

s Austria führte An- fang
eine umfassende Untersu-
g mit Forschungsdaten in Öster-
000 WissenschaftlerInnen von 20
täten sowie drei
Forschungseinrich-
reich beteiligten
frage2. In Bezug

E-INFRASTRUCTURES AUSTRIA DELIVERABLE Cluster D

raum von Juli
rsitätsbibliothek Wien unterschiedliche europäi-
olicies sowohl formell als auch inhaltlich analy-
dingungen für ein kompetentes Forschungsda-
zu Beginn 2016 die ExpertInnengruppe - Strate-

ter-
schungs-
licy an österreichi-
tungen zu erstellen. Das vorliegende Doku-
orschungseinrichtung lokalisiert und an die
eigenen Institution angepasst werden
n der Sitzung der ExpertInnengrup-

Kostenmodell für die Archivierung von Forschungsdaten

Work-Package-Cluster:	Cluster D: Aufbau Infrastruktur
Leitung des Clusters:	Raman Ganguly Universität Wien raman.ganguly@univie.ac.at
Datum:	11.08.2016
Version:	1.0
AutorInnen/ Sonstige Beteiligte:	Michael Birkner AK Bibliothek Wien michael.birkner@akwien.at David Mitterhuber Akademie der bildenden Künste Wien d.mitterhuber@akbild.ac.at Raman Ganguly Universität Wien raman.ganguly@univie.ac.at Gerhard Gonter Universität Wien gerhard.gonter@univie.ac.at Christoph Ladurner TU Graz christoph.ladurner@tugraz.at
Schlagwörter (Deutsch):	Repositories, Kosten, Archivierung
Keywords (English):	repositories, costs, archiving



Inhalt

1. Einleitung.....	4
2. Kosten umgelegt auf das Workflow-Modell.....	4
2.1 Kosten beim Ingest	5
2.2 Kosten beim Management	5
2.3 Kosten beim Re-Use.....	5
2.4 Zusätzlicher Entwicklungsaufwand	5
3. Beispiel	6

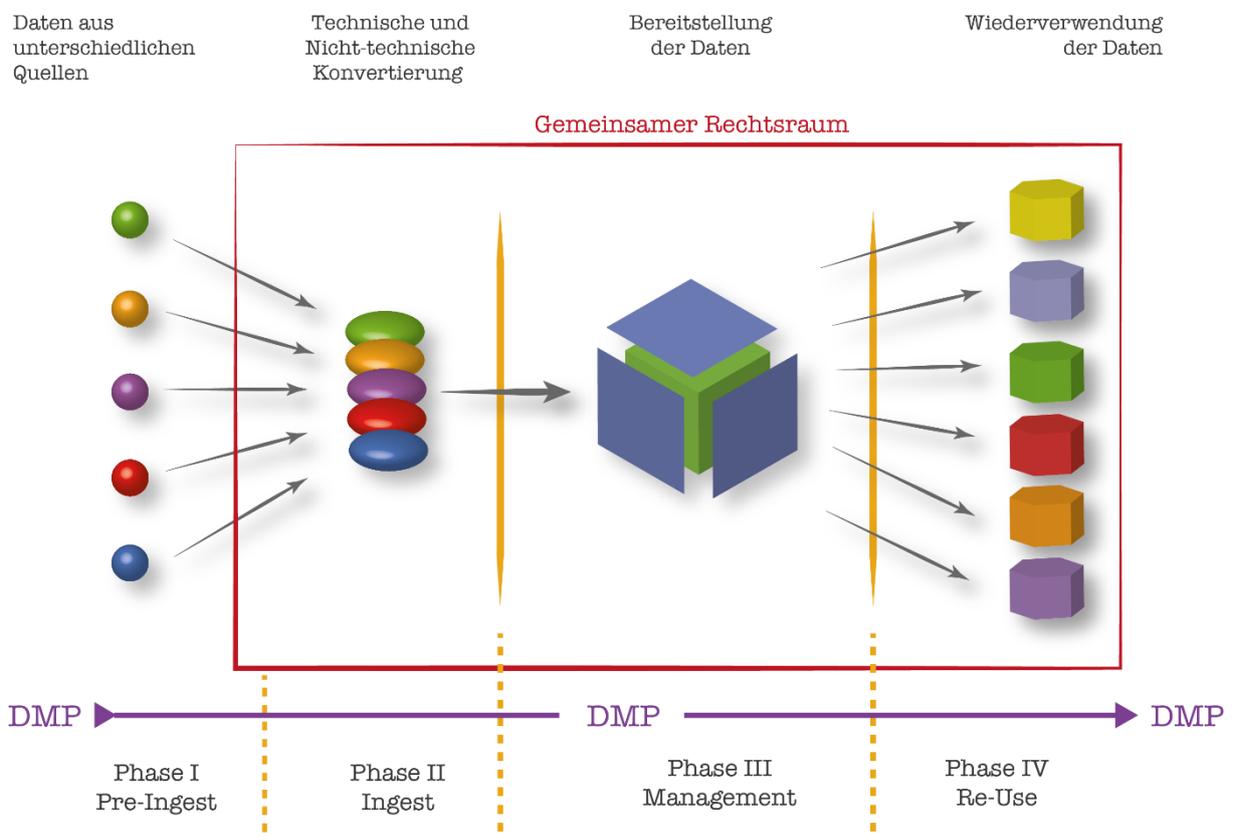
1. Einleitung

Dieses Dokument ist ein Modell, wie Kosten beim Research Data Management verrechnet werden können. Vorbild für dieses Modell ist ein Generationenmodell: Künftige Projekte sollen durch vergangene Projekte finanziert werden. Die Basis ist der Speicherbedarf für die Archivierung, zu dem Services für den Ingest und den Re-Use hinzugerechnet werden. Auch bei den Zusatzkosten ist der benötigte Speicherbedarf die Basis für die Berechnung der Mehrkosten und wird als Faktor multipliziert.

Zusätzliche Services, die gut eingeschätzt werden können, wie die Aufbereitung der OAI-PMH Schnittstelle, können pauschal berechnet werden. Der Speicherbedarf wird herangezogen, da er einerseits einfach zu erheben ist, andererseits aber auch einen Aufwand widerspiegelt, da die archivierten Daten über die Zeit hinweg betreut werden müssen.

2. Kosten umgelegt auf das Workflow-Modell

Die Berechnung der Kosten basiert auf einem Workflow-Modell, das den Weg der Daten von der Entstehung bis zur Wiederverwendung beschreibt. In diesem Modell werden vier Phasen unterschieden, wobei bei den drei Phasen (Ingest, Datenmanagement und Re-Use) Kosten aus Sicht des Repository Managements entstehen.



2.1 Kosten beim Ingest

Die Kosten richten sich nach der Komplexität der Formate beziehungsweise der Aufgabenstellung des Projekts. Je nach Komplexität braucht es mehr oder weniger Support vom Repository Management, um die Daten in das entsprechende Archivsystem zu überführen. Es ergibt sich ein Faktor, der mit dem Datenvolumen multipliziert wird.

In der nachfolgenden Tabelle sind Beispiele von möglichen Faktoren dargestellt.

Beschreibung	Faktor
einfacher technischer Support, Upload über die zur Verfügung gestellten Module	1
Dateien (z.B. Videodateien), die über einen einfachen Upload nicht mehr verwaltet werden können (< 700 MB)	1,2
Daten, bei denen Qualitätskontrollen während des Ingests benötigt werden	1,4
Dateien über > 700 MB, gilt auch für Videodateien über der genannte Größe	1,6
Daten, bei denen der Workflow der Archivierung noch nicht geklärt ist.	1,8
Übernahme aus einem abgeschlossenen Projekt	2

Soll es beim Ingest zu Automatisierungen über Skripte kommen bzw. werden zusätzliche Software Applikationen benötigt, bietet das Repository Management Beratungen in diesem Bereich an. Diese Faktoren werden auch über die Zeit gerechnet, da beim Management auch hierfür ein Mehraufwand entsteht.

2.2 Kosten beim Management

Die Kosten berechnen sich nach dem jährlichen Bedarf des Speichers pro Terabyte (TB). Da für die Langzeitarchivierung keine Dauer angegeben werden kann, wird hier ein Faktor von zehn Jahren angenommen. Die zehn Jahre beziehen sich auf die Richtlinie der Agentur für wissenschaftliche Integrität, die vorschreibt, dass Primärdaten, Laborbücher usw. an den Institutionen zehn Jahre lang aufbewahrt werden müssen. Aus Sicht der Langzeitarchivierung wird aber trotzdem eine Archivierung über die Zeit hinaus garantiert.

2.3 Kosten beim Re-Use

Diese Kosten ergeben sich aus Services, wie die Daten für die Nachnutzung bereitgestellt werden sollen (z.B. Streaming Services für Videos oder Aufbereitung großer Bilder).

2.4 Zusätzlicher Entwicklungsaufwand

Bei den Schritten Ingest und Re-Use kann es bei der Übergabe der Daten der Fall sein, dass hier Automatismen gebaut werden sollen, um einen wiederkehrenden Workflow abzubilden. Diese Automatisierung kann ein Script und/oder eine Software Applikation sein. Dies muss auch als Kostenfaktor berücksichtigt werden und kann nur von Fall zu Fall berechnet werden.

3. Beispiel

Bei einem Projekt mit der Laufzeit von drei Jahren fallen 700 GB an Daten an. 500 GB der Daten sollen langfristigarchiviert werden und der Rest 5 Jahre nach dem Projekt gelöscht werden. 300 GB der Daten sind Video und Audiodaten, die gestreamt werden sollen. Dieser Teil der Daten gehört zu den Daten, die im Langzeitarchiv sind. Der Rest sind Bilder und Dokumente. Die Metadaten sollen über die OAI-PMH Schnittstelle von einem anderen Repository geharvestet werden.

Beispielshalber wird hier von einer mittelfristigen Archivierung (bis 10 Jahre) mit einem Preis von 400 Euro/Jahr ausgegangen sowie von einer Langzeitarchivierung mit 800 Euro/Jahr. Das Streaming wird mit einem Faktor von 1,5 angenommen.

Berechnung der Kosten für den Datamanagement Plan:

Daten für die Langzeitarchivierung ohne Streaming	$0,2 \text{ TB} * 10 \text{ Jahre} * 800 \text{ Euro}$	1.600 Euro
Daten für die Langzeitarchivierung mit Streaming	$0,3 \text{ TB} * 10 \text{ Jahre} * 800 \text{ Euro}$	2.400 Euro
Aufschlag für Ingest von Videodaten	$2.400 \text{ Euro} * 0,2$	480 Euro
Aufschlag für Re-Use – Streaming von Videodaten	$2.400 \text{ Euro} * 0,5$	1.200 Euro
Daten für die mittelfristige Archivierung	$0,2 \text{ TB} * 8 \text{ Jahre} * 400 \text{ Euro}$	640 Euro
Gesamt		6.320 Euro

e-Infrastructures Austria

Nachhaltige Datensicherung und das Bereitstellen von Daten für Dritte ist eine zentrale Aufgabe der Wissenschaft. e-Infrastructures Austria ist ein vom Bundesministerium für Wissenschaft, Forschung und Wirtschaft (MBWF) gefördertes Hochschulraumstrukturmittel-Projekt für den koordinierten Ausbau und die Weiterentwicklung von Repositorien in ganz Österreich. Dadurch wird die sichere Archivierung und dauerhafte Bereitstellung von elektronischen Publikationen, Multimedia-Objekten und anderen digitalen Daten aus Forschung und Lehre gewährleistet. Eng damit zusammenhängend werden Themen im Bereich Forschungsdatenmanagement und Workflows von digitaler Archivierung bearbeitet.

Cluster A	Monitoring und Austausch zum Aufbau von Dokumentservern in den lokalen Einrichtungen <i>Patrick Danowski (IST Austria)</i>
Cluster B	Planung und Durchführung einer österreichweiten Umfrage zu Forschungsdaten <i>Christian Gumpenberger (Universität Wien)</i>
Cluster C	Aufbau eines Wissensnetzwerks: Erarbeitung eines Referenzmodells für den Aufbau von Repositorien <i>Paolo Budroni (Universität Wien)</i>
Cluster D	Aufbau Infrastruktur <i>Raman Ganguly (Zentraler Informatikdienst Universität Wien)</i>
Cluster E	Legal and Ethical Issues <i>Seyavash Amini (Rechtsberater Universität Wien)</i>
Cluster F	Open Access <i>Andreas Ferus (Akademie der bildenden Künste Wien)</i>
Cluster G	Visuelle Datenmodellierung – Generierung von Wissenschaftsräumen <i>Martin Gasteiner (Universität Wien)</i>
Cluster H	Life Cycle Management <i>Andreas Rauber (Technische Universität Wien)</i>
Cluster I	Metadatenkomplex <i>Susanne Blumesberger (Universität Wien)</i>
Cluster J	Dauerhafte Sicherung der Daten (aus nicht-technischer & technischer Sicht) <i>Adelheid Mayer (Universität Wien)</i>
Cluster K	Daten aus wissenschaftlichen und künstlerisch-wissenschaftlichen Forschungsprozessen (Entwicklung und Erschließung der Künste) <i>Bernhard Haslhofer (Austrian Institute of Technology)</i>
Cluster L	Projektübergreifende Fragen (aus nicht-technischer & technischer Sicht) <i>Andreas Jeitler (Universität Klagenfurt)</i>