

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

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 ExpertInnengrupp-

# Technische Funktionen beim Research Data Management

Work-Package-Cluster:	Cluster D: Aufbau Infrastruktur
Leitung des Clusters:	Raman Ganguly Universität Wien <a href="mailto:raman.ganguly@univie.ac.at">raman.ganguly@univie.ac.at</a>
Datum:	09.08.2016
Version:	1.0

AutorInnen/ Sonstige Beteiligte:	Michael Birkner AK Bibliothek Wien <a href="mailto:michael.birkner@akwien.at">michael.birkner@akwien.at</a>
	David Mitterhuber Akademie der bildenden Künste Wien <a href="mailto:d.mitterhuber@akbild.ac.at">d.mitterhuber@akbild.ac.at</a>
	Raman Ganguly Universität Wien <a href="mailto:raman.ganguly@univie.ac.at">raman.ganguly@univie.ac.at</a>
	Gerhard Gonter Universität Wien <a href="mailto:gerhard.gonter@univie.ac.at">gerhard.gonter@univie.ac.at</a>
	Christoph Ladurner TU Graz <a href="mailto:christoph.ladurner@tugraz.at">christoph.ladurner@tugraz.at</a>

Kurzbeschreibung (Deutsch):	Research Data Management basiert auf einer technischen Infrastruktur, die betrieben werden muss. Hier werden die wichtigsten technischen Funktionen dargestellt und deren Tätigkeiten beschrieben. Ausgehend von einem Vierphasenmodell für das Research Data Management, wird ein Workflow beschrieben, von dem die Funktionen abgeleitet werden.
--------------------------------	--

Description (English):	Research Data Management is based on technical infrastructures which have to be maintained by a technical department. Starting form a workflow model consisting of four phases, both function and tasks will be explained.
------------------------	--



Schlagwörter (Deutsch):	RDM, technische Funktionen, Support, Beratung, Software Entwicklung, Operationen, Workflowmodell
Keywords (English):	RDM, technical functions, support, consultant, software development, operations, workflow model
Verwandte Dokumente/ Related Documents:	Abschlussbericht Cluster H – Lifecycle Management



## Inhalt

---

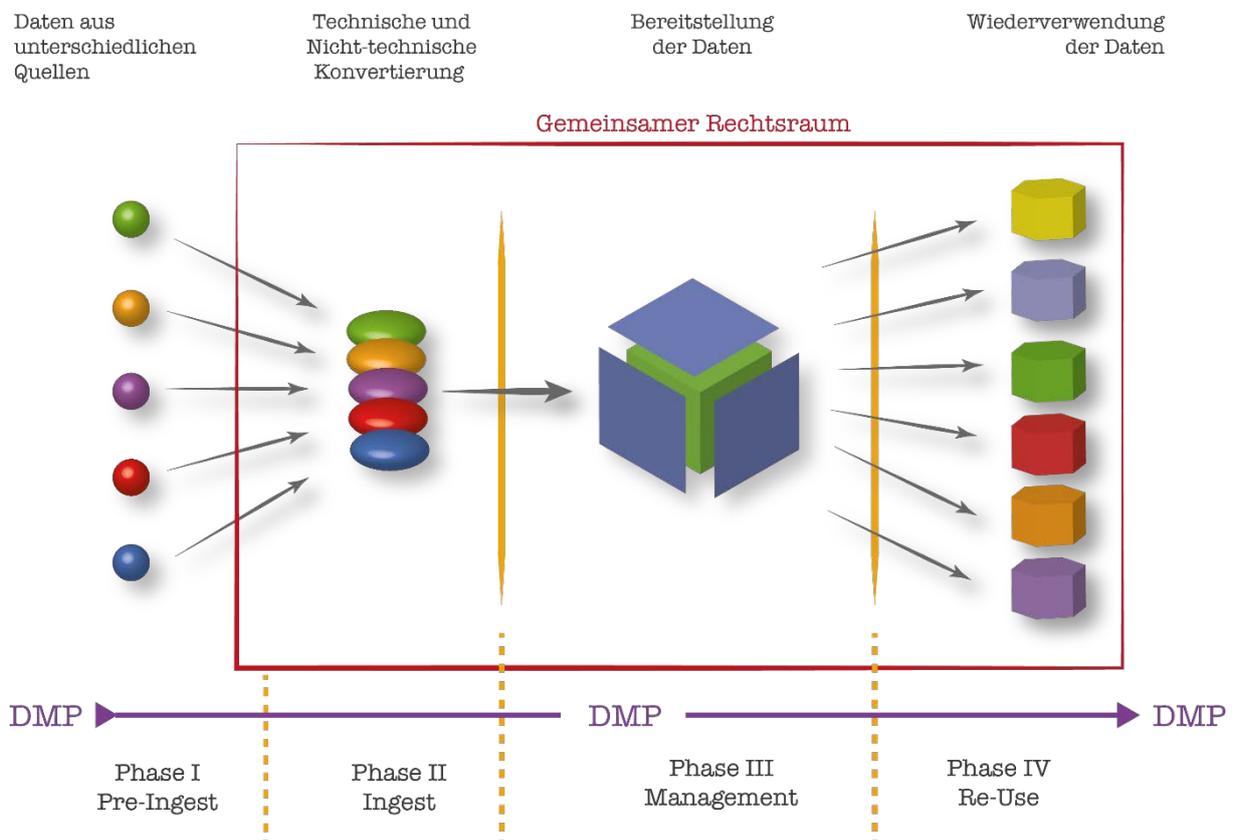
1. Einleitung.....	5
2. Vierphasenmodell beim Research Data Management.....	5
2.1 Pre-Ingest.....	5
2.2 Ingest.....	6
2.3 Management.....	6
2.4 Re-Use .....	7
2.5 Überblick über die Funktionen pro Phase .....	7
3. Funktionen und deren Aufgaben .....	7
3.1 Überblick über die Tätigkeiten .....	8

## 1. Einleitung

Research Data Management basiert auf einer technischen Infrastruktur, die betrieben werden muss. Hier werden die wichtigsten technischen Funktionen dargestellt und deren Tätigkeiten beschrieben. Ausgehend von einem Vierphasenmodell für das Research Data Management, wird ein Workflow beschrieben, von denen die Funktionen abgeleitet werden.

Die nicht-technischen Funktionen sind für den Betrieb gleichwertig, werden aber in diesem Dokument nicht berücksichtigt.

## 2. Vierphasenmodell beim Research Data Management



Wichtig für die Beschreibung der Funktionen sind, wie an der Grafik zu sehen, die vier Phasen Pre-Ingest, Ingest, Management und Re-Use.

### 2.1 Pre-Ingest

In dieser Phase werden die Daten erzeugt. Es gibt noch wenig Berührungspunkte zur Infrastruktur des Repository Managements, in dem die Daten dauerhaft bereitgestellt werden. Hier steht die Datengenerierung im Vordergrund. Es soll die bestmögliche Qualität der Daten erzeugt und nur wenig Rücksicht auf das eigentliche Management genommen werden. Die Qualität der Erzeugung und nicht die Qualität der Archivierung ist im Fokus.

Es macht aber auch aus technischer Sicht durchaus Sinn möglichst früh eingebunden zu werden, da es zum Teil notwendig ist, auch die Methode der Datengenerierung und/oder Software, die zum Erzeugen oder Sammeln der Daten verwendet wird, zu archivieren. Diese Abhängigkeiten sollten zu Beginn der Datenerhebung geklärt werden.

Frühzeitige Beratung kann auch die spätere Ingest-Phase erleichtern, da oft die Daten unterschiedlich in gleicher Qualität erzeugt werden können. Manche Wege sind einer späteren Aufbewahrung dienlicher als andere. Auch können die Datenerzeuger hier bereits auf zentrale Erfahrungen des Research Data Management zurückgreifen.

Wesentliche Funktionen sind hier: Consultant und Support

## 2.2 Ingest

Beim Ingest erfolgt die Übergabe der Daten an das Repository Management. Bei den Übergaben sind Konvertierungen notwendig, um eine dauerhafte Sicherung zu garantieren. Diese Konvertierungen können von rein technischer Natur sein, wie z.B. die Konvertierung der Formate. Sie kann aber auch nicht-technische Klärungen betreffen, wie die Klärung der Reche und/oder der Metadaten.

Da es sehr viele Metadatenstandards gibt, beziehungsweise nicht in jeder wissenschaftlichen Disziplin Standards existieren, sind hier immer wieder softwaretechnische Anpassungen der Archivsysteme notwendig, was wiederum eine Software-Entwicklung notwendig macht. Oft braucht es auch für projektspezifische Anforderungen die Anpassung beim Ingest in einem Repository.

Bereits bestehende Systeme, die für den Ingest der Daten dem Repository Management zur Verfügung stehen, müssen betrieben werden. Während dieser Phase ist eine enge Zusammenarbeit zwischen dem Repository Management und dem Datenproduzenten notwendig. Technischer Support ist hier wesentlich, um einen reibungslosen Ablauf zu gewährleisten.

Wesentliche Funktionen sind hier: Consultant und Support, Operation und Software Entwicklung.

## 2.3 Management

Es ist die zentrale Phase aus Sicht des Research Data Managements. Die Daten werden in dieser Phase in der Qualität, in der sie angenommen wurden, über eine vom Datenerzeuger bestimmte Dauer einem bestimmten Benutzerkreis zur Verfügung gestellt. Die Dauer kann im Sinne der Langzeitarchivierung auch unbestimmt sein. Der Benutzerkreis umfasst bei Open Access die gesamte Öffentlichkeit, die nach Bedarf auf bestimmte Benutzerkreise beschränkt werden können.

Zentral ist hier der Betrieb der Infrastruktur für die Datenarchivierung, wie Server, Storage mit Backup und der notwendigen Software. Da es beim Research Data Management auch der Fall sein kann, dass die Datenerfassung nicht abgeschlossen ist, müssen hier auch Systeme betrieben werden, die Versionierung der Daten zur Verfügung stellen und die Änderungen bei den Daten nachvollziehbar machen. Die Anforderungen sind auf Grund der heterogenen Datenlandschaft sehr unterschiedlich und es ist daher nicht möglich ein System für alle Daten zu betreiben. Wichtig ist dabei, dass die Daten eindeutig identifizierbar sind, untereinander in Bezug gestellt werden können und im Beschreibungssystem gleich sind. Persistente Identifier und Linked Data sind Techniken, die bei den Archivierungssystemen angewendet werden müssen.

Wesentliche Funktionen sind hier: Operation und Software Entwicklung

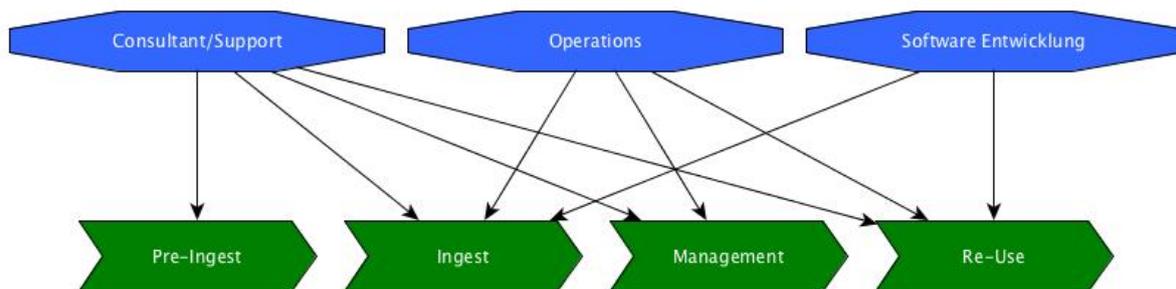
## 2.4 Re-Use

In der Re-Use Phase werden die Daten an die autorisierten Benutzer übergeben. Diese ist die Wichtigste der vier Phasen, da sie dem gesamten Research Data Management Sinn verleiht. Die Übergabe ist ähnlich wie bei den Übergaben zum Repository Management. Die enge Zusammenarbeit mit dem Nutzer ist notwendig. Hier gibt es aber den Unterschied, dass der Nutzer speziell im Bereich Citizen Science oder Open Science keinen direkten Kontakt mit dem Repository Management hat. Daher sind Applikationen, die an die Bedürfnisse zur Nachnutzung angepasst sind, notwendig. Auch ein technischer Support, der für eine große Zielgruppe ausgerichtet ist, muss geschaffen werden.

Wesentliche Funktionen sind hier: Consultant und Support, Operation und Software Entwicklung.

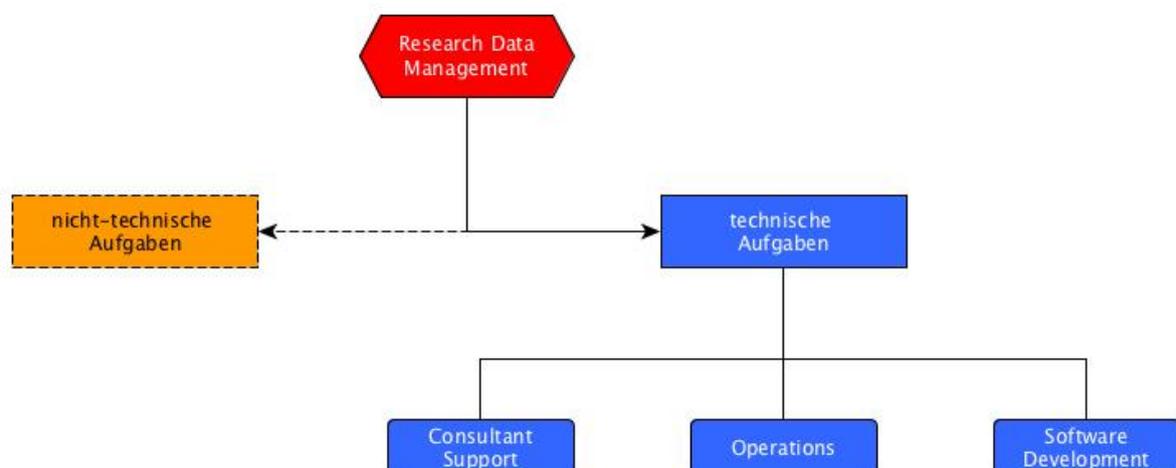
## 2.5 Überblick über die Funktionen pro Phase

Hier ein grafischer Überblick, welche Funktionen in welcher Phase wirken.



## 3. Funktionen und deren Aufgaben

Wie schon in der Einleitung beschreiben, beschäftigt sich dieses Dokument nur mit den Funktionen und Aufgaben aus technischer Sicht. Einen strukturellen Überblick bietet die folgende Grafik.



### 3.1 Überblick über die Tätigkeiten

Die folgende Aufzählung wurde exemplarisch in der Arbeitsgruppe gesammelt und spiegelt den Erkenntnisstand von den derzeitigen Erfahrungen aus der täglichen Arbeit des Research Data Managements wieder. Die Liste ist nicht vollständig und ist im Laufe der Zeit sicher auch Änderungen unterworfen.

#### **Consultant und Support:**

- Beratung beim Umsetzen von Projekten
- Beratung bei der Verwendung von Datenbanken
- Beratung bei der IT-Security
- Beratung bei der Form Konvertierung
- Schulungen und Trainings
- Helpdesk für allgemeine Fragen und Fehlermeldungen
- Dokumentation
- Qualitätskontrolle

#### **Operation:**

- Konfiguration von Archivsystemen
- Inventarisierung
- Monitoring von Applikationen, Servern, Storage und Backup
- IT-Security
- Betrieb von Datenbanken
- Betrieb von Applikationen
- Prozessmanagement
- Systemarchitektur aus Serversicht

#### **Software Entwicklung:**

- Applikationsadministration
- Implementierung von Metadaten
- Datenmanipulation
- Schnittstellenentwicklung
- Entwicklung von Massenimporten
- Systemarchitektur aus Applikationssicht
- Customizing der Applikationen

## 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.

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