechtzeitige Erneuerung des Zeitstempels, d.h. bevor die verwendeten kryptografischen Verfahren in der Schweiz nicht mehr als sicher gelten, und ermöglicht dabei den Integritätsnachweis bis zum Zeitpunkt des ersten Zeitstempels.

- P10. Selbstverantwortlichkeit:** Die Archivdienstleistung für Primärdaten ermöglicht es den Forschenden, die von ihnen deponierten Daten jederzeit und ohne Angabe von Gründen wieder zu löschen oder durch eine neue Versionen zu ersetzen. Die Archivdienstleistung bewahrt keine Kopien gelöschter Daten auf, protokolliert aber den Löschvorgang und bewahrt dieses Protokoll dauerhaft auf.
- P11. Keine Rechteübertragung:** Die Archivdienstleistung für Primärdaten erwirbt kein Eigentum und keine Nutzungsrechte an den Daten, welche die Forschenden in ihr deponieren. Das Eigentum an den Daten sowie Nutzungsrechte, Urheberrecht, Copyrights etc. bleiben, soweit vorhanden, unverändert bei dem/der Forschenden bzw. der Forschungsinstitution, welche die Forschung finanziert hat. Die Hinterlegung erfolgt somit auch nicht fiduziarisch, sondern „treuhänderisch“<sup>34</sup>.
- P12. Haftungsausschluss:** Die Archivdienstleistung für Primärdaten führt ihre Aufgaben nach bestem Wissen und Gewissen und hohen Standards aus, haftet jedoch bei einem Verlust der deponierten Daten in keiner Weise für den Schaden, welche den Nutzer/innen der Archivdienstleistung oder Dritten möglicherweise daraus entsteht. Auch haftet sie nicht für den Inhalt der deponierten Daten.
- P13. Verzeichnung:** Forschende dokumentieren jedes Depot in einem standardisierten Verzeichniseintrag, der neben einer inhaltlichen Beschreibung der Daten und Dokumentation auch administrative Informationen enthält (z.B. Kontaktinformationen, Dateneigentümer, Aufbewahrungszeit, Nutzungs- und Urheberrechte, Zugriffseinschränkungen etc.).
- P14. Zitierbarkeit:** Die Archivdienstleistung für Primärdaten ermöglicht den Forschenden, die von ihnen deponierten Daten sofort nach der Deponierung in wissenschaftlichen Publikationen mittels eines Digital Object Identifiers (DOI) zu zitieren<sup>35</sup>, und zwar so, dass Dritte über Internet mindestens auf einen bibliografischen Eintrag (gemäss P13) sowie Informationen zu einer Kontaktperson des Forschungsprojekts zugreifen können.
- P15. Zugangsgarantie:** Die Archivdienstleistung für Primärdaten ermöglicht es Forschenden, jederzeit innert nützlicher Frist eine Kopie der von ihnen deponierten Daten zu beziehen, je nach Datenmenge über Internet oder per Versand eines Datenträgers.

<sup>33</sup> [http://www.admin.ch/ch/d/sr/c943\\_03.html](http://www.admin.ch/ch/d/sr/c943_03.html)

<sup>34</sup> Dies hat wesentliche Auswirkungen z.B. im Fall eines durch Forschungsinstitutionen gegen eine/n Forschende/n angestregten Disziplinarverfahrens wegen Verletzung der „Guten wissenschaftlichen Praxis: Die Archivdienstleistung gibt Daten nur mit Einwilligung des *tatsächlichen* Eigentümers der deponierten Daten heraus (was jedoch in der Regel dieselbe Forschungsinstitution sein wird, die das Verfahren anstrengt.)

<sup>35</sup> Hier können Lösungen des von der ETH-Bibliothek geführten Projekts «DOI-CH» (<http://www.e-lib.ch/doi.html>) genutzt werden, das bis 2011 für alle Schweizer Hochschulen und Hochschulbibliotheken eine zentrale Infrastruktur für die Vergabe, Registrierung und Verwaltung von Digital Object Identifiers (DOIs) für digitale Primär- und Sekundärdaten aufbaut.



- P16. **Institutionelle Verfügbarkeit:** Daten aus öffentlich finanzierter Forschung sind in der Regel Eigentum der Institution, welche die Forschung finanziert. Die Archivdienstleistung für Primärdaten ermöglicht es auf geeignete Weise, dass die teilnehmenden Institutionen (siehe P3) jederzeit ein Inventar der sie betreffenden deponierten Daten einsehen und nach definierten Regeln *lesend* auf die deponierten Daten zugreifen können (z.B. wenn die/der deponierende Forschende nicht mehr kontaktiert werden kann), sofern sie im Depot als Eigentümerin der Daten eingetragen sind.
- P17. **Sekundärnutzung:** Die Archivdienstleistung für Primärdaten ermöglicht es den Forschenden, einem Dritten den einmaligen lesenden Zugriff auf von ihnen deponierte Daten zu gewähren (je nach Datenmenge online über ein „One-time“-Passwort oder als schriftlicher Auftrag an die Archivdienstleistung, die Daten per Kurier an eine bestimmte Person zu liefern).
- P18. **Kein Zwang zu OpenAccess:** Die Archivdienstleistung diskriminiert keine Forschenden, welche ihre deponierten Daten *nicht* nach den Grundsätzen des OpenAccess für Dritte frei zugänglich machen wollen oder können (z.B. wegen Urheberrechten bei mit privaten Drittmitteln mitfinanzierter Forschung). Die Archivdienstleistung ermöglicht es den Forschenden aber, die von ihnen deponierten Daten für den Online-Download durch bzw. den Versand per Datenträger an Dritte sowohl generell wie auch selektiv (d.h. nur für einzelne Personen mit Passwort) freizuschalten (siehe P17).
- P19. **Freiwillige Katalogisierung:** Die Archivdienstleistung für Primärdaten ermöglicht es Forschenden zu wählen, dass für die von ihnen deponierten Daten die bibliografische Beschreibung (die bereits der Zitierbarkeit dienen, siehe P14 oben) in einem über Internet öffentlich zugänglichen Gesamtkatalog erscheint. Dieser Gesamtkatalog wird von der Archivdienstleistung unterhalten ist nach Fachgebieten sowie Forschungsinstitutionen strukturiert und nach Stichwörtern durchsuchbar. Ein lesender Zugriff auf die eigentlichen Daten zu einem Katalogeintrag ist nur möglich, wenn dieser Zugang vom Forschenden bei der Deponierung der Daten freigeschaltet wurde (siehe P18).
- P20. **Maximale Aufbewahrungsfrist:** Die Archivdienstleistung für Primärdaten ermöglicht es dem/der Forschenden bzw. der Forschungsinstitution, die Aufbewahrungsfrist, während der die Archivdienstleistung die deponierten Daten aufbewahren soll, selber zu bestimmen und jederzeit zu verlängern oder zu verkürzen. Sie beträgt aber maximal 15 Jahre. Zu einem angemessenen Zeitpunkt vor Ablauf der Aufbewahrungsfrist informiert die Archivdienstleistung die angegebene Kontaktperson und die betreffende Institutionen per eingeschriebenem Brief und E-Mail über den anstehenden Fristablauf und mögliche Handlungsoptionen.
- P21. **Automatische Löschung:** Die Archivdienstleistung für Primärdaten löscht deponierte Daten nach Ablauf der Aufbewahrungsfrist automatisch, wenn die vorher informierten Kontaktpersonen dies nicht durch eine Fristverlängerung vorher verhindert haben. Die Löschung wird protokolliert und die bibliografische Beschreibung (P14) und ein allfälliger Eintrag im Gesamtkatalog (P19) bleiben bestehen.
- P22. **Bewertungskompetenz:** Der endgültige Entscheid über eine *Verlängerung* der Aufbewahrungsfrist oder einer dauerhaften Archivierung liegt bei der Forschungsinstitution, über welche das Depot beantragt und die in P20 und P21 informiert wurde.

- P23. Dauerhafte Archivierung:** Eine *dauerhafte* Archivierung von deponierten Daten durch die Archivdienstleistung erfolgt nur auf Grund eines fachlichen Bewertungsgutachtens (P6 und P22). Sollen die deponierten Daten auf Grund der Bewertung dauerhaft mit der Archivdienstleistung archiviert werden, so wird dies in der Datenverwaltung der Archivdienstleistung eingetragen, und der Zugang zu den und die Verantwortung für die Daten wird aus der Abhängigkeit der Person des/der Forschenden gelöst und ganz der zugehörigen Forschungsinstitution übertragen. Diese entscheidet auch über eine allfällige Konversion und/oder nachträgliche Erschliessung der Daten zugunsten einer langzeit-tauglichen Aufbewahrungsform und Verständlichkeit sowie über einen allfälligen Transfer der Daten in ein Staatsarchiv. Dabei kann die Archivdienstleistung auf Wunsch der Institution beratend und/oder ausführend tätig werden. Solche Arbeiten sind jedoch kostendeckend zu vergüten. Die darauf folgende dauerhafte Aufbewahrung durch die Archivdienstleistung erfolgt dagegen kostenlos, wobei die Archivdienstleistung jedoch weiterhin keine Verantwortung für die dauerhafte Interpretierbarkeit und Verständlichkeit der Daten bezüglich deren Datenformaten und Dokumentation übernimmt.
- P24. Kompetenznetzwerk:** Die Archivdienstleistung für Primärdaten baut zur Erreichung der Ziele ein Netzwerk von fachspezifischen Kontaktpersonen aus Forschung und Archivstellen auf. Es besteht einerseits aus Personen, welche deponierte Daten im Rahmen von P6 auf einen dauerhaften Wert für die Sekundärnutzung bewerten können, entsprechend qualifiziert und möglichst von Forschungsinstitutionen oder -gesellschaften mandatiert sind. Es umfasst andererseits Personen, welche über Erfahrungen und Wissen zur Aufbewahrung und Langzeitarchivierung von Primärdaten in den einzelnen Forschungsbereichen haben oder solches Wissen aufbauen wollen und im Rahmen von P7 und P24 für die Archivdienstleistung beratend und unterstützend tätig sein können.

Obwohl das hier vorgeschlagene Modell primär darauf ausgerichtet ist, dass Forschende in der Schweiz die Richtlinien zur Aufbewahrung von Primärdaten im Rahmen der „Integrität der Forschung“ freiwillig, eigenverantwortlich, unabhängig vom Forschungsgebiet und mit einer zentralen Archivdienstleistung möglichst einfach und effizient erfüllen können, so bietet es doch einige Synergien und Schnittpunkte mit anderen Nutzenzielen. Dazu gehört die Zitierfähigkeit und erleichterte Sekundärnutzung von Primärdaten, die Identifikation und Aufbewahrung von dauerhaft wertvollen Daten sowie mögliche Kooperationen mit staatlichen Archiven des Bundes und der Kantone.

#### 5.1.3.1 Kosten und Finanzierung

Die Kosten für die Realisierung eines Modells gemäss den Prinzipien in 5.1.2 halten wir für eher gering. Nach einer Investition für die Detailspezifikation, die Beschaffung oder Entwicklung einer Software und die Realisierung der Infrastruktur fallen nur geringe Betriebskosten an. Die Archivdienstleistung gemäss P1 – P24 kann weitgehend ohne wissenschaftliches und archivfachliches Personal weitgehend automatisiert betrieben werden, und das Kompetenznetzwerk (P24) funktioniert auf Milizbasis.

Es bleiben langfristig im Wesentlichen Betriebskosten für Hard- und Software-Infrastruktur sowie die Kosten für Bereitstellung und Unterhalt der (wachsenden) Speicherkapazitäten. Eine Untersuchung zur Situation von Primärdaten in Deutschland aus dem Jahre 2004 [19] nennt für sprachwissenschaftliche, geisteswissenschaftliche und soziologische Institutionen insgesamt 100 GB pro Jahr und für die Gesamtmenge der für die Aufbewahrung relevanten Primärdaten aus naturwissenschaftlicher und sonstiger Forschung in Deutschland insgesamt rund 1000 – 2000 TB pro Jahr, mit „stark steigender Tendenz“.

Eine ähnliche Erhebung für die Schweiz gibt es nicht. Über den Daumen gepeilt, und unter Berücksichtigung des Alters der deutschen Studie, darf man für die Schweiz bei den für die Aufbewahrung relevanten Primärdaten von einer Gesamtmenge von 50 – 300 TB pro Jahr ausgehen, wenn wenige „Data Hot Spots“ wie z.B. Messdaten aus den Beschleunigerdetektoren des PSI und gewissen digitalen bildgebenden Anlagen nicht mitgerechnet werden.

Hier bietet sich an, drei Rechenzentren (z.B. der ETHZ, des PSI und das Swiss National Supercomputing Centre) als Partner und Träger zu gewinnen, die entsprechende Speicherkapazitäten zur Verfügung stellen, die dann in ein Speichernetzwerk zusammengefasst und an zentraler Stelle verwaltet werden<sup>36</sup>, wobei SWITCH als weiterer Partner die Konnektivität beisteuert.

Die Finanzierung des Modells gemäss P1 – P24 teilt sich in eine Investition (Detailkonzept, Pilotprojekt, Beschaffung/Entwicklung Software, Realisierung, Einführung) sowie laufende Betriebskosten. Für den Investitionsteil schätzen wir grob 3 Mio. CHF, exklusive der Einrichtung der initialen Speicher- und Kommunikations-Infrastruktur, welche durch die oben genannten vier Partner kostenlos zur Verfügung gestellt werden sollte. Diesen Betrag sehen wir als direkten, einmaligen Projektbetrag von Bund und Kantonen, den Organen der Forschungsförderung sowie von Stiftungen und möglicherweise privater Unternehmen der IT-Branche (Sponsoring).

Für die Deckung der laufenden Betriebskosten – die Archivdienstleistungen P1 – P24 sollen nach P2 für die Forschenden prinzipiell und für Forschungsprojekte grundsätzlich kostenlos sein –, sind nach unserer Meinung primär die Organe der Forschungsförderung und die Universitäten bzw. Hochschulkantone sowie für den ETH-Bereich der Bund<sup>37</sup> verantwortlich. Die Kosten könnten z.B. über einen Schlüssel aufgeteilt werden, der sich aus der Statistik der deponierten Datenmengen durch Forschende an den unterschiedlichen Institutionen oder in den unterschiedlichen Forschungsbereich errechnet.

Wie bereits erwähnt, dürften die externalisierten Betriebskosten jedoch gering ausfallen, wenn vor allem der Bereich der Speicherung und Datenkommunikation direkt und ohne Entgelt durch universitäre Partner übernommen wird.

---

<sup>36</sup> Dafür gibt es Standardlösungen, auch solche, die für Bildungsinstitutionen kostenlos sind, wie z.B. den «Storage Resource Broker SRB» des San Diego Supercomputer Center ([http://www.sdsc.edu/srb/index.php/Main\\_Page](http://www.sdsc.edu/srb/index.php/Main_Page))

<sup>37</sup> Wir weisen darauf hin, dass der Bundesrat ab 2010 mit 0.9 Mio. CHF und ab 2011 mit 1.6 Mio. CHF pro Jahr die beiden Datenzentren „Euro-Climhist“ (Universität Bern) und „World Glacier Monitoring Service“ (Universität Zürich) unterstützt, die unter anderem auch für die langfristige Archivierung der Messdaten verantwortlich sind. (Siehe dazu Fussnote 12 auf Seite 25.)

### 5.1.4 Erläuterungen

Das durch P1 – P24 charakterisierte Modell hat allerdings auch einige wesentliche Einschränkungen. Erstens handelt es sich um reines *Depositorium*, das z.B. mit „Swiss Scientific Data Deposit“ betitelt werden könnte, und nicht um ein *Repository*. Wörtlich übersetzt heisst Repositoryum „Büchergestell“ oder „Aktenschrank“, während Depositorium „Aufbewahrungsort“ oder „Hinterlegungsstelle“ bedeutet.

Ein Repositoryum ist somit darauf ausgelegt, dass damit Daten fachspezifisch und häufig genutzt sowie thematisch geordnet und die *Dateninhalte* selber durchsucht und mehrere Datenbestände für Auswertungen kombiniert werden können. Dagegen liegt das Gewicht bei einem Depositorium auf der Aufbewahrung/Archivierung und langfristigen Verfügbarkeit, der eher seltenen und nicht fachspezifischen Nutzung der Daten sowie einer eher homogenen, nicht-thematischen Ordnung und Verzeichnung. Auch müssen die Datensätze in der Regel zuerst aus dem Depositorium exportiert werden, bevor die Dateninhalte selber direkt durchsucht, kombiniert oder mit externen Tools ausgewertet werden können.

Die «Alliance for Permanent Access to the Records of Science in Europe» (Alliance PARSE, siehe Kapitel 8.4) geht in ihrer Strategie [24] ebenso wie die «Allianz-Initiative ‚Digitale Information‘» [25] der deutschen Forschungsgesellschaften (siehe Kapitel 8.5) davon aus, dass wissenschaftliche Repositorien (für Primärdaten oder Auswertungsdaten) immer fachspezifisch sind und von den jeweiligen Forschungsbereichen selber eingerichtet und betrieben werden. Eine nationale oder internationale Strategie muss sich deshalb primär mit der Förderung von fachspezifischen Repositorien sowie der Etablierung von Interoperabilitäts-Standards und der Koordination und Vernetzung der Repositorien innerhalb derselben Fachbereiche konzentrieren.

Diese Ansicht ist aus unserer Sicht richtig, weshalb wir auch kein Repositoryum vorschlagen, denn dieses wäre nur fachspezifisch. Ein fachübergreifendes Repositoryum würde heissen, den fachspezifischen Anforderungen aller Fachgebiete gerecht werden zu wollen, was erstens technisch unrealistisch wäre und zweitens den Aufbau von entsprechenden fachwissenschaftlichen Kenntnissen innerhalb der Archivdienstleistung bedingen würde. Dies scheint uns bei einer zentralen Basisdienstleistung mit nationalem Anspruch keine sinnvolle Aufgabe.

Im Rahmen der freiwilligen Eintragung in den Gesamtkatalog des Depositoriums (P19) können mehrere Forschende oder ein verteiltes Projektteam sich darauf einigen, diese Funktion zu nutzen, um über einen bibliografisch identifizierbaren Bereich systematisch fachspezifische „virtuelle“ Depositorien ihrer Primärdaten zu bilden (wobei die Dateninhalte deswegen trotzdem nicht direkt durchsuchbar und nutzbar sind). Dies reicht aus unserer Sicht.

Die zweite Einschränkung des durch P1 – P24 vorgeschlagenen Depositorium-Modells ist weitaus schwerwiegender und prinzipieller Natur: Gemäss dem Self-Deposit-Prinzip (P5) können die Forschenden die Daten in einer von ihnen selber gewählten Form (Datenformate und Dokumentation) deponieren. Zwar wird ihnen die Verwendung von für die Langzeitaufbewahrung geeigneten Standardformaten empfohlen und über P7 eine entsprechende Beratungsmöglichkeit angeboten, doch beides wird der Eigeninitiative überlassen. Dies ist unserer Meinung nach aus folgenden Gründen notwendig:

- Aufbewahrungspflichten zur „Integrität der Forschung“ beziehen sich auf die originalen Primärdaten (bei der ETHZ sogar auf die «nicht weiter bearbeiteten Originaldaten» [14]), die Grundlage der Auswertung bzw. der publizierten Forschungsergebnisse sind. Eine Konvertierung der Daten und Datenformate kann mit einem Verlust an Information (z.B. Datentypen) und Originalität verbunden sein und ist deshalb heikel.
- Eine Konvertierung der Daten in für die Langzeitaufbewahrung geeignete Standardformate als Voraussetzung für die Nutzung der Archivdienstleistung würde für die Forschenden einen erheblichen Aufwand bedeuten und damit die Hemmschwelle zur Nutzung des Angebots stark erhöhen.
- In vielen Forschungsgebieten existieren gar keine speziell für die Langzeitaufbewahrung geeigneten Standardformate<sup>38</sup>.
- Wenn die Archivdienstleistung für die Langzeitaufbewahrung geeignete Standardformate als Voraussetzung fordert statt nur empfehlen würde, so müsste sie die Daten auch auf Einhaltung dieser Anforderung prüfen, was einen grossen manuellen Aufwand bedeuten würde.

Aus diesem Grund ist mit P3 auch „nur“ die Konformität mit RFC 4810 «Long-Term Archive Service Requirements» (2007, [23]) vorgesehen, nicht aber mit ISO 14721:2003 «Open Archival Information System Reference Model» (OAIS) [27]: Der OAIS-Standard enthält vier wesentliche Anforderungen („OAIS Responsibilities“), damit eine Lösung OAIS-konform sein kann [27]:

- Mit der Übernahme der Daten übernimmt das OAIS-Archiv auch die volle Verantwortung für die langfristige Verfügbarkeit und Benutzbarkeit der Daten für seine Kunden. Dies schliesst auch die regelmässige Konversion und Migration der Informationen in neue Formate ein, falls die benutzten Formate obsolet werden zu drohen. Das OAIS-Archiv muss also das Recht haben, die format- und strukturbezogenen Charakteristika der Daten nach eigenem Ermessen zu verändern.
- Das OAIS-Archiv muss bis zu dem Grade die technische und organisatorische Kontrolle über die Daten erlangen, die eine erfolgreiche Langzeitarchivierung garantiert.
- Das OAIS-Archiv muss fähig sein, die übernommenen Daten selber auch unabhängig von den ursprünglichen Datenproduzenten verstehen können.
- Das OAIS-Archiv muss mit den Datenproduzenten über die Kriterien der Übernahme (inkl. Datenformate und -strukturierung und Dokumentation) verhandeln können.

Mit den Zielen P1 – P24 kann die Archivdienstleistung keine dieser vier Anforderungen erfüllen, und realistischere sind diese in der Realität auch nicht zu erfüllen, solange die Archivdienstleistung leicht benutzbar sein, für alle Forschenden effizient nutzbar und primär dem Zweck der Aufbewahrung im Rahmen der „Integrität der Forschung“ dienen soll.

---

<sup>38</sup> Eine Erhebung in Deutschland ergab [17], dass umgerechnet auf die produzierten Datenvolumina 97,8% der Daten in binären, fachspezifischen Formaten gespeichert werden, nur 0,3% in XML und 1,9% in reinen Textformaten. Die fachspezifischen binären Formate sind zwar in der Regel offen dokumentiert. Trotzdem ist bei den strukturierten Forschungsdaten in binären Formaten eine Konversion dieser Formate in einheitliche Standardformate, wie es zum Beispiel bei der Langzeitarchivierung von dokumentenzentrierten Unterlagen üblich ist, in der Regel illusorisch.

### 5.1.5 OAIS-Erweiterung der operativen Ziele

Das Modell kann bezüglich der im Abschnitt 5.1.4 oben diskutierten Einschränkungen – Depositorium statt Repositorium sowie fehlende OAIS-Konformität – in einem gewissen Rahmen erweitert werden, was uns allerdings ausschliesslich für die gemäss P23 dauerhaft aufzubewahrenden Daten sinnvoll erscheint:

- P25. Langzeitverantwortung:** Die Archivdienstleistung für Primärdaten *kann* für die gemäss P22 als dauerhaft aufbewahrungswürdig beurteilte Daten statt der in P8 vorgesehenen „Bitstream Preservation“ und der in P23 angebotenen, rein unterstützenden Tätigkeit auch eine umfassende Verantwortung für die Daten übernehmen. Dies erfolgt im Auftrag der betreffenden Forschungsinstitution oder einer staatlichen Archivstelle. Dazu erstellen die Archivdienstleistung und die Institution eine Archivierungsvereinbarung und bereiten die Daten dementsprechend und gemeinsam in eine für die dauerhafte Archivierung taugliche Form auf. Kommt keine Vereinbarung zustande, kann die Archivdienstleistung die Übernahme der Daten ablehnen. Eine Vereinbarung mit einzelnen Forschenden oder Forschungsprojekten ist nicht möglich.
- P26. Kostenpflicht:** Für die dauerhafte Archivierung von Daten im Sinne von P25 (und nur für diese) hat die beauftragende Institution einen angemessenen Beitrag an die Kosten zur Aufbereitung sowie einen kostendeckenden jährlichen Beitrag für die dauerhafte Archivierung zu bezahlen. Wird der jährliche Beitrag nicht mehr bezahlt, muss die Institution die Daten zurücknehmen.
- P27. «Preservation Levels»:** Zur Realisierung des Angebots P25 kann die Archivdienstleistung mehrere Erfüllungsgrade (Service Level Agreement oder Preservation Levels) definieren, in denen die von beiden Seiten jeweils zu erfüllenden Anforderungen an die Daten und die Datenhaltung detailliert definiert sind.
- P28. OAIS-Konformität:** Die zur Erfüllung der Aufgabe P25 (und nur für diese) nötigen organisatorischen und technischen Komponenten der Archivdienstleistung genügen den Grundsätzen und Anforderungen des internationalen Standards ISO 14721:2003 «Open Archival Information System Reference Model» (OAIS) [27] sowie dem «nestor Kriterienkatalog vertrauenswürdige digitale Langzeitarchive»<sup>39</sup> [28] oder der «Trustworthy Repositories Audit & Certification: Criteria and Checklist»<sup>40</sup> [29] des «Center for Research Libraries» (CRL). Wenn möglich soll die Archivdienstleistung für diesen Bereich über eine entsprechende Zertifizierung verfügen.
- P29. Fachkompetenz:** Die Archivdienstleistung für Primärdaten kann die Schweiz in internationalen Gremien und Initiativen zur Langzeitarchivierung von wissenschaftlichen Primärdaten vertreten, an entsprechenden Forschungsprojekten teilnehmen und eigene Projekte initiieren.
- P30. Wissenstransfer:** Die Archivdienstleistung für Primärdaten soll mit Forschenden aller Forschungsbereiche, die in der Schweiz eigene Repositorien oder Depositorien mit wissenschaftlichen Primärdaten betreiben oder aufbauen, den Wissensaustausch pflegen und bezüglich der Anforderungen der Langzeitaufbewahrung beratend wirken.

<sup>39</sup> <http://nbn-resolving.de/urn:nbn:de:0008-2006060710>

<sup>40</sup> <http://www.crl.edu/PDF/trac.pdf>

### 5.1.5.1 Kosten und Finanzierung

Die Dienstleistungen P25 – P30 sind mit wesentlich mehr personellen und fachlichen Aufwand verbunden als die eher rein infrastrukturellen in P1 – P24. Die Übernahme der Verantwortung für den Erfolg einer dauerhaften Archivierung (P25) muss für die beauftragende Institution jährlich kostenpflichtig und für die Archivdienstleistung dauerhaft kostendeckend finanziert sein, weil sonst die Beauftragung leichtfertig erfolgt und die Kostenentwicklung dieser Dienstleistung langfristig nicht kontrolliert werden kann.

Eine Schätzung der Kosten für die Erweiterung der Ziele um P25 – P30 ist ohne genaue Definition der Preservation Levels in P27 und ohne Kenntnis der quantitativen Nachfrage nach einer solchen „Vollarchivierung“ nicht möglich.

### 5.1.6 Übersicht

Abbildung 1 zeigt die Leistungskomponenten dieses Modells gemäss den in den Abschnitten 5.1.3 und 5.1.5 beschriebenen operativen Zielen.

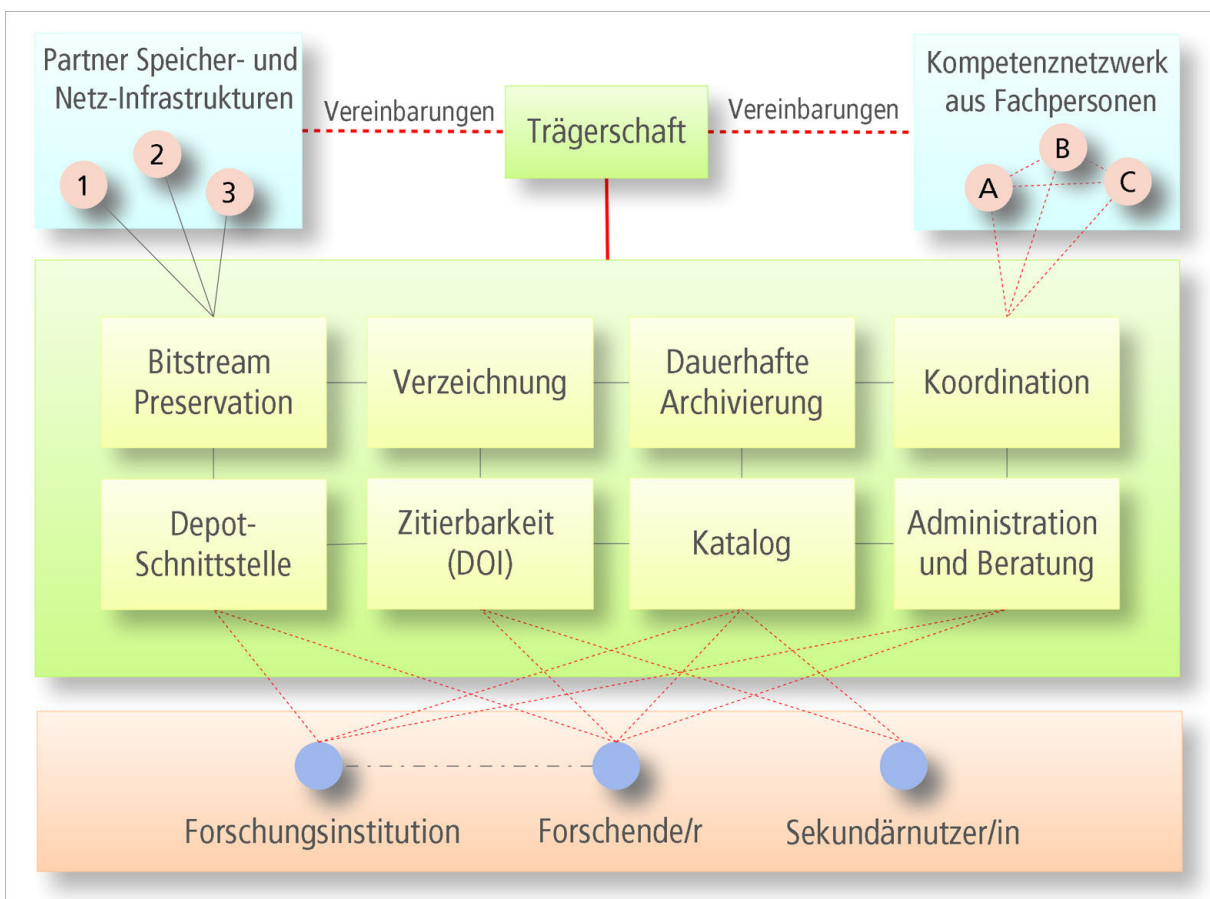


Abbildung 1: Leistungskomponenten im Modell «Aufbewahrung mit Selbstregulierung»

- **Trägerschaft:** Sie definiert und kontrolliert den Leistungsauftrag, sichert die Finanzierung der Archivdienstleistung und schliesst Vereinbarungen mit den Infra-

strukturpartnern und den Fachexperten des Kompetenznetzwerks ab. Die Trägerschaft muss gut und breit abgestützt sein, um langfristig Bestand haben zu können. Als Organisationsform kommt zum Beispiel eine Stiftung des Bundes, der Universitäten oder der Organe der Forschungsförderung in Frage. Dies würde der langfristig angelegten Dienstleistung die nötige Vertrauenswürdigkeit verleihen.

- **Infrastrukturpartner:** Sie stellen kostenlos oder gegen Entschädigung Netzwerk- und Speicherressourcen zur Verfügung.
- **Bitstream Preservation:** Verwaltung des Speichernetzwerks und Speicher-managements, sichere physische Langzeitspeicherung, Qualitätsüberwachung und dauerhafte Sicherstellung der ungebrochenen Nachweiskette für die Integrität der Daten durch qualifizierte elektronische Zeitstempel gemäss Signaturgesetz.
- **Depot-Schnittstelle:** Funktionalität für die Datenübergabe und Verwaltung der Depots.
- **Verzeichnung:** Standardisierte Verzeichnung der Datendepots, sowohl inhaltlich (z.B. über DublinCore) wie auch administrativ (Aufbewahrungszeit, Dateneigentümer, Nutzungs- und Urheberrechte, Zugriffsbedingungen etc.)
- **Zitierbarkeit (DOI):** Erzeugung von Digital Object Identifier (DOI) für eindeutige Identifizierung und Zitierbarkeit der Daten in Publikationen. Dazu können ab 2009 die Lösungen des Projektes «DOI-CH»<sup>41</sup> genutzt werden. Über einen entsprechenden DOI Resolving Service ermöglicht die Archivdienstleistung Dritten den Internet-Zugriff auf die öffentlich zugänglichen Teile des betreffenden Depots.
- **Katalog:** Öffentlicher Katalog, in dem Forschende freiwillig Verzeichnungseinträge zu ihren Depots publizieren und gegebenenfalls auch Daten zugänglich machen können.
- **Kompetenznetzwerk:** Fachpersonen, die über Erfahrung und Wissen zu langfristigen Aufbewahrung von Forschungsdaten in ihren Forschungsbereichen verfügen und andere Forschende diesbezüglich beraten und unterstützen können. Ausserdem Fachpersonen, die den langfristigen Wert von Primärdaten für die Sekundärnutzung in ihren Forschungsbereichen bewerten und entsprechende Gutachten erstellen können.
- **Koordination:** Koordination des Kompetenznetzwerks; Suche, Motivation und Betreuung von Fachpersonen des Netzwerks; Vermittlung von Fachpersonen des Netzwerks an Forschende, die Unterstützung suchen, oder Forschungsinstitutionen, die Daten auf ihren langfristigen Sekundärnutzen bewerten lassen wollen; Organisation von Fort-bildungsveranstaltungen; Redaktion eines Wissensportals zur Langzeitarchivierung etc.
- **Administration und Beratung:** Kontakt zu teilnehmenden Forschungsinstitutionen und Forschenden mit Depots; Einrichtung und Verwaltung der Depots; Beratung von Forschenden in grundsätzlichen Fragen der Langzeitaufbewahrung von Primärdaten.
- **Dauerhafte Archivierung:** Übernahme und Aufbereitung von Depots mit dauerhaft wertvollen Primärdaten. Entsprechende Zusammenarbeit mit Forschungsinstitutionen, Dateneigentümern, Forschenden und Fachpersonen des Kompetenznetzwerks. Danach Übernahme der Verantwortung für die dauerhafte Archivierung der Daten.

<sup>41</sup> Das von der ETH-Bibliothek geführte Projekt «DOI-CH» (<http://www.e-lib.ch/doi.html>) ist daran, bis 2009 für alle Schweizer Hochschulen und Hochschulbibliotheken eine zentrale Infrastruktur für die Vergabe, Registrierung und Verwaltung von Digital Object Identifiers (DOI) für digitale Primär- und Sekundärdaten aufzubauen.



Die folgende Abbildung 2 zeigt die Abläufe in diesem Modell gemäss den in den Abschnitten 5.1.3 und 5.1.5 beschriebenen operativen Zielen. Unterschieden werden vier Anwendungsfälle:

- **1. Deponierung** (Rot): Die/der für die Daten verantwortliche Forschende (Projektleiter/in) beantragt bei ihrer/seiner Forschungsinstitution die Einrichtung eines neuen Depots. Der/die Forschende erstellt einen standardisierten Verzeichniseintrag, in dem die Daten und ihre Herkunft beschrieben werden. Für die Aufbereitung der Daten für die Aufbewahrung können seitens einer Fachperson des Kompetenznetzwerks fachspezifische Empfehlungen erfolgen, wenn entweder der/die Forschende oder die Forschungsinstitution dies wünscht. Auf Wunsch des/der Forschenden kann eine Fachperson des Kompetenznetzwerks Unterstützung bei der Aufbereitung der Daten bieten. Die/der Forschende übergibt die aufbereiteten Daten (inkl. Dokumentation) der Archivdienstleistung (per Internet-Upload oder Datenträger) und bestimmt die Aufbewahrungsfrist sowie weitere Modalitäten des Zugriffs (z.B. Nutzungslizenz, allfällige Urheberrechte etc.) und der Administration des Depots. Die Archivdienstleistung lagert danach die Daten ins Langzeitarchiv ein und erzeugt einen qualifizierten Zeitstempel und einen Digital Object Identifier (DOI) zur Zitierbarkeit und publiziert optional den Verzeichniseintrag in einem öffentlichen Gesamtkatalog.
- **2. Zugriff** (Grün): Auf Grund des in einer Publikation zitierten DOI oder nach einer Suche im Katalog möchte eine Person Daten zur Sekundärnutzung beziehen. Hat der/die deponierende Forschende für die betreffenden Daten den freien Zugriff ermöglicht, können die Daten direkt bezogen werden (Download oder Versand Datenträger). Andernfalls erhält die deponierende Person eine entsprechende Anfrage per E-Mail (oder auf andere Weise gemäss der im Verzeichniseintrag zum DOI hinterlegten Informationen) und kann den Zugriff durch Setzen eines Einmal-Passworts für den Download oder durch Erteilen eines Auftrags zum Versenden eines Datenträgers ermöglichen.
- **3. Inventar der Institution** (Grün): Die an der Archivdienstleistungen teilnehmende Forschungsinstitution kann jederzeit Einsicht in das Inventar bzw. die Verzeichniseinträge und Statusinformationen der von Ihnen genehmigten Depots (siehe Anwendungsfall 1 oben) nehmen. Sie kann auch Einsicht in die Daten nehmen, sofern sie im Depot als Eigentümerin der Daten eingetragen ist.
- **4. Ablauf der Aufbewahrungszeit** (Blau): Ist die gesetzte Aufbewahrungszeit des Depots abgelaufen, entscheidet die Forschungsinstitution bzw. die Eigentümerin der Daten nach Rücksprache mit dem Depot-Inhaber, ob die Daten gelöscht, die Aufbewahrungszeit verlängert oder die Daten dauerhaft archiviert werden sollen. Sie kann dazu den Wert der Daten für die Sekundärnutzung durch eine Fachperson des Kompetenznetzwerks bewerten lassen. Bei einer dauerhaften Archivierung bereitet die Archivdienstleistung auf Kosten der Forschungsinstitution bzw. der Daten-Eigentümerin sowie mit Unterstützung des Depot-Inhabers und der Personen des Kompetenznetzwerkes die Daten für die Langzeitarchivierung auf und übernimmt danach die Verantwortung für die dauerhafte Archivierung der Daten.

## 5 Modelle für Primärdaten

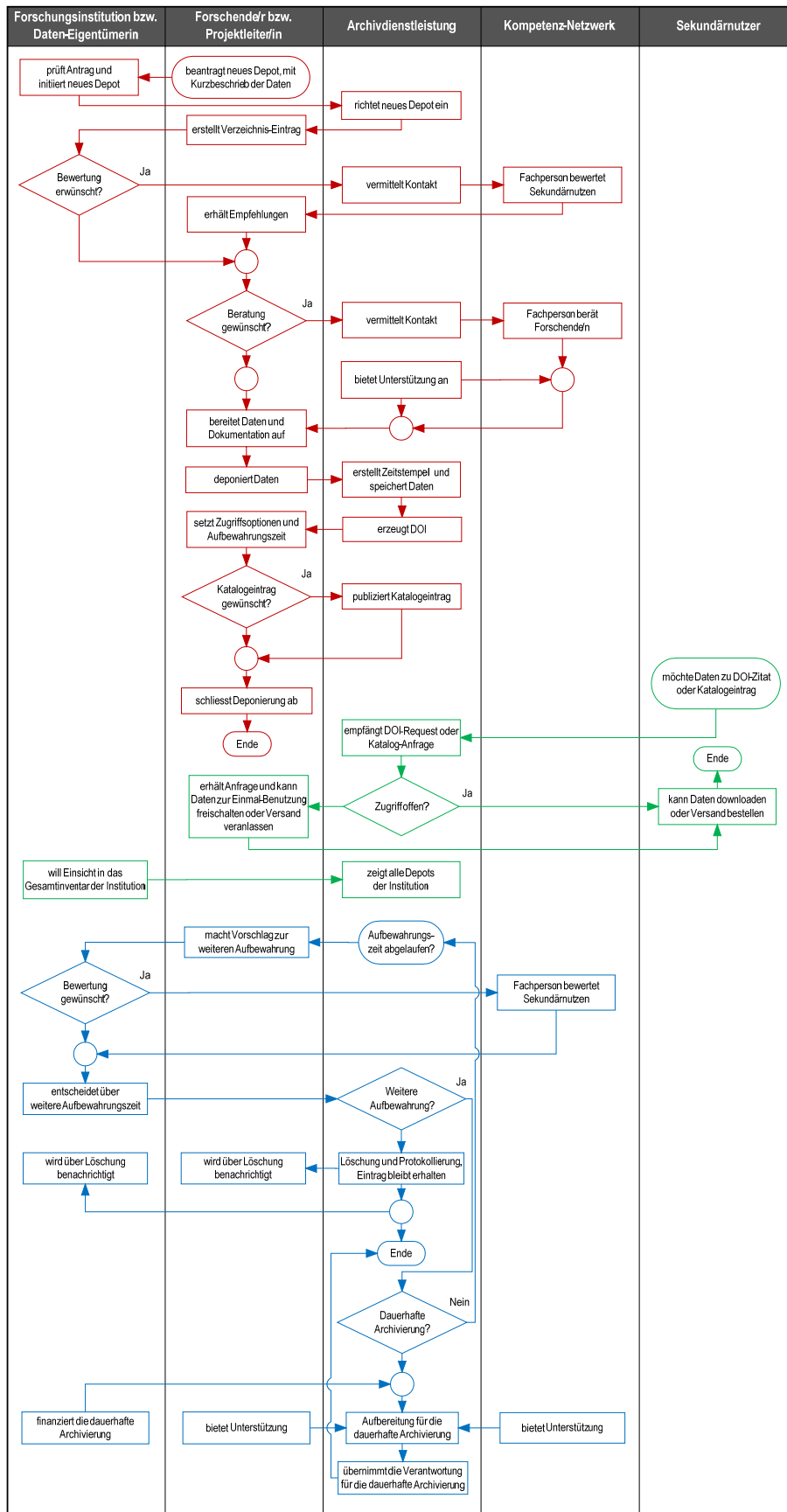


Abbildung 2: Abläufe im Modell «Aufbewahrung mit Selbstregulierung».

## 5.2 Modell « Archivierung mit Anbietepflicht »

Dieses zweite Modell für Primärdaten verfolgt im Gegensatz zum ersten (siehe Kapitel 5.1, Modell «Aufbewahrung mit Selbstregulierung») einen regulierenden Ansatz, der von der generellen Anbietepflicht für Forschungsunterlagen auf Grund der staatlichen Archivierungsgesetze und einer nationalen Archivdienstleistung für Primärdaten ausgeht. In diesem Modell werden alle Primärdaten und die dazugehörige Dokumentation systematisch einer Beurteilung der Archivwürdigkeit zugeführt.

Die Umsetzung der systematischen Anbietepflicht und der archivischen Bewertung setzt dort an, wo Forschungsprojekte ohnehin von Fachgremien beurteilt werden: bei der Bewilligung bzw. Finanzierung und bei der Begutachtung bzw. Evaluation der Ergebnisse sowie beim Reporting bzw. beim Abschlussbericht. Die jeweiligen Gutachter werden angehalten, die Archivwürdigkeit (entweder zeitlich beschränkt oder dauernd) aus der Sicht der Fachdisziplin zu beurteilen. Die Forschenden selber beurteilen den langfristigen Wert der Daten, die sie produzieren, sowohl im Projektantrag wie auch im Abschlussbericht.

Diese Vorschläge werden zusammen mit allfälligen Archivierungsverpflichtungen, welche der Schweizerische Nationalfonds an die Finanzierung geknüpft hat, systematisch der jeweils zuständigen Archivstelle zugeleitet, welche darüber entscheidet bzw. ihren Entscheid gegebenenfalls mit den Gutachtern und Forschenden bereinigt. Bestimmte Kategorien von Forschungen und Forschungsprojekten, bspw. studentische Forschungsarbeiten bis zur Stufe Diplom- oder Masterarbeiten können über prospektive Bewertungsentscheide von der Anbietepflicht ausgenommen werden.

Hauptsächliche Kriterien zur Beurteilung der Archivwürdigkeit sind das Potential der Primärdaten für eine Sekundärauswertung und die Nachvollziehbarkeit der durchgeführten Forschung, insbesondere dort, wo in einem ethisch und politisch heiklen Gebiet geforscht wurde (z.Bsp. Forschungen am Menschen), wo ein beträchtliches Potential besteht, dass die Forschungen entweder grosse Auswirkungen auf Mensch und Umwelt (z.Bsp. Freisetzungsversuche) haben oder Forschungsergebnisse politische oder gesetzgeberische Entscheidungsprozesse beeinflussen. Bei dieser Gelegenheit werden auch die sonstigen Forschungsunterlagen des Projekts archivisch bewertet und ggf. deren Ablieferung an die zuständige Archivstelle geregelt.

Für Primärdaten, die nach dem Bewertungsentscheid der zuständigen staatlichen Archivstelle als *nicht* archivwürdig gelten, stellt die nationale Archivdienstleistung ein Angebot gemäss dem ersten Modell (Kapitel 5.1) bereit.

Neben der Anbietepflicht und systematischen archivischen Bewertung stellt die systematische Aufbereitung der Daten zum Zwecke der Archivierung einen zweiten wesentlichen Unterschied zum ersten Modell (Kapitel 5.1) dar: Archivierung unter dem Titel der staatlichen Archivgesetze erfolgt dauerhaft und muss den Kriterien der staatlichen Archive bezüglich Benutzbarkeit sowie Ordnung und Verzeichnung genügen, da diese Archive ihren gesetzlichen Auftrag zu erfüllen haben, für die fach- und sachgerechte Archivierung zu sorgen.

Die Wahl der Datenformate und die Form der Erschliessung können somit nicht wie im ersten Modell alleine den Forschenden überlassen werden. Vielmehr muss sichergestellt sein, dass die verwendeten Formate langzeittauglich sind, damit sie auch nach langen Zeiträumen noch lesbar und interpretierbar sind oder wenn nötig in der Zukunft in neue Formate konvertiert werden können. Dies bedeutet unter anderem, dass die Formate vollständig und offen dokumentiert sein müssen und nicht von einer spezifischen Software abhängig sind, die nach einigen Jahren unter Umständen nicht mehr verfügbar ist.

In der Regel werden die originalen Primärdaten diese Anforderungen nicht erfüllen, müssen also für die Langzeitarchivierung entsprechend aufbereitet und z.B. in offen dokumentierte Standardformate konvertiert werden. Es ist vor allem bei komplexen Binärformaten wahrscheinlich, dass eine solche Konversion zu Verlusten an Originalität und Information (z.B. Präzision, Datentypen, Struktur etc.) führt, die aus der Sicht der Archivierung zwar akzeptabel sind, jedoch das exakte Überprüfen der Forschungsergebnisse, die aus den Primärdaten in ihrer *ursprünglichen* Form gewonnen wurden, verunmöglichen.

Eine Erhebung in Deutschland ergab [4], dass umgerechnet auf die produzierten Datenvolumina 97,8% der Primärdaten in binären, fachspezifischen Formaten gespeichert werden, nur 0,3% in XML und 1,9% in reinen Textformaten. In einer Umfrage der ETH Zürich [3] gaben 49% der Befragten an, dass die aus ihrer Sicht *dauerhaft* zu archivierenden Primärdaten in ihrem Fachbereich *nicht* in einem Standardformat vorliegen, «*das in ihrem Fach oder allgemein anerkannt und gut dokumentiert ist*».

Die Vielfalt an Datenformaten für strukturierte Primärdaten – Formate von proprietärer Software, produktspezifische Formate zu Messinstrumenten sowie von den einzelnen Forschenden selber definierte Formate – ist bereits innerhalb eines einzelnen Forschungsbereichs weitaus grösser als im generischen Bereich der Dokumentformate.

Weder für die nationale Archivdienstleistung noch für die staatlichen Archivstellen ist es möglich, zu jedem Forschungsbereich archivtaugliche Standardformate zu identifizieren und zu prüfen sowie Richtlinien für die Konversion der grossen Vielfalt proprietärer Formate in diese Standardformate zu definieren.

Das Modell sieht deshalb vor, dass die Trägerschaft der Archivdienstleistung in Zusammenarbeit mit den einzelnen Wissenschaftsakademien in den einzelnen Fachbereichen Archivkommissionen rekrutiert und mandatiert, welche die archivtauglichen Standardformate für den jeweiligen Fachbereich identifizieren oder definieren. Die von diesen fachspezifischen Archivkommissionen definierten Standards werden durch die staatlichen Archivstellen akzeptiert und gelten damit als verbindliche Vorgaben für die Aufbereitung archivwürdiger Primärdaten.

Archivwürdige Primärdaten werden von der nationalen Archivdienstleistung für Primärdaten in Zusammenarbeit mit den Forschenden, der Forschungsinstitution sowie der zuständigen staatlichen Archivstelle gemäss den Standards der Archivkommissionen für die dauerhafte Archivierung aufbereitet und von der nationalen Archivdienstleistung *im Auftrag der staatlichen Archivstelle* übernommen und dauerhaft betreut.

Die Kosten für diese Aufbereitung werden je nach den gesetzlichen Grundlagen entweder von der anbietepflichtigen Stelle (Forschungsinstitution) oder dem staatlichen Archiv übernommen.

Ist bereits bei der Bewilligung eines Forschungsprojekts klar, dass die Daten langfristig archiviert werden müssen, bspw. wenn der Schweizerische Nationalfonds dies in der Zuspracheverfügung verlangt, müssen diese Kosten im Forschungsbudget mitberücksichtigt werden.

Müssen archivwürdige Primärdaten bei der Aufbereitung für die fachgerechte Langzeitarchivierung so verändert (z.B. konvertiert) werden, dass den betroffenen Forschenden dadurch die Aufbewahrungspflicht im Rahmen der „Integrität der Forschung“ (siehe Kapitel 3.1.3) nicht geboten werden kann, so werden die Primärdaten *zusätzlich* auch in der Originalform nach dem ersten Modell (siehe Kapitel 5.1) aufbewahrt.

Dieses Modell orientiert sich (wie bereits die Erweiterung des ersten Modells in Abschnitt 5.1.5) bei dauerhaft zu archivierenden Primärdaten an ISO 14721:2003 «Open Archival Information System Reference Model» (OAIS) [27]. Die fachlich-technische Gesamtverantwortung für die erfolgreiche Langzeit-Archivierung der als archivwürdig bestimmten Primärdaten tragen die nationale Archivdienstleistung und die von der Trägerschaft eingesetzten fachbereichsspezifischen Archivkommissionen gemeinsam (und sorgen für das notwendige fachbereichsspezifische Knowhow).

### 5.2.1 Strategieziele

Das in diesem Unterkapitel dargelegte Modell folgt den nachstehenden strategischen Zielen:

- H1. **Systematisch:** Die Archivdienstleistung baut auf einer systematischen Umsetzung der Anbietepflicht und archivischen Bewertung auf, die in den jeweiligen staatlichen Archivgesetzen und -verordnungen von Bund bzw. Kantonen festgelegt ist. Die Verantwortung für die Umsetzung der Anbietepflicht soll jedoch nicht zentral, sondern lokal, d.h. innerhalb des Rechtskreises und innerhalb der Verfügungsgewalt der einzelnen Forschungsinstitution wahrgenommen werden.
- H2. **National:** Die Archivierungsdienstleistung für Primärdaten soll grundsätzlich allen für die gesetzliche Archivierung zuständigen Stellen sowie den Hochschulen und Forschungsanstalten des Bundes und der Kantone zur Verfügung stehen.
- H3. **Zentral:** Die Archivdienstleistung soll, mit Ausnahme der Anbietepflicht und archivischen Bewertung (siehe H1), zentral erbracht werden, jedoch in ein enges Netzwerk von fachbereichsspezifischen Archivkommissionen eingebunden sein. Die Zentralisierung soll Synergien bei Infrastruktur und Knowhow ermöglichen.
- H4. **Ergänzend:** Die Archivdienstleistung soll für die im Sinne der staatlichen Archivgesetze als archivwürdig bezeichneten Primärdaten dienen und betrifft deshalb nur eine Teilmenge aller produzierten Forschungsunterlagen. Für die nicht archivwürdigen Primärdaten soll zusätzlich eine Lösung für die Aufbewahrung bis maximal 10 Jahre nach dem in Kapitel 5.1. dargestellten Modell angeboten werden. Die Archivdienstleistung soll in einzelnen Fachbereichen bestehende Datenarchivierungslösungen<sup>42</sup> nicht ersetzen, sofern diese den allgemein anerkannten Grundsätzen der Langzeitarchivierung folgen.

<sup>42</sup> Als Beispiel sei der Schweizerische Informations- und Daten-Archivdienst für die Sozialwissenschaften (SIDOS) erwähnt.

Auch in diesen Fällen sollen diese Lösungen jedoch in die systematische Umsetzung der Anbietepflicht und der archivischen Bewertung einbezogen werden.

- H5. **Forschungspolitisch:** Investitionen in die Produktion von Primärdaten sollen geschützt und Dritten für die Sekundärnutzung langfristig zur Verfügung gestellt werden, sofern nicht wichtige Schutzgründe und Eigentumsvorbehalte dagegen sprechen.

### 5.2.2 Begründung

Dieses Modell schafft die systematischen Voraussetzungen zur Umsetzung der Archivierungsgesetze auch auf Primärdaten der Forschung. Für im Sinne dieser Gesetze nicht archivwürdige Primärdaten kommt das erste Modell (Abschnitt 5.1) zur Anwendung, damit Forschende auch in diesen Fällen die Aufbewahrungsvorschriften im Rahmen der wissenschaftlichen Integrität effizient umsetzen können.

Dieses Modell entspricht insbesondere auch den Erwartungen der Öffentlichkeit, dass Forschung in ethisch und politisch heiklen Gebieten und dort, wo die Forschungen selber grosse Auswirkungen auf Mensch und Umwelt haben oder den politischen und gesetzgeberischen Prozess beeinflussen können, längerfristig oder dauerhaft nachvollziehbar bleibt.

Die strategischen Ziele bauen darauf auf, dass die Anforderungen an eine geordnete und systematische Aktenführung erfüllt werden und dass für als nicht archivwürdig beurteilte Unterlagen Aufbewahrungsmöglichkeiten existieren, mit welchen die Forschenden die Anforderungen der Reglemente zur Integrität der Forschung erfüllen können.

Neben der Forderung nach Nachvollziehbarkeit und nach Integrität von Forschungsarbeiten wird von Forschungsorganisationen auch die Förderung der Sekundärnutzung von wichtigen Primärdaten verlangt. In einem Projekt erhobene Primärdaten werden oft nicht in vollem Umfang ausgewertet und bieten Ausgangsmaterial für zusätzliche Forschungen. Die Archivierung bietet hier auch einen entsprechenden Investitionsschutz für Primärdaten.

### 5.2.3 Operative Ziele

Das hier vorgeschlagene Modell deckt nur den klassischen Archivierungsprozess ab, d.h. die Bewertung, Aufbereitung, Aufbewahrung und Vermittlung von archivwürdigen Primärdaten.

- Q1. **Institutionalisierung:** Die Archivdienstleistung wird durch eine eigene Organisationseinheit erbracht, die administrativ einer Institution der schweizerischen Hochschul- und Forschungslandschaft angegliedert ist. Die strategische und fachliche Aufsicht wird durch eine Trägerschaft ausgeübt.
- Q2. **Trägerschaft:** Die Trägerschaft besteht aus Hochschulen, Forschungsanstalten, Hochschulbibliotheken und den gesetzlich zuständigen Archiven sowie den Forschungsförderungsorganen zusammen. Die Trägerschaft wird in einer Vereinbarung zwischen den beteiligten Stellen bzw. den ihnen vorgesetzten Behörden gegründet. Die Trägerschaft delegiert ihre Kompetenzen an einen Ausschuss, der die Aufgaben und die Verantwortung der Trägerschaft gegenüber der nationalen Archivdienstleistung wahrnimmt.

- Q3. **Archivkommissionen:** Die Trägerschaft bestimmt zusammen mit den jeweiligen Wissenschaftsakademien und ev. anderen Organisationen des Fachbereichs die fachbereichsspezifischen Archivkommissionen, welche für ihre Fachbereiche Richtlinien und Standards für die Langzeitarchivierung von Primärdaten aus dem betreffenden Forschungsbereich definieren und unterhalten.
- Q4. **Differenzierte Anbietepflicht:** Die gesetzlich für die Hochschulen und Forschungsanstalten zuständigen Archive setzen die geforderte Anbietepflicht in ihrem Zuständigkeitsbereich systematisch um. Sie richten sich dabei im Detail an den für sie geltenden gesetzlichen Vorschriften aus. In Absprache mit ihren Aufsichtsstellen (Kantonsarchive, Bundesarchiv) schliessen sie Primärdaten aus bestimmten Kategorien von Forschungsarbeiten von der Anbietepflicht aus, zum Beispiel alle studentischen Forschungen bis auf Stufe Diplom- und Masterarbeit, mit Ausnahme derjenigen, die in grössere Forschungsprojekte eingebunden sind und einen wesentlichen Beitrag dazu liefern.
- Q5. **Kooperative Bewertung:** Die zuständigen Archive organisieren die Ermittlung der Archivwürdigkeit der Primärdaten in ihrem Zuständigkeitsbereich gemäss den gesetzlichen Vorgaben. Das hier dargelegte Modell ist nicht davon abhängig, welches Verfahren gewählt wird. Einzig der Zeitpunkt der Bewertung spielt eine Rolle. Je länger man mit einem Entscheid über die Archivwürdigkeit zuwartet, umso schwieriger und teurer wird die Aufbereitung der Daten und ihre Dokumentation für die Langzeitarchivierung. Im Rahmen dieses Modells stützen wir uns darauf, dass Forschungsgesuche und Forschungsergebnisse in der Regel evaluiert bzw. begutachtet werden. Man kann deshalb von den Gutachtern verlangen, dass sie standardmässig in den jeweiligen Evaluationsberichten bzw. Gutachten einen begründeten Vorschlag zur Archivwürdigkeit machen und dieser Vorschlag in einem organisierten Verfahren dem Archiv zum Entscheid vorgelegt wird. Ausserdem soll in Abschlussberichten, die von grösseren Forschungsprojekten in der Regel verlangt werden, die gewählte Lösung für die Archivierung gemäss den Archivierungsgesetzen und den Standards der Archivkommissionen dazulegen. Diese Abschlussberichte sollten ebenfalls in einem organisierten Verfahren an die zuständigen Archive geleitet werden.<sup>43</sup>
- Q6. **Wiederbewertung:** Die Archivwürdigkeit kann mit einer Frist versehen werden, nach deren Ablauf die Archivwürdigkeit erneut vom gesetzlich zuständigen Archiv in Zusammenarbeit mit verantwortlichen Fachvertretern (Archivkommissionen) beurteilt wird. Damit soll darauf Rücksicht genommen werden, dass sich der Wert von Primärdaten aufgrund von neuen Erkenntnissen im Laufe der Zeit verändern kann.
- Q7. **Delegation der Bewertungskompetenz:** Um Synergien und Fachwissen besser zu nutzen, kann die zuständige staatliche Archivstelle die archivistische Bewertungskompetenz (unter Beibehaltung der Aufsichtsfunktion) an die von der Trägerschaft der nationalen Archivdienstleistung eingesetzten Archivkommissionen delegieren.
- Q8. **Schutzfristen und Eigentumsvorbehalt:** Die archivierten Primärdaten und die zugehörige Dokumentation sind grundsätzlich frei benutzbar. Allfällige Schutzfristen und Eigentumsvorbehalte müssen die Projektverantwortlichen im Rahmen der Übergabe der

---

<sup>43</sup> Abschlussberichte von Forschungsprojekten dürften ohnehin in den meisten Hochschularchiven generell als archivwürdig beurteilt werden und den Archiven zugeleitet werden.

Unterlagen an die nationale Archivdienstleistung melden. Sie müssen begründen, dass die Daten aus wichtigen persönlichen oder öffentlichen Gründen während einer bestimmten Frist vor einem Zugriff durch Dritte geschützt werden müssen und Eigentumsvorbehalte (z.Bsp. bei durch private Drittmittel mitfinanzierten Forschungsprojekten) und Vergütungsansprüche bestehen.

- Q9. **Aufbereitung:** Die Aufbereitung von als archivwürdig bestimmten Primärdaten (mit zugehöriger Dokumentation) für die Langzeitarchivierung erfolgt nach den Standards und Richtlinien der jeweils betroffenen fachbereichsspezifischen Archivkommission. Nach Möglichkeit wird sie unmittelbar nach Abschluss des Forschungsprojektes in Zusammenarbeit mit am Projekt beteiligten Personen durchgeführt.
- Q10. **Kosten Budget:** Die einmaligen Kosten für die spezifische Aufbereitung von als archivwürdig bestimmten Primärdaten (mit zugehöriger Dokumentation) für die Archivierung werden je nach den anwendbaren gesetzlichen Grundlagen entweder von der anbietepflichtigen Stelle (Forschungsinstitution) oder dem zuständigen staatlichen Archiv übernommen. Die Trägerschaft der nationalen Archivdienstleistung stattet diese mit einem entsprechenden jährlichen Budget aus. (Dabei ist zu berücksichtigen, dass zum Zeitpunkt der Budgetierung eines Forschungsprojektes die Archivwürdigkeit der darin erzeugten Primärdaten oft noch nicht geklärt werden kann.) Ist bereits bei der Bewilligung eines Forschungsprojektes klar, dass die Daten langfristig archiviert werden müssen, bspw. wenn der Schweizerische Nationalfonds dies in der Zuspracheverfügung verlangt, müssen diese Kosten im Forschungsbudget mitberücksichtigt werden.
- Q11. **Kostenmanagement:** Die nationale Archivdienstleistung erstellt für jeden Datenbestand, der als archivwürdig bewertet wurde, ein Budget für die fachgerechte Aufbereitung der Daten für die Langzeitarchivierung. Dieses Budget wird von der Stelle genehmigt, welche die Kosten gemäss den gesetzlichen Grundlagen zu tragen hat (d.h. entweder die anbietepflichtige Stelle, die staatliche Archivstelle oder die Trägerschaft).
- Q12. **Kostenkontrolle:** Übersteigen die voraussichtlichen Kosten für die Aufbereitung der Daten eines Forschungsprojektes die Möglichkeiten, welche das laufende Budget der nationalen Archivdienstleistung bietet, so stellt sie einen ausserordentlichen Finanzierungsantrag an die Trägerschaft. Kommt keine Finanzierung zustande, werden die Daten durch die nationale Archivdienstleistung nicht archiviert.
- Q13. **Betriebskosten:** Nach der einmaligen Aufbereitung der archivwürdigen Primärdaten für die Archivierung fallen bei der nationalen Archivdienstleistung jährliche Betriebskosten für den weiteren Unterhalt und langfristig auch für die Migrationen zur Erhaltung der weiteren Lesbarkeit und Verstehbarkeit der Daten an. Diese Kosten steigen mit der Anzahl der archivierten Datenbestände sowie auch mit dem Alter der Daten. Die Trägerschaft der nationalen Archivdienstleistung sorgt für die entsprechende Finanzierung, welche den langfristigen Erhalt der Archivdaten sicherstellt.
- Q14. **Erschliessung:** Die archivwürdigen Primärdaten werden von der nationalen Archivdienstleistung gemäss einschlägigen archivischen Grundsätzen erschlossen und über die Zuteilung von Persistent Identifiern (DOI oder URN) zitierfähig gemacht. Die Erschliessung



auf höherer Ebene ist einheitlich und standardisiert.<sup>44</sup> Die entsprechenden Verzeichnungsdaten werden den jeweils gesetzlich zuständigen über eine standardisierte Schnittstelle zum Import in deren Findmittelsysteme bereitgestellt. Gegebenenfalls notwendige fachspezifische *Detailerschliessungen* erfolgen durch die zuständigen staatlichen Archivstellen in deren eigenen Erschliessungssystemen.

- Q15. Zugang:** Der Zugriff auf Primärdaten in der nationalen Archivdienstleistung erfolgt über die Benutzersysteme der zuständigen staatlichen Archivstellen. Diese nutzen dazu die von der Archivdienstleistung bereitgestellten Identifikatoren (DOI oder URN), um die Benutzer ihrer Benutzersysteme im Benutzungsfall (d.h. Zugriff auf die Primärdaten) über eine Web-Schnittstelle an die nationale Archivdienstleistung weiterzuleiten. Diese bietet dazu eine einheitliche Benutzerschnittstelle, über welche die Daten und die standardisierte Verzeichnung einsehbar sind. Für die Kontrolle der personenbezogenen Zugriffsberechtigung sind alleine die staatlichen Archivstellen verantwortlich: Zur Kommunikation zwischen den Benutzersystemen der Archivstellen und der nationalen Archivdienstleistung werden nur technische, d.h. nicht personenbezogene Identifikationen eingesetzt, d.h. die Archivdienstleistung führt selber keine personenbezogenen Benutzerkonti. Einzig für Primärdaten, welche die zuständige staatliche Archivstelle als frei zugänglich qualifiziert hat, kann die nationale Archivdienstleistungen einen direkten Benutzungszugang anbieten. Da ein auf die einzelnen staatlichen Archive aufgeteilter Zugang für die Benutzung durch die Forschenden (Sekundärnutzung) sehr umständlich ist, sollte als Alternative die Einrichtung eines zentralen Portals durch die nationale Archivdienstleistung in Betracht gezogen werden.
- Q16. Verantwortung:** Die nationale Archivdienstleistung übernimmt zusammen mit den von ihrer Trägerschaft eingesetzten Archivkommissionen die Verantwortung für die sachgemässe Aufbereitung und Langzeitarchivierung. Die Archivdienstleistung kann diese Verantwortung in Einzelfällen ablehnen, wenn ihr die dazu nötigen Mittel, Verfahren und Rechte fehlen. In diesen Fällen entscheidet die Trägerschaft über das weitere Vorgehen.
- Q17. Aufbewahrungsplanung:** Die Aufbewahrungsplanung, insbesondere die mit der Zeit notwendig werdenden Migrationen in neue technologische Umgebungen, liegt in der gemeinsamen Verantwortung des nationalen Archivdienstleisters und der fachbereichsspezifischen Archivkommissionen. Können sie sich nicht einigen, entscheidet die Trägerschaft aufgrund einer eingeholten Drittmeinung.

### 5.2.3.1 Kosten und Finanzierung

Die Kosten für die Realisierung und den Betrieb eines zentralen Archivdienstleisters gemäss diesem Modell lassen sich im Gegensatz zum Modell in Kapitel 5.1 im Rahmen dieser Studie nicht sinnvoll abschätzen, da Menge und Komplexität der als archivwürdig beurteilten Primärdaten nicht vorauszusagen sind. Ein wesentlicher Kostenfaktor dieses Modells ist die fach- und sachgerechte Aufbereitung der Primärdaten und der dazugehörigen Dokumentation für die dauerhafte Archivierung. Ein zweiter Hauptfaktor dürfte die Arbeit der Archivkommissionen

<sup>44</sup> Unter „höheren Ebenen“ werden hier gemäss ISAD(G) die Verzeichnungsstufen Bestand, Serie und Dossier verstanden.

darstellen. Hinzu kommen die Kosten für die Organisation und Institutionalisierung der systematischen Anbietepflicht und archivischen Bewertung in den Forschungsinstitutionen und staatlicher Archivstellen.

Wenn allerdings die gesetzlich zuständigen Archive alle Aufgaben der Archivierung von Primärdaten selber übernehmen müssten, entstünde in der Summe ein wesentlich höherer Aufwand bei mutmasslich geringerer Qualität der Ergebnisse. Eine nationale Archivdienstleistung, welche Hochschul- und Forschungsinstitutionen in der verantwortlichen Trägerschaft hat, und welcher die Trägerschaft fachbereichsspezifische Archivkommissionen zur Seite stellt, kann insgesamt effizienter und besser arbeiten.

Die zuständigen Archive haben im Bereich der Forschungsdaten bisher wenig unternommen. Das hier dargestellte Modell füllt eine wesentliche Lücke nicht nur in der systematischen Archivarbeit, sondern auch in der systematischen Forschungsförderung. Die Sekundärnutzung, die über kurz- und mittelfristige Horizonte hinausgeht, ist in der Regel nur über eine fachgerechte Langzeitarchivierung, wie sie dieses Modell vorsieht, erreichbar.

Das hier besprochene Modell deckt allerdings nur die Archivierung von jenen Primärdaten ab, welche durch die staatlichen Archivstellen auf der Grundlage der Archivgesetze als archivwürdig bezeichnet werden. Eine *generelle* Aufbewahrung von Primärdaten, mit welcher die Forschenden ihre Pflichten bezüglich der Integrität der Forschung erfüllen können, wird dadurch nicht geboten, weder für *nicht* archivwürdige Primärdaten noch für alle archivwürdige Primärdaten, die unter Umständen durch die Archivierung so aufbereitet werden müssen, dass sie nicht mehr genügend den Originaldaten im Sinne der Aufbewahrungspflichten genügen.

## 6 Modell für Sekundärdaten

Die Konferenz der Universitätsbibliotheken der Schweiz (KUB) verfolgt mit dem vorliegenden Projekt die «Entwicklung eines Modells für eine zentrale Langzeitarchivierung von digitalen Primär- und Sekundärdaten aus der Forschung für die Schweiz» (Projekttitle<sup>45</sup>).

Wir schlagen hier ein Modell für Verlagspublikationen (eJournals, eBooks und wissenschaftliche Datenbanken) vor, ausgehend von den Projektzielen und -vorgaben, unserer Analyse in den Kapiteln 3 (Bedarf und Nutzen), 4 (Anspruchsgruppen) und 8 (Europäisches Umfeld) sowie der im einleitenden Abschnitt 2.3 diskutierten Schwerpunktsetzung.

Das Modell hat die *Langzeitarchivierung* von Sekundärdaten zum Inhalt. Herausforderungen, Strategien und technische Lösungskomponenten für die langfristige *Zugänglichmachung* von elektronischen Publikationen wurden dagegen 2005 – 2008 im Rahmen des Projektes «E-Archiving»<sup>46</sup> des Konsortiums der Schweizer Hochschulbibliotheken erarbeitet, wozu auch der Aufbau eines verteilten Dokumentenservers gehört.

Das Modell entspricht weitgehend der Strategie «Allianz-Initiative ‚Digitale Information‘» der Forschungsgesellschaften in Deutschland (vgl. Kapitel 8.5). Es soll die dauerhafte Verfügbarkeit von digitalen Verlags-Zeitschriften (eJournals) und -Büchern (eBooks) über den Ablauf von Lizenzen und die Verfügbarkeit von Online-Diensten der Verlage hinaus sicherstellen.

### 6.1 Strategieziele

Die folgenden drei strategischen Ziele sollen mit dem Modell erreicht werden:

- G7. **National:** Die Archivdienstleistung für Sekundärdaten soll allen im «Konsortium der Schweizer Hochschulbibliotheken»<sup>47</sup> vertretenen Bibliotheken dienen können, also insbesondere auch den kantonalen Universitäten und Fachhochschulen.
- G8. **Unterstützend:** Die Archivdienstleistung für Sekundärdaten (Publikationen) soll die langfristige Verfügbarkeit von elektronischen Verlagspublikationen für die Wissenschaft

<sup>45</sup> Auf die Bedeutungsunterschiede zwischen *Archivierung* und *Aufbewahrung* bzw. *Langzeitarchivierung* und *Langzeitaufbewahrung*, die sich aus staatlichen Archivgesetzen und archivfachlichen Gepflogenheiten herleiten, wurde in den Kapiteln 2.1 eingegangen.

<sup>46</sup> <http://lib.consortium.ch>. Der Schlussbericht ist noch nicht öffentlich zugänglich, weshalb wir auf die Resultate einer Vorstudie [17] aus dem Jahr 2005 verweisen.

<sup>47</sup> <http://lib.consortium.ch>. Dem in 2000 gegründeten Konsortium gehören alle kantonalen Universitäten, den ETH-Bereich, alle Fachhochschulen, mehrere Pädagogische Hochschulen, die Schweizerische Nationalbibliothek und weitere sekundäre Partner aus öffentlich finanzierten Einrichtungen in der Schweiz an. Bis 2005 wurde das Konsortium zu 50% vom Bund finanziert, seither vollständig durch die Partnerbibliotheken. Die zentrale Aufgabe des Konsortiums besteht in der Lizenzierung von elektronischen Informationsprodukten für die Partnerbibliotheken.

über den Ablauf von Lizenzen und die Verfügbarkeit von Online-Diensten der Verlage hinaus sicherstellen und damit den Versorgungsauftrag der einzelnen Universitätsbibliotheken effizient unterstützen.

- G9. **Integral:** Die Archivdienstleistung für Sekundärdaten soll aus Gründen der Transparenz und wegen ihrer organisatorischen und technischen Komplexität sowie Kostenintensität von einer einzelnen, geeigneten Institution angeboten werden, die vom «Konsortium der Schweizer Hochschulbibliotheken» als vertrauenswürdig anerkannt und entsprechend mandatiert und mitgetragen wird.

## 6.2 Begründung

Die Problematik der dauerhaften Archivierung von elektronischen Verlagszeitschriften wurde in den Kapiteln 3.2 und 8.3 diskutiert. Die Papierexemplare einer abonnierten Zeitschrift bewahrt die Bibliothek dauerhaft in ihren Magazinen auf, während die elektronischen Ausgaben nur auf dem Dokumentenserver des betreffenden Verlages lagern und über eine Internetverbindung benutzt werden. Der Zugriff ist zudem auf Computer-Arbeitsplätze in der Internet-Domäne der Universität beschränkt, während die interuniversitäre Zugreifbarkeit auf die Papierexemplare über den interbibliothekaren Ausleihverkehr (auch für Privatpersonen) möglich ist, in der Regel über einen Kopier- oder Scanning-Service für Einzelartikel.

Die langfristige Verfügbarkeit der bereits erschienenen elektronischen Ausgaben einer Zeitschrift kann durch den Verlag aus mehreren Gründen nicht garantiert werden:

- Der Verlag erlaubt keinen Zugriff mehr auf alte Ausgaben, wenn die Bibliothek die Lizenz kündigt.
- Der Verlag stellt seine Geschäftstätigkeiten ein oder geht Konkurs.
- Der Verlag stellt einen Titel ein und bietet keinen Zugriff mehr auf die alten Ausgaben.
- Durch einen technischen Datenverlust oder eine Katastrophe werden die Daten beim Verlag vernichtet.

Das Ende der Verfügbarkeit ist für die Bibliothek und die Forschenden dann besonders gravierend, wenn die Zeitschrift ausschliesslich elektronisch publiziert oder lizenziert wurde, was langfristig die Regel werden wird.

Die Bibliotheken sind somit gezwungen, physische Kopien der elektronischen Ausgaben bei sich zu archivieren, wenn sie die langfristige Verfügbarkeit<sup>48</sup> gewährleisten wollen. Dazu müssen die Verlage dieses Archivrecht in den Lizenzverträgen zugestehen und die elektronischen Ausgaben zusammen mit den bibliografischen Metadaten in Datenformaten liefern, welche erstens durch die Archivlösungen der Bibliotheken importiert werden können und zweitens für die langfristige bzw. dauerhafte Archivierung tauglich sind. In vielen Fällen sind diese drei Voraussetzungen (Archivrecht, Importformate, Archivtauglichkeit) nicht erfüllt.

---

<sup>48</sup> Bei der Archivierung entfallen allerdings alle spezifischen Recherchefunktionalitäten, die ein Verlag im Rahmen einer Lizenz über Internet bereitstellt. Es werden nur die Dokumente und bibliografischen Daten archiviert, und das Archiv kann nur Recherchefunktionen anbieten, die für alle Zeitschriften gleich sind, z.B. durch standardisierte Katalogeinträge und Volltextsuche.

Ausserdem fällt der technische und logistische Aufwand für die Archivierung bei jeder Bibliothek an, auch wenn derselbe Zeitschriftentitel von mehreren Bibliotheken lizenziert ist und die elektronische im Gegensatz zur papiergebundenen Ausgabe nicht an einen bestimmten Ort gebunden sein müsste. Vor allem kleineren Bibliotheken fehlen die Ressourcen, um eine eigene Archivlösung für eJournals aufzubauen, und die bezogenen Archivexemplare, soweit dieses Recht überhaupt besteht und genutzt wird, werden in fragwürdigen Datenformaten und auf einzelnen Datenträgern gelagert, die eine dauerhafte Aufbewahrung nicht garantieren können.

Einige Bibliotheken haben damit begonnen, ihre Papierausgaben von häufig benutzen Zeitschriftentiteln retrospektiv zu digitalisieren („Scanning“). Diese elektronischen Abbilder werden zwar nicht als eJournals bezeichnet, sind hier aber bezüglich Archivierung ebenfalls mitgemeint. Der wesentliche Unterschied zu eJournals ist, dass die Bibliothek die technischen Details der Formate (auch der bibliografischen Metadaten) unter eigener Kontrolle hat.

Für von Verlagen publizierte und von Bibliotheken erworbene wissenschaftliche elektronische Bücher (eBooks) besteht grundsätzlich dieselbe Problematik wie bei den eJournals, denn auch hier werden in der Regel nur Zugangsrechte<sup>49</sup> zu den Servern der Verlage erworben.

Sinn und Nutzen einer zentralen Archivdienstleistung, die im Auftrag von mehreren oder allen wissenschaftlichen Bibliotheken der Schweiz deren eJournals und eBooks an einem Ort archiviert, liegt somit auf der Hand. Dazu nötig ist, dass die Bibliotheken das nötige Archivierungsrecht und das Recht, auch nach Ablauf einer Lizenz die archivierten Exemplare weiterhin nutzen zu können, von den Verlagen erwerben. Dieses rechtliche Ziel wird in der Schweiz jedoch bereits zentral durch das «Konsortium der Schweizer Hochschulbibliotheken»<sup>50</sup> verfolgt, das im Namen seiner Partnerbibliotheken von den Verlagen die Lizenzen für elektronische Zeitschriften erwirbt. Eine zentrale Archivierung elektronischer Zeitschriften wird rechtlich und logistisch durch sogenannte Nationallizenzen<sup>51</sup> für Zeitschriften erleichtert.

Es gibt jedoch auch Alternativen zu einer zentralen Archivierung von eJournals in der Schweiz, die grundsätzlich auch vertretbare Lösungen darstellen können. Dies sind einerseits vertragliche Vereinbarungen mit Bibliotheken im Ausland, welche dieselben Zeitschriften bereits archivieren. So hat zum Beispiel die niederländische Nationalbibliothek (Koninklijke Bibliotheek, KB) bereits Archivierungsvereinbarungen mit 13 internationalen Zeitschriftenverlagen und archiviert 5'000 elektronische Zeitschriftentitel (siehe Kapitel 8.3 und 8.6.3.1). Bei einer solchen, rein vertraglichen Lösung jedoch folgende Problematik:

- Keine Bibliothek kann alle Zeitschriftentitel aller Verlage archivieren. Es wird deshalb immer in der Schweiz relevante Titel geben, deren Archivierung nicht durch Verträge mit ausländischen Bibliotheken erreicht werden kann.

<sup>49</sup> Bspw. umfasst die vom Springer-Verlag angebotene «eBook Collection» bereits 25'000 eBooks mit 3'500 Neuerscheinungen pro Jahr, darunter wissenschaftliche Monographien, Lehrbücher, Buchreihen und Nachschlagewerke.

<sup>50</sup> <http://lib.consortium.ch>. Dem in 2000 gegründeten Konsortium gehören alle kantonalen Universitäten, den ETH-Bereich, alle Fachhochschulen, mehrere Pädagogische Hochschulen, die Schweizerische Nationalbibliothek und weitere sekundäre Partner aus öffentlich finanzierten Einrichtungen in der Schweiz an. Bis 2005 wurde das Konsortium zu 50% vom Bund finanziert, seither vollständig durch die Partnerbibliotheken. Die zentrale Aufgabe des Konsortiums besteht in der Lizenzierung von elektronischen Informationsprodukten für die Partnerbibliotheken.

<sup>51</sup> <http://www.nationallizenzen.de>

- Es werden Verträge mit mehreren und zum Teil wechselnden ausländischen Bibliotheken nötig sein, was einen relativ hohen Aufwand bedeutet.
- Die Verträge könnten nicht direkt zwischen Bibliotheken abgeschlossen werden, sondern müssten Verträge zwischen den Staaten sein, um eine genügende Rechtssicherheit zu bieten (da die Ausführung der Archivaufgaben durch die ausländische Archivbibliothek von deren gesetzlichen Aufträgen und staatlichen Geldern abhängt).
- Die Verträge werden in der Regel zeitlich begrenzt sein, womit die Kontinuität der Archivierung von nicht durch die Schweiz beeinflussbaren Entscheidungen der ausländischen Vertragspartner abhängt.
- Die Sicherheit und technische Qualität der Archivierung ist alleine durch den ausländischen Partner bestimmt bzw. lässt sich nur schwer beurteilen.
- Für die von Bibliotheken in der Schweiz retrospektiv digitalisierten Zeitschriften müsste eine eigene Archivlösung gefunden werden, oder sie müssten ebenfalls der ausländischen Archivbibliothek übergeben werden. Letzteres wäre jedoch nicht mehr nur eine reine Mitnutzungsvereinbarung, sondern eine Dienstleistung.
- Unwahrscheinlich, jedoch grundsätzlich nicht ausgeschlossen sind internationale politische Konflikte oder kriegerische Ereignisse, die einen weiteren Zugriff auf die Archivdaten der ausländischen Archivbibliothek verunmöglichen oder sogar zur Zerstörung der Archivdaten führen.

Die zweite Alternative zu einer zentralen schweizerischen Infrastrukturlösung ist, die durch Universitäten von Verlagen bezogenen elektronischen Archivexemplare einem Dritten zu übergeben, der die dauerhafte Archivierung im Auftrag übernimmt. Ein solches Angebot bietet die US-Bildungstiftung Portico (siehe Kapitel 8.3 und 8.6.3.2) an, die auch im Auftrag von Verlagen archiviert. Portico enthält zur Zeit rund 8.3 Mio. Artikel aus 5'600 Zeitschriftentiteln von 66 teilnehmenden Verlagen – für weitere 8'085 Titel bestehen Vereinbarungen – und 476 teilnehmenden Bibliotheken. Die Kosten des Angebots sind relativ moderat und betragen pro Jahr maximal 0.1 – 0.5% der jährlichen totalen Sachmittelausgaben der beauftragenden Bibliothek. Die rechtlichen und sicherheitsrelevanten Risiken der Portico-Lösung sind für die beauftragende Bibliothek allerdings eher noch höher als bei der Kooperation mit europäischen Bibliotheken, da Portico eine private Stiftung mit Sitz in den USA ist und ihre Existenz wesentlich von der jährlichen finanziellen Unterstützung durch andere Stiftungen in den USA abhängt.

Das Konsortium der Hochschulbibliotheken und die Konferenz der Universitätsbibliotheken der Schweiz streben deshalb eine zentrale schweizerische Infrastrukturlösung an, damit die Kontrolle über die Archivdaten und sicherheitsrelevanten Faktoren in der Schweiz bleibt. Dies halten wir aus organisatorischen, logistischen, rechtlichen und technischen Gründen für sinnvoll.

Wie in Kapitel 5.1.2 bei Primärdaten halten wir auch hier ein Depositorium für sinnvoll, also eine reine Archivlösung, die keine Benutzeroberfläche und -funktionen für *Endbenutzer* (der beteiligten Bibliotheken) bietet, sondern sich auf die dauerhafte Verfügbarkeit der eJournal-Ausgaben und eBooks inklusive bibliografische Metadaten beschränkt. Die dauerhafte Verfügbarkeit schliesst, gleich wie bei den erweiterten Zielen für dauerhaft zu archivierende Primärdaten (siehe 5.1.5), auch die Konversion und Migration der Daten mit ein, falls solche in der Zukunft zur Sicherung der Kontinuität der Archivierung nötig werden.

Dieses zentralistische Modell entspricht weitgehend der Strategie «Allianz-Initiative ‚Digitale Information‘» [25] der Forschungsgesellschaften in Deutschland (siehe Kapitel 8.5).

### 6.3 Operative Ziele für eJournals und eBooks

Im Rahmen der strategischen Ziele G7 – G9 halten wir folgende operativen Ziele für sinnvoll.

- Z1. **Langzeitverantwortung:** Die Archivdienstleistung für eJournals und eBooks übernimmt die operative und technische Verantwortung für die dauerhafte Archivierung der eJournals/eBooks der im Konsortium der Schweizer Hochschulbibliotheken vertretenen Bibliotheken. Sie stellt den Zugang der Bibliotheken zu den entsprechenden Archivdokumenten sicher, wenn die Online-Benutzung der „Original“-Exemplare durch den Verlag über längere Zeit oder dauerhaft nicht mehr möglich ist.
- Z2. **Vertragliche Voraussetzungen:** Das Konsortium der Schweizer Hochschulbibliotheken regelt mit den Verlagen die zur Ausführung der Archivierung durch die Archivdienstleistung nötigen lizenzrechtlichen Bestimmungen. Dazu gehören das Recht zur Aufbewahrung und das Recht, die Daten in Zukunft nach eigenem Ermessen zu konvertieren, sofern dies für die Sicherung der Archivierung nötig ist.
- Z3. **Kosten:** Die Kosten für die Gesamtleistung Z1 werden kostendeckend und dauerhaft durch das Konsortium der Schweizer Hochschulbibliotheken getragen.
- Z4. **Kein Eigentum:** Die Archivdienstleistung für eJournals und eBooks erwirbt kein Eigentum an den durch sie archivierten Publikationen.
- Z5. **Haftungsausschluss:** Die Archivdienstleistung für eJournals und eBooks führt ihre Aufgaben nach bestem Wissen und Gewissen und hohen Standards aus, haftet jedoch bei einem Verlust der deponierten Daten in keiner Weise für den Schaden, welche den Nutzer/innen der Archivdienstleistung oder Dritten möglicherweise daraus entsteht. Auch haftet sie nicht für den Inhalt der archivierten Daten.
- Z6. **OAIS-Konformität:** Die zur Erfüllung der Aufgaben Z1 nötigen organisatorischen und technischen Komponenten der Archivdienstleistung genügen den Grundsätzen und Anforderungen des internationalen Standards ISO 14721:2003 «Open Archival Information System Reference Model» (OAIS) [27] sowie dem «nestor Kriterienkatalog vertrauenswürdige digitale Langzeitarchive» [28] oder der «Trustworthy Repositories Audit & Certification: Criteria and Checklist»<sup>52</sup> [29] des «Center for Research Libraries» (CRL). Wenn möglich soll die Archivdienstleistung für diesen Bereich über eine entsprechende Zertifizierung verfügen.
- Z7. **Ingest:** Die Archivdienstleistung für eJournals und eBooks übernimmt von den Bibliotheken oder direkt von der Verlagen die Archivexemplare der digitalen Publikationen (inkl. bibliografische Metadaten), validiert diese bezüglich Formaten und Vollständigkeit und lagert sie im Archiv ein. Mängel an den gelieferten Daten, welche nicht durch Nachlieferung der Verlage behoben werden können, werden durch die Archivdienstleistung oder durch beauftragte Dritte soweit behoben, dass eine fach-

<sup>52</sup> <http://www.crl.edu/PDF/trac.pdf>

gerechte, dauerhafte Archivierung anschliessend möglich ist. Ist eine solche Behebung aus technischen oder finanziellen Möglichkeiten unmöglich, so kann die Archivdienstleistung nach Rücksprache mit dem Konsortium die Übernahme ins Archiv ablehnen.

- Z8. **Importschnittstellen:** Die Archivdienstleistung für eJournals und eBooks implementiert, betreibt und pflegt zur Ausführung des Ingest (Z7) entsprechende Importschnittstellen pro Zeitschriftentitel bzw. eBook-Verleger. Über die Schnittstellen sollen die Daten automatisiert geprüft und wenn nötig konvertiert sowie fehlerfreie Objekte ins Archiv eingelagert und fehlerhafte Objekte ausgeschieden und isoliert werden. Die Schnittstellen ermöglichen eine sachgerechte Qualitätssicherung und Stichproben-Kontrolle.
- Z9. **Formate:** Die Archivdienstleistung für eJournals und eBooks verwendet, soweit möglich, standardisierte, für die Langzeitarchivierung taugliche Datenformate.
- Z10. **Verzeichnung:** Die Archivdienstleistung für eJournals und eBooks verwaltet mindestens jene Teilmenge der zu den Publikationen gehörigen bibliografischen Metadaten in strukturierter Form (Datenbank), die eine eindeutige Identifikation der einzelnen Publikation zulässt. Sie erzeugt und verwaltet zudem eigene Metadaten, welche die Zuordnung zum Lizenznehmer sowie die Nachvollziehbarkeit des Archivierungsvorganges und der Archivierungsgeschichte ermöglichen.
- Z11. **Sicherheit:** Die Archivdienstleistung für eJournals und eBooks sorgt für eine sichere und fachgerechte und langfristige Speicherung der Daten und schützt sie gegen unberechtigte Zugriffe und Verlust durch Katastrophen.
- Z12. **Preservation Planning und Migration:** Die Archivdienstleistung für eJournals und eBooks stellt die dauerhafte Integrität und maschinelle Interpretierbarkeit der übernommenen Dokumente und dazugehörigen bibliografischen Metadaten sicher. Dazu vereinbart sie, wenn möglich, mit den Verlagen entsprechende Standardformate für die gelieferten Daten oder konvertiert diese, wenn möglich, selber in Standardformate (siehe Z7). Während der Archivierung plant sie, insbesondere bei der Verwendung von schlecht für die Archivierung geeigneten Formaten, frühzeitig die Konversion der Daten in neue Formate, bevor die alten obsolet und nicht mehr handhabbar werden.
- Z13. **Integritätsnachweis:** Die Archivdienstleistung für eJournals und eBooks ermöglicht die Nachweisbarkeit der Integrität (Unverändertheit) der Dokumente ab dem Archivierungszeitpunkt, in dem sie für jedes Dokument beim Import (siehe Z8) automatisch einen qualifizierten Zeitstempel gemäss dem schweizerischen Bundesgesetz über die elektronische Signatur ZertES<sup>53</sup> [26] erzeugt. Sie sorgt für eine rechtzeitige Erneuerung des Zeitstempels, d.h. bevor die verwendeten kryptografischen Verfahren in der Schweiz nicht mehr als sicher gelten, und ermöglicht dabei den Integritätsnachweis bis zum Zeitpunkt des ersten Zeitstempels. Werden Archivdokumente im Rahmen von Z12 konvertiert, so beginnt eine neue Nachweiskette.
- Z14. **Herkunftsnachweis:** Die Archivdienstleistung für eJournals und eBooks versieht jedes Dokument beim Import (siehe Z8) mit einer fortgeschrittenen Signatur, durch welche sich später die Herkunft des Dokuments als „aus der Archivdienstleistung“ nachweisen lässt.

<sup>53</sup> [http://www.admin.ch/ch/d/sr/c943\\_03.html](http://www.admin.ch/ch/d/sr/c943_03.html)



Die Archivdienstleistung sorgt für eine rechtzeitige Erneuerung der Signatur, d.h. bevor die verwendeten kryptografischen Verfahren in der Schweiz nicht mehr als sicher gelten.

- Z15. **URN/UID/DOI:** Die Archivdienstleistung für eJournals und eBooks erzeugt für jede Publikation beim Import (siehe Z8) einen Uniform Resource Name (URN) oder einen technischen (d.h. nicht-sprechenden) eindeutigen Identifikator (UID) bzw. Digital Objekt Identifier (DOI).
- Z16. **Zugriff:** Die Archivdienstleistung für eJournals und eBooks lässt keinen Zugriff von aussen auf die Archivdaten zu, mit Ausnahme von Z18.
- Z17. **Exportschnittstellen:** Ist ein eJournal oder eBook eines Verlages dauerhaft nicht mehr online zugänglich, so exportiert die Archivdienstleistung die archivierten Publikationen dieses Bestandes derart, dass die bibliografischen Metadaten in die Katalogsysteme der Bibliotheken (bzw. die Verbundkataloge NEBIS etc.) und die Dokumente in einen zentralen Dokumentenserver der Bibliotheken importiert werden können. Die dazu nötigen Exportschnittstellen hängen vom Format der Archivdaten in der Archivdienstleistung sowie vom Format der Katalogsysteme und des Dokumentenservers ab und werden erst bei Bedarf erstellt. Als Alternative dient Z18.
- Z18. **Integrationschnittstelle/Hosting<sup>54</sup>:** Die Archivdienstleistung für eJournals und eBooks stellt eine Webservice-Schnittstelle bereit, durch welche ein normaler Web-Browser über das Internet und mit Hilfe der URN/UID (siehe Z15) der Publikation das dazugehörige Archivdokument (ohne bibliografische Metadaten) als Datei aus dem Archiv beziehen kann. Damit diese Integration funktionieren kann, muss vorher die entsprechende Bibliothek die URN/UID in den bibliografischen Eintrag der Publikation im Katalogsystem integriert haben. Ausserdem muss das Archivdokument ein Format haben, das ein Web-Browser anzeigen kann (z.B. PDF). Diese Schnittstelle der Archivdienstleistung überprüft die Zugriffsberechtigung anhand der Internet-Domäne des Computers, von dem aus der Aufruf erfolgt (z.B. aus einer Universität). Die zugriffberechtigten Domänen-Namen werden im Rahmen der Funktion Z10 verwaltet<sup>55</sup>, da es sich dabei gleichzeitig um die Lizenznehmer der Zeitschrift bzw. des eBooks handelt.

Anmerkung: Das mit Z1 – Z18 vorgeschlagene Modell für eJournals und eBooks bezieht sich ausschliesslich auf die von Mitgliedern des Konsortium der Schweizer Hochschulbibliotheken (unter anderen alle kantonalen Universitäten und Fachhochschulen sowie der ETH-Bereich) lizenzierten Verlagszeitschriften und -eBooks. Das Modell lässt sich natürlich leicht auf die sogenannten «OpenAccess»-Zeitschriften<sup>56</sup> ausdehnen. Da diese nicht lizenziert werden, sondern im Internet kostenlos verfügbar sind, entfallen dabei die Lizenz als Selektionskriterium und mit dem Investitionsschutz ein wichtiges Archivierungskriterium.

<sup>54</sup> Dies ist aus unserer Sicht nicht eine zwingende Funktion. Die Hosting-Funktionalität ist Bestandteil der Strategie «Allianz-Initiative ‚Digitale Information‘» [22] der Forschungsgesellschaften in Deutschland.

<sup>55</sup> Die Möglichkeit einer Integration der Zugriffsberechtigung in die von SWITCH betriebene «Authentifizierungs- und Autorisierungs-Infrastruktur (AAI)» (<http://www.switch.ch/aai/>) hängt von den Details der Implementierungen dieser Integrationschnittstelle ab. Grundsätzlich ist davon auszugehen, dass bereits der Zugriff des Benutzers auf den externen Katalog über die AAI erfolgt.

<sup>56</sup> Es gibt weltweit bereits gegen 4000 wissenschaftliche «OpenAccess»-Zeitschriftentitel (<http://open-access.net>, <http://www.doaj.org>, bei denen es sich in der Regel um normale Verlagszeitschriften handelt, die aber nicht lizenziert werden müssen, sondern kostenlos im Internet verfügbar sind.

Es dürfte schwierig sein, für die dauerhafte Archivierung in der Archivdienstleistung eine sinnvolle Auswahl für «OpenAccess»-Zeitschriften zu treffen *und* dabei die Kosten kontrollieren zu können. Insbesondere verlangen Verlage auch bei «OpenAccess»-Zeitschriften zumindest eine Aufwandsentschädigung für die separate Lieferung der Archivexemplare und bibliografischen Metadaten. Wir halten es deshalb für sinnvoller, auf eine systematische Archivierung von «OpenAccess»-Zeitschriften im Rahmen der Archivdienstleistung zu verzichten bzw. dies anderen Projekten zu überlassen.

### 6.3.1 Kosten und Finanzierung

Die in Z1 – Z18 umrissene Funktionalität lässt sich grundsätzlich mit einem marktgängigen Dokumenten-Management-System (DMS) aus dem mittleren Preissegment realisieren. Wichtiges Auswahlkriterium ist dabei die Flexibilität, welche das Produkt bezüglich der Implementierung der Import- und Exportprozesse, der Unterstützung von qualifizierten elektronischen Zeitstempel (insbesondere des Nachstempelns) und fortgeschrittenen Signaturen sowie der verlustfreien Migration des gesamten Archivinhalts in ein künftiges, neues Archivsystem bietet. Zentral, aber unabhängig vom Produkt, sind die gewählten Datenformate für die Archivierung.

Die Anschaffungskosten für DMS-Lösungen sind stark von der Anzahl der benötigten Benutzerlizenzen abhängig. Die Archivdienstleistung benötigt jedoch (ausser für Z18) nur 2 – 3 Lizenzen, für Administratoren, die Archivdaten importieren und unterhalten.

Da die Nachweisbarkeit der Integrität über qualifizierte Zeitstempel erfolgen soll, können die Archivobjekte auf beliebigen, für die Langzeitspeicherung tauglichen Speichersystemen abgelegt werden. Für diese Speicherung wird sinnvollerweise dieselbe Lösung wie bei den Primärdaten (siehe Kapitel 5.1.5.1) verwendet.

Die Finanzierung des Modells gemäss Z1 – Z18 teilt sich in eine Investition (Detailkonzept, Beschaffung und Konfiguration der Software, Realisierung, Einführung) sowie laufende Betriebskosten (inkl. Implementierung der Import- und Exportschnittstellen).

Für den Investitionsteil schätzen wir grob 0.6 Mio. CHF, exklusive Import- und Exportschnittstellen sowie der Einrichtung der initialen Speicher- und Kommunikations-Infrastruktur. Letztere sollte durch Partner aus dem universitären Bereich kostenlos zur Verfügung gestellt werden. Nicht berücksichtigt sind mögliche Synergien mit der Lösung für die Archivdienstleistung „Primärdaten“ (Kapitel 5.1.2).

Diesen Betrag sehen wir als direkten, einmaligen Projektbetrag von Bund und Kantonen, den Organen der Forschungsförderung sowie von Stiftungen und möglicherweise privater Unternehmen der IT-Branche (Sponsoring).

Hinzu kommen die Import- und Exportschnittstellen (Z8 und Z17) die man entweder zu den Investitions- oder den Betriebskosten rechnen kann, da grundsätzlich für jeden neuen Verlag eine eigene Importschnittstelle und für jedes in Z17 zu beliefernde externe System eine individuelle Schnittstelle erstellt werden muss. Die Kosten schätzen wir (inkl. Testphase) auf ungefähr 20 – 40'000 CHF pro Importschnittstelle, inklusive Analyse, Detailspezifikation, Implementierung und Testphase. Dies hängt davon ab, ob das gewählte Archivprodukt eine vernünftige Umgebung für die Implementierung von Importschnittstellen bietet, welche

Ansprüche an die Qualitätssicherung und allfällige Formatkonversionen gestellt werden und ob der Verlag die Daten in einem einigermaßen vernünftigen Format liefert.

Die laufenden Betriebskosten dürften dagegen im Normalbetrieb klein sein, da der Import weitgehend automatisiert ablaufen sollte. Höhere Kosten fallen singular an, etwa bei der Implementierung neuer Importschnittstellen (siehe oben) oder etwa alle 5 – 10 Jahre für die Migration des gesamten Archivs in eine neue Lösung oder die Konversion von einzelnen Datenbeständen.

Gemäss Z3 sollen die Betriebskosten durch das Konsortium der Schweizer Hochschulbibliotheken bzw. dessen Mitglieder getragen werden.

Aus unserer Sicht gibt es möglicherweise auch eine Alternative zu einer neuen Lösung: Die Schweizerische Nationalbibliothek (NB) verfügt bereits über eine Lösung zur Archivierung von *in der Schweiz* publizierten elektronischen Publikationen. Es sollte geprüft werden, ob diese Lösung den Anforderungen an eine gesamtschweizerische Lösung für alle eJournals aller Bibliotheken der Hochschulen genügt (oder entsprechend ausgebaut werden könnte), und es die rechtlichen Grundlagen der NB erlauben, im Auftrag des Konsortiums (und gegen Bezahlung) die Archivdienstleistung im Sinne von Z1 – Z18 zu übernehmen.

## 6.4 Operative Ziele für weitere Datentypen

### 6.4.1 Wissenschaftliche Datenbanken

Über das Konsortium der Schweizer Hochschulbibliotheken haben zahlreiche Mitglieder auch den Online-Zugang zu wissenschaftlichen Datenbanken lizenziert<sup>57</sup>. Hierbei handelt es sich um Literaturhinweis-Sammlungen, thematische Nachschlagewerke, Fakten- und Abstract-Sammlungen, Wörterbücher und Volltext-Datenbanken (z.B. Zeitungsartikel) etc. Diese Datenbanken stellen ebenfalls Publikationen dar, auf die sich Forschende in einigen Wissenschaftsbereichen in hohem Masse stützen. Diese Datenbanken können jedoch aus den gleichen Gründen wie bei elektronischen Zeitschriften plötzlich nicht mehr verfügbar sein, wenn der Verlag seinen Dienst oder seine Geschäftstätigkeit einstellt.

Die Archivierung von Datenbanken (d.h. strukturierten Daten) ist allerdings wesentlich komplizierter und aufwendiger als die Archivierung von dokumentenzentrierten Daten wie eJournals und eBooks.

- **Komplexe Struktur:** Während die eJournals Einzeldateien sind, die bei den meisten Verlagen im relativ leicht zu handhabenden Format PDF (Portable Document Format) vorliegen, ist jede Datenbank individuell, und viele Datenbanken haben eine komplexe technische Struktur.
- **Software-Abhängigkeit:** In jedem Fall ist eine Datenbank von der verwendeten, produktspezifischen Datenbank-Software abhängig. Zusätzlich können viele Datenbanken nur in Zusammenhang mit einer zusätzlichen Spezialsoftware genutzt werden, welche für die Benutzer eine Bedienungsoberfläche mit spezifische Such- und Auswertungsfunktionen bereitstellt.

<sup>57</sup> [http://lib.consortium.ch/products\\_categories\\_lizenzen.php](http://lib.consortium.ch/products_categories_lizenzen.php)

- **Dynamik:** Während die publizierte Ausgabe einer elektronischen Zeitschrift statisch ist und sich nicht mehr ändert, sind Datenbanken dynamisch, d.h. sie verändern sich laufend durch Inhalte, die der Verlag neu hinzufügt oder löscht. Der Inhalt einer Datenbank lässt sich somit überhaupt nur „im Zustand zum Zeitpunkt X“ archivieren, also nur als „Schnapschüsse“ einer zeitlichen Entwicklung.

Da die Mitarchivierung der gesamten technischen Datenbank- und Anwendungssoftware (zusammen mit den Daten) in der Regel auch mit Emulationsansätzen nicht realisierbar und handhabbar ist, kommt nur eine Extraktion der Daten zu einem bestimmten Zeitpunkt in Frage. Dabei müssen die Daten jedoch in ein Format und eine Struktur transformiert werden, die später für Endbenutzern noch sinnvoll verwendet und verstanden werden kann. Diese Transformation in ein Archivformat ist für jede Datenbank individuell zu konzipieren und zu implementieren, was entsprechend aufwendig ist.

Zur effizienten Langzeitarchivierung von Datenbanken ist deshalb eine spezielle Archivierungssoftware nötig, welche die Daten kontrolliert aus der originalen Datenbank des Verlegers extrahiert, in ein für die Langzeitarchivierung taugliches und einheitliches Standardformat transformiert und schliesslich alle auf diese Weise archivierten Datenbanken kollektiv verwalten und über dieselbe Benutzeroberfläche, unabhängig von der originalen technischen Software-Umgebung des Verlages, benutzbar macht.

Uns sind nur zwei Lösungen bekannt<sup>58</sup>, die dies Funktionalität in einigermaßen vernünftiger Weise bieten können. Sie benötigen allerdings eine direkte Verbindung zur Datenbank (während normale Benutzer nur Kontakt mit der Web-Benutzeroberfläche haben). Ist der Verlag jedoch nur bereit, von ihm selber hergestellte Export-Dateien zu liefern, so ist die Situation bezüglich Archivierung in der Regel hoffnungslos.

Eine systematische Archivierung von vielen Datenbanken durch Bibliotheken ist, soweit es die einzelnen Verlage überhaupt zulassen, aus unserer Sicht nicht realistisch, in ausgewählten Einzelfällen aber machbar. Genauere Aussagen liessen sich nur im Rahmen einer detaillierten Analyse der Bedürfnisse und Möglichkeiten machen, die jedoch ausserhalb des Rahmens dieser Konzeptstudie liegt.

- D1. **Datenbanken:** Die Archivdienstleistung für wissenschaftliche Datenbanken kann in Einzelfällen von besonders wertvollen, von Bibliotheken lizenzierten Datenbanken, deren Zugänglichkeit nicht mehr gesichert ist, Projekte zur dauerhaften Archivierung der Daten durchführen. Die Kosten müssen durch Partner gedeckt werden. Eine systematische Archivierung der von Bibliotheken lizenzierten wissenschaftlichen Datenbanken erfolgt nicht.

## 6.4.2 eLearning-Objekte

Die Problematik der dauerhaften Archivierung von elektronischen Lehr- und Lernunterlagen (eLearning-Objekte) wurde in den Kapitel 3.3 diskutiert. Zu eLearning-Objekten gehören zum

<sup>58</sup> CHRONOS ([www.datenbankarchivierung.de](http://www.datenbankarchivierung.de)) der deutschen Firma CSP, das unter anderem mit staatlichen Fördermitteln des Staates Bayern zusammen mit dem Fachbereich Informatik der Fachhochschule Landshut entwickelt wurde, sowie SIARD des Schweiz. Bundesarchivs (<http://arxiv.org/abs/cs.DL/0408054>).

Beispiel Videoaufzeichnungen von Vorlesungen sowie Online-Versionen von Vorlesungsskripten, PowerPoint-Folien und Tutorials der Dozierenden, Selbsttests und Prüfungsfragen, aber auch komplexe Multimedia-Präsentationen, vorlesungsbezogene Blogs und „chat rooms“ von Dozierenden etc.

Auch die ETH Zürich und die Hochschulstiftung SWITCH führen Projekte zur Erstellung, Verwaltung und Zugänglichmachung von eLearning-Unterlagen. Zumindest bei SWITCH – sie bietet den Universitäten auch Infrastruktur zur Erstellung von eLearning-Unterlagen an – ist die Langzeitaufbewahrung dieser Unterlagen noch kein Thema [18]. Auch liegen die eLearning-Unterlagen mehrheitlich in sehr proprietären, für die Langzeitaufbewahrung nicht tauglichen Benutzerformaten vor. Auch scheinen uns die Strategien der Bildungsinstitutionen bezüglich des Einsatzes von eLearning noch in einem relativ experimentellen Stadium. Und schliesslich ist ihre Erschliessung (systematische Ordnung und Verzeichnung) auf Grund der Vielfältigkeit und Heterogenität sehr aufwendig.

Aus diesen Gründen könnten wir bezüglich der Langzeitarchivierung von eLearning-Objekten in einer zentralen, nationalen Archivdienstleistung präzisere Aussagen nur nach einer detaillierten Analyse machen, die ausserhalb des Rahmens dieser Konzeptstudie liegt.

### 6.4.3 Scanning-Masterkopien

Die Problematik der dauerhaften Archivierung von Master- bzw. Roh-Exemplaren von retrospektiv digitalisierten Archivunterlagen wurde in den Kapitel 3.4 diskutiert. Die von Bibliotheken retrospektiv digitalisierten Zeitschriften sind hier nicht mitgemeint, da diese im Rahmen der Archivdienstleistung für eJournals und eBooks (siehe Kapitel 6) archiviert werden können.

Gemeint sind hier die optisch hochauflösenden Master- bzw. Roh-Kopien anderer Digitalisate, zum Beispiel von Fotografien, Filmen, Tonaufnahmen oder historischen Büchern. Aus den Master-Kopien werden qualitativ schlechtere, jedoch schlankere Kopien für die einfache Benutzung hergestellt. Benutzungskopien werden nicht archiviert: gehen Sie verloren oder genügen sie nicht mehr den Benutzeransprüchen, werden aus den Master-Kopien neue Benutzungskopien erzeugt. Archiviert werden müssen also die Master-Kopien. Allerdings ist auch dies heute nicht mehr immer der Fall. Die technische Entwicklung der Digitalisierungstechnologie schreitet so schnell voran, dass nach einigen Jahren die analogen Originale nochmals digitalisiert werden.

Wesentlich ist die dauerhafte Archivierung der Master-Kopien jedoch in den Fällen, in denen die analogen Originale vom physischen oder chemischen Zerfall bedroht sind, so dass ihre Verfügbarkeit für eine spätere, neue Digitalisierung nicht mehr sichergestellt werden kann.

Solche Master-Kopien fallen heute an zahlreichen Orten an, vor allem in den staatlichen Archiven von Bund und Kantonen im Rahmen der Bestandserhaltung, aber auch in Universitätsbibliotheken, z.B. im Projekt «e-rara»<sup>59</sup> zur Digitalisierung von alten Drucken. Erfahrungsgemäss werden in diesen (zeitlich befristeten) Projekten jedoch die langfristigen Kosten für die sichere Archivierung bzw. Speicherung der grossen Datenmengen für die Master-Kopien nicht

---

<sup>59</sup> <http://www.e-rara.ch>

eingerechnet, wodurch die Langzeitarchivierung und damit auch der Investitionsschutz nicht gewährleistet sind.

Die «Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen» (KOST) der Kantonsarchive, das Projekt «e-rara» und weitere Digitalisierungsprojekte haben in Bezug auf die hier besprochene nationale Archivdienstleistung Interesse vor allem an einer reinen „Bitstream Preservation“ bzw. günstigem Speicherplatz angemeldet. Ein solches Angebot für öffentlich finanzierte Digitalisierungsprojekte kann im Rahmen dieser Archivdienstleistung organisatorisch und technisch leicht realisiert werden. Dies kann innerhalb der Lösung für die Primärdaten (siehe Kapitel 5.1.2) geschehen. Die Unterschiede sind lediglich:

- Ein solches Angebot macht erst für grössere Datenmengen (pro Kunde) Sinn.
- Die Anlieferung der Daten erfolgt nicht online, sondern über per Kurier zugesandte Magnetbänder oder Harddisks.
- Es erfolgt keine bibliografische oder archivische Verzeichnung.

Wir schlagen deshalb folgende operativen Ziele vor:

- S1. **Digitalisate:** Die Archivdienstleistung für Digitalisate kann im Auftrag von öffentlich finanzierten Stellen Digitalisate nach dem „Schliessfachprinzip“ aufbewahren.
- S2. **Menge und Kosten:** Das Angebot für Digitalisate gilt nur ab einer Datenmenge von 200 GB pro Institution / Projekt. Der Aufwand der Archivdienstleistung muss von der Institution jährlich und kostendeckend bezahlt werden. Will die Institution den Aufwand nicht mehr kostendeckend bezahlen, so muss sie ihre Daten zurücknehmen.
- S3. **«Bitstream Preservation»:** Die Archivdienstleistung für Digitalisate übernimmt im Auftrag der beauftragenden Institution die Verantwortung für die Verfügbarkeit und sichere Erhaltung des Bitstroms der ihr übergebenen Originaldaten. Dies schliesst auch den Schutz gegen unberechtigte Zugriffe und Katastrophen sowie die fachgerechte langfristige Speicherung mit ein.
- S4. **Keine Rechteübertragung:** Die Archivdienstleistung für Digitalisate erwirbt kein Eigentum und keine Nutzungsrechte an den Daten, welche die Institution bei ihr deponiert. Das Eigentum an den Daten sowie Nutzungsrechte, Urheberrecht, Copyrights etc. bleiben, soweit vorhanden, unverändert bei der Institution.
- S5. **Haftungsausschluss:** Die Archivdienstleistung für Digitalisate führt ihre Aufgaben nach bestem Wissen und Gewissen und hohen Standards aus, haftet jedoch bei einem Verlust der deponierten Daten in keiner Weise für den Schaden, welche der beauftragenden Institution oder Dritten möglicherweise daraus entsteht. Auch haftet sie nicht für den Inhalt der deponierten Daten.
- S6. **Anlieferung / Keine Prüfung:** Die Archivdienstleistung für Digitalisate nimmt die Daten nicht online, sondern nur durch per Kurier zugesandte Magnetbänder oder Harddisks entgegen. Sie führt eine Überprüfung der Anzahl und der Hashwerte der Dateien durch, jedoch keine Überprüfung der Dateiformate.
- S7. **Integrität:** Die Archivdienstleistung für Digitalisate erzeugt für jede in einer Lieferung der Institution enthaltene Datei einen qualifizierten Zeitstempel gemäss dem

schweizerischen Bundesgesetz über die elektronische Signatur ZertES<sup>60</sup>, wodurch die Integrität der Datei ab dem Archivierungszeitpunkt beweisbar ist. Sie sorgt zudem für eine rechtzeitige Erneuerung des Zeitstempels, d.h. bevor die verwendeten kryptografischen Verfahren in der Schweiz nicht mehr als sicher gelten, und ermöglicht dabei den Integritätsnachweis bis zum Zeitpunkt des ersten Zeitstempels.

58. **UUID:** Die Archivdienstleistung für Digitalisate identifiziert jede Datei eindeutig über einen technischen (nicht-sprechenden) „Universally Unique Identifier“ (UUID)<sup>61</sup> und liefert diese UUIDs zusammen mit den dazugehörigen Dateinamen und Hashwerten an die beauftragende Institution.
59. **Zugriff:** Archivdienstleistung für Digitalisate ermöglicht der beauftragenden Institution, die deponierten Daten jederzeit innert nützlicher Frist zu beziehen. Dies ist Offline und Online möglich: Offline schickt die Institution die UUIDs der gewünschten Dateien an die Archivdienstleistung, welche diese dazugehörigen Dateien auf Datenträgern per Kurier an die Institution sendet. Online: Die Institution erhält über ein temporär freigeschaltetes Benutzerkonto per SSH (Secure Shell) Zugriff auf ihren Datenbestand.

---

<sup>60</sup> [http://www.admin.ch/ch/d/sr/c943\\_03.html](http://www.admin.ch/ch/d/sr/c943_03.html)

<sup>61</sup> Universally Unique Identifiers (UUID) sind 128 Bit lange Zahlen, die gemäss international standardisierten Methoden erzeugt werden (ITU-T Rec. X.667, ISO/IEC 9834-8:2005, RFC 4122). Sie werden in sog. Hexadezimal-Schreibweise als Folge von 36 Zeichen dargestellt, z.B. 550e8400-e29b-41d4-a716-446655440000. Eine einmal erzeugte UUID kann nach mathematischer Wahrscheinlichkeit kein zweites Mal erzeugt werden, sie ist also universell eindeutig.

## 7 Alternativmodelle

Zusätzlich zu den zwei in Kapitel 5 beschriebenen Modellen für Primärdaten skizzieren wir hier folgend vier grundsätzlich abweichende Alternativen für eine zentrale, national ausgerichtete Dienstleistung im Bereich der Langzeitarchivierung von Primärdaten aus der Forschung. Wir haben diese Alternativmodelle nicht im Detail ausgearbeitet, weil sie aus unserer Sicht der Dringlichkeit der Problematik, auf die das vorliegende Projekt der Konferenz der Universitätsbibliotheken der Schweiz (KUB) direkt Bezug nimmt, nur ungenügend gerecht werden können.

Die beiden in Kapitel 5 beschriebenen Modelle haben bezüglich Infrastruktur und operativer Ausführung der Dienstleistung einen zentralistischen Ansatz. Die folgenden Alternativmodelle verfolgen deshalb alle eine im Vergleich zum Primärmodell betont föderalistischen Ansatz, bei dem das *eigentliche* Anliegen, dass nämlich in absehbarer Zeit *tatsächlich* operative Lösungen für die Aufbewahrung / Archivierung von Primärdaten entstehen, der Eigeninitiative der Universitäten überlassen wird. Die nationale Dienstleistung besteht dabei „nur“ aus der Motivation, Beratung und Koordination solcher Eigeninitiativen.

Natürlich lassen sich auch einzelne Elemente dieser Alternativmodelle in die beiden Modelle aus Kapitel 5 einbauen. Wir sehen dabei allerdings die Gefahr, die Modelle zu „überladen“ und dadurch ihre Akzeptanz zu reduzieren.

Bei der dauerhaften Archivierung von digitalen Verlags-Publikationen der Bibliotheken sehen wir keine sinnvolle Alternative zu einer zentralen, gesamtschweizerischen Lösung. Bei anderen Sekundärdaten wie zum Beispiel grauer Literatur sind dagegen auch andere Lösungen denkbar.

### 7.1 Strategieträger / nationale Task Force

Dieses Alternativmodell besteht aus der Bildung eines nationalen Gremiums oder einer Task Force aus Vertreter/innen der Organe der Forschung, Forschungsförderung, Universitäten, Hochschulen, Bibliotheken und staatlichen Archive zur „Langzeitarchivierung von Primär- und Sekundärdaten der Forschung in der Schweiz“ (LZA von Forschungsdaten). Solche Gremien und Task Forces gab und gibt es in zahlreichen Ländern (vgl. die Beispiele in Kapitel 8.6.1), jedoch in der Regel mit geringer praktischen Auswirkungen.

Das Gremium erarbeitet die strategischen Ziele, Rahmenbedingungen und Richtlinien, die bei der LZA von Forschungsdaten schweizweit erreicht und umgesetzt werden sollen.

Die Umsetzung der nationalen Strategie wird dabei den einzelnen Forschungsinstitutionen überlassen bzw. den Mitglieder des Gremiums, welche die Initiative in die einzelnen Institutionen und Organe tragen. Insbesondere erarbeitet das Gremium :



- Strategische Ziele der LZA für Forschungsdaten im Rahmen der Forschungsinfrastrukturen der Schweiz.
- Klärung der rechtlichen Stellung von Forschungsdaten im Rahmen staatlicher Archivierungsgesetze von Bund und Kantonen.
- Richtlinien, Kriterien, Rahmenbedingungen und Controlling-Mechanismen für die *praktische Umsetzung* der Richtlinien der Akademien Schweiz, des SNF und der Universitäten und Hochschulen bezüglich der Aufbewahrung von Primärdaten im Rahmen der „Integrität der Forschung“. Insbesondere zur Rolle, welche neben den Forschenden die Institutionen dabei spielen.
- Empfehlungen zur Finanzierung bzw. Förderung von LZA-Lösungen und -Projekten für Primärdaten an den Universitäten und Hochschulen.

Dieses Gremium bzw. diese nationale Task Force führt ein Backoffice, welches die die Arbeiten koordiniert und fachlichen Diskussionsgrundlagen zuhanden des Gremiums aufbereitet.

## 7.2 Nationales Kompetenz-Zentrum

Dieses Alternativmodell sieht die Bildung einer mit eigenen wissenschaftlichem Personal ausgestatteten Institution vor, welche Forschende und Forschungsinstitutionen in der ganzen Schweiz durch Beratung, ein Fachportal im Internet, Projektunterstützung, selber initiierte Forschungsprojekte und ggf. auch durch finanzielle Förderbeiträge bei der langfristigen Aufbewahrung bzw. Langzeitarchivierung von Forschungsdaten unterstützt.

Das Zentrum kann auch Standards und Tools zur LZA von Primärdaten entwickeln und fördern. Es erbringt aber selber keine Archivdienstleistungen.

Als Vorbild für dieses Modell kann das «Digital Curation Centre» (DCC) in Grossbritannien dienen (siehe Kapitel 8.6.1.2.3).

## 7.3 Nationales Kompetenz-Netzwerk

Dieses Alternativmodell sieht die Bildung eines Vereins vor, der durch direkte Förderbeiträge des Bundes sowie Spenden ein informelles Netzwerk aus Personen und Institutionen aufbaut, welche an der digitalen Langzeitarchivierung von Forschungsdaten interessiert oder in entsprechende Projekte involviert sind.

Das Netzwerk organisiert Arbeitsgruppen und regelmässige Treffen für die Diskussion, den Wissensaustausch und die Erarbeitung von Empfehlungen zur LZA von Primärdaten.

Als Vorbild für dieses Modell können die entsprechenden Kompetenz-Netzwerke in den Nachbarländern Deutschland und Frankreich dienen: NESTOR (Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen in Deutschland, siehe Kapitel 8.6.1.2.2) und PIN (Groupe Pérennisation des Informations Numériques, siehe Kapitel 8.6.1.2.5).

## 7.4 Register, Inventar und Portal

Dieses Alternativmodell sieht einen redaktionellen Dienst vor, der in der Schweiz vorhandene und geplante Repositorien und Depositorien für wissenschaftliche Primärdaten recherchiert und mit Kurzbeschreibungen und Kontaktpersonen in einer Datenbank charakterisiert und registriert.

Diese Informationen werden über ein Web-Portal öffentlich als Inventar publiziert und in Fällen, wo die Repositorien-Betreiber einen öffentlichen Online-Zugang ermöglichen, mit entsprechenden Gatewayservices versehen. Das Portal ermöglicht es Forschenden und Institutionen, ihre Repositorien und Depositorien selber online zu registrieren.

Mit dieser Dienstleistung wird nur die Sekundärnutzung und die Kontaktaufnahme mit Projekten erleichtert.

## 8 Europäisches Umfeld

Teil des Auftrages zu dieser Studie war auch die Darstellung ähnlicher Aktivitäten im europäischen Umfeld. In diesem Kapitel führen wir zuerst eine kurze Typisierung und Abgrenzung verschiedener Vorhaben ein und verdeutlichen diese exemplarisch anhand der EU-finanzierten Forschungsprojekte zur digitalen Langzeitarchivierung im Allgemeinen, der Projekte zu institutionellen Repositorien und der nationalen Depot-Bibliotheken.

Anschliessend gehen wir auf eine prominente gesamteuropäische und eine gewichtige nationale Initiative ein, die sich direkt den Primärdaten der Forschung widmen und sich als Hintergrund zur Positionierung der beiden in Kapitel 5 vorgestellten Modelle eignen.

Schlussendlich präsentieren wir in Form eines beschreibenden Katalogs die Resultate unserer Recherche zu nationalen Strategien und Einzelvorhaben zur Langzeitarchivierung von Primär- und Sekundärdaten, die für diese Studie und die weitere Entwicklung dieses «E-lib.ch»-Teilprojektes relevant sein können.

Alleine in Europa gibt es eine kaum noch überschaubare Anzahl und Vielfalt von öffentlich finanzierten Forschungsprojekten und operativen Dienstleistungen, die sich auf die breite Verfügbarkeit und langfristige Aufbewahrung von flüchtigen digitalen Informationsressourcen aus wissenschaftlichen, kulturellen und staatlichen Tätigkeiten beziehen. Diese Aktivitäten lassen sich grob in folgende fünf Typen unterteilen:

1. Von der Europäischen Kommission über ihre Forschungsrahmenprogramme FP5 (1998-2002), FP6 (2002-2006) und FP7 (2007 - 2013) finanzierte Projekte, die per Definition eine breite internationale (europäische) Beteiligung und Ausrichtung aufweisen.
2. Fachspezifische, jedoch länderübergreifende Strategien, Projekte und Dienstleistungen internationaler Fachverbände, Forschungsorganisationen und Forschungseinrichtungen.
3. Fachübergreifende und auf einen nationalen Nutzerkreis ausgerichtete Initiativen, Projekte und Dienstleistungen, meist getragen von National- und grossen Hochschul-Bibliotheken (zum kleineren Teil auch von Nationalarchiven) sowie nationalen Forschungsgesellschaften und Forschungseinrichtungen.
4. Fachspezifische und auf einen nationalen Nutzerkreis ausgerichtete Projekte und Dienstleistungen von Forschungsgesellschaften oder universitärer Forschungsinstitute.
5. Fachspezifische und auf eine einzelne Institution oder Organisationseinheit oder den Nutzerkreis eines einzelnen Datenbestandes beschränkte Projekte und Dienstleistungen.

## 8.1 EU-Projekte im Rahmen von FP6 und FP7

Die überwiegende Zahl aller Projekte und Dienstleistungen fallen in die Kategorien 4 und 5, haben also entweder einen lokal oder fachlich sehr eingeschränkten Anspruch. Seit 2004 haben die EU-finanzierten Projekte massiv an Bedeutung gewonnen. Alleine im Forschungsbereich «*IST-2005-2.5.10: Access to and preservation of cultural and scientific resources*» des FP6 (2002-2006) wurden bzw. werden 18 Projekte mit einem Gesamtvolumen von über 70 Mio. Euro durchgeführt, darunter die beiden Schwergewichte PLANETS<sup>62</sup> und CASPAR<sup>63</sup> mit je rund 15 Millionen Euro. Die Forschung zur Erhaltung und Zugänglichmachung von digitalen Ressourcen aus Kultur und Wissenschaft werden im FP7 (2007 – 2013) unter dem Bereich «*ICT-2007.4.1: Digital libraries and technology-enhanced learning*» mit einem Gesamtvolumen von rund 200 Millionen Euro<sup>64</sup> weitergeführt, darunter auch das Projekt SHAMAN<sup>65</sup> mit rund 13 Millionen Euro. Hinzu kommt im FP7 der Sonderbereich «e-Infrastructures»<sup>66</sup> mit einem Budget von rund 600 Millionen Euro, aus dem auch Projekte zur Förderung der länderübergreifenden Zugänglichkeit von Forschungsdaten unterstützt werden.

Das relativ hohe Engagement der Europäischen Kommission bzw. der Europäischen Union im Bereich „Preservation and Access“ bzw. der Bewahrung von und des Zugangs zu digitalen Ressourcen aus Kultur und Wissenschaft sollte allerdings nicht darüber hinwegtäuschen, dass Grossprojekte wie PLANETS und CASPAR (wie grundsätzlich alle FP-Projekte) erstens Forschungs- und keine Infrastrukturvorhaben und zweitens bezüglich ihren Zielsetzungen wie auch den Projektpartner ausgesprochen diversifiziert sind. Damit ist ihre Wirkung auf die Bildung nationaler Infrastrukturlösungen relativ gering. Die Fördergelder teilen sich die Teilprojekte der in der Regel rund 15 öffentlichen und privaten Projektpartner, und die Projektziele reichen von allgemeinem Informationsaustausch und „Awareness Raising“ über die Erarbeitung allgemeiner Richtlinien und Konzeptgerüste („Frameworks“) und Testumgebungen („Testbeds“) bis zur Entwicklung von kleineren Software-Tools.

Die Idee ist, dass die in solchen Projekten erarbeiteten „Frameworks“ anschliessend auf nationaler Ebene konkretisiert und in Infrastrukturlösungen umgesetzt werden. Auf der nationalen Ebene fehlt dazu dann aber in der Regel die entsprechende Finanzierung und eine Verankerung in der nationalen Forschungsförderung und -politik, da die EU-Projektpartner nicht die nationalen Organe der Forschungsförderung und -politik sind, sondern Universitäten bzw. einzelne Universitätsinstitute.

<sup>62</sup> PLANETS (Preservation and long-term access to our cultural and scientific heritage), 14 Mio. € (EU-Beitrag: 8.6 Mio. €), 2006 – 2010. Die öffentlich-rechtlichen Projektpartner sind die National- und Staatsbibliotheken von Österreich, Dänemark und den Niederlanden, die Nationalarchive der Niederlande und der Schweiz (Bundesarchiv) sowie vier Universitäten aus England, Österreich und Deutschland. Zu den Industriepartnern gehören u.a. IBM und Microsoft.

<sup>63</sup> CASPAR (Cultural, artistic and scientific knowledge preservation, for access and retrieval), 15.1 Mio. € (EU-Beitrag: 8.8 Mio. €), 2006 – 2009. Die öffentlich-rechtlichen Projektpartner sind die European Space Agency (ESA), das französische Institut National de l'Audiovisuel sowie zehn Universitäten und Forschungseinrichtungen aus Frankreich, England, Italien, Griechenland und der Tschechischen Republik. Zu den Industriepartnern gehört unter anderen IBM.

<sup>64</sup> <http://cordis.europa.eu/fp7/ict/programme/>

<sup>65</sup> SHAMAN (Sustaining heritage access through multivalent archiving), 12.72 Mio. € (EU-Beitrag: 8.4 Mio. €), 2008 – 2011. Die öffentlich-rechtlichen Projektpartner sind die deutsche Nationalbibliothek, die Staats- und Universitätsbibliothek Göttingen, das flämische Parlament Belgiens sowie neun Universitäten aus Portugal, Grossbritannien, Deutschland, Schweden und den USA. Zu den Industriepartnern gehören unter anderem das PHILIPS Innovation Lab und das Xerox Research Centre Europe.

<sup>66</sup> <http://cordis.europa.eu/fp7/ict/e-infrastructure/>

## 8.2 Institutionelle Repositorien

Die einleitend aufgeführte Projekttypisierung orientiert sich an Breite des Wirkungsanspruches der Projekte. In Bezug auf diese Konzeptstudie sind aber zusätzlich auch zwei weitere, inhaltliche Unterscheidungsmerkmale relevant:

- In den meisten Projekten geht es nicht um die Aufbewahrung und Zugänglichmachung von Primärdaten, sondern um Sekundärdaten im weiteren Sinn, also neben Publikationen (inkl. Preprints und grauer Literatur) und digitalen Lehr- und Lernunterlagen auch schwergewichtig um Daten aus *Auswertungen* und *Aggregationen* von Primärdaten sowie um die *Föderation* oder *Vernetzung* von Auswertungsdaten aus heterogenen Quellen<sup>67</sup>.
- Bei den meisten Projekten liegt der Schwerpunkt nicht auf der Langzeitaufbewahrung von Unterlagen, sondern auf deren Zugänglichmachung zur breiten Nutzung über Internet, z.B. im Rahmen der Initiative OpenAccess<sup>68</sup>.

Beide Tendenzen sind nachvollziehbar, da primär die leichte und breite *Zugänglichkeit* und *Benutzbarkeit* von Ressourcen mit *ausgewerteten* und *publizierten* Informationen die wissenschaftliche Tätigkeit fördert und somit Projekten einen hohen Publizitätsgrad verschafft. Der langfristige Zugang („Access“) zu Informationen bedingt entsprechende Lösungen für deren langfristige Bewahrung und Aufbewahrung („Preservation“). Somit nennen alle Projekte beide Aspekte in einer einzelnen Zielvorgabe „Access and Preservation“. Tatsächlich ist aber bei den meisten Aktivitäten (und bei zeitlich begrenzten Projekten sowieso) die Aufbewahrung der vermittelten digitalen Ressourcen weder technisch noch finanziell langfristig gesichert.

Alleine in Europa gibt es Hunderte institutionelle Repositorien, häufig auch etwas despektierlich als „Dokumentenserver“ betitelt, die sich der Sammlung und Online-Zugänglichmachung von vor allem dokumentorientierten (textuellen) Sekundärdaten widmen, also ausgewählten Zeitschriften, Preprints, Dissertationen und grauer Literatur. Die Projekte DRIVER<sup>69</sup> und DRIVER-II<sup>70</sup> (Digital Repository Infrastructure Vision for European Research), die im Rahmen von FP6 (IST) und FP7 («e-infrastructures») von der EU mit insgesamt 5.2 Mio. Euro dotiert sind, verfolgen eine Inventarisierung und Vernetzung der institutionellen, wissenschaftlichen Repositorien mit dem Ziel «*to establish a cohesive, pan-European infrastructure of Digital Repositories, offering sophisticated functionality services to both researchers and the general public*». <sup>71</sup>

<sup>67</sup> Eines unter vielen Beispielen ist die weltweit genutzte Protein-Wissensdatenbank UniProtKB (<http://www.uniprot.org>) des Schweizerischen Instituts für Bioinformatik.

<sup>68</sup> Siehe z.B. <http://open-access.net>

<sup>69</sup> DRIVER (Digital Repository Infrastructure Vision for European Research), 2.5 Mio. € (EU-Beitrag: 1.84 Mio. €), 2006 – 2007. Die öffentlich-rechtlichen Projektpartner sind das französische Centre National de la Recherche Scientifique (CNRS), die niederländische Hochschulstiftung SURF, die Staats- und Universitätsbibliothek Göttingen, der italienische Forschungsrat (CNR) sowie fünf Universitäten aus Deutschland, Belgien, England und Polen.

<sup>70</sup> DRIVER-II (Digital Repository Infrastructure Vision for European Research), EU-Beitrag 2.7 Mio. €, 2008 – 2010. Die öffentlich-rechtlichen Projektpartner sind das französische Centre National de la Recherche Scientifique (CNRS), die niederländische Hochschulstiftung SURF, die Staats- und Universitätsbibliothek Göttingen, die Universitätsbibliothek Ghent, der italienische Forschungsrat (CNR) sowie neun Universitäten aus Griechenland, Deutschland, England, Polen, Dänemark, Portugal und Slowenien.

<sup>71</sup> <http://www.driver-repository.eu> und <http://www.driver-support.eu>, siehe auch [20].

### 8.3 Nationale Depot-Bibliotheken

Unzählige institutionelle Dokument-Repositoryn einzelner Universitäten und Forschungseinrichtungen („Dokumentenserver“, siehe Abschnitt 8.2 oben) machen Dissertationen, Preprints, graue Literatur, OpenAccess-Zeitschriften, Zitiersammlungen und weitere Informationen aller Couleur über Internet verfügbar. Darüber hinaus haben jedoch verschiedene Nationalbibliotheken und grössere Universitätsbibliotheken weltweit auch Archivlösungen für wissenschaftliche Verlagszeitschriften aufgebaut, die von den Universitäten und Hochschulen zwar immer noch in Papierform abonniert und benutzt werden, von denen die Verlage aber zusätzlich auch elektronische Exemplare anbieten. Dies ist heute die Regel, und langfristig dürften viele Verlagszeitschriften sogar nur noch in elektronischer Form erscheinen. Daraus ergibt sich eine spezifische Archivproblematik, die in Kapitel 3.2 beschrieben ist.

Neben der technischen ergibt sich jedoch vor allem eine organisatorische Problematik: Für die Aufbewahrung der Papierexemplare von abonnierten Verlagszeitschriften sind die Universitätsbibliotheken im Rahmen ihres Versorgungsauftrages für die Wissenschaften verpflichtet. Dies gilt grundsätzlich auch für die elektronischen Exemplare, doch sind vor allem kleinere Universitäten von der fachgerechten und langfristigen Aufbewahrung technisch und finanziell überfordert.

Bei den Nationalbibliotheken ist die rechtliche Situation von Land zu Land verschieden. Mindestens sind sie aber verpflichtet, die im jeweiligen Land publizierten Verlagszeitschriften im Rahmen eines Depot-Auftrages zu archivieren, der je nach Land auch ausdrücklich unabhängig vom Informationsträger gilt. Zum Beispiel archiviert die Schweizerische Nationalbibliothek neben den Publikationen der schweizerischen Hochschulen auch die in der Schweiz verlegten wissenschaftlichen Zeitschriften in Papier- und elektronischer Form als Teil der eHelvetica. Dies ist allerdings nur eine verschwindend kleine Zahl der von Schweizer Universitäten und Hochschulen abonnierten Zeitschriften.

Andere Nationalbibliotheken haben bezüglich elektronischer wissenschaftlicher Zeitschriften einen breiteren Auftrag (oder interpretieren ihren Auftrag breiter) und archivieren auch im Ausland verlegte Zeitschriften, die an den Universitäten des Landes lizenziert sind. Zum Beispiel umfasst das «e-Depot»<sup>72</sup> der niederländischen Nationalbibliothek (Koninklijke Bibliotheek, KB) über 10 Millionen digitale Objekte, von denen die meisten Publikationen von internationalen Verlagen sind. Das Ziel der KB ist es, mit den zwanzig grössten internationalen Verlagen Archivierungsvereinbarungen abzuschliessen, im Moment sind es dreizehn.

Unabhängig vom gesetzlichen Auftrag der Nationalbibliotheken stellt sich in jedem Land die Frage, welche Institutionen bereit *und* fähig sind, die Verfügbarkeit der an den Universitäten und Hochschulen des Landes lizenzierten elektronischen Zeitschriften-Exemplare langfristig zu garantieren, also auch über den Ablauf einer Lizenz oder das Verschwinden eines Verlages vom Markt hinaus (siehe Kapitel 3.2). Sinnvollerweise macht dies eine einzige Institution zentral für alle Universitäten. Sie muss dazu nur Archiv-Funktionalitäten erfüllen und keinen Online-Zugriff bieten, solange der Zugriff durch die Benutzer in den Universitäten per Internet auf die Dokumenten-Server der Verlage möglich ist.

---

<sup>72</sup> <http://www.kb.nl/dnp/e-depot/e-depot-en.html>

Eine zentrale Archivierung elektronischer Zeitschriften wird durch Nationallizenzen<sup>73</sup> erleichtert, da eine nationale Stelle Vertragspartner der Verlage ist und Einschränkungen beim inter-universitären Online-Zugriff auch auf die im Land zentral archivierten Exemplare nicht mehr bestehen.

Erwähnt sei hier auch die 2005 gegründete US-Stiftung Portico<sup>74</sup> (siehe Abschnitt 8.6.3.2), an der unter anderem auch die U.S. Library of Congress beteiligt ist. Sie bietet weltweit Verlagen und Bibliotheken an, gegen einen Unkostenbeitrag ihre digitalen Publikationen in der Portico-Infrastruktur zu archivieren und jederzeit Online darauf zuzugreifen. Dabei konvertiert Portico die Originalformate in „standardisierte Archivformate“ und übernimmt die langfristige Verantwortung für deren Migrationen in neue Formate, falls dies nötig wird. Portico enthält heute rund 8.3 Mio. Artikel und hat Archivvereinbarungen mit 66 Zeitschriftenverlagen für rund 14'000 Zeitschriftentitel.

## 8.4 Alliance for Permanent Access to the Records of Science

Zusammenfassend lässt sich sagen, dass die grosse Zahl der in den Kapitel 8.1 und 8.2 angesprochenen Aktivitäten zum Thema „Access to and preservation of cultural and scientific resources“ (wie auch der Titel des Schwerpunktbereiches IST des FP6 lautet) zwar viele allgemeine, aber nur sehr wenige direkte Bezüge zum Kernthema des vorliegenden Projektes der KUB haben, nämlich einer nationalen, fachübergreifenden Dienstleistung zur Langzeitaufbewahrung von wissenschaftlichen Primärdaten.

Es gibt allerdings auch einige Aktivitäten mit einem direkten Bezug, die jedoch finanziell auf kleinerer Flamme gekocht werden als die grossen, publizitätsträchtigen EU-Grossprojekte wie PLANETS, CASPAR oder SHAMAN.

Eine wichtige Rolle spielt die offiziell 2006 gegründete «Alliance for Permanent Access to the Records of Science in Europe» (Alliance PARSE, kurz: „Die Allianz“), die sich fachübergreifend auch mit der Aufbewahrung von Forschungsunterlagen, ausdrücklich auch von Primärdaten, befasst. Der Begriff „Records of Science“ wird dabei jedoch sehr weit gefasst [24]:

- *'Science' covers the natural sciences and engineering, mathematics and computer sciences, the life sciences and the medical sciences, and the social sciences and the humanities;*
- *'Science' is understood to include both science and technology;*
- *The 'records of science' include the 'document' record, in all its modern variety, and the 'data' record, again in all the variety in terms of data formats, media and origin, including observational data gathered primarily for operational purposes;*
- *The focus will be on public, non-military data, and where privacy and ethics pose considerable additional conditions on the preservation of medical data, the latter will not be the primary focus.*

Die Allianz legt ihre Schwerpunkte auf die Ausbildung und Verbindung von nationalen Infrastrukturen und fachübergreifenden Standards in Zusammenarbeit nicht nur mit den Gedächtnisinstitutionen, sondern vor allem direkt mit den Datenproduzenten. Dabei sollen die Anspruchsgruppen auf strategischer Ebene und nicht über Projektkooperationen eingebunden werden.

<sup>73</sup> <http://www.nationallizenzen.de>

<sup>74</sup> <http://www.portico.org>

Das heisst, dass die Allianz nicht Projekte initiieren, sondern Einfluss auf und Koordination mit den Trägern von nationalen Infrastrukturen erreichen will. Die Allianz grenzt sich somit klar von den anderen, weiter oben erwähnten bisherigen Aktivitäten der Europäischen Union ab [24]:

- *Its focus is squarely on long-term preservation and access. Inevitably this will imply some involvement with ongoing efforts to build digital collections. But the Alliance will restrict itself to activities that follow from its perspective to establish a Europe-wide infrastructure that takes care of providing long-term access for the various user communities.*
- *Its focus is on infrastructure, and not simply on a multitude of projects. Infrastructure should be taken here to mean a virtual 'infrastructure' of a number of physical archives (or organisations that function as archives) together with a set of standards and conditions that enable them to function interoperably. [...] A vital task is therefore the establishment of a consensus in various communities as to which organisations will be the mainstays of a cost-effective infrastructure for a particular community.*
- *Its focus is on science and technology in a broad sense. It not only covers all fields of science and technology including the social sciences and the humanities, but also those areas where large data and information collections are needed for operational as well as for scientific use. This provides a direct link to the specific activities of libraries (especially research libraries) as well as, for example, archives (the collections of which serve for example historical research).*
- *It is crucial to involve not only traditional custodians of collections, such as libraries, but major stakeholders in the production ('first-hand' such as major laboratories, or 'secondhand' such as scientific publishers) or users of the records of science as well.*
- *It first of all involves these stakeholders at a strategic level, and not at the level of project cooperations. Such project-level cooperations are important; after all the real work of building digital repositories starts and ends at the project level. However what Europe has missed up until now [...] is a commitment of the major stakeholders to jointly create a digital European information infrastructure, as a direct consequence of their own strategies, i.e. as a 'board-level' issue.*

Die Allianz will auch ausdrücklich die Empfehlungen der OECD (siehe 3.1.3.1 und 3.1.3.4) über den offenen Zugang zu Primärdaten aus öffentlich finanzierter Forschung unterstützen [30]. Und schliesslich legt die Allianz grosses Gewicht darauf, dass nicht nur die institutionellen Bedürfnisse von internationalen Forschungszentren, sondern auch die Situation der einzelnen Forschenden und fachlich diversifizierten Universitäten berücksichtigt und die Frage der Verantwortlichkeiten für die Aufbewahrung thematisiert wird [24]:

- *Universities, or even individual departments, seldom have explicit deposit policies. It is usually left (and mostly implicitly) to researchers or research teams, or the communities these researchers belong to, to decide whether and how to store and maintain data. Scientific journals impose certain obligations, but the incentives for researchers to spend time and money on maintaining data collections are often absent. Yet, there is a growing feeling among universities (albeit seldom at management levels) that this will require greater attention. Some are considering university repositories, and discussions are currently being held with some research councils as to who should be responsible for providing access to (in the first place) publicly-funded research. Yet it would seem unlikely – in contrast to the world of paper publications where university libraries have often served and still serve as major collections of past research output in many areas – that this will continue in the same way in the digital world.*

Die Allianz ist breit abgestützt. Zu den Mitgliedern gehören heute bereits über zwanzig namhafte Institutionen, darunter zum Beispiel die European Science Foundation, die European Space Agency, das CERN, die Max Planck Gesellschaft, das Science and Technology Facilities Council,



die British Library, die Deutsche Nationalbibliothek, die niederländische Koninklijke Bibliotheek, das britische Joint Information Systems Committee und die Digital Preservation Coalition, die International Association of Scientific, Technical and Medical Publishers und das deutsche Kompetenznetzwerk Langzeitarchivierung nestor. Die Schweiz ist in der Allianz nicht vertreten.

### 8.4.1 PARSE.insight

Einige Mitglieder der «Alliance for Permanent Access to the Records of Science in Europe» (PARSE) haben PARSE.insight (2008 – 2010) initiiert<sup>75</sup>, das von der EU im Rahmen des FP7 «e-infrastructures» mit rund 2 Mio. Euro finanziert wird. Die Ziele des Projekts<sup>76</sup>:

- Fokussierung auf Persistenz und langfristige Verstehbarkeit von Forschungsunterlagen
- Diesbezügliche Bedürfnisse der unterschiedlichen Forschungsbereiche sollen genauer untersucht werden.
- Erstellen einer einheitlichen Roadmap, welche die nationalen, europäischen und globalen Sichtweisen und Strategien integriert.
- Erstellen eines detaillierten Inventars existierender und geplanter Infrastrukturen.
- Erstellen einer Gap-Analyse aus dem Vergleich zwischen Inventar und Roadmap.
- Erstellen von Anforderungen und eines Tools zur Einflussanalyse (Impact Analysis) bezüglich der langfristigen Verfügbarkeit von Forschungsunterlagen.
- Etablieren eines Katalogs von Best Practices und von Leistungs-Benchmarks für Repositorien mit Forschungsunterlagen.

## 8.5 Deutsche Allianz-Initiative Digitale Information

An einem Rundgespräch der Deutschen Forschungsgemeinschaft (DFG) Anfang 2008 stellten die Beteiligten fest, dass [31]

- derzeit ein unkontrollierter Verlust von Primärdaten als Fakt festzustellen ist,
- diesbezüglich Handlungsbedarf besteht und die Selbstorganisation der Wissenschaft nicht überall funktioniert,
- zwischen den Fachdisziplinen eine erhebliche Heterogenität der Daten, der Metadaten, der Forschungseinrichtungen und der bestehenden Standards besteht,
- die Wissenschaftler eine zentrale Rolle für die Qualität der Daten und die Verbesserung
- der Dokumentation der notwendigen Kontextinformationen spielen und für sie daher sinnvolle Anreize zu schaffen sind.

<sup>75</sup> Science and Technology Facilities Council (UK), Koninklijke Bibliotheek (NL), Deutsche Nationalbibliothek (DE), Max Planck Gesellschaft (DE), International Association of Scientific, Technical and Medical Publishers (NL), European Space Agency ESRIN (FR), Fernuniversität in Hagen (DE), European Organization for Nuclear Research CERN (CH), Niedersächsische Staats- und Universitätsbibliothek Göttingen (DE).

<sup>76</sup> <http://www.parse-insight.eu>

Im Juni 2008 beschliesst die Allianz von neun deutschen Forschungsgesellschaften<sup>77</sup> daher die «Allianz-Initiative ‚Digitale Information‘» [25] für den Zeitraum 2008 – 2012. Als ein Schwerpunkt in dieser Initiative wird der Umgang mit Forschungsprimärdaten identifiziert [25]:

- *Es ist unbestreitbar, dass viele dieser Daten nach einer relativ kurzen Phase der Auswertung durch Einzelne oder kleine Gruppen dem Vergessen oder gar dem Verfall ausgesetzt sind. Hier sehen alle Wissenschaftseinrichtungen einen dringenden Handlungsbedarf hinsichtlich der systematischen Sicherung, Archivierung und Bereitstellung dieser Daten für die Nachnutzung durch Dritte.*
- *Ziel ist es, für alle geeigneten Fachdisziplinen, für die noch keine Infrastruktur für Forschungsprimärdaten besteht, Strukturen aufzubauen, die es ermöglichen, Forschungsprimärdaten zu sammeln, zu archivieren und für eine Weiternutzung bereitzustellen – sei es, um wissenschaftliche Ergebnisse im Sinne einer guten wissenschaftlichen Praxis überprüfbar zu machen, sei es für eine Nachnutzung im Kontext anderer Forschungsfragen. Das zentrale Element dieses Prozesses, die *Conditio sine qua non* für den Erfolg, ist die eng Kooperation zwischen Fachwissenschaftlern und Informationsdienstleistern.*

Die Allianz-Initiative ist in gewissem Sinne das deutsche Pendant zur internationalen Allianz PARSE (siehe Kapitel 8.4), nur das hier nicht einzelne Forschungseinrichtungen, sondern die Forschungsgesellschaften selber das Zepter in die Hand nehmen wollen.

Die deutsche Allianz-Initiative verfolgt drei Ziele:

- Die Entwicklung einer gemeinsamen Primärdaten-Policy, um bei Wissenschaftlerinnen und Wissenschaftlern das Bewusstsein für den Handlungsbedarf und für den Nutzen von Primärdaten-Infrastrukturen zu schärfen.
- Die Zusammenarbeit zwischen Fachwissenschaftlern und Informationswissenschaftlern anzustossen und zu begleiten sowie Förderangebote für Modellprojekte anzubieten, welche fachspezifische Standards und Prozesse definieren, um zuverlässige und gut zugängliche Archive für Forschungsprimärdaten aufzubauen und diese mit international und interdisziplinär interoperablen Zugangsschnittstellen zu versehen.
- An den Bedürfnissen der einzelnen Disziplinen orientierte und international vernetzte Repositorien- und Archivstrukturen für Forschungsprimärdaten aufzubauen.

Das dritte Ziel könne und solle gemäss der Allianz *«jedoch erst dann angegangen werden, wenn ausreichend Erfahrungen aus der Förderung und Evaluation der Modellprojekte vorliegen, um sichergehen zu können, dass neue Strukturen den Bedarf der einzelnen Fachdisziplinen aufgreifen und nicht an diesem vorbeigehen»*.

Die Allianz-Partnerorganisationen haben sich darauf verständigt, ihre Förderangebote im Bereich der Forschungsprimärdaten zu koordinieren und zu gegebener Zeit den Aufbau *gemeinsamer* Infrastrukturen für Forschungsprimärdaten zu prüfen.

<sup>77</sup> Alexander von Humboldt-Stiftung, Deutscher Akademischer Austauschdienst, Deutsche Forschungsgemeinschaft, Fraunhofer-Gesellschaft, Helmholtz-Gemeinschaft Deutscher Forschungszentren, Hochschulrektorenkonferenz, Leibniz-Gemeinschaft, Max-Planck-Gesellschaft, Wissenschaftsrat

## 8.6 Weitere ähnliche Aktivitäten

In den folgenden drei Abschnitten stellen wir weitere Aktivitäten in der Schweiz und im Ausland kurz vor, die uns für das aktuelle Projekt der KUB relevant scheinen, also einen nationalen und fachübergreifenden Anspruch in Bezug auf Primär- und Sekundärdaten der Forschung haben. Offen bleibt dabei, ob diese Aktivitäten *tatsächlich* zielführend im Sinne des KUB-Projektes sind. Dies würde eine Detailanalyse der einzelnen Aktivitäten bedingen, welche jedoch ausserhalb des Rahmens dieser Konzeptstudie liegt.

Die Liste enthält (trotz des Titels dieses Kapitels 8) auch einige Aktivitäten in den USA, Kanada und Australien. Sie umfasst jedoch prinzipiell nur solche Aktivitäten ...

- mit Fokus auf langfristiger (mind. 10 Jahre) oder dauernden *Aufbewahrung* bzw. Archivierung und ggf. Zugänglichmachung von strukturierten wissenschaftlichen Primärdaten und publizierten Sekundärdaten (Publikationen) aus *öffentlich* finanzierter Forschung;
- mit einem nationalen und inter-institutionellen Anspruch, also nicht beschränkt auf die Daten einer einzelnen Institution;
- mit einem multidisziplinären Anspruch mindestens innerhalb derselben Wissenschaft, also nicht beschränkt auf einen einzelnen, fachspezifischen Datentyp.

Deswegen entfallen im Bereich der dokumentorientierten (textuellen) Sekundärdaten insbesondere alle institutionellen „Dokumentenserver“, die sich nur mit der Sammlung und Online-Zugänglichmachung von ausgewählten Zeitschriften, Preprints, Dissertationen und grauer Literatur beschäftigen. Im Bereich der strukturierten Daten entfallen Aktivitäten, welche entweder gar keine Primärdaten sondern vielmehr aggregierte Daten bzw. Auswertungsergebnisse oder aber Primärdaten nur einer Institution betreffen<sup>78</sup>.

Der erste Teil nennt nationale und internationale Strategieprojekte, Aktionspläne und Kompetenzzentren. Der zweite Teil umfasst fachspezifische, jedoch nationale Projekte und Dienstleistungen in den Sozialwissenschaften und Naturwissenschaften. Der dritte Teil bezieht sich auf nationale Lösungen und Projekte im Bereich der Archivierung elektronischer Zeitschriften.

Im zweiten Teil haben wir uns auf die Sozialwissenschaften und Naturwissenschaften beschränkt. Dies geschieht einerseits aus Gründen des Umfangs dieser Studie. Andererseits haben diese beiden Bereiche nach unserer Einschätzung einen im Vergleich etwa zu den Wirtschafts- und Ingenieurwissenschaften einen relativ hohen Entwicklungsstand bezüglich Aufbewahrung von Primärdaten. Die medizinische Forschung wurde ausgelassen, da die Fragestellung hier weitaus komplexer ist, weil in der Regel keine klare Trennung zwischen öffentlich und privat finanzierter Forschung möglich und der Umgang mit Forschungsdaten in hohem Masse durch Gesetze und Regulierungen bestimmt ist.

---

<sup>78</sup> Aus diesem Grund erscheinen auch die (sehr grossen und bekannten) Datenzentren von nationalen und internationalen Forschungsinstitutionen wie z.B. dem CERN, der ESA oder der NASA nicht in der Liste.

## 8.6.1 Strategien, Aktionspläne und nationale Infrastrukturen

### 8.6.1.1 Internationale Aktivitäten

#### 8.6.1.1.1 PARSE, PARSE.insight

*Ziele* Vgl. Kapitel 8.4. Die europäische Alliance for Permanent Access to the Records of Science in Europe (PARSE) befasst sich strategisch und fachübergreifend mit der Förderung und Koordination europäischer Infrastrukturen für Zugang und Aufbewahrung auch von wissenschaftlichen Primärdaten.

*Organisation* Die Allianz wurde 2006 als Stiftung gegründet und wird von über 20 namhaften Organisationen aus dem europäischen Forschungs- und Bibliotheksbereich getragen, darunter auch die European Science Foundation, die European Space Agency, das CERN, die British Library, die Deutsche Nationalbibliothek und die niederländische Koninklijke Bibliotheek.

*Finanzierung* Die Allianz wird durch Beiträge ihrer Mitglieder finanziert. Das Projekt PARSE.insight wird durch die Europäische Kommission im Rahmen des FP7 «e-infrastructures» mit rund 2 Mio. Euro finanziert.

*Website* <http://www.parse-insight.eu>

#### 8.6.1.1.2 Committee on Data for Science and Technology (CODATA)

*Ziele* CODATA wurde 1966 als Arbeitsgruppe des «International Council for Science (ICSU)» gegründet und hat die allgemeine Aufgabe, die Qualität, Verwaltung und Verfügbarkeit von wissenschaftlichen Daten zu verbessern und die Zusammenarbeit international zu fördern.

*Organisation* CODATA wird von einem Exekutivkomitee aus sechs Mitgliedern geleitet und verfügt über mehrere «Task Groups» und Arbeitsgruppen. Mitglied von CODATA sind nationale Wissenschaftsförderungen, internationale Wissenschaftsverbände und einzelne wissenschaftliche Institutionen.

*Finanzierung* Die Finanzierung von CODATA wird hauptsächlich über die Beiträge ihrer Mitglieder sichergestellt.

*Website* <http://www.codata.org>

### 8.6.1.2 Nationale Aktivitäten

#### 8.6.1.2.1 Deutschland: Allianz-Initiative ‚Digitale Information‘ 2008 – 2010

*Ziele* Siehe Kapitel 8.5. Die Allianz von neun deutschen Forschungsgesellschaften sieht dringenden Handlungsbedarf bei der langfristigen Aufbewahrung und Zugänglichmachung von wissenschaftlichen Primärdaten. Sie will diesbezüglich das Bewusstsein der Wissenschaftler/innen schärfen, Modellprojekte zur Definition von gemeinsamen Standards fördern und später auch gemeinsame Infrastrukturen für Primärdaten-Archive aufbauen.

*Organisation* Getragen von 7 nationalen Forschungsgesellschaften, darunter auch die Deutsche Forschungsgemeinschaft, sowie vom Wissenschaftsrat und der Hochschulrektorenkonferenz.

*Finanzierung* Förderung von Modellprojekten durch die Allianz

#### 8.6.1.2.2 Deutschland: NESTOR

*Ziele* NESTOR (Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen in Deutschland) ist eine Informations- und Kommunikationsplattform für Institutionen, Organisationen und Unternehmen, die im Bereich der digitalen Langzeitarchivierung tätig sind oder Projekte durchführen.

---

*Organisation* NESTOR wird von der Deutschen Nationalbibliothek zusammen mit Projektpartnern getragen, von einem wissenschaftlichen Beirat begleitet und verfügt über fünf Arbeitsgruppen, die sich jeweils mit einem bestimmten Teilthema der digitalen Langzeitarchivierung befassen.

---

*Finanzierung* NESTOR wird vom Bundesministerium für Bildung und Forschung (BMBF) sowie den Projektpartnern (unter ihnen die Deutsche Bibliothek und das Deutsche Bundesarchiv) finanziert. 2006 – 2009 erhält NESTOR vom BMBF einen Beitrag von 1.5 Mio. EUR.

---

*Website* <<http://www.langzeitarchivierung.de/>>

---

#### 8.6.1.2.3 Deutschland: KoLaWiss

---

*Ziele* Das Projekt KoLaWiss («Kooperative Langzeitarchivierung für Wissenschaftsstandorte») will am Beispiel des Wissenschaftsstandortes Göttingen ein Organisations- und Geschäftsmodell für eine Langzeitarchivierung von Primärdaten der Forschung entwickeln, wobei lokale IT-Einrichtungen und deren Vernetzung mit überregionalen und internationalen Institutionen hinsichtlich LZA-Anforderungen und Ausbaubarkeit hin evaluiert werden. Dazu werden auch die an den unterschiedlichen Einrichtungen in Göttingen anfallenden digitalen Datenbestände erfasst und hinsichtlich Technik, Recht, Kosten, Organisation und Fördermassnahmen untersucht sowie Konzepte für eine Kompetenz-, Zuständigkeits- und Aufgabenteilung entwickelt. Ferner sollen Kosten-, Abrechnungs- und Geschäftsmodelle entwickelt werden.

---

*Organisation* Das einjährige Projekt (März 2008 – März 2009) KoLaWiss wird von der Gesellschaft für wissenschaftliche Datenverarbeitung mbH (GWDG) geleitet. Beteiligt sind die Staats- und Universitätsbibliothek Göttingen (SUB) sowie der Geschäftsbereich Informationstechnologie und die Abteilung Medizinische Informatik des Bereichs Humanmedizin der Georg-August-Universität Göttingen. Erste Resultate werden voraussichtlich im Dezember 2008 veröffentlicht.

---

*Finanzierung* KoLaWiss wird von der Deutschen Forschungsgemeinschaft (DFG) gefördert.

---

*Website* <<http://kolawiss.uni-goettingen.de>>

---

#### 8.6.1.2.4 Grossbritannien: Digital Curation Centre (DCC)

---

*Ziele* Das DCC versteht sich als „Center of Excellence“ im Bereich der digitalen Langzeitarchivierung mit nationalem Fokus. Ausserdem definiert sich das DCC als Anbieter von «a wide range of resources, software, tools and support services».

---

*Organisation* Das DCC-Konsortium wird von drei Abteilungen an der Universität Edinburgh, dem «National e-Science Centre», den Insituten HATII an der Universität Glasgow und UKOLN an der Universität Bath sowie dem «Science and Technology Facilities Council (STFC)» getragen.

---

*Finanzierung* Das Budget des DCC von 2007 – 2010 in Höhe von ca. 3.4 Mio £ wird zu ungefähr 80% vom «Joint Information Systems Committee (JISC)» finanziert. Die restlichen 20% stammen vom Programm e-Science des «Research Councils UK (RCUK)», der staatlichen Forschungsförderung Grossbritanniens sowie den beteiligten Universitäten Edinburgh, Glasgow und Bath.

---

*Website* <http://www.dcc.ac.uk>

---

#### 8.6.1.2.5 Frankreich: Groupe Pérennisation des Informations Numériques (PIN)

---

*Ziele* Die Groupe PIN ist eine „groupe de réflexion“ von Vertretern aus ungefähr 20 Forschungseinrichtungen, Regierungsstellen und privaten Firmen, welche sich alle drei Monate zu einer Plenarsitzung zum Austausch von Informationen und der Diskussion der Probleme zur nachhaltigen Zugänglichkeit von digitalen Informationen treffen.

---

---

*Organisation* Die Gruppe PIN ist am Centre National d'Etudes Spatiales (CNES) in Toulouse lokalisiert. Zu ihr gehören unter anderen die Archives de France, die Bibliothèque nationale de France, das Institut National de l'Audiovisuel, die France Telecom, das französische Justiz- und das Landwirtschaftsministerium.

---

*Finanzierung* keine Informationen

---

*Website* <http://vds.cnes.fr/pin/>

---

#### 8.6.1.2.6 Grossbritannien: National Digital Archive of Datasets (NDAD)

---

*Ziele* Archivieren von Datenbeständen der britischen Regierung und den staatlichen Verwaltungsbehörden.

---

*Organisation* NDAD ist organisatorisch dem britischen Nationalarchiv unterstellt, die Daten werden aber beim «University of London Computer Centre (ULCC)» gespeichert und verwaltet.

---

*Finanzierung* NDAD wird komplett aus Mitteln der britischen Regierung finanziert.

---

*Website* <http://www.ndad.nationalarchives.gov.uk>

---

#### 8.6.1.2.7 Grossbritannien: National Data Repository (NDR)

---

*Ziele* Das NDR archiviert „Daten aller Art“ nach den Wünschen seiner Kunden aus dem akademischen und kulturellen Bereich sowie für non-Profit-Organisationen und öffentliche Verwaltungen und Institutionen. Die Dienstleistungen reichen vom einfachen Deposit mit Bitstream-Preservation über Beratung und Ausbildung, Digitalisierungen, Metadaten-Katalogisierung und Web-Archivierung bis zu Konzeption, Realisierung und Hosting von kompletten, kundenspezifischen Langzeit-Archivlösungen.

---

*Organisation* Das NDR ist eine Dienstleistung des «University of London Computer Centre (ULCC)».

---

*Finanzierung* Die Dienstleistungen des NDR werden kostendeckend erbracht und die Kosten den Kunden nach Aufwand in Rechnung gestellt.

---

*Website* <http://www.ulcc.ac.uk/digital-preservation/current-activities/ndr.html>

---

#### 8.6.1.2.8 USA: National Digital Information Infrastructure and Preservation Program

---

*Ziele* Das NDIIPP entwickelt seit 2000 eine nationale Strategie zur Sammlung, Archivierung und Erhaltung insbesondere von original-digitalen Informationen, die gar nie in analoger Form existieren. Dazu gehören auch Forschungsdaten. Unter dem strategischen und finanziellen Dach des NDIIP sammeln und archivieren die Partner-Institutionen digitale Inhalte und bauen und vernetzen Archivinfrastrukturen unter anderem auch im Bereich e-Science.

---

*Organisation* Im Auftrag des Kongresses hat die Library of Congress mit NDIIPP ein nationales Netzwerk von Kompetenzträgern zur Langzeitarchivierung aus über 130 öffentlichen und privaten Partnerinstitutionen aufgebaut. 2005 ist das NDIIP eine Partnerschaft mit der National Science Foundation eingegangen.

---

*Finanzierung* In 2000 hat der Kongress für NDIIPP 100 Mio. US\$ bewilligt, deren Einsatz von der Library of Congress geleitet wird.

---

*Website* <http://digitalpreservation.gov>

---

## 8.6.1.2.9 USA: DataNet

---

<i>Ziele</i>	Das Projekt «Sustainable Digital Data Preservation and Access Network Partners (DataNet)» ist ein erst im Planungsstadium befindliches Programm der amerikanischen Forschungsförderung «National Science Foundation (NSF)» zum Aufbau und Betrieb eines dezentralen Netzwerks zur Archivierung von wissenschaftlichen Daten ab 2010.
<i>Organisation</i>	Die Organisation von DataNet ist noch nicht festgelegt. Interessenten können sich bis Mitte 2009 bewerben. Die bis zu fünf Partner müssen sich stark miteinander vernetzen, um die langfristige Verfügbarkeit der ihnen anvertrauten Daten sicherstellen zu können.
<i>Finanzierung</i>	DataNet wird vollumfänglich von der NSF finanziert. Für das Programm stehen insgesamt etwa 100 Mio. US\$ zur Verfügung, die über eine Zeitdauer von fünf Jahren an bis zu fünf Institutionen ausgerichtet werden sollen.
<i>Website</i>	<a href="http://www.nsf.gov/pubs/2007/nsf07601/nsf07601.htm">http://www.nsf.gov/pubs/2007/nsf07601/nsf07601.htm</a>

---

## 8.6.1.2.10 Kanada: National Data Archive, Data Canada

---

<i>Ziele</i>	Eine Arbeitsgruppe des «Social Sciences and Humanities Research Council of Canada (SSHRC)» und des kanadischen Nationalarchivs hat im Jahr 2002 eine Studie zu einem «National Data Archive» durchgeführt. 2005 kam eine weitere Studie der «Task Force for National Consultation on Access to Scientific Research Data» zu ähnlichen Ergebnissen. Allerdings ist seit dem Erscheinen dieser beiden Berichte offenbar nichts Nennenswertes mehr geschehen.
<i>Organisation</i>	Vorgeschlagen wird im ersten Bericht entweder die Errichtung eines Netzwerks von Universitäten und Institutionen mit einer zentralen Koordinationsstelle oder die Schaffung einer nationalen Agentur unter der Leitung des kanadischen Nationalarchivs. Der zweite Bericht schlägt die Schaffung einer zentralen Institution namens «Data Canada» vor, die von allen wissenschaftlichen Institutionen Kanadas getragen werden soll.
<i>Finanzierung</i>	Beide Berichte sehen vor, dass die Finanzierung hauptsächlich mit Fördermitteln des kanadischen Staates sichergestellt werden müsste.
<i>Website</i>	<a href="http://www.sshrc.ca/web/about/publications/da_finalreport_e.pdf">http://www.sshrc.ca/web/about/publications/da_finalreport_e.pdf</a> <a href="http://ncasrd-cnads.scitech.gc.ca/finalrpt_05_e.shtml">http://ncasrd-cnads.scitech.gc.ca/finalrpt_05_e.shtml</a>

---

## 8.6.1.2.11 Australien: Australia National Data Service, Australian Data Commons

---

<i>Ziele</i>	ANDS plant den Aufbau eines nationalen Zentrums für die Speicherung und langfristige Verfügbarkeit von wissenschaftlichen Daten. ANDS wirkt dabei als zentrale Koordinationsstelle während die Daten weiterhin dezentral gespeichert werden. Die Daten sollen dabei in einem Australian Data Commons aufgehen und in diesem gemeinsamen Datenpool durchsuchbar, benutzbar und langfristig verfügbar gehalten werden.
<i>Organisation</i>	Die Organisationsstruktur von ANDS ist noch nicht definitiv festgelegt. Zurzeit sind die Monash University in Melbourne, die Australian National University (ANU) in Canberra und die Commonwealth Scientific and Industrial Research Organisation (CSIRO) am Aufbau von ANDS beteiligt. Später soll ein Steuerungsausschuss geschaffen werden, in dem alle Partner von ANDS einen Sitz haben. Geleitet wird ANDS von einem Direktor sowie je einem Vizedirektor an der Monash University und an der Australian National University. Sitz von ANDS ist die Monash University. Die Mitarbeiterzahl soll ungefähr 20 bis 22 Personen umfassen.

---

---

**Tätigkeit** Das Tätigkeitsgebiet von ANDS umfasst die folgenden vier Bereiche: *Developing Frameworks*: Aufbau eines gemeinsamen gesetzlichen und institutionellen Rahmens für wissenschaftliche Daten und deren Speicherung, Verwaltung und Nutzung. *Providing Utilities*: Entwickelt Werkzeuge für die Suche, den Zugriff und die Nutzung der gespeicherten wissenschaftlichen Daten (gemeinsames Portal, einheitliche Identifikation etc.). *Seeding the Commons*: Bietet den angeschlossenen Institutionen Unterstützung bei der Speicherung und langfristigen Verfügbarkeit ihrer wissenschaftlichen Daten. *Building Capabilities*: Aufbau von Wissen für die Generierung, Speicherung, Verwaltung und langfristige Verfügbarkeit von wissenschaftlichen Daten.

---

**Finanzierung** Zurzeit existiert nur ein Budgetentwurf, der bis ins Jahr 2011 Aufwendungen von insgesamt 24.5 Mio AU\$ vorsieht. Der grösste Teil der Kosten wird vom Programm «National Collaborative Research Infrastructure Strategy (NCRIS)» des australischen «Department of Innovation, Industry, Science and Research (DIISR)» bestritten. Es ist geplant, dass sich auch die Partner-Institutionen an der Finanzierung von ANDS beteiligen, sei dies durch Personal, Dienstleistungen oder Produkte. Es wird geschätzt, dass diese Zuwendungen ungefähr gleich hoch sein werden wie die direkt bezahlten Mittel durch das Programm NCRIS.

---

**Website** <http://ands.org.au>

---

## 8.6.2 Fachspezifische Aufbewahrung von Primärdaten der Forschung

### 8.6.2.1 Sozialwissenschaften

#### 8.6.2.1.1 International: Council of European Social Science Data Archives (CESSDA)

**Ziele** Hauptzweck ist der CESSDA-Katalog, über den Online in den Repositorien der nationalen Partner gesucht und Daten bezogen werden können. Zusätzlich organisiert der Verband jährliche „Expert Seminars“ zur Archivierungsproblematik in den Partnerländern.

---

**Organisation** CESSDA ist der Dachverband von 20 nationalen Datendiensten der Sozialwissenschaften.

---

**Finanzierung** CESSDA finanziert sich hauptsächlich durch Beiträge der Partner.

---

**Website** <http://www.cessda.org>

---

#### 8.6.2.1.2 Schweiz: Informations- und Datenarchivdienst für die Sozialwissenschaften (SIDOS)

**Ziele** SIDOS archiviert Daten der sozialwissenschaftlichen Forschung und führt ein Forschungsinventar.

---

**Organisation** Geründet 1992 ist SIDOS seit dem 1. Januar 2008 ein Teil der «Fondation suisse pour la recherche en sciences sociales (FORS)» und hat seinen Sitz an der Universität Lausanne.

---

**Finanzierung** SIDOS finanziert sich hauptsächlich durch Subventionen der Schweizerischen Akademie der Geisteswissenschaften (SAGW).

---

**Website** <http://www.sidos.ch>

---

#### 8.6.2.1.3 Deutschland: Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen (GESIS)

**Ziele** GESIS-ZA archiviert Primärmaterial (Daten, Fragebögen, Codepläne) und Ergebnisse empirischer Untersuchungen. Das Material wird für wissenschaftliche Sekundäranalysen aufbereitet und der interessierten Öffentlichkeit zugänglich gemacht.

---

**Organisation** Das frühere «Zentralarchiv für Empirische Sozialforschung» ist jetzt eine Abteilung des neu gegründeten Instituts GESIS. Es besteht eine enge Zusammenarbeit mit der Universität Köln.

---



---

*Finanzierung* Das GESIS finanziert sich durch staatliche Fördergelder, die der Bund und die Länder im Rahmen verfügbarer Haushaltsmittel und auf der Grundlage der zwischen Bund und Ländern getroffenen Vereinbarungen zur Verfügung stellen. Es werden keine Mitgliederbeiträge erhoben.

---

*Website* <http://www.gesis.org>

---

#### 8.6.2.1.4 Grossbritannien: UK Data Archive (UKDA)

---

*Ziele* Das UKDA wurde 1967 gegründet und ist ein Kompetenzzentrum für die Erstellung, die Erhaltung und Nutzung sozialwissenschaftlicher Daten und verfügt mit mehr als 5 000 Datenbeständen über das grösste Archiv in diesem Bereich in Grossbritannien. Das UKDA wird vom britischen Nationalarchiv als Aufbewahrungsort («Place of Deposit») für digitale Unterlagen anerkannt.

---

*Organisation* Das UKDA mit Sitz an der Universität von Essex in Colchester und ist eine von vier Partnerorganisationen des 2003 gegründeten «Economic and Social Data Service (ESDS)». Die dreizehn Abteilungen des UKDA mit rund 50 Mitarbeiter/innen werden von einem Direktor geleitet..

---

*Finanzierung* Das UKDA wird vom «Economic and Social Research Council (ESRC)», dem «Joint Information Systems Committee (JISC)» und der Universität von Essex finanziert.

---

*Website* <http://www.data-archive.ac.uk>

---

#### 8.6.2.1.5 USA: Data Preservation Alliance for the Social Sciences (Data-PASS)

---

*Ziele* Ziel des Projektes ist es, zusammen mit den Partnern archiwürdige, vom Verlust bedrohte Datenbestände aus den Sozialwissenschaften zu identifizieren, zu retten und mit Hilfe einer über die Partner verteilten Speicherplattform langfristig zu archivieren. Dazu gehören insbesondere Meinungsumfragen, Wahlresultate und breitangelegte Untersuchungen über Familienwachstum und –einkommen.

---

*Organisation* Das Projekt Data-PASS wird zurzeit von sieben Partnerinstitutionen in den USA getragen, darunter die «Library of Congress», das amerikanische Nationalarchiv («National Archives and Record Administration NARA») sowie das ICPSR (siehe 8.6.2.1.6).

---

*Finanzierung* Data-PASS finanziert sich über einen Beitrag des «National Digital Information Infrastructure and Preservation Program (NDIIPP)» der «Library of Congress» (siehe 8.6.1.2.8).

---

*Website* <http://www.icpsr.umich.edu/DATAPASS>

---

#### 8.6.2.1.6 USA: Inter-university Consortium for Political and Social Research (ICPSR)

---

*Ziele* Das ICPSR übernimmt und archiviert Daten aus der sozialwissenschaftlichen Forschung und stellt diese Daten der Öffentlichkeit wieder zur Verfügung. Ein Schwerpunkt des ICPSR ist dabei die aktive Unterstützung der Benutzer im effektiven Gebrauch der Daten, indem umfassende Dokumentationen der Daten erstellt und Trainings zu Kollektionen sowie zu analytischen Methoden zu ihrer Auswertung angeboten werden.

---

*Organisation* Das ICPSR wurde 1962 gegründet und hat heute weltweit ungefähr 660 Mitglieder (vornehmlich Universitäten). Geleitet wird ICPSR von einem Vorstand mit zwölf Mitgliedern, die ehrenamtlich für vier Jahre gewählt werden. Der Sitz von ICPSR befindet sich an der Universität von Michigan in Ann Arbor und beschäftigt ungefähr 27 vollamtliche Mitarbeiter/innen.

---

*Finanzierung* 22% der Einnahmen von 14.25 Mio US\$ im Jahre 2007 stammten aus Mitgliederbeiträgen, weitere 46% aus zweckgebundenen Forschungsgeldern für Projekte sowie aus Spenden.

---

*Website* <http://www.icpsr.umich.edu>

---

## 8.6.2.2 Naturwissenschaften

### 8.6.2.2.1 International: ICSU World Data Centers (WDC)

<i>Ziele</i>	Die World Data Centers (WDC) sind ein Verbund von dezentralen Einheiten, die wissenschaftliche Daten eines bestimmten Fachgebiets archivieren und der Wissenschaft zur Verfügung stellen.
<i>Organisation</i>	Die 52 WDC in 12 Ländern sind unter dem Dach des International Council for Science (ICSU) zusammengeschlossen, es herrscht allerdings nur eine lose Bindung. Die Informationen auf der Website der ICSU sind zudem mehrere Jahre alt und seit mindestens zwei Jahren nicht mehr aktualisiert worden.
<i>Finanzierung</i>	Es gibt keine zentrale Finanzierung, jedes WDC hat ein separates Budget, das meistens von einer Universität, einer Forschungsinstitution oder einem Forschungsprojekt im Gastland getragen wird.
<i>Website</i>	<a href="http://www.ngdc.noaa.gov/wdc/">http://www.ngdc.noaa.gov/wdc/</a>

### 8.6.2.2.2 USA: SDSC Data Central, San Diego Supercomputing Center (SDSC)

<i>Ziele</i>	Bereitstellung von Know-how, Dienstleistungen und Speicherkapazität für die Archivierung von Daten öffentlich und privat finanzierter wissenschaftlicher Forschungsprojekte, Bibliotheken und Archiven in den USA. Das SDSC archiviert zurzeit rund 120 wissenschaftliche Datensammlungen mit einem Gesamtumfang von 5 Petabytes, wobei die Gesamtkapazität 27 Petabytes beträgt.
<i>Organisation</i>	Das SDSC ist eine Forschungseinheit der University of California, San Diego und zählt ungefähr 400 Mitarbeiter. Die Data Central ist Teil des SDSC.
<i>Finanzierung</i>	Die Finanzierung des SDSC wird hauptsächlich durch Mittel der «National Science Foundation (NSF)» sichergestellt. Das Archivieren von Daten ist kostenlos, allerdings müssen die Interessenten dazu einen Antrag stellen und die Kriterien der National Science Foundation erfüllen.
<i>Website</i>	<a href="http://datacentral.sdsc.edu">http://datacentral.sdsc.edu</a>

### 8.6.2.2.3 Deutschland: World Data Center for Climate (WDCC)

<i>Ziele</i>	Das WDCC speichert keine Primärdaten, sondern übernimmt aggregierte, nach Fragestellungen ausgewertete Datenprodukte zu Klimamodellen und Messreihen und stellt sie Wissenschaftlern und Institutionen zur Verfügung. Lieferanten sind das Deutsche Klimarechenzentrum (DKRZ) sowie einige internationale Datenzentren zur Klimaforschung.
<i>Organisation</i>	Das WDCC befindet sich am Max-Planck-Institut für Meteorologie in Hamburg, das wiederum Teil des Zentrums für Meeres- und Klimaforschung an der Universität Hamburg ist. Das WDCC arbeitet zudem eng mit dem Deutschen Klimarechenzentrum (DKRZ) in Hamburg zusammen, dessen Gesellschafter wiederum die Max-Planck-Gesellschaft und die Universität Hamburg sind.
<i>Finanzierung</i>	Die Finanzierung des WDCC wird von den vier Gesellschaftern des DKRZ und vom Bundesministerium für Bildung und Forschung (BMBF) sichergestellt.
<i>Umfang</i>	Der Datenbestand des WDCC umfasste Ende 2007 ungefähr 340 Terabytes, die in mehr als 6 Billionen BLOBs (Binary Large Objects) in einer Datenbank abgespeichert sind. Neben selbst erzeugten Daten werden auch thematisch passende Daten aus externen Quellen übernommen und mit Metadaten versehen abgespeichert. Die Menge der externen Daten am gesamten Datenbestand beträgt ungefähr 5%.

---

*Dienste* Das WWDC bietet neben der Speicherung von Klimadaten weitere Dienstleistungen an. Die Integrierende Modell- und Daten-Infrastruktur (IMDI) ist eine Umgebung zur Kompilierung und Ausführung von Erdsystemmodellen und zur Analyse und Archivierung der Ergebnisse. Datenentitäten, deren Bearbeitung abgeschlossen ist, als eigenständige Entität gelten und zentrale Bedeutung im wissenschaftlichen Arbeitsprozess besitzen, werden mit einem persistenten Identifikator, dem sogenannten Digital Object Identifier (DOI) versehen und über den zentralen Bibliothekskatalog der Technischen Informationsbibliothek (TIB) in Hannover veröffentlicht.

*Website* <http://wdc-climate.de>

---

### 8.6.3 Archivierung elektronischer Verlagszeitschriften

Die meisten Nationalbibliotheken verfügen über Archivlösungen für elektronische Zeitschriften, digitalisierte Zeitschriften und Bildersammlungen sowie teilweise auch für die Websites der nationalen Internet-Domäne. (Siehe dazu Kapitel 8.3.) Wir haben nur drei europäische Beispiele sowie ein internationales Angebot ausgewählt. Die Liste ist bei weitem nicht vollständig.

#### 8.6.3.1 Niederlande: e-Depot Koninklijke Bibliotheek

---

*Ziele* Das e-Depot der niederländischen Nationalbibliothek (Koninklijke Bibliotheek, KB) ist eine Langzeitarchivlösung für digitale Publikationen und Objekte, basierend auf Hardware und Software (DIAS) der Firma IBM. Es enthält die nationale digitale Depot-Kollektion der KB sowie in Zukunft das niederländische Webseiten-Archiv und Master-Exemplare von Digitalisaten der KB.

---

*Mengen* Das e-Depot der KB umfasst heute 5 000 elektronische Zeitschriften-Titel (eJournals) mit 10 Mio. Artikeln. Die Datenmenge beträgt 11 TB; die Speicherkapazität von 25 TB kann bis 1,5 PB ausgebaut werden. Die KB hat bisher „Archiving Agreements“ mit 13 Zeitschriftenverlagen abgeschlossen, welche die digitalen Zeitschriften regelmässig für den Import in das e-Depot liefern.

---

*Finanzierung* Die Investitionen für das e-Depot sind nicht genannt. Die internen Kosten der KB für „Digital Archiving and R&D Digital Preservation“ werden mit jährlich 1.1 Mio. EUR für Personal und Systemunterhalt sowie 1.2 Mio. EUR für Forschung angegeben.

*Website* <http://www.kb.nl/dnp/e-depot/e-depot-en.html>

---

#### 8.6.3.2 Deutschland: Kopal

---

*Ziele* Kopal betreibt ein Archivsystem für die Langzeitverfügbarkeit digitaler Publikationen. Dies wird als kooperativer Service Bibliotheken, Archive und Museen sowie Universitäten und Forschungseinrichtungen angeboten. Die technische Grundlage von Kopal ist die Software DIAS von IBM, welche auch die niederländischen Nationalbibliothek verwendet (siehe 8.6.3.1).

---

*Organisation* Kopal wird von der Deutschen Nationalbibliothek (DNB), der Niedersächsischen Staats- und Universitätsbibliothek (SUB), der Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG) und der Firma IBM getragen. Der technische Betrieb des Langzeitarchivs ist bei GWDG angesiedelt. Die DNB und SUB nutzen Kopal auch zur Archivierung ihrer eigenen Publikationen.

---

*Finanzierung* Das Projekt (2004 – 2007) zum Aufbau des heutigen Archivsystems wurde vom Bundesministerium für Bildung und Forschung (BMBF) mit Mitteln in Höhe von 4.2 Mio EUR gefördert.

---

---

*Dienste* Kopal bietet prinzipiell drei unterschiedliche Nutzungsmodelle an: *Teilnehmer*: Eine Institution lässt ihre Daten kommissarisch durch einen Mandanten archivieren. *Mandant*: Eine Institution verwaltet selbstständig einen eigenen Bereich (Schliessfach) des Kopal-Archivsystems der DNB. *Eigenbetrieb*: Eine Institution betreibt unter Rückgriff auf Erfahrungen des kopal-Projekts ein eigenes vollständiges Archivsystem mit der Kopal-Software.

---

*Website* <http://kopal.langzeitarchivierung.de>

---

### 8.6.3.3 Schweiz: e-Helvetica

---

*Ziele* Im Rahmen des Projektes e-Helvetica archiviert die Schweizerische Nationalbibliothek (NB) elektronische Publikationen, welche unter eHelvetica fallen. Diese beinhaltet neben amtlichen Schriften, Hochschulschriften und Dissertationen sowie Webseiten aus der .ch-Domain auch elektronische Zeitschriften, die in der Schweiz verlegt werden (die allerdings nur einen verschwindend kleinen Teil der an den Universitäten des Landes genutzten Zeitschriften ausmachen).

---

*Finanzierung* Die Finanzierung des Projektes e-Helvetica erfolgt durch die Mittel der NB sowie Beiträgen aus den eGovernment-Mitteln des Bundes.

---

*Dienste* Die NB bietet Dritten die Aufbewahrung von digitalen Daten nach dem Schliessfachprinzip an.

---

*Website* [http://www.nb.admin.ch/slb/slb\\_professionnel/01693/index.html?lang=de](http://www.nb.admin.ch/slb/slb_professionnel/01693/index.html?lang=de)

---

#### 8.6.3.1 Frankreich: Hyper Article en Ligne (HAL)

---

*Ziele* Ziel ist der Aufbau eines nationalen und offenen Archivs für wissenschaftliche Publikationen.

---

*Organisation* HAL ist ein Projekt des «Centre pour la Communication Scientifique Directe (CCSD)» und als solches Teil des «Centre national de la recherche scientifique (CNRS)», der staatlichen französischen Forschungsförderung. HAL wird von zahlreichen wissenschaftlichen Organisationen innerhalb Frankreichs getragen, die auch den strategischen Ausschuss bilden (zusammen mit einigen Vertretern der Regierung). Betrieben wird HAL vom CCSD, das insgesamt sechs Personen dafür beschäftigt.

---

*Finanzierung* Die Finanzierung von HAL wird hauptsächlich durch das CCSD sichergestellt.

---

*Umfang* HAL verfügt zurzeit über mehr als 100'000 wissenschaftliche Publikationen.

---

*Website* <http://hal.archives-ouvertes.fr>

---

#### 8.6.3.2 International: Portico

---

*Ziele* Portico bietet weltweit Bibliotheken und Verlagen an, ihre digitalen Publikationen in der Portico-Infrastruktur zu archivieren und jederzeit Online darauf zuzugreifen. Dabei konvertiert Portico die Originalformate in „standardisierte Archivformate“ und übernimmt die langfristige Verantwortung für Migrationen in neue Formate, falls solche nötig werden. Dazu müssen die Kunden eine (nicht-exklusive) „Archivilizenz“ unterzeichnen, durch welche sie Portico zur Konversion und späteren Migration ihrer Inhalte berechtigen.

---

*Organisation* Portico ist eine 2005 in den USA gegründete non-profit-Stiftung der Bildungsstiftungen JSTOR und Ithaka sowie der U.S. Library of Congress und Andrew W. Mellon Foundation.

---

## 8 Europäisches Umfeld

---

*Finanzierung* Portico ist für die Kunden kostenpflichtig. Bei den Verlagen beträgt der jährliche Beitrag 0.03 – 0.1% vom gesamten Jahresumsatz des Verlags aus Zeitschriften. Die Beiträge von Bibliotheken betragen 0.1 – 0.5% ihrer jährlichen totalen Sachmittelausgaben. Weitere wesentliche finanzielle Beiträge erhält Portico von seinen Stiftern (siehe oben).

---

*Umfang* Portico enthält zur Zeit rund 8.3 Mio. Artikel aus 5'600 Zeitschriftentiteln von 66 teilnehmenden Verlagen und 476 teilnehmenden Bibliotheken. Für weitere 8'085 Titel bestehen Vereinbarungen mit Verlagen.

---

*Website* <http://www.portico.org>

---

## 9 Bibliographie

1. Wolfram Neubauer, *Antrag für Einzelvorhaben im Rahmen des nationalen Projektes Elektronische Bibliothek Schweiz „E-lib.ch“: Konzeptentwicklung eines Modells für eine zentrale Langzeitarchivierung von digitalen Primär- und Sekundärdaten für die Schweiz*, 20.08.2008.
2. *Bundesgesetz über die Archivierung (BGA, SR 152.1)*. 26.6.1998.
3. Pascalia Boutsouci, Daniel Metje, *Digitale Langzeitarchivierung. Erhebung an der ETH Zürich und der Georg-August-Universität Göttingen*, in *Deutscher Bibliothekartag 2008, Forum Langzeitarchivierung, Mannheim, 5. Juni 2008*. 2008.
4. Thomas Severiens, Eberhard R. Hilf, Institute for Science Networking Oldenburg GmbH and an der Carl von Ossietzky Universität Oldenburg, *Studie zum Stand vorhandener Forschungsdaten und Rohdaten aus wissenschaftlichen Tätigkeiten: Erfordernisse und Eignung zur Archivierung bzw. Zurverfügungstellung in Deutschland (Primärdaten)*. in *nestor - Materialien 6* (<http://nbn-resolving.de/urn:nbn:de:0008-20051114018>). 2006.
5. American Economic Review. *Data Availability Policy* ([http://www.aeaweb.org/aer/data\\_availability\\_policy.html](http://www.aeaweb.org/aer/data_availability_policy.html)). 2008.
6. European Science Foundation (ESF), *Good scientific practice in research and scholarship*. European Science Foundation Policy Briefing, 2000(10).
7. Organisation for Economic Co-Operation and Development (OECD), *OECD Principles and Guidelines for Access to Research Data from Public Funding* (<http://www.oecd.org/dataoecd/9/61/38500813.pdf>). 2007.
8. Deutsche Forschungsgemeinschaft (DFG), *Sicherung guter wissenschaftlicher Praxis: Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“* (<http://www.dfg.de/antragstellung/gwp/>). 1998, Weinheim: Wiley-VCH.
9. *Bundesgesetz über die Forschung (SR 420.1)*. 7. Oktober 1983 (Fassung vom 25. Februar 2008).
10. Akademien der Wissenschaften Schweiz, *Wissenschaftliche Integrität - Grundsätze und Verfahrensregeln* (<http://www.akademien-schweiz.ch/Publikationen/Richtlinien/>). 2008, Bern.
11. Schweizerischer Nationalfonds, *Beitragsreglement. Reglement des Schweizerischen Nationalfonds über die Gewährung von Beiträgen* ([http://www.snf.ch/SiteCollectionDocuments/allg\\_reglement\\_d.pdf](http://www.snf.ch/SiteCollectionDocuments/allg_reglement_d.pdf)). 14.12.2007 (vom Bundesrat genehmigt am 13. Februar 2008).
12. Schweizerischer Nationalfonds, *Reglement über die Information, die Valorisierung und die Rechte an Forschungsergebnissen*, N. Forschungsrat, Editor. 17.6.2008.
13. Senat der Universität Bern, *Reglement über die wissenschaftliche Integrität vom 1.5.2007* ([http://www.forschung.unibe.ch/content/richtlinien/integritaet/index\\_ger.html](http://www.forschung.unibe.ch/content/richtlinien/integritaet/index_ger.html)).
14. Schulleitung ETH Zürich, *Richtlinien für Integrität in der Forschung und gute wissenschaftliche Praxis an der ETH Zürich vom 14. November 2007* ([http://www.rechtssammlung.ethz.ch/pdf/414\\_Integrit%C3%A4t-Forschung.pdf](http://www.rechtssammlung.ethz.ch/pdf/414_Integrit%C3%A4t-Forschung.pdf)).
15. Tony Mayer, Nick Steneck, *Final Report to ESF and ORI: First World Conference on Research Integrity: Fostering Responsible Research (Lisbon, Portugal, 16-19 September 2007)*:

- [http://www.icsu.org/5\\_abouticsu/PDF/WC\\_final\\_report.pdf](http://www.icsu.org/5_abouticsu/PDF/WC_final_report.pdf). 2007, European Science Foundation, and U.S. Office of Research Integrity.
16. Peter Doorn, Heiko Tjalsma, *Introduction: archiving research data*. Archival Science, 2007. **7**(1): p. 1-20.
  17. Matthias Töwe unter Mitarbeit von Arlette Piguet, *Konzeptstudie E-Archiving, Version 1.2* ([http://lib.consortium.ch/external\\_files/Konzeptstudie\\_D\\_V1\\_2.pdf](http://lib.consortium.ch/external_files/Konzeptstudie_D_V1_2.pdf)). 2005, Konsortium der Schweizer Hochschulbibliotheken: Zürich.
  18. Peter Keller-Marxer, Niklaus Bütikofer, «E-lib.ch»-Projekt „Entwicklung eines Modells für eine zentrale Langzeitarchivierung von digitalen Primär- und Sekundärdaten für die Schweiz“: *Workshop II mit erweiterem Personenkreis, 26.08.2008, ETH Zürich. Bericht zum Workshop*. 2008.
  19. *Verordnung zum Bundesgesetz über die Archivierung (VBGA, SR 152.11)*. 8.9.1999.
  20. *Botschaft des Bundesrates über das Bundesgesetz über die Archivierung*. Bbl, 26.2.1997. **1997**(II).
  21. Archiv der ETH Zürich, *Richtlinien für die Bewertung von Unterlagen am Archiv der ETH Zürich* ETH-Bibliothek, Editor. 2002.
  22. Staatskanzlei Kanton Bern, *Gesetz über die Archivierung. Fassung vom 12. Juni 2008 für die Vernehmlassung*. 2008.
  23. C. Wallace et al., *IETF RFC 4810: Long-Term Archive Service Requirements* (<http://www.ietf.org/rfc/rfc4810.txt>). 2007, Internet Engineering Task Force (IETF).
  24. European Task Force Permanent Access, *PERMANENT ACCESS TO THE RECORDS OF SCIENCE: Strategic Action Programme 2006 - 2010* (<http://www.alliancepermanentaccess.eu/Strategic%20Action%20Programme.pdf>). 2005, Amsterdam: Koninklijke Bibliotheek.
  25. Deutsche Forschungsgemeinschaft, Hochschulrektorenkonferenz, Wissenschaftsrat, Max-Planck-Gesellschaft, Leibniz-Gemeinschaft, Fraunhofer-Gesellschaft, Helmholtz-Gemeinschaft Deutscher Forschungszentren, Alexander von Humboldt-Stiftung u.a., *Schwerpunktinitiative „Digitale Information“ der Allianz-Partnerorganisationen: Allianz-Initiative Digitale Information – Das Leitbild* ([http://www.dfg.de/aktuelles\\_presse/das\\_neueste/download/pm\\_allianz\\_digitale\\_information\\_details\\_080612.pdf](http://www.dfg.de/aktuelles_presse/das_neueste/download/pm_allianz_digitale_information_details_080612.pdf)).
  26. *SR943.03: Bundesgesetz vom 19. Dezember 2003 über Zertifizierungsdienste im Bereich der elektronischen Signatur (Bundesgesetz über die elektronische Signatur, ZertES)*. 2003.
  27. Consultative Committee for Space Data Systems (CCSDS), *ISO 14721:2003: Space data and information transfer systems - Open archival information system - Reference model*. 2003, Geneva: International Organization for Standardization (ISO).
  28. nestor-Arbeitsgruppe Vertrauenswürdige Archive - Zertifizierung, *Kriterienkatalog vertrauenswürdige digitale Langzeitarchive, Version 1*: <http://nbn-resolving.de/urn:nbn:de:0008-2006060710>. nestor Materialien. Vol. 8. 2006, Frankfurt am Main: nestor - Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland.
  29. Inc. (OCLC) The Center for Research Libraries (CRL) and Online Computer Library Center, *Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC), Version 1.0, February 2007*: <http://www.crl.edu/PDF/trac.pdf>. 2007, Chicago: The Center for Research Libraries (CRL) and Online Computer Library Center, Inc. (OCLC).
  30. Luigi Fusco, *Co-ordinating national efforts: Alliance for Permanent Access* (<http://www.wepreserve.eu/events/fp6-2007/presentations/Co-ordinating-national-efforts-luigi.ppt>), in *DPE, PLANETS, and CASPAR Second Annual Conference: Progress towards Addressing Digital Preservation Challenges, 5-6 September 2007*. 2007.
  31. Thekla Kluttig, *Bericht über das DFG-Rundgespräch „Forschungsprimärdaten“ am 17.01.2008* ([http://www.dfg.de/forschungsfoerderung/wissenschaftliche\\_infrastruktur/lis/download/forschungsprimaerdaten\\_0108.pdf](http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/forschungsprimaerdaten_0108.pdf)). 2008: Deutsche Forschungsgemeinschaft (DFG).