

## ■ E-INFRASTRUCTURES AUSTRIA: EINE REFERENZARCHITEKTUR ZUR DAUERHAFTEN BEREITSTELLUNG VON DATEN AUS DER FORSCHUNG ALS AUFGABE FÜR WISSENSCHAFTLICHE BIBLIOTHEKEN

von Paolo Budroni und Raman Ganguly

### **Inhalt**

1. Einleitung
2. Arbeitspakete und Arbeitscluster
3. Daten und Forschung – Die Welt der Daten
4. Aufbau einer Referenzarchitektur für das Forschungsdatenmanagement
5. Content Models und Fragen des Ingests
6. Management der Daten und Nachnutzung von Forschungsdaten
7. Re-use und Interoperabilität
8. Schlussfolgerungen

**Zusammenfassung:** Im Jänner 2014 wurde das nationale dreijährige HRSM-Projekt *e-Infrastructures Austria* initiiert. Gesamtziel ist der koordinierte Aufbau und die Weiterentwicklung von Repositorieninfrastrukturen für Forschung und Lehre in ganz Österreich sowie ein effizientes und nachhaltiges Forschungsdatenmanagement an allen teilnehmenden 20 Universitäten und fünf weiteren außeruniversitären Forschungseinrichtungen. Das Projekt gliedert sich in drei Teilprojekte, die thematisch ineinander übergreifen und/oder aufeinander aufbauen. Die Koordination erfolgt durch die Universität Wien.

*Bis Ende 2016 sollen drei Ziele realisiert werden:*

- Teilprojekt A (Ziel 1): Aufbau von Dokumentenservern auf lokaler Ebene an den Partneruniversitäten. Zweck: Aufstellung von institutionellen Repositorien an allen teilnehmenden Einrichtungen (alle österreichischen Universitäten mit Ausnahme der Medizinischen Universität Innsbruck)
- Teilprojekt B (Ziel 2): Erarbeitung eines strategischen Konzepts für das zukünftige Forschungsdatenmanagement in Österreich
- Teilprojekt C (Ziel 3): Aufbau einer Wissensinfrastruktur („knowledge infrastructure“) für alle 25 Projektpartner.

*In der jetzigen Phase des Projekts e-Infrastructures Austria erfolgt eine Fokussierung auf den Aufbau von Repositorien, darunter die Kategorie der sogenannten Dokumentenserver (unter denen sich bislang die unterschiedlichen institutional repositories*

österr. Einrichtungen aufzählen lassen). Nun verhält es sich jedoch so, dass für bestimmte Daten weitere Lösungen gefragt sind. Diese Lösungen können auch Formen jenseits der Repositorienlandschaft annehmen. Hierbei könnte man von „Informationsinfrastrukturen“ (e-Infrastructures) sprechen.

**Schlüsselwörter:** Open Access; Open University; Open Data; Forschungsdatenmanagement (FDM), Informationsspeicherung; Informationsverbreitung; Kulturpolitik; Kulturelles Erbe

## **E-INFRASTRUCTURES AUSTRIA: A REFERENCE ARCHITECTURE FOR THE PERMANENT PROVISION OF RESEARCH DATA AS A TASK FOR RESEARCH LIBRARIES**

**Abstract:** In January 2014, the national three-year HRSM project e-Infrastructures Austria was initiated. The overall objective is the coordinated design and development of repository infrastructures for research and teaching in Austria, as well as efficient and sustainable research data management at all participating 20 universities and five extramural research institutions. The project is divided into three sub-projects that thematically overlap each other and/or build on each other. Coordination is carried out by the University of Vienna.

By the end of 2016, three objectives are to be realized:

- Sub-project A (objective 1): Implementation of repositories at the local level at the partner universities. Purpose: establishment of institutional repositories at all participating institutions (all Austrian universities with the exception of the Medical University of Innsbruck)
- Sub-project B (objective 2): Developing a strategic approach for future research data management in Austria
- Sub-project C (objective 3): Establishing a knowledge infrastructure for all 25 project partners.

In the current phase of the e-Infrastructures Austria project, the focus is on building repositories, including the category of so-called document servers (which can include the different institutional repositories of Austrian institutions). More recently, certain data requires even different solutions. These solutions can also have forms that are not common to the repository world. This is where the term „information infrastructure“ (e-Infrastructures) becomes significant.

**Keywords:** Open Access; Open University; Open Data, Research Data Management (RDM), Information storage; Dissemination of Information; Cultural Policy, Cultural Heritage

## 1. Einleitung

Das Management, die Aufbewahrung sowie die Zugänglichmachung von digitalen Ressourcen sind wesentliche Aufgaben von wissenschaftlichen Einrichtungen. Neben Publikationen und digitalisierten Hochschulschriften gewinnen in den letzten Jahren auch komplexere digitale Objekte (Forschungsdaten, Multimedia-Inhalte, e-Learning-Inhalte u.a.) und ihre Nachnutzung vermehrt an Bedeutung. Durch die Weiterentwicklung von Technologien haben sich für die Forschung neue Ansätze und Methoden hinsichtlich der Darstellung und Wiederverwendung von Daten eröffnet. Zugänglichkeit geht heute weit über einen bloßen Download von Objekten hinaus und bietet vielfältige Formen der Nachnutzung. Auch die Wahrung der Rechtssicherheit und die Beachtung von ethischen Fragestellungen erlangen in diesem Kontext vorrangige Bedeutung. Sie sind, sozusagen, „*key issues*“.

Ein zentrales Anliegen von e-Infrastructures Austria ist es, WissenschaftlerInnen, Forschungsgemeinschaften, wissenschaftliche Bibliotheken, IT-Dienstleister und Forschungsförderer aller nationalen universitären und außeruniversitären Einrichtungen einzubeziehen, um gemeinsam Wissen zu bündeln, zu analysieren und zu teilen. Instrumente der ersten Wahl für die Implementierung des Projektes sind:

- a) Repositorien für einfache Inhalte (z.B. Texte und Bilder), die auf lokaler Ebene aufgebaut werden,
- b) Konzeptionen bzgl. eines kooperativen Umgangs mit komplexen Inhalten (z.B. Forschungsdaten, Multimedia-Inhalte),
- c) die Klärung der Fragestellungen rund um deren Sicherung und Bereitstellung,
- d) der Aufbau einer gemeinsamen Informationsinfrastruktur (z. B. ein kollaborativ genutztes, nationales Projekt-Wiki),
- e) eine tragfähige *Project Governance* (z.B. eine Konsensualvereinbarung), die es ermöglicht, die Geschäftsprozesse zwischen den 25 teilnehmenden Partnern und weiteren beobachtenden Institutionen sowie den Umgang mit Projektergebnissen zu regeln,
- f) die Bereitstellung von *Deliverables* für alle am Projekt Beteiligten und somit
- g) der Aufbau einer gemeinsam gebildeten Wissensinfrastruktur, die nachhaltig über das Ende der dreijährigen Projektarbeit hinaus wirken soll.

Repositorien	Konzept Forschungsdaten
Kooperativer Umgang	Sicherung, Bereitstellung
Project Governance	Konsensualvereinbarung
Informationsinfrastruktur	Nationales Wiki
Deliverables	Wissensinfrastruktur

Tab. 1: Instrumente, die aus der Sicht des Managements eingesetzt werden

## 2. Arbeitspakete und Arbeitscluster

Die Zusammenarbeit zu den konkreten Arbeitsbereichen erfolgt im Rahmen von 12 Arbeitspaketen (WP), die sämtliche Prozesse rund um den Aufbau von Repositorien und Forschungsdatenmanagement untersuchen. Sie werden dezentral bearbeitet. In diesen Packages werden diese Prozesse gebündelt (deshalb auch die im Projekt verwendete Definition Cluster). Zwischen den Arbeitsgruppen findet ein regelmäßiger Austausch statt.

Ein wesentlicher Bestandteil der derzeitigen Bemühungen (Stand: Sommer 2015) ist die Auswertung einer Befragung der WissenschaftlerInnen über den Status Quo zum Umgang mit Forschungsdaten an österreichischen Einrichtungen. Zentrales Anliegen ist es, Desiderata zu identifizieren, mögliche Incentives für Forschungsdatenmanagement und speziell für Open Data zu eruieren sowie Anforderungen an zukünftige forschungsunterstützende Services zu erfragen. Das wird ein wesentlicher Baustein sein, um ein Konzept für den zukünftigen Umgang mit Forschungsdaten zu erstellen.

Die Arbeitspakete im Einzelnen:

- Cluster A: Monitoring und Austausch zum Aufbau von Dokumentenservern in den lokalen Einrichtungen (*Koordination: IST Austria*)
- Cluster B: Planung und Durchführung eines „National Survey“ zu Forschungsdaten (*Koord.: Universität Wien*)
- Cluster C: Aufbau eines Wissensnetzwerks: Erarbeitung eines Referenzmodells für den Aufbau von Repositorien (*Koord.: Universität-Wien*)
- Cluster D: Aufbau Infrastruktur (*Koord.: Universität Wien*)
- Cluster E: Legal and Ethical Issues (*Koord.: Seyavash Amini, Rechtsberater d. Universitätsbibliothek Wien*)
- Cluster F: Open Access (*Koord.: Akademie der bildenden Künste Wien*)
- Cluster G: Visuelle Datenmodellierung – Generierung von Wissenschaftsräumen (*Koord.: Universität Wien*)

- Cluster H: Life Cycle Management (*Koord.: Technische Universität Wien*)
- Cluster I: Metadatenkomplex (aus nicht-technischer & technischer Sicht) (*Koord.: Universität Wien*)
- Cluster J: Dauerhafte Sicherung der Daten (aus nicht-technischer & technischer Sicht) (*Koord.: Universität Wien*)
- Cluster K: Daten aus wissenschaftlichen und künstlerisch-wissenschaftlichen Forschungsprozessen (Entwicklung und Erschließung der Künste) (*Koord.: Austrian Institute of Technology – AIT*)
- Cluster L: Projektübergreifende Fragen aus nicht-technischer & technischer Sicht (*Koord.: Universität Klagenfurt*)

### 3. Daten und Forschung – Die Welt der Daten

In letzter Zeit wurden interessante Beiträge zum Thema Definition von Forschungsdaten publiziert. Wir möchten in diesem Kontext von zwei Definitionen Gebrauch machen, die für unseren nationalen Kontext von Bedeutung sind. Beide sind von Ingrid Dillo (DANS) im Rahmen einer Arbeit zu Zertifizierungsprozessen von Repositorien veröffentlicht worden<sup>1</sup> und gehen von einem weit gefassten Begriff von Forschungsdaten aus:

- 1) The Australian Griffith University presents the following definition:
 

„*Research data are factual records, which may take the form of numbers, symbols, texts, images or sounds, which used as primary sources for research, which are commonly accepted in the research community as necessary to validate research findings.*“
- 2) The University of Minnesota about research data (\*): „*Research data are data in any format or medium that relates to or support research, scholarship, or artistic activity. They can be classified as:*
  - *Raw or primary data: information recorded as notes, images, video footage, paper surveys, computer files, etc.*
  - *Processed data: analyses, descriptions, and conclusions prepared as reports or papers*
  - *Published data: information distributed to people beyond those involved in data acquisition and administration.*“

Für unsere Tätigkeit im Rahmen des Projekts und im Hinblick auf die Erstellung eines Konzepts für den zukünftigen Umgang mit Forschungsdaten, ergeben sich aus diesen beiden Definitionen – und vor allem aus der letztgenannten – interessante Schlussfolgerungen.

Die zweite Definition beinhaltet bei näherer Betrachtung der bis jetzt angewandten Logik unterschiedliche Ebenen bei der Verarbeitung („processing“) von Daten. Diese können aus technischer Sicht auch „layers“ und aus nicht-technischer Sicht „levels“ genannt werden. Unser Rückschluss definiert also mindestens 5 Ebenen (layers oder levels), die sich wie folgt aufzählen lassen:

- Die Welt der (Forschungs)Daten (*raw data, primary data*)
- *processed data* (verarbeitete Daten, aufgeteilt in positive, negative und inconclusive results)
- *shared data*
- *published data*
- *Open Access published data*

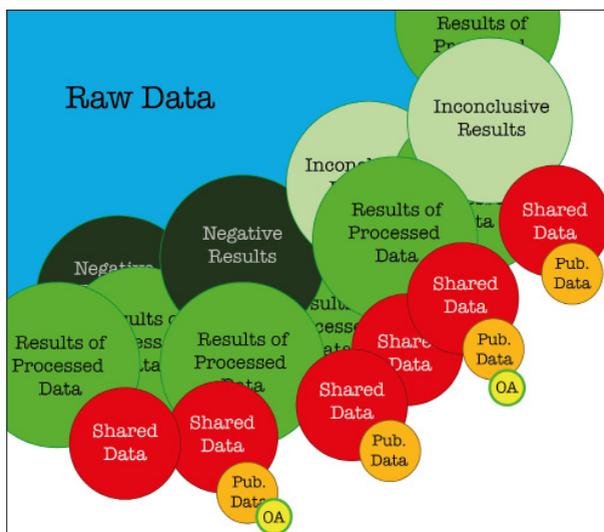


Abb. 1: Ebenen der Verarbeitung von Daten im Modell für digitale Archivierung (© Ganguly)

Die Grundlage für das Arbeiten im digitalen Bereich ist – bei Einhaltung von vordefinierten Regelwerken und Usancen (nach Klärung von rechtlichen und ethischen Fragen, policies und codes of conduct) der möglichst „problemlose“ Zugriff mittels geeigneter Tools und Anwendungen auf die sogenannte „Welt der Daten“. Die eingeleiteten Prozesse bei der Verarbeitung dieser Daten bilden an sich den Mehrwert diesen Vorgangs: Es entstehen so die sogenannten „Processed Data“. Für das wissenschaftliche Arbeiten und die Förderung des Forschungsprozesses bilden die dabei neuentstandenen Daten (darunter auch die Beschreibung der Entwicklung und Anwendung von Me-

thoden und Verfahren bei der Verarbeitung der Daten) den wesentlichen Mehrwert dieser Forschungsarbeit. Um genau diese Daten und die ihnen zugrunde liegenden Prozesse geht es, wenn man ein Konzept für den Umgang mit Forschungsdaten definieren möchte. Diese processed data können zu sogenannten *positive results*, *inconclusive results* und *negative results* führen.

Derzeit werden vor allem Daten, die zu einem gewünschten Forschungsergebnis geführt haben, und Daten, die einer Publikation zugrunde liegen, aufbewahrt bzw. zur Nachnutzung zur Verfügung gestellt. Der Zugang zu Daten und damit verbundenen Prozessen, die nicht zu einem gewünschten bzw. zu keinem Ergebnis geführt haben, hier *inconclusive* oder *negative results* genannt, ist hingegen meist nur erschwert möglich. Derzeit werden diese Daten und die ihnen zugrundeliegenden Entstehungsprozesse an unseren Forschungseinrichtungen weder beachtet noch archiviert, wodurch auch die dauerhafte Bereitstellung nicht gewährleistet ist.

Somit ist die dauerhafte Bereitstellung dieser Forschungsdaten im Bereich *inconclusive and negative results* in einer dafür geeigneten digitalen Umgebung von eminenter Bedeutung. Damit korrelierende Themen sind: policies für die Erhaltung dieser Daten, Vergabemechanismen von persistenten digitalen Signaturen zur Wiederauffindbarkeit, standardisierte Vergabe von fachlich kompetenter Seite von Metadaten (darunter auch Angaben zum Forschungskontext, der Provenienz, technischer und rechtliche Natur sowie das sogenannte „Datenmanifest“), die Klärung der sogenannten legal and ethical issues und bereits an dieser Stelle die Darlegung des Forschungsprozesses sowie die Entscheidung, in welcher geeigneten Form zukünftig die Datenmodellierung erfolgen wird („Welche Form der Visualisierung dieser Daten wird angestrebt?“).

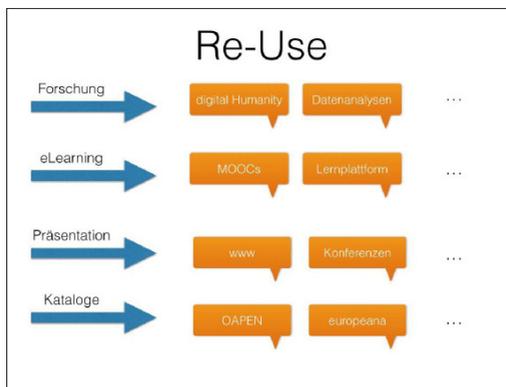


Abb. 2: Digitaler Workflow (© Ganguly)

#### 4. Aufbau einer Referenzarchitektur für das Forschungsdatenmanagement

Ein mögliches Modell für den digitalen Workflow zur Bildung einer Referenzarchitektur zur dauerhaften Bereitstellung von Daten aus der Forschung sollte Aspekte beinhalten, welche die Anforderungen von mindestens 5 Stakeholdern in diesem Prozess berücksichtigen. Diese sind:

- a) die ForscherInnengemeinschaft (Datenlieferant),
- b) das Repository Management (mit besonderer Berücksichtigung der technischen Anforderungen),
- c) die Institution (Forschungseinrichtung),
- d) die sogenannten Funding Bodies (Forschungsförderer, inkl. der Forschungsservices), und zuletzt
- e) die qualifizierte Öffentlichkeit (darunter wiederum Teile der ForscherInnengemeinschaft

Diese Aspekte sollten im Wesentlichen auch in den sogenannten Data Management Plans (DMP)<sup>2</sup> und in den neuauftkommenden Process Management Plans (PMP) berücksichtigt werden.<sup>3</sup>

Ihre „Makro-Darstellung“ kann durch folgende Graphik besser verdeutlicht werden.

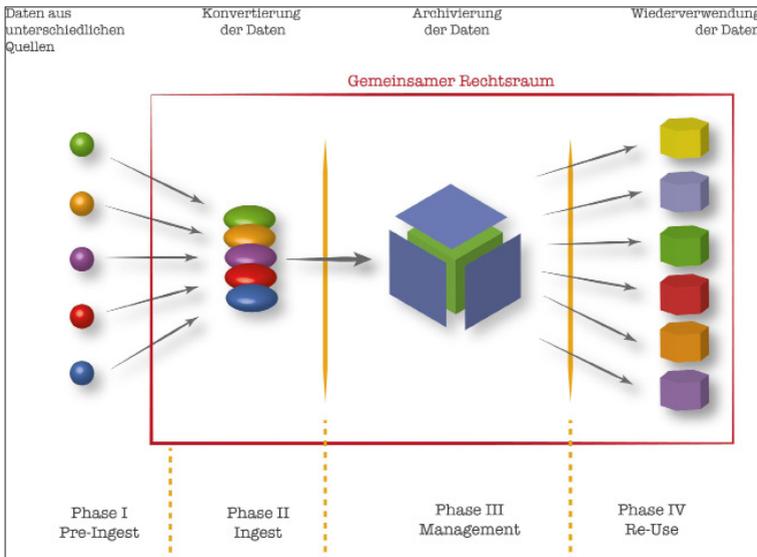


Abb. 3: Digitales Workflowmodell (© Ganguly)

Die Referenzarchitektur zur dauerhaften Bereitstellung von Daten aus der Forschung sieht also in ihrem Workflowmodell die Unterscheidung von vier Segmenten vor, die wie folgt genannt werden:

- a) *pre-ingest*
- b) *ingest*
- c) *management*
- d) *re-use*

Dieser Workflow ergibt sich aus der Verarbeitung der Daten („Forschungsdaten“).

Es bietet sich jedoch auch eine weitere Sichtweise an, die nicht nur eine Aufteilung technischer Prozesse vorsieht. Diese weitere Sicht entsteht aufgrund der Unterscheidung der jeweiligen Verantwortungsbereiche bei der Verarbeitung der Daten und der dahinterliegenden prozessorientierten und gesteuerten Handlungen. Es ergeben sich somit auch Definitionen im Bereich Rollen und Kompetenzen. Zusammengefasst handelt es sich hierbei um die Sicht aus dem „Managementbereich“ heraus.

Welche Rollen und Verantwortungen sollen und werden von den Hauptakteuren des digitalen Workflows wahrgenommen?

- a) Im Segment *pre-ingest* hat den lead in Sachen Verantwortung der Datenerzeuger.
- b) Beim *ingest* liegt die Ergebnisverantwortung zunächst beim Datenerzeuger und erst bei Abgabe des Datenpakets wechselt diese über das Prozessmanagement zum Repository Management.
- c) Im Bereich *management* sind die Kompetenzen klar verteilt, hier liegt die Verantwortung im Management der Arbeitsprozesse beim *Repository Management*.
- d) Im Bereich *re-use* wechselt die Verantwortung nach erfolgter Datenmodellierung und Übergabe des Contents wieder zum Datenerzeuger (in seiner Rolle als Owner des Contents).

Aus Sicht des Rollenmodels ergeben sich vier Rollen:

- a) *Qualitätssicherung*: Der Datenerzeuger (Datenerzeuger) sichert zunächst die Qualität des Contents
- b) *Sicherstellung*: Das Repository Management sichert die Daten (auch zeitlich gesehen), zunächst in der Art wie sie geliefert werden
- c) *Wahrung der Rechtssicherheit*: Die Institution sorgt für die Wahrung der Rechtssicherheit
- d) *Wahrnehmung von gesellschaftlicher Verantwortung*: Der Forschungsförderer nimmt diese Funktion wahr. Somit entsteht auch ein weiterer Akteur.

In weiterer Folge wäre zu untersuchen, inwieweit ein weiterer Player zu den bereits vier genannten zählen könnte, nämlich Verlage. Das Verlagswesen spielt auf den ersten Blick eine große Rolle im Bereich Re-use, vor allem im Bereich von Publikationen (meist in Form von Artikeln). Es stellt sich die Frage, ob Verlage nicht auch beim Management der Daten Einfluss haben könnten. Der Verantwortungsbereich läge im Bereich der Bereitstellung der Daten, was auf einen möglichen Interessenskonflikt mit dem Archiv und dessen Management hinweisen könnte bzw. mit dem Betreiber des Archivs, die Institution. Sollte dies zutreffen, ergäbe sich für den Verlag auch eine Teilverantwortung im Bereich Wahrung der Rechtssicherheit.

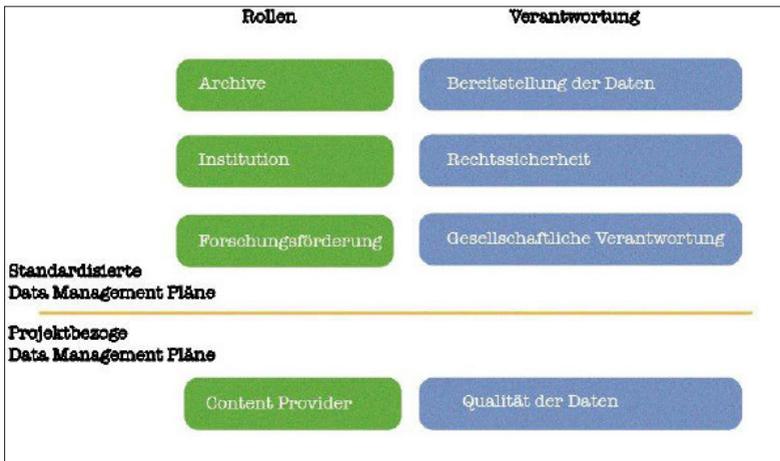


Abb. 4: Rollenmodell (© Ganguly)

## 5. Content Models und Fragen des Ingests

Eine möglichst genaue Definition der Contentmodelle und -typen sollte bereits während der *pre-ingest*-Phase stattfinden. Diese Definition sichert von Beginn an die korrekte und rasche Arbeit bei den nachfolgenden Schritten, vor allem bei der Vergabe der Metadaten wie auch bei der Klärung von rechtlichen Aspekten. Die vielleicht wichtigste Frage, die es zu klären gilt, ist folgende: welche Form der Visualisierung der Daten ist (z. B. nach Projektende) intendiert? Eine Erkenntnis ist auch, dass sich je nach Content Model/Type ein unterschiedlicher Workflow in das digitale Archiv ergibt. Es macht einen Unterschied, ob es sich um Dokumente, Datenbanken oder um *Big Data* handelt.

In der Folge die Auswirkungen:

- a) Es wird immer einen unterschiedlichen Workflow beim Ingestprozess geben.
- b) Dieser Umstand muss bei jeder Planung der Ressourcen berücksichtigt werden. Am vorteilhaftesten ist es, wenn dies *bereits in der Projektantragstellungsphase* erfolgt. Data Management Pläne können nur zum Teil standardisiert werden. Der Datenerzeuger definiert mit der jeweils gewählten Qualität die Struktur der Daten, daher lässt sich dieser Bereich nicht standardisieren. Die Teile, die standardisiert werden können, sind jene, die das Management im Repository betreffen bzw. auch jene, die der Forschungsförderer bestimmt (funding body und jene, die von einer bestimmten policy betroffen sind). Alle anderen Prozesse rund um DMPs unterliegen der Beschaffenheit und Struktur der Daten (Forschungsdaten).

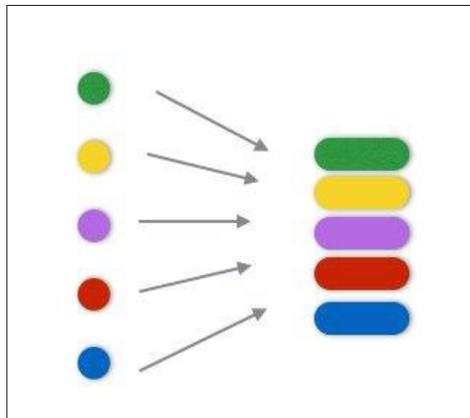


Abb. 5: Heterogene Contenttypen bedingen unterschiedliche Workflows im Bereich des Ingests, dies hat einen Impact auf die Gestaltung der DMPs (© Ganguly)

## 6. Management der Daten und Nachnutzung von Forschungsdaten

Das Management der Daten in den Repositorien verfolgt im Allgemeinen diese Ziele:

- a) Suchen und Finden, Identifizieren, Auswählen, Bereitstellen von Daten
- b) Identität und Authentizität der Daten bewahren
- c) Möglichkeit zum Teilen der Daten zur Verfügung stellen.

In diesem Zusammenhang darf jedoch nie vergessen werden, dass eine weitere Fragestellung vorhanden ist, die stets die Geschäftsprozesse begleitet: „Welche Form der Visualisierung wird nach Extraktion der Daten gewünscht?“ (*Wie sollen die Daten modelliert werden?*)

Bei der Frage bezüglich der Integrität und Identität der Daten (vor allem der digitalen Objekte) spielen folgende Aspekte eine Rolle: Die konsequente Wahrung der *Integrität* und *Konsistenz* ermöglichen letzten Endes die *Authentizität* dieser Daten.

Ihrerseits bedingt die Wahrung der Integrität eine Reihe von Prozessen rund um die Vollständigkeit der Daten (*completeness*). Die Integrität hingegen basiert auf der korrekten Berücksichtigung von *Konsistenz* (aufbauend vornehmlich auf: content Information, contextual information, provenance). Die Faktoren, die hierbei eine Rolle spielen sind:

- a) Aspekte rund um die Langzeitarchivierung
- b) Aspekte rund um die Mittel- und Kurzzeitarchivierung (*mid and short-term storage*)
- c) Sicherheitsaspekte
- d) *Big Data*
- e) source code
- f) Fragen der Publikationstypen
- g) Fragen des Open Access
- h) Fragen rund um Zugriff, Identifikation, Suche, Rechte und Qualität
- i) Fragen hinsichtlich der vom Gesetzgeber vorgesehenen Aspekte der e-Accessibility
- j) Rechtliche und ethische Fragestellungen

## 7. Re-use und Interoperabilität

Beim letzten Abschnitt des Workflows wird der Mehrwert des Prozesses bei der Verarbeitung der Forschungsdaten ersichtlich. Hier findet die Nachnutzung in effizienter und geeigneter Form statt. Bevorzugte Szenarien der Wiederverwendung seien an dieser Stelle genannt: Digital Humanities, Datenanalysen, MOOCs, Lernplattformen, World Wide Web, Konferenzen, OAPEN, Europeana, OpenAIRE, um nur einige beispielhaft zu nennen.

Die vielleicht wichtigste Voraussetzung für die reibungslose und effiziente Wiederverwendung dieser Daten, vor allem dann, wenn eine möglichst breite Streuung der Forschungsdaten und deren Ergebnisse intendiert wird, ist die Wahrung der Rechtssicherheit. Sowohl im Bereich des Managements dieser Daten, wenn es um Zugriffsrechte oder Nutzungs-

bedingungen geht, als auch im Bereich standardisierter Nutzungslizenzen und der Wahrung der vom Gesetzgeber vorgesehenen rechtlichen Rahmenbedingungen. Auf diese Weise wird ein wesentlicher Teil der Interoperabilität *innerhalb der Informationsinfrastruktur der Institution* wie auch nach außen gesichert (die Behandlung weiterer Thematiken, wie z. B. die Vergabe von permanent identifiers, den Komplex rund um die Metadaten oder Fragen der e-accessibility soll an dieser Stelle nicht stattfinden). Ein weiterer Themenkomplex, der in weiterer Folge und an anderer Stelle beleuchtet werden sollte, ist die Nachnutzung im Bereich der forschungsorientierten Lehre, denn in diesem Bereich spielt nicht nur die Langzeit- sondern vor allem auch die Kurzzeitarchivierung eine Rolle.



Abb. 6: Digitaler Workflow – Management der Daten (© Ganguly)

## 8. Schlussfolgerungen

Wir haben eingangs die vier Sichtweisen dargestellt, die besonders beim DMP zum Tragen kommen – die der Institution, des Datenerzeugers, des Repository Managements und des Forschungsförderers – und daraus wesentliche Rollen und Kompetenzen abgeleitet.

Versucht man nun alles in Bezug auf die vier genannten Akteure auf einen gemeinsamen Nenner zu bringen, ergibt sich folgende Liste an Anforderungen, die für e-Infrastructures Austria von Bedeutung sind:

1. Die Schaffung eines gemeinsamen Rechtsraums innerhalb der Institution für die Behandlung sämtlicher digitaler Objekte, Ressourcen und Assets. Die Schaffung dieses Raumes wird von geeigneten policies begleitet (wie z. B. einer policy zum Workflow oder zum Datenexport). Dieser Rechtsraum, der auch die Berücksichtigung von ethischen Aspekten vorsieht, ist vor allem für das Management von Forschungsdaten von Bedeutung. Und erst dadurch sind der Austausch und die Wechselwirkung von Rollen und Kompetenzen im digitalen Workflow möglich.
2. Dieser rechtliche Rahmen sollte sämtliche Phasen im Lebenszyklus der Daten, der digitalen Objekte, Ressourcen und Assets sowie die dahinter liegenden Daten in allen vier Phasen begleiten und absichern: nämlich *pre-ingest-*, *ingest-*, *management-* und *re-use-Phase*.
3. Die Stärkung der Rolle der Institution bei der Wahrnehmung der Verantwortung im Bereich «Forschungsdatenmanagement». Die Institution muss sich dieser Rolle bewusst werden, um dann eine Befähigung (empowerment) der wichtigsten Akteure in diesem Bereich zu ermöglichen, nämlich Repository Management und Forschungsservice.
4. Einführung von Data Management Plans (DMPs) an allen Einrichtungen. Die Möglichkeiten, die sich durch die Anwendung der DMPs (in der Projektantragsstellungsphase) ergeben, würden dazu beitragen, Ressourcen – vor allem finanzieller Natur – optimal einzusetzen, um die digitalen Inhalte langfristig auffindbar, verständlich und nachnutzbar zu machen (und kommen schon in der Projektantragsstellungsphase institutionell zur Anwendung).
5. Automatisierung der DMPs vorantreiben. Daten sollen nach Möglichkeit nur einmal erfasst werden, damit sie auch im Rahmen der digitalen Wertschöpfungskette nachgenutzt werden können (so z. B. in der Phase der Evaluation).
6. DMPs (und in späterer Folge PMPs) sind ein wesentlicher Bestandteil sogenannter *archiving policies*. Die archiving policy sollte den gesamten Zyklus abdecken und Hand in Hand mit der Schaffung des gemeinsamen Rechtsraums innerhalb der Institution gehen.
7. Involvierung sämtlicher wichtiger Akteure (Bibliotheken, IT-Services, Forschungsförderungseinrichtungen, WissenschaftlerInnen) in die Geschäftsprozesse des digitalen Workflows sowie die Wahrnehmung der jeweiligen Rolle und Verantwortung im jeweiligen Bereich des digitalen Workflows. Hierbei sollte geprüft werden, welche Rolle das Verlagswesen bei der Bereitstellung von Daten wahrnehmen könnte.

8. Ein „User“ kann mehrere unterschiedliche Rollen annehmen. Das ergibt die Möglichkeit unterschiedlicher Arbeitsabfolgen auf unterschiedlicher Ebene.
9. Die Leitfrage stets beachten: Welche Form der Visualisierung wird nach Extraktion der Daten gewünscht?
10. Ein Schlusswort: In der jetzigen Phase des Projekts e-Infrastructures Austria erfolgt eine Fokussierung auf den Aufbau von einfachen Repositorien, darunter die Kategorie der sogenannten Dokumentenserver (unter denen sich bislang die unterschiedlichen institutional repositories österr. Einrichtungen aufzählen lassen). Nun verhält es sich jedoch so, dass für bestimmte Daten weitere komplexere und artikuliertere Lösungen gefragt sind. Diese Lösungen können auch Formen jenseits der üblichen Repositorienlandschaft annehmen. Hierbei könnte man von „Informationsinfrastrukturen“ sprechen => e-Infrastructures.

Dr. Paolo Budroni  
Universitätsbibliothek der Universität Wien  
e-Infrastructures Austria, Projektleitung  
E-Mail: [paolo.budroni@univie.ac.at](mailto:paolo.budroni@univie.ac.at)

Dipl.-Ing. (FH) Raman Ganguly  
Zentraler Informatikdienst der Universität Wien  
e-Infrastructures Austria, Technische Projektleitung  
E-Mail: [raman.ganguly@univie.ac.at](mailto:raman.ganguly@univie.ac.at)  
Website: [www.e-infrastructures.at](http://www.e-infrastructures.at)

- 1 Ingrid Dillo – Data Archiving and Networked Services (DANS): Certification as a means of providing trust. Florence, Fondazione Rinascimento Digitale, 2012. Online unter: [http://www.rinascimento-digitale.it/conference2012/slides\\_ic\\_2012/dillo\\_slides.pdf](http://www.rinascimento-digitale.it/conference2012/slides_ic_2012/dillo_slides.pdf)
- 2 Zu den unterschiedlichen Views auf DMPs vgl.: COAR 2012, Annual Meeting, Data Management Plans – How to Treat Digital Resources. The imminent future for repositories and their management, Paolo Budroni, Tomasz Miksa, Andreas Rauber. Online unter: [https://www.coar-repositories.org/files/7\\_DMP\\_Vienna.pdf](https://www.coar-repositories.org/files/7_DMP_Vienna.pdf)
- 3 Vgl. dazu: Tomasz Miksa, Stephan Strodl, Andreas Rauber: Process Management Plans. International Journal of Digital Curation 9 (2014) 1, pp. 83–97. doi:10.2218/ijdc.v9i1.303 sowie Tomasz Miksa, Andreas

Rauber: Increasing preservability of research by process management plans. In: Proceedings of the 1st International Workshop on Digital Preservation of Research Methods and Artefacts (DPRMA'13). doi:10.1145/2499583.2499591



Dieses Werk ist lizenziert unter einer [Creative-Commons-Lizenz Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/)