

# Report on the nestor PERICLES School 2016

---

DOC:	February 15 <sup>th</sup> – 17 <sup>th</sup> nestor / PERICLES School 2016		
WP/TASK:	WP7 / TASK7.3	DATE:	April 2016
PARTNER:	UGOE	AUTHOR:	Stefan Strathmann

## nestor / PERICLES School 2016: Langzeitarchivierung von Forschungsdaten" („Long term preservation of research data“)

---

The nestor / PERICLES School 2016 took place from 15<sup>th</sup> to 17<sup>th</sup> of February 2016 in the Landesmusikakademie Hessen Schloss Hallenburg <http://www.landesmusikakademie-hessen.de/> in Schlitz in the central area of Germany. The three-day training event was organised together with representatives from several FE and HE institutions in Germany, Austria and Switzerland (the nestor qualification group).

The event was entitled "nestor / PERICLES School 2016: "Langzeitarchivierung von Forschungsdaten" („Long term preservation of research data“)" and introduced a broad range of aspects related to digital preservation. The main area, digital preservation of research data, was covered by several sessions.

38 individuals from Germany and Switzerland - including students, researchers, representatives from different companies, technical staff from archives, museums and libraries – used the opportunity to exchange their knowledge in the field of digital preservation and research data management and to consult with their peers.

The school intended to introduce participants to the first steps of the implementation of digital research data management and preservation. The event hosted four sessions; each consisted of an introductory talk, followed by group work and presentations of the group-results and discussions summing up the topic.

The first session dealt with an introduction into research data and their management and preservation. At the evening of the first day an informal poster session allowed the attendees to present their current work and to get together in an easy way.

In the second session some practical aspects of research data management were discussed, based on experiences which were presented by an embedded data manager of one of the Collaborative Research Centres (Sonderforschungsbereiche) founded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG).

Selection and appraisal of research data for preservation and reuse were the topics for the third session. The second day was closed with a get together with open end.

During the last session, the attendees got an introduction into the PERICLES Sheer Curation Tools and hands on exercise in how to use the PERICLES Extraction Tool (PET) and PERICLES Content Aggregation Tool (PeriCAT).

Finally the attendees got a certificate which can earn 2 ECTS Credit Points in a HE context (which are at least accepted by the 12 HE institutions from Austria, Germany and Switzerland organised in the training section of nestor).



Table: nestor / PERICLES School 2016: Long term preservation of research data

Course title: nestor / PERICLES School 2016: Long term preservation of research data	
<b>Location and date:</b> Landesmusikakademie Hessen Schloss Hallenburg in Schlitz (Germany), 15th to 17th of February 2016	<b>Duration:</b> 3 days
<b>Number of attendees:</b> 31 students and 7 trainers	
<b>Involved Partners:</b> Goettingen State and University Library (UGOE) / nestor / PERICLES	
<b>Sessions delivered:</b> <ul style="list-style-type: none"> <li>• What are Research Data and what does it mean to manage and preserve them?</li> <li>• Research Data Management in practice: Embedded Data Manager</li> <li>• Selection and Appraisal of research data for preservation and reuse</li> <li>• PERICLES Sheer Curation Tools – The preservation of digital objects starts in the origin environment</li> </ul>	
<b>Teaching method:</b> Mixture of lectures, practical activities and discussions	
<b>Trainers:</b> <ul style="list-style-type: none"> <li>• Fabian Cremer, Max Weber Foundation, Bonn</li> <li>• Anna Eggers, Goettingen State and University Library (SUB), Goettingen</li> <li>• Jens Ludwig, Staatsbibliothek zu Berlin – Preußischer Kulturbesitz, Berlin</li> <li>• Reiner Mauer, GESIS – Leibniz Institute for the Social Sciences, Cologne</li> <li>• Achim Oßwald, TH Köln – Cologne University of Applied Sciences, Cologne</li> <li>• Jonas Recker, GESIS – Leibniz Institute for the Social Sciences, Cologne</li> <li>• Stefan Strathmann, Goettingen State and University Library (SUB), Goettingen</li> </ul>	

An evaluation sheet has been developed to assess the impact of the nestor PERICLES school 2016. The evaluation form was designed to determine the effectiveness of the following:

- organisation of event;
- structure of event;
- content of event;
- speakers;
- learning and teaching methodologies;
- supporting material and documentation
- communication channels and promotion of event.

Based on the analysis of returned evaluation forms, an evaluation report was created outlining the short-term impact of the training programme and identifying areas where improvement is recommended to meet participants' needs in an optimised way.

The overall feedback confirms that this training event was a great success. The organisation and structure were highly rated. Participants particularly appreciated the informal exchange of ideas and group discussions with other researchers, international experts, and practitioners. The course met the participants' expectations, providing them with useful knowledge to take back and utilise within their own institutions.

The speakers and teaching methods received very high praise. However participants expressed a wish for more real-life examples and would have appreciated more hands-on exercises involving

specific preservation and curation tools. These comments clearly demonstrate the value of providing hands-on exercises and indicate that more would be appreciated.

Finally, the fact that participants felt inspired to become more deeply involved in the field of digital preservation is evidence of the positive effect of the course - both in terms of the quality of training materials and the enjoyment of the event itself. By providing students with the possibility of meeting colleagues and exchanging information and experiences about digital preservation, the organisers hope to have helped them to establish and develop a new community of future collaborators that will endure for years to come.

The most negative feedback was remarked regarding the instable WLAN access and the a bit unpleasant vegetarian food.

## Annex

---

### **Annex A: PERICLES School 2016, 15. – 17. February 2016: Reader**

### **Annex B: nestor PERICLES School 2016: Assignment of tasks**

- **nestor / PERICLES School 2016: Übung 1**
- **nestor / PERICLES School 2016: Übung 2 Gruppe 1**
- **nestor / PERICLES School 2016: Übung 2 Gruppe 2**
- **nestor / PERICLES School 2016: Übung 2 Gruppe 3**
- **nestor / PERICLES School 2016: Übung 2 Gruppe 4**
- **nestor / PERICLES School 2016: Übung 3 Gruppe 1**
- **nestor / PERICLES School 2016: Übung 3 Gruppe 2**
- **nestor / PERICLES School 2016: Übung 3 Gruppe 3**
- **nestor / PERICLES School 2016: Übung 3 Gruppe 4**
- **nestor / PERICLES School 2016: Übung 4**



## Langzeitarchivierung von Forschungsdaten

Landesmusikakademie Hessen  
Schloss Hallenburg

Vom 15. bis 17. Februar 2016

nestor PERICLES School 2016



# nestor



nestor School 2016

„Langzeitarchivierung von Forschungsdaten“

Landesmusikakademie Hessen  
Schloss Hallenburg in Schlitz  
vom 15. bis 17. Februar 2016

## **Herausgeber**

Stefan Strathmann, SUB Göttingen  
Achim Oßwald, FH Köln

Jonas Recker, GESIS - Leibniz-Institut für Sozialwissenschaften

## **im Auftrag von**

nestor – Kompetenznetzwerk Langzeitarchivierung und Langzeit-  
verfügbarkeit digitaler Ressourcen für Deutschland

nestor – Network of Expertise in Long-Term Storage of Digital Resources  
<http://www.langzeitarchivierung.de>

## **Kontakt**

Niedersächsische Staats- und Universitätsbibliothek Göttingen  
Stefan Strathmann  
Forschung und Entwicklung  
Papendiek 14  
37073 Göttingen  
[strathmann@sub.uni-goettingen.de](mailto:strathmann@sub.uni-goettingen.de)  
Tel. +49 (0) 551 39 78 06

Technische Hochschule Köln  
Institut für Informationswissenschaft  
Fakultät für Informations- und Kommunikationswissenschaften  
Prof. Dr. Achim Oßwald  
Claudiusstr. 1  
50678 Köln  
Tel.: +49 (0)221/8275-3375  
E-Mail: [achim.osswald@th-koeln.de](mailto:achim.osswald@th-koeln.de)

Der Inhalt steht unter folgender Creative Commons Lizenz:  
<http://creativecommons.org/licenses/by-nc-sa/3.0/de/>



# Inhaltsverzeichnis

<b>Vorwort</b>	<b>6</b>
<b>Referenten</b>	<b>7</b>
<b>Programm</b>	<b>10</b>
<b>Veranstaltungsort</b>	<b>12</b>
<b>Vorträge</b>	<b>15</b>
Einführung	18
Lektion I	
Forschungsdaten managen und archivieren	30
Lektion II	
Forschungsdatenmanagment (FDM) in der Praxis: Embedded Data Manager	54
Lektion III	
Auswahl und Bewertung von Forschungsdaten für die Archivierung und Nachnutzung	62
Lektion IV	
PERICLES Sheer Curation Tools	
Das Erhalten digitaler Objekte beginnt in ihrer Entstehungsumgebung	78

## Vorwort

Herzlich willkommen zur nestor PERICLES School 2016!

Die nestor School – diesmal zum Thema

„Langzeitarchivierung von Forschungsdaten“

– findet in dieser Form seit 2007 bereits zum neunten Mal statt. Wir freuen uns, damit die bewährte Fort- und Weiterbildungsreihe der nestor Schools fortsetzen zu können. Die Veranstaltung wendet sich an Praktikerinnen und Praktiker sowie Interessierte aus dem Bereich Forschung und Lehre. Als spezielle Form einer Schulungsveranstaltung bietet sie auch einen idealen Rahmen, Kontakte und Netzwerke auf- und auszubauen und damit Synergieeffekte zu ermöglichen.

Wie schon bei den letzten nestor Schools verzichten wir auch in diesem Jahr auf eine allgemeine Einführung in die digitale Langzeitarchivierung. Stattdessen greifen wir direkt sowohl ganz praktische als auch konzeptionelle Aspekte des School-Themas auf. Nach dem gemeinsame Versuch zu klären, was unter Forschungsdaten verstanden wird bzw. werden kann und welche Fragen sich ergeben, wenn Forschungsdaten zum Gegenstand kuratierender Aktivitäten werden, stellt ein Referent seine ganz praktischen Erfahrungen beim Management von Forschungsdaten vor und zur Debatte. Welche Forschungsdaten überhaupt und weshalb für die Langzeitarchivierung ausgewählt werden, hat vielfältige und weitreichende Folgen. Erfahrungen und grundsätzliche Überlegungen hierzu werden in der dritten Session der School angesprochen. Im EU-Projekt PERICLES, das die diesjährige nestor School mitträgt, stehen Fragen rund um die Sicherung des mittel- und langfristigen Zugriffs auf Forschungsdaten in veränderten Anwendungsumgebungen im Mittelpunkt. Werkzeuge und Verfahren, die hierfür entwickelt werden, sind Gegenstand der abschließenden Session der School.

Erneut ist es uns hierfür erfreulicherweise gelungen, eine Reihe renommierter Experten als Referentinnen und Referenten zu gewinnen. Ihnen wie auch unseren Mitorganisatorinnen und Mitorganisatoren der nestor School 2016 gilt unser spezieller Dank!

Herzlich bedanken möchten wir uns an dieser Stelle auch für die Unterstützung durch das EU-Projekt PERICLES (Promoting and Enhancing Reuse of Information throughout the Content Lifecycle taking account of Evolving Semantics; <http://pericles-project.eu/>), das erfreulicherweise zur Realisierung der Veranstaltung beigetragen hat.

Ein weiterer Dank geht an die Hochschulen und anderen Qualifizierungseinrichtungen, die sich seit 2007 im Rahmen von nestor der Qualifizierung im Bereich Langzeitarchivierung widmen und die die Organisation der nestor Schools mit tragen.

Wir freuen uns auf eine interessante und inspirierende Veranstaltung!

Februar 2016 Stefan Strathmann – Jonas Recker – Achim Oßwald  
SUB Göttingen – gesis – TH Köln

---

## Referenten

**Fabian Cremer****Max Weber Stiftung:**

Fabian Cremer studierte Kunstgeschichte und Kunstpädagogik in Frankfurt und Wien. Seit 2011 unterstützt und berät er WissenschaftlerInnen im Bereich der digital gestützten Forschung - beim Kunstgeschichtlichen Institut der Goethe-Universität Frankfurt, in der Abteilung Forschung und Entwicklung an der SUB Göttingen sowie am Max-Planck-Institut zur Erforschung multireligiöser und multiethnischer Gesellschaften. Seit 2015 koordiniert er die digitalen Quellen- und Datenbankprojekte der Max Weber Stiftung - Deutsche Geisteswissenschaftliche Institute im Ausland. Neben den Arbeitsschwerpunkten der Digital Humanities und Forschungsinfrastrukturen verfolgt er Strategien und Methoden zur Integration von Datenmanagement in Forschungsprojekten.

**Anna Eggers****Niedersächsische Staats- und  
Universitätsbibliothek Göttingen (SUB)****Jens Ludwig****nestor/DINI AG Forschungsdaten & Staatsbibliothek zu Berlin,  
Stiftung Preußischer Kulturbesitz:**

Jens Ludwig hat Philosophie und Informatik studiert und begann 2005 an der SUB Göttingen als Entwickler in einem Projekt zur Langzeitarchivierung digitaler Daten zu arbeiten. 2008 schloss er ein Studium in Library and Information Science ab. Er hat in verschiedenen Projekten in den Bereichen Langzeitarchivierung, Forschungsdaten, e-Science und Digital Humanities gearbeitet und Arbeitspakete in EU- und BMBF-Projekten geleitet. 2012-2014 war er stellvertretender Leiter der Abteilung Forschung & Entwicklung der SUB Göttingen und seit 2015 arbeitet er an der Staatsbibliothek zu Berlin. Mit dem Thema Forschungsdaten beschäftigt sich Jens Ludwig seit 2009 und er ist zurzeit als einer der Sprecher der nestor/DINI AG Forschungsdaten und im Projekt DMP-werkzeug aktiv.



**Reiner Mauer**

**GESIS – Leibniz-Institut für Sozialwissenschaften:**

Reiner Mauer studierte Volkswirtschaft und Soziologie an der Universität Köln. Seit 1998 ist er in verschiedenen Projekten und Positionen am Datenarchiv für Sozialwissenschaften mit der Kuratierung, Langzeitarchivierung und Bereitstellung von Forschungsdaten befasst. Seit 2006 ist er verantwortlich für die zentralen Prozesse der Datenarchivierung bei GESIS. Neben der operativen Steuerung, gehört die konzeptionelle und praktische Weiterentwicklung der Datenarchivierung und damit verbundener Services zu seinen Aufgaben. Er ist Sprecher der AG Forschungsdaten der Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen.



**Achim Oßwald**

**Technische Hochschule Köln, Institut für Informationswissenschaft:**

Achim Oßwald studierte Geschichte und Germanistik in Stuttgart und Freiburg i.Br. sowie Informationswissenschaft in Berlin und Konstanz. Er arbeitete mehr als 10 Jahre im Bereich Bibliothek, Information und Dokumentation - als Anwender, Vertriebsmitarbeiter eines Softwareanbieters, Dozent und Leiter einer Weiterbildungseinrichtung (Lehrinstitut für Dokumentation, Frankfurt) sowie als Consultant.

Seit 1994 Professor an der heutigen TH Köln, Fakultät für Informations- und Kommunikationswissenschaften. Im Institut für Informationswissenschaft u.a. verantwortlich für den Bereich „IT-Anwendungen im Informationswesen“. Aktuelle Arbeitsgebiete sind u.a. Aspekte geschäftsgangorientierter Bibliothekssysteme, des Digitalen Publizierens und der Qualifizierung im Bereich Langzeitarchivierung.

Studiengangsleiter des berufsbegleitenden Masterstudiengangs „Bibliotheks- und Informationswissenschaft“ (MALIS) der TH Köln. Vertreter der im Memorandum of Understanding zusammengeschlossenen Hochschulpartner im nestor-Kompetenzwerk.



**Jonas Recker**

**GESIS – Leibniz-Institut für Sozialwissenschaften:**

Jonas Recker studierte Englisch, Philosophie und Erziehungswissenschaften an der Universität zu Köln und absolvierte den Masterstudiengang „Library and Information Science (MALIS)“ an der TH Köln. Seit 2012 ist er im Datenarchiv für Sozialwissenschaften am GESIS - Leibniz Institut für Sozialwissenschaften tätig. Dort ist er - unter anderem im Rahmen der EU-Projekte DASISH, CESSDA SaW und SERISS - mit den Themen Forschungsdatenmanagement in den Sozialwissenschaften, Langzeitarchivierung von Forschungsdaten und vertrauenswürdige Archive befasst. Für das Consortium of European Social Science Data Archives (CESSDA) ist er zudem an der Durchführung und Koordination von Weiterbildungsangeboten in diesen Themenbereichen beteiligt.

**Stefan Strathmann - Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB):**

Stefan Strathmann studierte Deutsche Philologie, Publizistik und Kommunikationswissenschaften und Politische Wissenschaften an der Georg August Universität Göttingen. Er ist und war in eine Vielzahl von Projekten und Initiativen zur digitalen Langzeitarchivierung eingebunden und koordiniert die Aktivitäten der SUB Göttingen im Bereich der digitalen Langzeitarchivierung. In nestor koordiniert und vertritt er die Aktivitäten im Bereich Qualifizierung und hat in diesem Zusammenhang die nestor Schools mitorganisiert sowie als Herausgeber am nestor Handbuch und der nestor Bestandsaufnahme zur Langzeitarchivierung von Forschungsdaten mitgewirkt. Nach langjähriger Tätigkeit in der Abteilung Forschung und Entwicklung der SUB Göttingen ist er jüngst in die Abteilung Digitale Bibliothek der SUB gewechselt um künftig stärker die Umsetzung der LZA an der SUB zu begleiten.

## Programm

<b>Montag, 15. Februar 2016</b>	
11:45 – 12:15	Registrierung und Kaffee
12:15 – 13:00	Willkommen und Einführung in die nestor School 2016 <i>Stefan Strathmann, Achim Oßwald und Jonas Recker</i>
13:00 – 14:00	<i>Mittagessen</i>
14:00 – 15:00	<b>Vortrag I: Was sind Forschungsdaten und was bedeutet es, sie zu managen oder zu archivieren?</b> <i>Jens Ludwig</i>
15:00 – 15:30	<i>Kaffeepause</i>
15:30 – 17:00	<b>Übung I: Was sind Forschungsdaten und was bedeutet es, sie zu managen oder zu archivieren?</b>
17:00 – 18:00	Präsentation (ca. 5 min pro Arbeitsgruppe) und Diskussion der Ergebnisse der <b>Übung I</b>
ab 18:00	<i>Abendessen</i>
19:00	Poster-Session

<b>Dienstag, 16. Februar 2016</b>	
ab 08:00	<i>Frühstück</i>
09:00 – 10:00	<b>Vortrag II: Forschungsdatenmanagement (FDM) in der Praxis: Embedded Data Manager</b> <i>Fabian Cremer</i>
10:00 – 11:30	<b>Übung II: FDM in der Praxis: Embedded Data Manager</b>
11:30 – 12:00	<i>Kaffeepause</i>
12:00 – 13:00	Präsentation (ca. 5 Minuten pro Arbeitsgruppe) und Diskussion der Ergebnisse der <b>Übung II</b>
13:00 – 14:00	<i>Mittagessen</i>
14:00 – 15:00	<b>Vortrag III: Auswahl und Bewertung von Forschungsdaten für die Archivierung und Nachnutzung</b> <i>Reiner Mauer</i>
15:00 – 15:30	<i>Kaffeepause</i>
15:30 – 17:00	<b>Übung III: Auswahl und Bewertung von Forschungsdaten für die Archivierung und Nachnutzung</b>
17:00 – 18:00	Präsentation (ca. 5 Minuten pro Arbeitsgruppe) und Diskussion der Ergebnisse der <b>Übung III</b> und Zusammenfassung des Tages
18:00	<i>Abendessen</i>
19:30	Gemütliches Beisammensein bei einem Glas Wein

<b>Mittwoch, 17. Februar 2016</b>	
ab 08:00	<i>Frühstück</i>
09:00 – 10:00	<b>Vortrag IV: PERICLES Sheer Curation Tools - Die Erhaltung Digitaler Objekte beginnt in ihrer Entstehungsumgebung</b> <i>Anna Eggers</i>
10:00 – 11:30	<b>Übung IV: PERICLES Sheer Curation Tools - Die Erhaltung Digitaler Objekte beginnt in ihrer Entstehungsumgebung</b>
11:30 – 12:00	<i>Kaffeepause</i>
12:00 – 13:00	Präsentation (ca. 5 min pro Arbeitsgruppe) und Diskussion der Ergebnisse der <b>Übung IV</b>
13:00 – 14:00	<i>Mittagessen</i>
14:00 – 14:30	Abschluss und Evaluation





## Veranstaltungsort

Die nestor PERICLES School findet vom 15. bis 17. Februar 2016 in der Landesmusikakademie Hessen Schloss Hallenburg in Schlitz statt. Die Landesmusikakademie Hessen befindet sich in den historischen Gebäuden des Schloss Hallenburg in Schlitz.

Schlitz ist eine Kleinstadt im Vogelsbergkreis in Osthessen. Das Schloss Hallenburg war die Residenz der Grafen von Schlitz, liegt südöstlich der Altstadt von Schlitz und ging aus einem befestigten Hofgut hervor.



# Vorträge



---

## Vorträge

Einführung	18
Lektion I	
Forschungsdaten managen und archivieren	30
Lektion II	
Forschungsdatenmanagement (FDM) in der Praxis: Embedded Data Manager	54
Lektion III	
Auswahl und Bewertung von Forschungsdaten für die Archivierung und Nachnutzung	62
Lektion IV	
PERICLES Sheer Curation Tools	
Das Erhalten digitaler Objekte beginnt in ihrer Entstehungsumgebung	78

# Einführung

Stefan Strathmann - Jonas Recker - Achim Oßwald

Kompetenznetzwerk  
Langzeitarchivierung  
- Qualifizierung -

Willkommen bei der  
nestor PERICLES School 2016

**Stefan Strathmann**

SUB Göttingen  
strathmann@sub.uni-goettingen.de

SUB | NIEDERSÄCHSISCHE STAATS- UND  
UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

**Jonas Recker**

gesis  
jonas.recker@gesis.org

gesis  
Leibniz-Institut  
für Sozialwissenschaften

**Achim Oßwald**

TH Köln  
achim.osswald@th-koeln.de

Technology  
Arts Sciences  
TH Köln

nestor  
www.langzeitarchivierung.de

pericles  
FP7 Digital Preservation

## Willkommen in Schlitz



...und zur nestor PERICLES School 2016

nestor  
www.langzeitarchivierung.de

nestor PERICLES School 2016  
Einführung und Organisatorisches  
Strathmann/Recker/Oßwald – Folie 2

pericles  
FP7 Digital Preservation

# Überblick

- **PERICLES**
- nestor
- nestor Qualifizierung
- nestor PERICLES School 2016

# PERICLES

- PERICLES = **P**romoting and **E**nhancing **R**euse of **I**nformation throughout the **C**ontent **L**ifecycle taking account of **E**volving **S**emantics
- [pericles-project.eu](http://pericles-project.eu)
- Gefördert von 02/2013 – 01/2017 im 7. Rahmenprogramm der Europäischen Kommission
- Partner: King's College London, University of Borås, Information Technologies Institute, Dotsoft SA, Georg-August-Universität Göttingen, The University of Liverpool, Space Applications Services NV, Xerox SAS, The University of Edinburgh, TATE, Belgian User Support and Operations Centre

## PERICLES - Ziele

- Das PERICLES-Projekt erforscht verschiedene Arten von Änderungen, die einen Einfluss auf den Lebenszyklus digitaler Objekten haben.
- Der Fokus liegt auf Veränderungen in Semantik, Prozessen, Richtlinien und Nutzergruppen und nicht auf dem technologischen Wandel von Hard- und Software.
- Zusätzlich werden eine Reihe von Programmen entwickelt, mit denen nicht-technischer Wandel verwaltet werden kann.
- Die Entwicklungen des Projekts werden anhand von Anwendungsbeispiele und Daten aus den Bereichen digitale Kunst und Weltraumforschung erprobt.

## Überblick

- PERICLES
- **nestor**
- nestor Qualifizierung
- nestor PERICLES School 2016

## Was ist nestor?

- Network of Expertise in long-term STOrage and availability of digital Resources in Germany
- Projektlaufzeiten:  
Mai 2003 - Juni 2006 & Juli 2006 - Juni 2009
- Seit 2009: Kooperationsvertrag
- nestor betreibt selber kein Archiv
- nestor selber ist kein Archivierungsprojekt,  
nestor versteht sich als Kompetenznetzwerk!

## nestor - Partner

- Bayerische Staatsbibliothek
- Bibliotheksservice-Zentrum Baden-Württemberg
- Bundesarchiv
- Computerspiele Museum Berlin
- Deutsche Kinemathek – Museum für Film und Fernsehen
- Deutsche Nationalbibliothek
- FernUniversität Hagen
- GESIS - Leibniz-Institut für Sozialwissenschaften
- Goportis - Leibniz-Bibliotheksverbund Forschungsinformation
- Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen
- Georg-August-Universität Göttingen / Niedersächsische Staats- und Universitätsbibliothek Göttingen
- Humboldt-Universität zu Berlin
- Institut für Deutsche Sprache
- Institut für Museumsforschung (Stiftung Preußischer Kulturbesitz)
- Konrad-Zuse-Zentrum für Informationstechnik Berlin
- Landesarchiv Baden-Württemberg
- Landesarchiv Nordrhein-Westfalen
- PDF Association
- Rechenzentrum der Universität Freiburg
- Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden

## nestor - Ziele

- Beitrag zur Sicherung des kulturellen Erbes durch Organisation und Koordination der langfristigen Archivierung digitaler Objekte in Deutschland
- Strategieentwicklung
- Identifizieren und stimulieren von notwendiger Forschung und Entwicklung
- Nationale und internationale Zusammenarbeit

## nestor - Schwerpunkte

### *Qualifizierung*

- nestor entwickelt in Zusammenarbeit mit Hochschulpartnern Aus-, Fort- und Weiterbildungsangebote im Bereich der digitalen Langzeitarchivierung in Deutschland

### *Vernetzung*

- nestor ist das nationale Kompetenznetzwerk für digitale Langzeitarchivierung
- nestor ist eingebunden in ein Netz europäischer Partner

### *Standardisierung*

- nestor bündelt Standardisierungsaktivitäten und vermittelt Standards in die Anwender-Communities

# nestor - Angebote

- **Information**
  - Homepage, News, Kalender
  - Informationsdatenbank
  - Wer Wo Was
  - Veröffentlichungen
  - ...
- **Kommunikation**
  - Arbeitsbereiche
  - E-Mail Listen
  - Twitter
  - ...
- **Organisation**
  - Workshops
  - Fortbildungsseminare
  - Konferenzen
  - ...

# nestor - Angebote



■ **Home**

- Arbeitsgruppen
- **Qualifizierung**
- Standardisierung
- nestor-Siegel
- Publikationen
- LZA-Projekte und -Aktivitäten
- Veranstaltungen und Termine
- Expertenkontakte
- Webressourcen
- Mailingliste und Newsletter

## WILLKOMMEN

Willkommen bei nestor, dem deutschen Kompetenznetzwerk zur digitalen Langzeitarchivierung. In nestor arbeiten Bibliotheken, Archive, Museen sowie führende Experten gemeinsam zum Thema Langzeitarchivierung und Langzeitverfügbarkeit digitaler Quellen. nestor ist ein Kooperationsverbund mit Partnern aus verschiedenen Bereichen, die alle mit dem Thema "Digitale Langzeitarchivierung" zu tun haben.

### INFORMATION

 nestor bietet eine Reihe von eigenen Publikationen zur digitalen Langzeitarchivierung und verweist auf vielfältige Webressourcen. Eine kurze Einführung zum Thema findet sich im nestor-Wiki.

### QUALIFIZIERUNG

nestor kooperiert mit Hochschulpartnern, die Aus...

## NEWS

**Tweets** [Folgen](#)

-  nestor @nestorNetzwerk 24 Nov  
Neue die 33. Ausgabe des nestor-Newsletters: [tinyurl.com/pj4tsxg](http://tinyurl.com/pj4tsxg)  
Öffnen
-  nestor @nestorNetzwerk 18 Nov  
nestor-Praktikertag 2016: Schwerpunktthema: „Migration und Emulation“ 14.06.2016 – SLUB Dresden  
Öffnen
-  nestor @nestorNetzwerk 18 Nov

# Überblick

- PERICLES
- nestor
- **nestor Qualifizierung**
- nestor PERICLES School 2016

# nestor Qualifizierung

- Kooperation von nestor (vertreten durch SUB Göttingen) mit Hochschulen in D, A und CH
- Zielsetzungen und Aktivitäten:
  - Ausbau des curricularen Anteils von LZA
  - Entwicklung von eTutorials
  - Schools (wie diese)
  - Seminare
  - Workshops/Konferenzen

# nestor Qualifizierung: Partner

**SUB** | Niedersächsische Staats- und  
Universitätsbibliothek Göttingen

**FRP**

**Archivschule Marburg**  
Institut für Archivwissenschaft  
Fachhochschule für Archivwesen

**TU  
WIEN**

**h\_da**  
HOCHSCHULE DARMSTADT  
UNIVERSITY OF APPLIED SCIENCES

**Technology  
Arts Sciences  
TH Köln**

**H. M. H. ROLDT-UNIVERSITÄT  
ZU BERLIN**

**HTWK  
Leipzig**

**HTW Chur**  
Hochschule für Technik und Wirtschaft  
Fachhochschule Ostschweiz  
University of Applied Sciences

**STAATLICHE  
AKADEMIE  
BILDENDEN  
STUTTART** | **DER  
KUNSTE**

**gesis**  
Leibniz-Institut  
für Sozialwissenschaften

**HOCHSCHULE DER MEDIEN**

**htw**  
Hochschule für Technik  
und Wirtschaft Berlin  
University of Applied Sciences

**nestor**  
www.langzeitarchivierung.de

nestor PERICLES School 2016  
Einführung und Organisatorisches  
Strathmann/Recker/Oßwald – Folie 15

**pericles**  
FP7 Digital Preservation

# nestor Qualifizierung: Website

**nestor**

Kompetenznetzwerk  
Langzeitarchivierung  
- Qualifizierung -

Home

Qualifizierung

- › nestor Handbuch
- › Forschungsdatenbuch
- › School 2014
- › School 2012
- › Winter School 2011
- › Summer School 2010
- › Spring School 2009
- › Summer School 2008
- › Winter School 2007
- › Spring School 2007

Kontakt

## Qualifizierung

Im Projekt nestor wurde ein umfangreicher Aus- und Weiterbildungsbedarf erkannt und daher damit begonnen, ein Konzept für ein Aus- und Fortbildungsangebot in Zusammenarbeit mit bestehenden Einrichtungen zur Aus- und Fortbildung (Bibliothek, Archiv, Museum) zu entwerfen. Die Zusammenarbeit der Hochschulpartner mit nestor hat 2007 durch ein gemeinsames **Memorandum of Understanding** einen Rahmen erhalten. Dieses MoU wurde durch eine **Neuaufgabe** nach dem nestor Projektende kontinuierlich erweitert und erneuert.

Es werden unterschiedliche Veranstaltungsarten durchgeführt, die den Bedürfnissen der Aus- und Weiterbildung gerecht werden. Insbesondere die mehrtägigen Schools (**nestor/DPE Spring School 2007**, **nestor/DPE Winter School 2007**, **nestor/DPE Summer School 2008**, **nestor/DPE Spring School 2009**), aber auch eintägige Seminare und Workshops treffen auf ein großes Interesse bei den mit der digitalen Langzeitarchivierung befaßten Mitarbeitern von Archiven, Bibliotheken, Museen und anderen Informationseinrichtungen sowie Studenten verschiedener Hochschuleinrichtungen.

Eine zusätzliche Besonderheit ist die Entwicklung von Lehr- und

**Archivschule Marburg**  
Hochschule für Archivwissenschaft

**HTW Chur**  
Hochschule für Technik und Wirtschaft  
Fachhochschule Ostschweiz  
University of Applied Sciences

**TU  
WIEN**

**Fachhochschule Köln**  
Cologne University of Applied Sciences  
Institut für Informationswissenschaft  
Institute of Information Science

**nestor**  
www.langzeitarchivierung.de

nestor PERICLES School 2016  
Einführung und Organisatorisches  
Strathmann/Recker/Oßwald – Folie 16

**pericles**  
FP7 Digital Preservation

## nestor Handbuch & Bestandsaufnahme zur Langzeitarchivierung von Forschungsdaten



## Überblick

- PERICLES
- nestor
- nestor Qualifizierung
- **nestor PERICLES School 2016**

# nestor PERICLES School 2016

- 38 TeilnehmerInnen und ReferentInnen
  - ReferentInnen
  - PraktikerInnen aus Gedächtnisinstitutionen
  - VertreterInnen aus Wirtschaft & Gewerbe
  - Studierende & Lehrende

## Organisation



Stefan Strathmann



Jonas Recker



Achim Oßwald

Kompetenznetzwerk  
Langzeitarchivierung  
- Qualifizierung -

**Herzlichen Dank!**

Fragen & Ergänzungen ?

**nestor**  
www.langzeitarchivierung.de

**pericles**  
FP7 Digital Preservation

# Forschungsdaten managen und archivieren – Lektion I

Jens Ludwig

# Was sind Forschungsdaten und was bedeutet es, sie zu managen oder zu archivieren?

Jens Ludwig

nestor/DINI AG Forschungsdaten &  
Staatsbibliothek zu Berlin - Stiftung  
Preußischer Kulturbesitz

## Inhaltsübersicht

- Was sind Forschungsdaten?
- Was ist Langzeitarchivierung?
- Forschungsdatenmanagement und Unterschiede zur Langzeitarchivierung
- Einige Überlegungen zur Informationsinfrastruktur

# Inhaltsübersicht

- Was sind Forschungsdaten?
- Was ist Langzeitarchivierung?
- Forschungsdatenmanagement und Unterschiede zur Langzeitarchivierung
- Einige Überlegungen zur Informationsinfrastruktur

## Beispiele für Forschungsdaten

Statistiken, Interviews, Simulationen, Messdaten aus Experimenten, Beobachtungsdaten aus Instrumenten, Text mit semantischen Annotationen, 3D-Scans, Video, Audio, Bilder, Tabellen, Dokumente, Binärdaten, Software, Textdateien, ...

Der erste Eindruck ist, dass fast alles als Forschungsdaten angesehen werden kann.

## Abgrenzung von analogen Forschungsdaten

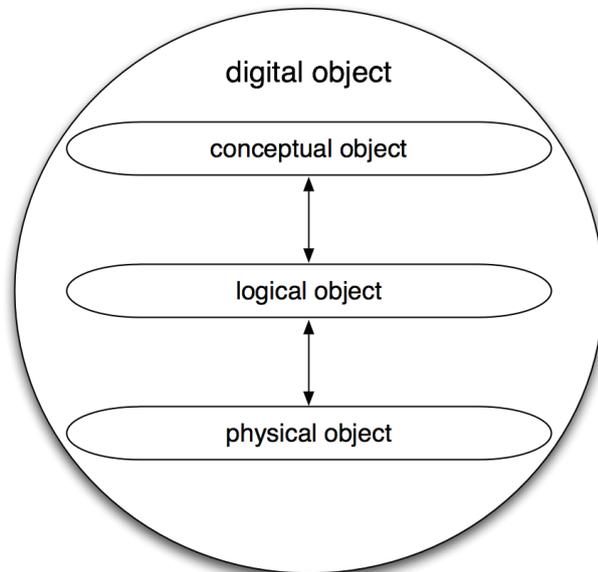
- Auch mit Bleistift und Papier notierte Messdaten sind eigentlich Forschungsdaten.
- Analoge Forschungsdaten werden im Folgenden ignoriert.
- Viele Fragestellungen des Forschungsdatenmanagements stellen sich nicht in gleicher Form und im gleichen Maße für analoge Forschungsdaten.
- Hauptgrund: Massenproduktion und -verarbeitung von Forschungsdaten ist digital effizienter, ermöglicht neue Verfahren und ist der neue Standard.

## Was ist ein digitales Objekt? (nach Thibodeau 2002)

conceptual object:  
für Menschen bedeutsames Objekt

logical object:  
von Software und Technik  
verstehbares Objekt

physical object:  
materielle Zeichen und  
Trägermedium

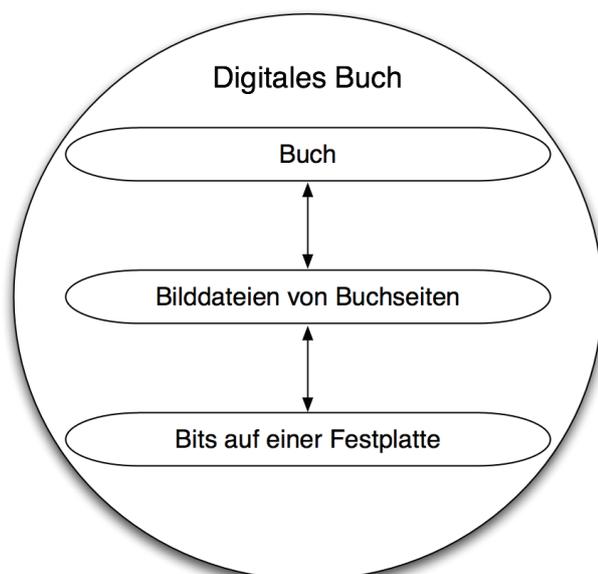


## Beispiel: Ebenen eines digitalen Buchs

conceptual object:  
Buch mit z.B. navigierbaren  
Verzeichnissen, Abbildungen, ...

logical object:  
Dateien in verschiedenen Formaten  
(TIFF, TXT, ...)

physical object:  
Bits auf einer Festplatte, die ein  
bestimmtes Dateisystem hat



# Definitionsversuch 1

- Forschungsdaten sind einfach alle Daten, die im Forschungsprozess benutzt werden.
- Das wird dem Phänomen gerecht, dass ein und dieselben digitalen Daten je nach Kontext Forschungsdaten sein können oder nicht. Z.B. Urlaubsfoto eines Denkmals in der privaten Fotosammlung versus Foto als Teil eines Trainingsdatensatzes für Bilderkennungsverfahren.
- Es werden aber viele Dinge als Forschungsdaten erfasst, die nicht als Forschungsdaten bezeichnet werden sollten. Z.B. Literatur in Form von PDF-Dateien, die von Forscherinnen und Forscher während des Forschungsprozesses konsultiert wird.
  - Natürlich können digitale Texte auch Forschungsdaten sein, wenn sie z.B. Teil eines Textkorpus sind, der in der Computerlinguistik benutzt wird.
  - Forschungsdaten können auch in Texten enthalten sein, z.B. chemische Strukturformeln in Fachartikeln.

## Zwei Arten von Forschungsdatenbezeichnungen und Definitionsversuch 2

- Beobachtung: Wenn man FachvertreterInnen befragt, was Forschungsdaten in ihrer Disziplin sind, erhält man grob zwei Gruppen von Antworten (Ludwig, in: Neuroth et al 2012, S. 299ff).
  1. Forschungsdaten werden durch ihren Medientyp (z.B. Video, Markup, Tabellen, 3D-Modelle)
  2. oder durch die erzeugende Forschungsmethode bezeichnet (z.B. Simulationsdaten, Beobachtungsdaten, Statistikdaten).
- Definitionsversuch 2: Forschungsdaten sind alle diejenigen Daten, die durch die Anwendung einer wissenschaftlichen Methode im Forschungsprozess entstehen.
  - Ähnlich: „Forschungsdaten sind Daten, die im Zuge wissenschaftlicher Vorhaben z.B. durch Digitalisierung, Quellenforschungen, Experimente, Messungen, Erhebungen oder Befragungen entstehen.“ (<http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten.html>)
- Aber nicht alle Forschungsdaten entstehen in Forschungsprozessen und durch wissenschaftliche Methoden. Z.B. Behördendaten, die später für sozialwissenschaftliche Forschung benutzt werden. Oder: „Help scientists recover Arctic and worldwide weather observations made by United States ships since the mid-19th century by transcribing ships' logs.“ (<http://www.oldweather.org/>)

## Vielfalt der Forschungsdaten und Definitionsversuch 3

- Forschungsdaten sind so vielfältig wie die Forschung selbst.
  - Es kann (fast?) alles Objekt wissenschaftlicher Untersuchung sein und es kann Forschungsdaten über fast alles geben.
  - Die Forschungsmethoden sind nicht beliebig, aber die dabei entstehenden Forschungsdaten sind sehr vielfältig.
- Forschungsdaten sind alle Daten, die mit einer wissenschaftlichen Methode über ein Forschungsobjekt erzeugt oder verarbeitet werden.

## Abgrenzungsversuch

- Die Klausel „über ein Forschungsobjekt“ soll von reinen digitalen Stellvertretern des Forschungsobjekts abgrenzen.
- Beispiel: Digitalisate einer mittelalterlichen Handschrift sind Daten, aber meist nicht „über ein Forschungsobjekt“, sondern eher eine Repräsentationsform des Forschungsobjekts.
  - Das hängt natürlich von der Forschung ab. Ein Literaturwissenschaftler erfährt durch das Digitalisat vermutlich nichts Neues „über sein Forschungsobjekt“, den Text, während das für eine Handschriftenforscherin, die das materielle Objekt untersucht, anders sein kann.

## Wozu überhaupt eine Definition?

- Definitionen sind ein Instrument und damit nicht zweckfrei und neutral.
- Definitionen können zwar in der Kommunikation hilfreich sein. Aber auch ohne oder mit sehr unscharfen Definitionen kann Kommunikation erfolgreich sein und werden Fortschritte gemacht.
- Wozu dient uns die Definition? Wozu dient uns der Forschungsdaten-Begriff?

## Position: Forschungsdatenmanagement als der Sinn des Forschungsdatenbegriffs

- Der Definitionsvorschlag 3 ist sinnvoll, weil er hilft die neuen und bisher in klassischen Informationseinrichtungen nicht behandelten Aufgaben zu verstehen.
- Für Digitalisate, digitale Texte und Textpublikationen gibt es sehr viele etablierte Verfahren und Methoden, um sie zu managen. Diese Aufgaben sind aber sehr verschieden von vielen anderen Aufgaben des Forschungsdatenmanagements.
- Die kürzeste Definition wäre deshalb, dass Forschungsdaten die Objekte des Forschungsdatenmanagements sind.

## Inhaltsübersicht

- Was sind Forschungsdaten?
- Was ist Langzeitarchivierung?
- Forschungsdatenmanagement und Unterschiede zur Langzeitarchivierung
- Einige Überlegungen zur Informationsinfrastruktur

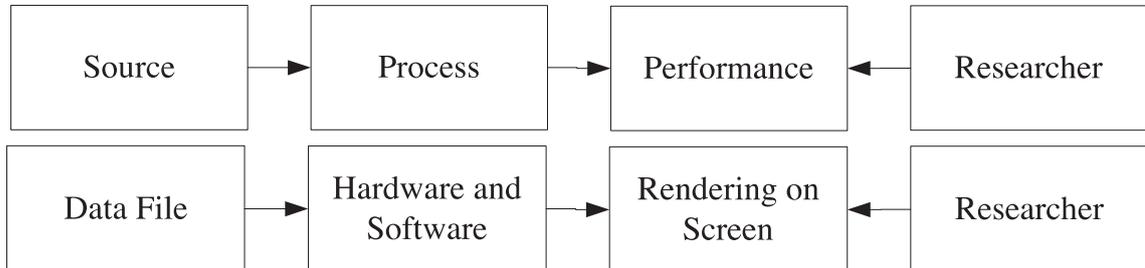
## Langzeitarchivierung

„Langzeit“ ist die Umschreibung eines nicht näher fixierten Zeitraumes, währenddessen wesentliche, nicht vorhersehbare technologische und soziokulturelle Veränderungen eintreten [...]

„Archivieren“ bedeutet [...] mehr als nur die dauerhafte Speicherung digitaler Informationen auf einem Datenträger. Vielmehr schließt es die Erhaltung der dauerhaften Verfügbarkeit und damit eine Nachnutzung und Interpretierbarkeit der digitalen Ressourcen mit ein.

Aus: Hans Liegmann, Heike Neuroth, Einführung, in: nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung hg. v. H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, K. Huth urn:nbn:de:0008-2010030508

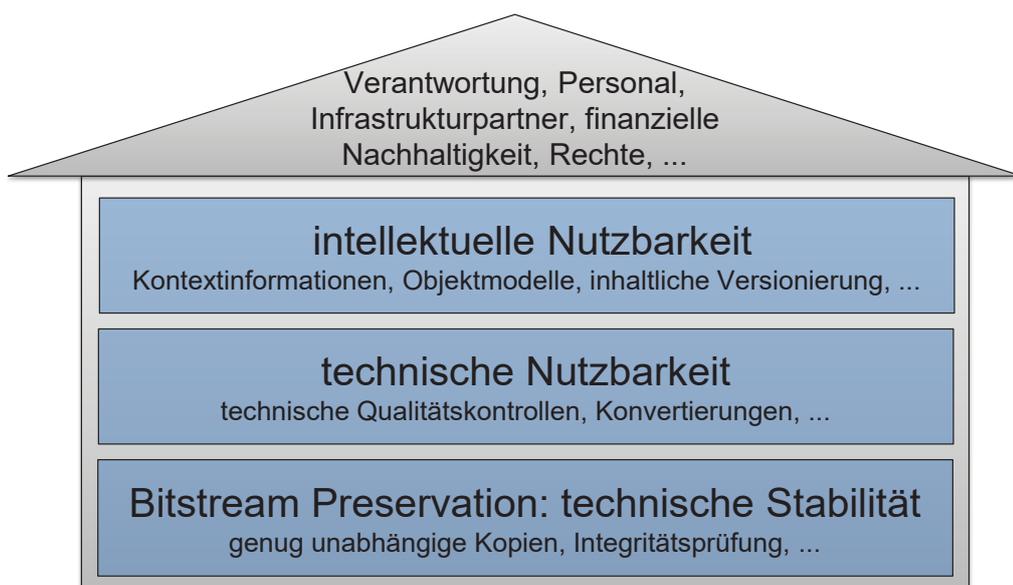
## Digitale Objekte existieren nur in der Ausführung



Quelle: Heslop, Helen; Davis, Simon; Andrew Wilson: An Approach to the Preservation of Digital Records. National Archives of Australia, 2002

- Ein Grundproblem der LZA: Das gleiche digitale Objekt kann abhängig vom Kontext und der Umgebung zu ganz unterschiedlichen bzw. gar keinen Ergebnissen führen.
- Vom Drei-Ebenen-Modell digitaler Objekte von Thibodeau kann man unterschiedliche Aufgaben der LZA ableiten.

## Aspekte der Langzeitarchivierung

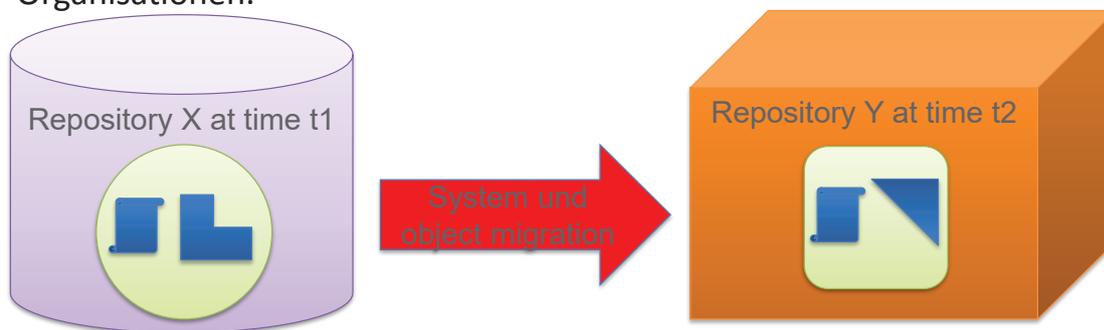


# Was ist Langzeitarchivierung nicht?

„Preservation is not a Place“ (Stephen Abrams et al., doi:10.2218/ijdc.v4i1.72)

Auch Archivsysteme veralten und nicht immer ist ein separates Archivsystem zusätzlich zum existierenden Repository sinnvoll.

Deshalb: Kein technisches System macht Langzeitarchivierung, sondern Organisationen.



## Inhaltsübersicht

- Was sind Forschungsdaten?
- Was ist Langzeitarchivierung?
- Forschungsdatenmanagement und Unterschiede zur Langzeitarchivierung
- Einige Überlegungen zur Informationsinfrastruktur

# Wieso Forschungsdatenmanagement?



# Service-Klasse: Dokumentation



## DFG, Sicherung guter wissenschaftlicher Praxis, 2013

„Ein in der Öffentlichkeit im In- und Ausland breit diskutierter Fall **wissenschaftlichen Fehlverhaltens** hat das Präsidium der Deutschen Forschungsgemeinschaft (DFG) veranlasst, eine international zusammengesetzte Kommission unter Vorsitz des Präsidenten zu berufen [...]“ (Vorwort zur ersten Auflage, 1997)

„Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, für zehn Jahre aufbewahrt werden.“



## Service-Klasse: Dokumentation

Ziel: Nachvollziehbarkeit für Verantwortungszwecke

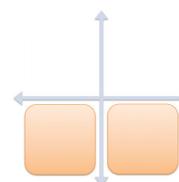
Zielgruppe: Institutionen (weniger Wissenschaftler)

Anbieter: lokale Infrastruktur z.B. Rechenzentren, Bibliotheken

Hohes Volumen, sehr seltene Datenanfragen, begrenzte Dauer

- einfacher und schneller Zugriff wird Kostenersparnis geopfert
- voraussichtlich Bitstream Preservation + basale Metadaten + Hard-/ Software-Museum

Faktisch werden Aufwand und Kosten für Dokumentation mit Kosten des Verantwortungsfalls abgewogen.



## Service-Klasse: Nachnutzung



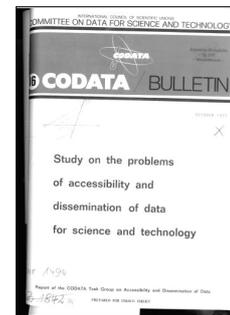
## Geschichte der organisierten Nachnutzung digitaler Forschungsdaten



# Geschichte der organisierten Nachnutzung digitaler Forschungsdaten

„One of the reasons for the ever-increasing efficiency of science and technology is that, whenever a new venture is undertaken, the results of past efforts are systematically retrieved and utilized. Such useful and important results are to be found, in concentrated form, in compilations of reliable data.“

Nachnutzbarkeit ist kein neues Aufgabengebiet:  
CODATA, „Study on the problems of accessibility and dissemination of data for science and technology“, 1975

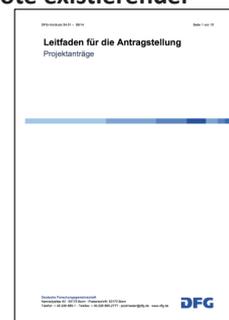


# DFG-Anträge seit 2010

„2.4 Umgang mit den im Projekt erzielten Forschungsdaten

[...] Die DFG ist daher bestrebt, durch ihre Förderung auch zur Sicherung, Aufbewahrung und **Nachnutzbarkeit von Forschungsdaten** beizutragen. [...] Wenn aus Projektmitteln systematisch Forschungsdaten oder Informationen gewonnen werden, die für die Nachnutzung durch andere Wissenschaftlerinnen und Wissenschaftler geeignet sind, legen Sie bitte dar, ob und auf welche Weise diese für andere zur Verfügung gestellt werden. **Bitte berücksichtigen Sie dabei auch - sofern vorhanden - die in Ihrer Fachdisziplin existierenden Standards und die Angebote existierender Datenrepositorien oder Archive.** [...]

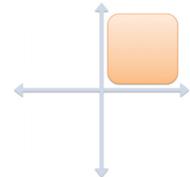
Die für die Nachnutzung der Forschungsdaten anfallenden projektspezifischen **Kosten können Sie im Rahmen des Projekts beantragen.** Stellen Sie in diesem Fall bitte auch dar, in welcher Form eine Unterstützung beim Daten- und Informationsmanagement durch die am Projekt beteiligten Institutionen geleistet wird.“ DFG, Leitfaden für die Antragstellung, Version 06/2014



## Service-Klasse: Nachnutzung

### Ziele:

- zitierfähige Datenpublikation
- erneute wissenschaftliche Nutzung von Daten oder Datenproduktion als Dienstleistung
- Förderersicht: erhöhte Effizienz
- Bewahrung nicht reproduzierbarer Daten



Zielgruppe: Fach-Communities

Anbieter: spezialisierte, in den Disziplinen verankerte Zentren

Daten werden selten benutzt, aber ohne klares Enddatum. Geringeres Datenvolumen, da Aufwand pro Datensatz hoch

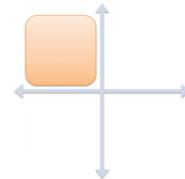
## Service-Klasse: verbesserte Nutzung



## Service-Klasse: verbesserte Nutzung

### Ziele:

- Erleichterung und Absicherung der Datennutzung
- Ermöglichung neuer Methoden/Funktionen



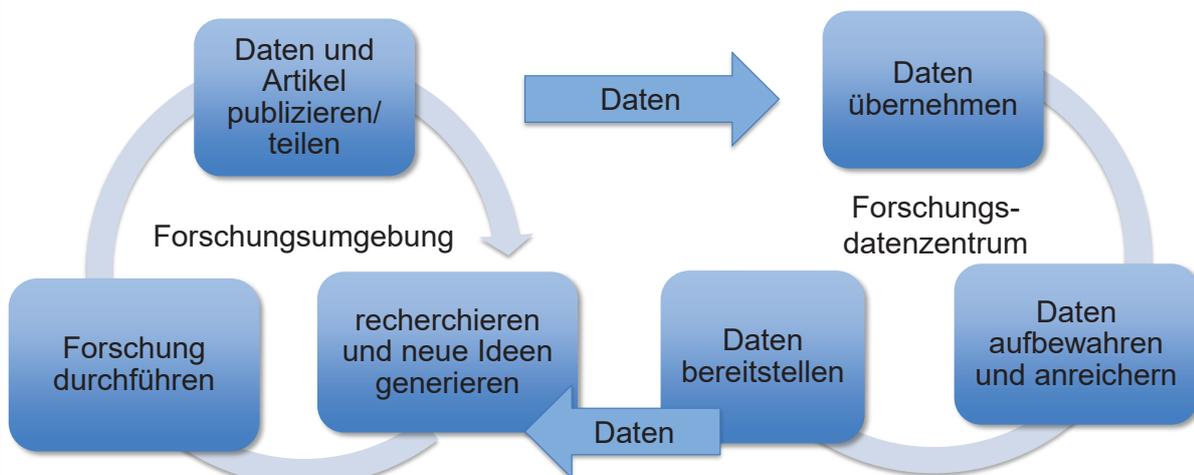
Zielgruppe: wissenschaftliche Arbeitsgruppen, z.B. SFBs

Anbieter: lokale Forschungsdaten-Support-Teams

Daten werden ständig benutzt und verändern sich, Aufbewahrungszeit durch Projektlaufzeit definiert.

Verknüpfung mit kollaborativen Forschungsumgebungen und Werkzeugen

## Forschungsdatenmanagement betrifft den gesamten (idealisierten) Forschungs- und Forschungsdatenzyklus



## Unterschiede zwischen Langzeitarchivierung und Forschungsdatenmanagement

- Langzeitarchivierung hat eine abstraktere Fragestellung (langfristigere Nutzbarkeit beliebiger digitaler Objekte) und sieht von dem Management des weiteren Anwendungsgebietes ab.
- Forschungsdatenmanagement beschränkt sich auf Forschungszwecke, aber kümmert sich viel stärker um spezielle Anforderungen und Prozesse der Zielgruppe.
- Tendenz: Forschungsdaten werden als instrumenteller Wert für die Forschung betrachtet. In der Langzeitarchivierung ist eher das digitale Objekt der Maßstab und wird als historische Überlieferung behandelt.

## Beispiel: Integrität und Authentizität

- Im Digitalen müssen wir das Identitätskriterium festlegen. Z.B. Identität...
  - des physical objects: dasselbe Trägermedium
  - des logical objects: dieselbe Bitfolge
  - des conceptual objects: dieselben wesentlichen Eigenschaften, unabhängig vom Dateiformat
- Bei Forschungsdaten sind dieselben wesentlichen Eigenschaften eher die Aussagen über das Forschungsobjekt für einen Forschungszweck. Oft ist es daher legitim und sinnvoll Forschungsdaten zu verändern, um sie zu korrigieren, zu optimieren oder zu erweitern.

## Beispiel: Dateiformate

- In der Langzeitarchivierung sind Dateiformate ein großes Aufgabenfeld:
  - Was sind die besten Dateiformate für die Langzeitarchivierung?
  - Wie bewerkstelligt man die Konvertierung von Dateiformaten?
  - ...
- Bei Forschungsdaten gibt es diese Fragen auch und z.T. sehr spezielle Dateiformate. Aber oft sind Dateiformate auch selbst entwickelte und etablierte Community-Standards und Werkzeuge werden von der Community selbst entwickelt. Dann stellen sich viele typische Fragen der Langzeitarchivierung nicht.

## Beispiel: disziplinspezifische Anforderungen des Forschungsdatenmanagements

- aufwändige Metadaten und Aufbereitung von Datensätzen
- inhaltliche und technische Qualitätskontrolle von Forschungsdaten
- Pflege langfristiger Zeitreihen
- Bereitstellung der notwendigen Werkzeuge zur Nachnutzung
- Unterstützung des Forschungsprozesses
- Interaktion mit Zielgruppe und Begleitung der Disziplinentwicklung (Anforderungen, Terminologie etc.)

# Inhaltsübersicht

- Was sind Forschungsdaten?
- Was ist Langzeitarchivierung?
- Forschungsdatenmanagement und Unterschiede zur Langzeitarchivierung
- Einige Überlegungen zur Informationsinfrastruktur

## Wieso kann nicht jede Infrastruktureinrichtung einfach das Forschungsdatenmanagement übernehmen?

- Die Aufgabe ist zu groß.
- Die Aufgabe wird zu pauschal betrachtet.
- Nicht jede Teilaufgabe ist richtig für jede Infrastruktureinrichtung.

# Daten- und Kostentrends

Verschiedene Quellen, mit Vorsicht zu genießen, aber Trend scheint klar:

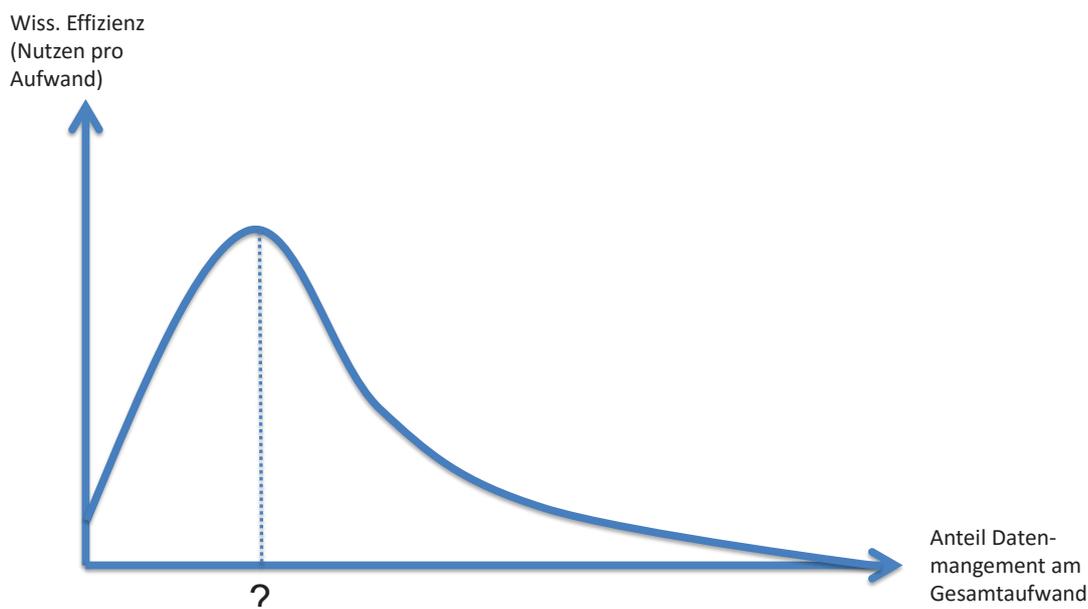
- Speicherplatzbedarf steigt jedes Jahr um 60%
- Die Speicherdichte wird nicht mehr als 20% pro Jahr steigen für die nächsten fünf Jahre. (Speicherdichte und Kosten sind eng verwandt.)
- IT-Budgets steigen ... quasi nicht.

David Rosenthal, Talk "Costs: Why Do We Care?", 2014.  
<http://blog.dshr.org/2014/11/talk-costs-why-do-we-care.html>

Der Speicherplatz ist aber nur überlicherweise nur ein Bruchteil der Gesamtkosten des Datenmanagement.

Neil Beagrie, Brian Lavoie and Matthew Woollard, Keeping Research Data Safe 2, 2010.  
<http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf>

In der Forschungsdaten-Diskussion implizit, aber offene Frage:  
Nicht nur Kosten, sondern Investition und Nutzen.  
Aber wieviel wäre optimal oder sinnvoll?



## Eine Konsequenz: Auswahl und Bewertung von Daten

- Nicht alles kann aufbewahrt werden, weder technisch noch ökonomisch.
  - Bedauerlich, aber stellt nicht das Gesamtvorhaben in Frage.
- Es müssen auch nicht alle Daten aufbewahrt werden.
  - Manche Daten aber schon.
- Scheindebatte: Wer entscheidet?
  - Meist von Besitz- und Autoritätsansprüchen überschattet.
  - Vorüberlegungen zu einem rationalen Auswahlprozess klären Vieles.

## So wie Kooperation notwendig ist bei der Gewinnung von Forschungsdaten, ...

Wissenschaft ist oft kompetitiv.

Kooperation z.B. aufgrund von:

- Aufwand der Datenerhebung/Instrumente
- Größe des Untersuchungsgegenstand
- Interdisziplinärität der Fragestellung, Spezialisierung der Fachexpertise



## ... so ist Kooperation notwendig beim Forschungsdatenmanagement.

### Datenvolumen

### Technik

- technischer Aufwand einzeln zu groß (aber economy of scale?)
- technische Expertise zu divers

### Disziplinanforderungen

- disziplinspezifische Expertise zu divers
- Disziplinverankerung notwendig

### Service-Struktur

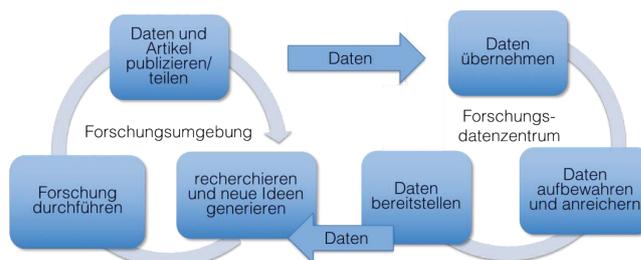
- Zielgruppe und Beratungs-/Betreuungsaufwand zu groß
- auch (kooperierende) Akteure mit sehr ähnlichem Angebot können sinnvoll sein (z.B. Ausfallsicherheit)

## Häufiger Ansatz generischer Infrastruktureinrichtungen

Bibliotheken und Rechenzentren fokussieren meist auf klassisches Publikations- und Archivparadigma und ignorieren den Forschungsprozess (links):

- Wie können wir Forschungsdaten mit unseren bestehenden Speichermöglichkeiten übernehmen?
- Wie können wir Forschungsdaten mit unseren bestehenden Recherche-Instrumenten bereitstellen?

Das wird oft den disziplinspezifischen Anforderungen und dem Forschungsprozess nicht gerecht.



## Optionen für generische Infrastruktureinrichtungen im Bereich Forschungsdaten

- eingeschränkte Basis-Infrastruktur (Dokumentation, dark archive)
- Spezialisierung zu überregionalem disziplinspezifischem Forschungsdatenzentrum (um Disziplinanforderungen für Nachnutzung zu erfüllen)
- Publikations- und Archivparadigma verlassen und in Forschungsprozess gehen. Vermittler und lokaler Basis-Support zwischen Wissenschaftler und Forschungsdatenzentrum.



Fotos: Hochspannung, Thomas Kohler, cc-by <https://www.flickr.com/photos/mecklenburg/4968331341/>, Archaeology Data Service, SUB Göttingen

## Weiterführende Literatur

- Kenneth Thibodeau. Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years. Council on Library and Information Resources, 2002. – URL: <http://www.clir.org/pubs/reports/pub107/thibodeau.html>
- Heike Neuroth, Stefan Strathmann, Achim Oßwald, Regine Scheffel, Jens Klump, Jens Ludwig (Hrsg.). Langzeitarchivierung von Forschungsdaten. Boizenburg 2012, urn:nbn:de:0008-2012031401
- Jens Klump, Jens Ludwig. Forschungsdatenmanagement. S. 257 – 276, in: Heike Neuroth, Norbert Lossau, Andrea Rapp (Hg.). Evolution der Informationsinfrastruktur – Kooperation zwischen Bibliothek und Wissenschaft. Vwh-Verlag, Glückstadt 2013, dx.doi.org/10.3249/webdoc-39006

# Forschungsdatenmanagement (FDM) in der Praxis: Embedded Data Manager– Lektion II

Fabian Cremer

Kompetenznetzwerk  
Langzeitarchivierung  
- Qualifizierung -

## Forschungsdatenmanagement (FDM) in der Praxis: Embedded Data Manager

Fabian Cremer  
Max Weber Stiftung

**nestor**  
www.langzeitarchivierung.de

**pericles**  
FP7 Digital Preservation

**ein|betonieren** *tr. V.* concrete in; **etw. in etw. (Akk.)** ~: concrete sth. into sth.  
**ein|betten** *tr. V.* **(A)** embed (**in** + *Akk.* in); **das Haus liegt eingebettet in ein Tal** the house nestles in a valley; **(B)** (*Sprachw.*) **ein-gebettete Sätze** embedded sentences  
**Einbett-:** ~**kabine** *die* single-berth cabin; ~**zimmer** *das* single room

The Oxford-Duden German Dictionary, Oxford University Press, 1999

**nestor**  
www.langzeitarchivierung.de

nestor PERICLES School 2016  
Vortrag II: FDM in der Praxis  
Fabian Cremer – Folie 2

**pericles**  
FP7 Digital Preservation

## Rahmenbedingungen

1. Neue Formen der wissenschaftlichen Kommunikation, Publikation und Kooperation basieren auf früher, breiter und vollständiger Verfügbarkeit von Forschungsdaten.
2. Die erhöhte Produktion und die gestiegene Bedeutung digitaler Forschungsdaten erfordert deren Management.
3. Die Forschungsförderer fordern die Nutzbarkeit der erzeugten (und geförderten) Forschungsdaten.
4. Das Management der Forschungsdaten erfordert zusätzliche Ressourcen und fachfremde Kompetenzen.
5. Digitale Forschungsinfrastrukturen und Technologien ermöglichen Speicherung, Austausch und Management der Forschungsdaten.

## Grundidee

- „Full Lifecycle Support“:  
Aspekte des Forschungsdatenmanagements werden in allen Arbeitsphasen der Forschungsprozesses berücksichtigt und die WissenschaftlerInnen entsprechend personell und infrastrukturell unterstützt.
- Überwindung der Paradigmen der Infrastrukturdienstleister:  
Sonderwünsche, Individualisierung, ganzheitliche Betreuung statt Basisinfrastruktur, Literaturbeschaffung, Archivierung.

# Arbeitssituationen

„Embedded Data Managers“ arbeiten ...

- als Vollzeitmitarbeiter in Großprojekten.
- mit einem Teilzeitanteil in mittelgroßen Verbundprojekten.
- als zuständiger Mitarbeiter in Forschungseinrichtung.
- als Dienstleister oder Gast in bestimmten Projekt- oder Arbeitsphasen.

## INF

(Teilprojekt Informationsinfrastruktur in  
Sonderforschungsbereichen)

... aus dem DFG Merkblatt 50.06 (06/14), geförderte Ziele:

- Aufbau einer Datenbank
- Implementierung und Entwicklung von Techniken und Verfahren
- Implementierung oder Entwicklung von Interoperabilität bzw. Schnittstellen
- Virtuelle Forschungsumgebungen, Blogs, Wikis, etc ...

Erwähnt werden: Technologien, Systeme, Infrastruktur

Nicht erwähnt: Personen, Management, Unterstützung

## Aufgaben

- Datenaustauschvereinbarung aushandeln
- Datenbanksoftware evaluieren
- Digitales Feldbuch konzipieren
- Entity-Relationship-Modelle erstellen
- Excel-Skript zur Datentransformation schreiben
- Ingest der Daten begleiten und überwachen
- Maßnahmen zum Datenschutz eines Datensatzes ermitteln
- Virtuelle Infrastruktur besorgen
- Zentrales Projektrepository betreiben
- ...

## Integrationsstrategien

- Projekttreffen besuchen
- Arbeitsplatz kennen
- Prozesse beobachten
- Alltag erleichtern
- Proaktiv
- Individualisierung
- Kooperation

## Strukturelle Voraussetzungen

- Die Unterstützung durch Autoritäten innerhalb des Projektes bestätigt die eigene Autorität und Handlungsmöglichkeit.
- Zugriff auf Ressourcen der Heimateinrichtung erleichtern die Infrastrukturdienstleistungen für das Projekt.
- Der Austausch mit einem Team der Heimateinrichtung erweitert die einsetzbaren Kompetenzen bei den Aufgaben.

## Kompetenzen

- Forschungsdatenmanagement
- (grundlegende) IT-Kenntnisse
- Anforderungen erfassen und beschreiben
- Konzepte von Metadaten und -standards
- Datenstrukturen erkennen und modellieren
- Suche und Recherche
- Beobachtung und Abstraktion von Arbeitsabläufen
- Moderation und Begleitung von Diskussionen
- Kenntnis der Arbeitsbedingungen im Forschungsbetrieb
- Dienstleistungsverständnis
- ...

# Strahlkraft

„Embedded Data Managers“ bieten Gelegenheit für

- Eigene Forschungsfragen der Informations- und Infrastrukturdienstleister
- Engen Kontakt zu und Kooperationen mit den Fachwissenschaften
- Projektbasierte Drittmittelförderung über Sachpauschalen hinaus
- Praxiserprobung für Dienstleistungskonzepte im Bereich Forschungsdaten
- Aufbau von Kompetenzen und Schulungsmöglichkeiten im Bereich Forschungsdaten

# Lesezeichen

- David Shumaker and Mary Talley: Models of embedded librarianship: Final report, Special Libraries Association 9 (2009)
- Engelhardt, Claudia: Forschungsdatenmanagement in Sonderforschungsbereichen. Teilprojekt Informationsinfrastruktur (INF-Projekte). In: LIBREAS. Library Ideas, 23 (2013) S. 106-130, urn:nbn:de:kobv:11-100212741
- Fabian Cremer, Claudia Engelhardt, Heike Neuroth: Embedded Data Manager – Integriertes Forschungsdatenmanagement: Praxis, Perspektiven und Potentiale, in: Bibliothek Forschung und Praxis, 39 (2015) S. 13–31, DOI: [10.1515/bfp-2015-0006](https://doi.org/10.1515/bfp-2015-0006), 2015



# Auswahl und Bewertung von Forschungsdaten für die Archivierung und Nachnutzung – Lektion III

Reiner Mauer

# Auswahl und Bewertung von Forschungsdaten für die Archivierung und Nachnutzung

Reiner Mauer

GESIS – Leibniz-Institut für  
Sozialwissenschaften

## Gliederung

- Einleitung: Auswahl und Bewertung von Forschungsdaten
- Auswahl – warum?
- Auswahl und Bewertung – wie?
- Kriterien in der Anwendung: Auswahlprozess implementieren
- Herausforderungen

# Bewertung

- Dient der Auswahl von digitalen Objekten für die Archivierung
- Informiert über den Zustand der digitalen Objekte
- Kann Grundlage für Kuratierungsentscheidungen, die Zuweisung von Service-/Kuratierungslevels und Aufbewahrungsfristen sein

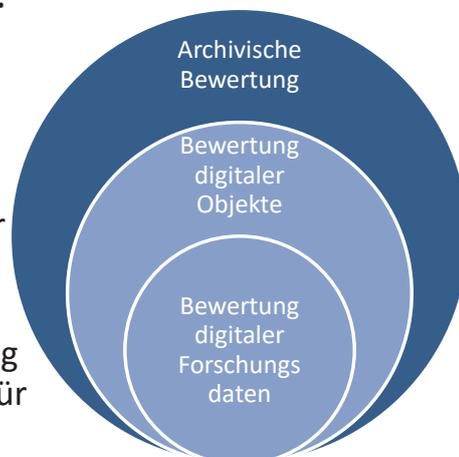
*„Was soll für wen zu welchem Zweck für wie lange archiviert und in welchem Umfang kuratiert werden?“*



Bild: CC-0

# Sonderfall Forschungsdaten?

- Unterscheiden sich Daten in Auswahl und Bewertung von anderen digitalen Objekten?
  - Eher nein!
  - Tendenziell Archivierung von Daten weniger zur „reinen“ Überlieferungsbildung, sondern als Grundlage der Nutzung in zukünftiger Forschung
  - Beispiel statistische Daten: einzelne beispielhafte Fälle für die Archivierung auswählen vs komplette Datenbank für die spätere Nutzung bewahren



## Gliederung

- Einleitung: Auswahl und Bewertung von Forschungsdaten
- **Auswahl – warum?**
- Auswahl und Bewertung – wie?
- Kriterien in der Anwendung: Auswahlprozess implementieren
- Herausforderungen

## Auswahl – warum?

- Alle Daten sind wertvoll
- Technologischer Fortschritt
  - immer leistungsfähigere Speicherlösungen (z. B. „die Cloud“)
  - Speicherplatz wird billiger



*Müssen oder können nicht alle Daten aufgehoben bzw. archiviert werden?*



Bild: CC-0

## Auswahl - warum?

- Technische Einschränkungen
  - Speicherkapazität bei extrem großen Datenmengen nicht ausreichend
  - Beispiel CERN, Large Hadron Collider: ca. 99,9 % der Daten, die in den Detektoren anfallen, wird unmittelbar ‚aussortiert‘, 0,0025 % werden vollständig analysiert und am Ende werden nur 0,0005 % dauerhaft gespeichert
- Forschungsethische und rechtliche Einschränkungen
  - z.B. keine informierte Einwilligung, sensible Daten, Datenschutz

## Auswahl - warum?

- Kosten
  - einfaches Abspeichern nicht ausreichend – Auffindbarkeit, Interpretierbarkeit und langfristige Zugänglichkeit müssen gewährleistet sein
  - Maßnahmen sind (je nach Datentyp) kostenintensiv
- Informationsüberflutung
  - Je mehr (nur gering erschlossene) Informationen verfügbar sind, desto schwieriger wird es, Relevantes aus dem „Rauschen“ herauszufiltern

# Gliederung

- Einleitung: Auswahl und Bewertung von Forschungsdaten
- Auswahl – warum?
- **Auswahl und Bewertung – wie?**
- Kriterien in der Anwendung: Auswahlprozess implementieren
- Herausforderungen

# Auswahl und Bewertung – wie?

- Auswahlprozess sollte sich auszeichnen durch
  - Konsistenz der Bewertung (sowohl über Bewerter\_innen als auch über die Zeit hinweg)
  - Nachvollziehbarkeit
  - Transparenz
- Erreichbar durch
  - definierten, dokumentierten Prozess
  - klar formulierte Auswahl-/Bewertungskriterien

## Kriterien – Beispiele

Kategorie	Beispiel
Technisch	<ul style="list-style-type: none"> <li>• Dateiformat</li> <li>• Volumen (Größe der Datei)</li> <li>• technische Restriktionen (DRM)</li> <li>• Integrität/Authentizität</li> </ul>
Inhaltlich-methodisch	<ul style="list-style-type: none"> <li>• Disziplin</li> <li>• Forschungsgegenstand</li> <li>• Methode</li> <li>• (methodische) Qualität</li> <li>• Verständlichkeit (Qualität der Dokumentation)</li> <li>• Vollständigkeit</li> <li>• Einzigartigkeit</li> <li>• wissenschaftliche Güte</li> </ul>

## Kriterien – Beispiele

Kategorie	Beispiel
Rechtlich	<ul style="list-style-type: none"> <li>• gesetzlicher Auftrag</li> <li>• Aufbewahrungspflichten und -fristen</li> <li>• Urheberrecht</li> <li>• Datenschutz</li> </ul>
Ethisch	<ul style="list-style-type: none"> <li>• (informierte) Einwilligung von Teilnehmer_innen</li> </ul>
Institutioneller Rahmen	<ul style="list-style-type: none"> <li>• Passung mit Sammelrichtlinie, institutioneller Mission</li> <li>• institutionelle Zugehörigkeit der Datengeber_in</li> <li>• Kosten der Archivierung</li> <li>• Bedürfnisse der Zielgruppen</li> </ul>

## Kriterientypen

- Formale vs inhaltliche Kriterien
  - Formal: Benutzbarkeit/Archivierbarkeit (Format, Größe, Virenfreiheit, Vollständigkeit); Rechtliche Situation (Urheberrecht, Datenschutz); Affiliation Primärforscher\_in; sind Daten schon in einem anderen Archiv/ Repository
  - Inhaltlich: Untersuchungsthema, Methodik, wissenschaftlicher Wert, Qualität
- K.O./Ausschluss-Kriterien vs qualifizierende Kriterien
  - Ausschluss: rechtl. Restriktionen (Nutzungsrechte, Datenschutz); außerhalb des Sammelgebiets
  - Qualifizierende Kriterien: „Ausgangszustand“ der Daten; wissenschaftlicher Wert (thematisch, methodisch), Einzigartigkeit (Messung kann nicht wiederholt werden)

## Kriterienbildung

- Operationalisierung
  - relativ einfach bei formalen Kriterien (Dateiformat, Volumen, Erhebungs-/Meßmethode, Urheberrechte...)
  - abstraktere Konstrukte (Qualität, Nachnutzungspotenzial, wissenschaftlicher Wert) müssen „messbar“ und Bewertung für Dritte nachvollziehbar gemacht werden

## Kriterienbildung

- Wie bestimme ich die für meine Organisation relevanten Kriterien?
- ⇒ Rahmenbedingungen und Kontext betrachten und Kriterien hieraus ableiten

## Relevante Rahmenbedingungen

- Aufgabe/Zweck der Institution
  - z.B. Sammelrichtlinie, Leitbild, Satzung
- Zur Verfügung stehende Ressourcen
  - finanzielle, technische und personelle
- Zielgruppen
  - Bedürfnisse, Erwartungen
- Gesetzliche Vorgaben
  - z.B. Aufbewahrungsfristen für bestimmte Datentypen,
- Markt/Mitbewerber

## Ableitung von Kriterien: Beispiel GESIS Datenarchiv

- Für wen sollen wir sammeln?

Auszug GESIS Leitbild

- Unsere Angebote richten sich primär an Forscher der empirischen Sozialforschung mit einem Schwerpunkt auf den Fachgebieten Soziologie und Politikwissenschaft sowie an solche der Sozialwissenschaften insgesamt.

Auszug GESIS Satzung (§2 Zweck)

- e) Schaffung eines benutzerfreundlichen und hochqualitativen Zugangs zu allen für die empirische Sozialforschung relevanten Informationen und Daten

Abzuleitendes Kriterium:

- Daten müssen für die Zielgruppe gut zu benutzen sein

## Ableitung von Kriterien: Beispiel GESIS Datenarchiv

- Was sollen wir sammeln?

Auszug GESIS Satzung (§2 Zweck)

- b) Beschreibung und Erklärung gesellschaftlicher Entwicklungen in nationaler, international vergleichender und historischer Perspektive einschließlich der Datenerhebung, statistischen Modellierung und Dauerbeobachtung,

- c) Archivierung, Dokumentation und Langzeitsicherung sozialwissenschaftlicher Daten, einschließlich ihrer Erschließung sowie qualitativ hochwertigen Aufbereitung besonders relevanter Daten für Sekundäranalysen.

Abzuleitende Kriterien:

- Daten sind geeignet, Aussagen über genannte Themen zu treffen
- Daten, die einen Vergleich in zeitlicher, räumlicher bzw. kultureller Perspektive erlauben, sind von besonderer Bedeutung
- Relevanz der Daten (wissenschaftlicher Wert) für die Zielgruppe
- Eignung für Sekundäranalysen

## Ableitung von Kriterien: Beispiel GESIS Datenarchiv

Kriterien (Auszug)	Inhaltliche Passung	Eignung für Sekundäranalysen	wiss. Relevanz
Technisch	Dateiformat ist mit gängiger Software zu verarbeiten		
	Dateien überschreiten max. Volumen nicht		
	Dateien unterliegen keinen technischen Zugriffsbeschränkungen		
	Dateien sind virenfrei und nicht beschädigt		
Inhaltlich-methodisch	Die Untersuchung ist geeignet, sozialwissenschaftlich relevante Fragestellungen zu beantworten.		
	Die Daten sprechen voraussichtlich einen breiten sozialwissenschaftlichen Nutzerkreis an.		
	Die Untersuchung ist hinsichtlich des Fragenprogramms bzw. der erhobenen Daten innovativ oder einzigartig.		
	Die Daten sind umfassend dokumentiert		
	a. Es ist ein detaillierter Projekt- bzw. Methodenbericht vorhanden.		
	b. Der Fragebogen sowie alle weiteren Untersuchungsmaterialien sind dokumentiert.		
Ethisch-rechtlich	Die Untersuchung ermöglicht Vergleiche über die Zeit oder über Länder hinweg.		
	Es liegt eine Einverständniserklärung der Teilnehmer_innen vor		
	Die Urheberrechte der Daten sind geklärt (insbesondere Rechteinhaber/Rechtsnachfolger).		

## Gliederung

- Einleitung: Auswahl und Bewertung von Forschungsdaten
- Auswahl – warum?
- Auswahl und Bewertung – wie?
- Kriterien in der Anwendung: Auswahlprozess implementieren
- Herausforderungen

## Kriterien in der Anwendung: Auswahlprozess implementieren

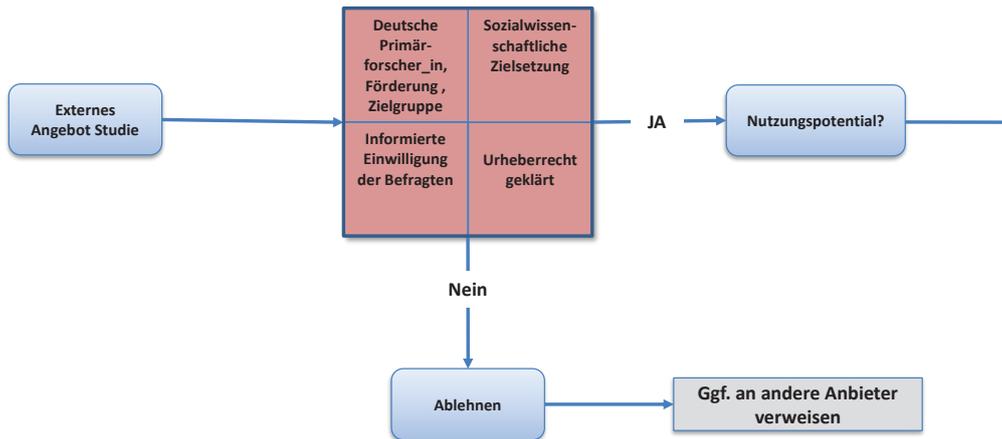
- Rollen und Verantwortlichkeiten müssen geklärt werden (Archiv, Datenproduzent\_in, Zielgruppe/Community, Förderer, Politik)
- Konsistenz und Nachvollziehbarkeit gewährleisten
  - Verständliche Beschreibung der Kriterien,
  - Schulung von Mitarbeiter\_innen,
  - Checkliste oder Leitfragenkatalog,
  - Vier-Augen-Prinzip oder Auswahlkonferenzen
- Transparenz gewährleisten
  - Dokumentation des Prozesses / der Entscheidung
- Effizienz gewährleisten
  - Bewertung muss auf möglichst hohem Abstraktionsniveau getroffen werden (insbesondere wenn große Volumina bewertet müssen; im Extremfall kostet Selektion und Bewertung ansonsten mehr als die Archivierung)

## Kriterien in der Anwendung: Auswahlprozess implementieren

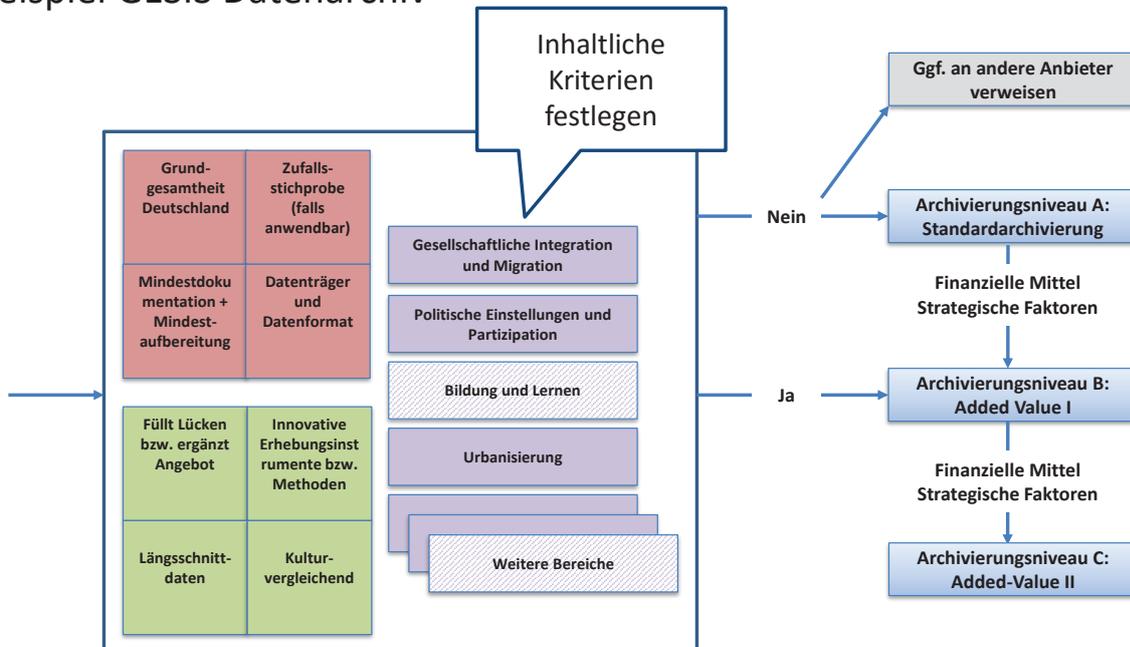
- Typischerweise: mehrstufiger Auswahlprozess
  - Passen angebotene Daten grundsätzlich ins Profil? Sind alle „Ausschlusskriterien“ positiv erfüllt?
  - Wenn dies der Fall ist, beginnt die Bewertung der Daten mit Blick auf Qualität etc.
  - Es folgt eine Entscheidung über Übernahme oder Kassation beziehungsweise bei Übernahme ggf. Entscheidung für ein Kuratierungslevel

# Beispiel GESIS Datenarchiv

Aufzunehmende Studien müssen bestimmte Mindestanforderungen erfüllen



# Beispiel GESIS Datenarchiv



## Kriterien in der Anwendung: Auswahlprozess implementieren

- Wie lehnt man eingereichte Forschungsdaten ab?

Nicht in jedem Fall endgültig...

... aber immer mit Begründung!

- ⇒ Je nach Ablehnungsgrund kann Einreicher\_in Nachbesserungen vornehmen (z.B. fehlende Dokumentation vervollständigen)
- ⇒ Bewertungs-/Entscheidungsprozess so dokumentieren, dass Datenproduzent\_innen und andere verstehen können, warum Daten aufbewahrt bzw. zurückgewiesen wurden.

## Gliederung

- Einleitung: Auswahl und Bewertung von Forschungsdaten
- Auswahl – warum?
- Auswahl und Bewertung – wie?
- Kriterien in der Anwendung: Auswahlprozess implementieren
- Herausforderungen

## Herausforderungen (1)

- Beurteilung des Werts von Daten für zukünftige Forschung:
  - Welche Themen, welche Arten von Daten sind zukünftig relevant?
  - Können Daten, die aus Forschungsprojekten stammen, die bereits mehrfach bewertet wurden (Förderantrag, auf Daten basierende Publikationen, Evaluationen ...), "wertlos" sein?
- ⇒ Bewertung kann nicht nur von Archiven, Repositorien, Infrastrukturen vorgenommen werden. Alle Stakeholder (Fachgesellschaften, Forschungsförderer, Forschungspolitik und Datenprofis müssen in Auswahlprozess einbezogen werden
- ⇒ Es lässt sich nicht vermeiden, dass 'Fehlentscheidungen' im Auswahlprozess getroffen werden

## Herausforderungen (2)

- Um Nachvollziehbarkeit und Transparenz von Wissenschaft zu unterstützen, sollen auch Forschungsdaten offengelegt werden
  - Zunehmend auch verpflichtende Auflagen von Forschungsförderern oder Zeitschriften
- ⇒ D.h. eigentlich müssen \*alle\* Daten aufbewahrt werden (zumindest für einen bestimmten Zeitraum)

## Einführende Literatur

- Keitel, Ch. (2015). Auswahl und Bewertung digitaler Forschungsdaten aus archivwissenschaftlicher Perspektive. DINI/nestor-Workshop Forschungsdaten. Essen, 17.11.2015.  
<http://www.forschungsdaten.org/index.php/Datei:04--keitel--Bewertung-Forschungsdaten.pdf>
- Tjalsma, H., & Rombouts, J. (2010). Selection of Research Data. Guidelines for Appraising and Selecting Research Data.  
<https://www.surf.nl/en/knowledge-base/2010/research-report-selection-of-research-data.html>
- Whyte, A. & Wilson, A. (2010). How to Appraise and Select Research Data for Curation. DCC How-to Guides. Edinburgh: Digital Curation Centre.  
<http://www.dcc.ac.uk/resources/how-guides/appraise-select-data>
- Präsentationen des 4. DINI/nestor Workshops „Forschungsdaten auswählen und bewerten“. Essen, 17.11.2015.  
<http://www.forschungsdaten.org/index.php/DINI-nestor-WS4>

PERICLES Sheer Curation Tools  
Das Erhalten digitaler Objekte beginnt in ihrer  
Entstehungsumgebung – Lektion IV

Anna Eggers

# PERICLES Sheer Curation Tools

## Das Erhalten digitaler Objekte beginnt in ihrer Entstehungsumgebung

Anna Eggers  
Niedersächsische Staats- und  
Universitätsbibliothek Göttingen

## Übersicht

- 'Lebende' Digitale Objekte
- Die Entstehungsumgebung digitaler Objekte
  - Bottom-up: Signifikante Umgebungsinformationen
  - Top-down: Das digitale Ökosystem
- Sheer Curation
- Die PERICLES Tools
  - PET (Pericles Extraction Tool)
  - PeriCAT (Pericles Content Aggregation Tool)
- Praktischer Teil

# Verwendung digitaler Objekte

Ausgangssituation:

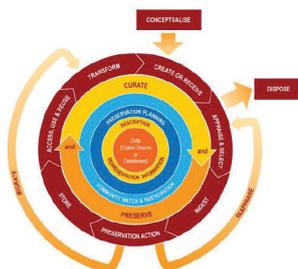
- Erhaltenswerte digitale Objekte entstehen
- Oder: Objekte werden verändert

Zweck und Ziel:

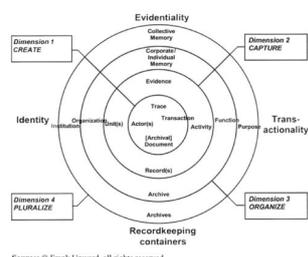
- Sofortiger/primärer Anwendungszweck
- Sekundäre Anwendungen / Wiederverwendung
  - Noch nicht alle zukünftigen Anwendungen zur Entstehungszeit bekannt
  - > Erhaltung

# Das Leben eines digitalen Objektes

Angerissen: Theoretische Modelle...



The DCC Curation Lifecycle Model  
<http://www.iijdc.net/index.php/iijdc/article/viewFile/69/48>



The Records Continuum Model  
<http://www.tandfonline.com/doi/pdf/10.1080/00379810120037522>

→ **Essenz:**

**Nicht nur der Archivierungszeitpunkt ist wichtig!**

# Signifikante Eigenschaften eines digitalen Objektes

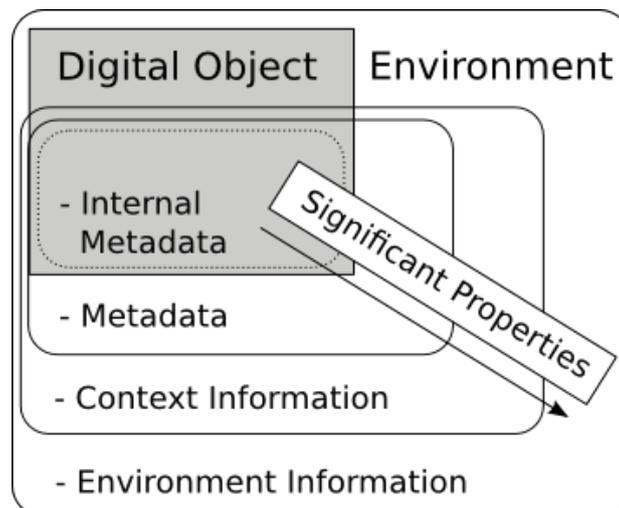
*Diejenigen Eigenschaften des zu übernehmenden Informationsobjekts, die durch den gesamten Archivierungsprozess hindurch erhalten werden sollen.*

\* nestor - Leitfaden zur digitalen Bestandserhaltung

→ Anwendungsabhängig

Was könnte für zukünftige Anwendungen erhaltenswert sein?  
Welche Einflüsse wirken auf diese Eigenschaften?

# Von Metadaten zu signifikanten Umgebungsinformationen

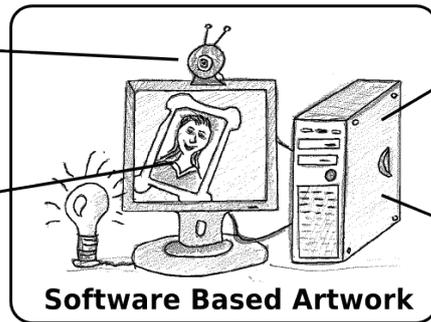


# Einfluss auf die signifikanten Eigenschaften

## SEI influence on SP of DO

**Interaction experience, usability**  
Peripheral usage and driver

**Appearance**  
Used graphic driver and libraries



**Execution quality**  
Installed programming language related software

**Behaviour patterns, performance**  
Available system resources

# Die Entstehungsumgebung



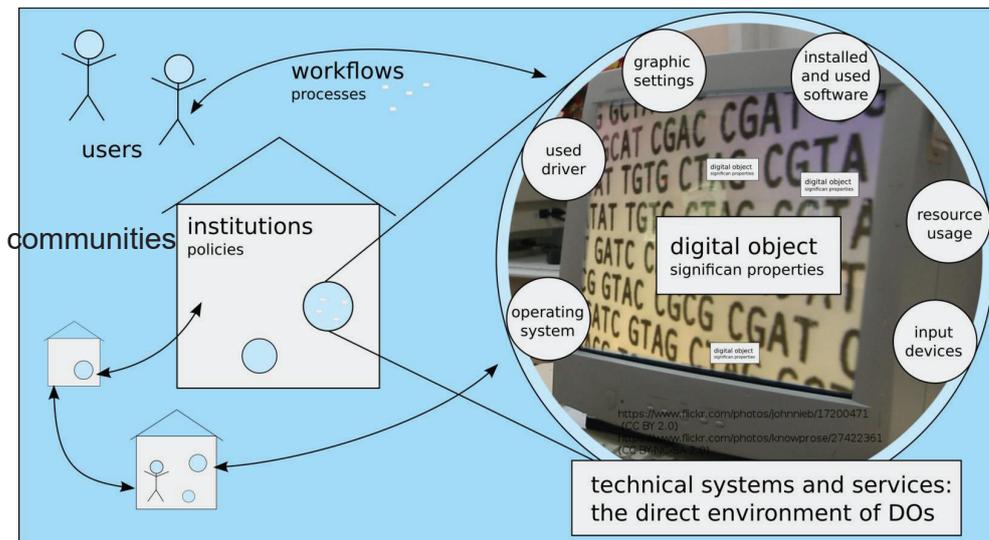
## Bottom-up Perspektive: Signifikante Umgebungsinformationen

- Die direkte Umgebung ist alles was ein digitales Objekt "sehen" kann
  - (entspricht meist einem Computersystem)
- Signifikanz bezieht sich auf einen Zweck / eine Anwendung
- Informationen über Abhängigkeiten bezogen auf diese Anwendung

## Top-down Blickwinkel: Das digitale Ökosystem

- Ein Konzept das digitale Objekte in ihren Entstehungsumgebungen betrachtet.
- Dabei werden alle Entitäten berücksichtigt, die auf die digitalen Objekte einwirken, bzw. die für eine bestimmte Analyse wichtig sind
  - Zustände analysieren
  - Änderungen simulieren

# Umgebungsinformationen und Einflüsse auf ein digitales Objekt



## Sammeln signifikanter Umgebungsinformationen

- Festlegen was signifikant ist
  - manuell
  - automatisch (Analyse ähnlicher Informationen)
  - halb automatisch (Templates)
- Flaschenhals Mensch
  - generell Interesse an Wiederverwendung
  - hat aber kaum Zeit
- → Automatisierung
- Gesammelte Informationen bewerten
- Sensible Daten aussortieren/verschlüsseln

## Welche Informationen?

Stark abhängig vom Szenario.

Generell zwei Arten von Informationen:

### 1. Objektabhängige Informationen

- a. Ort des Objektes (im Dateisystem / DB)
- b. Verwendete Software
- c. Workflow

### 2. Objektunabhängige Informationen

- a. Hardware Ressourcen
- b. Installierte Software, Treiber

## Sheer Curation

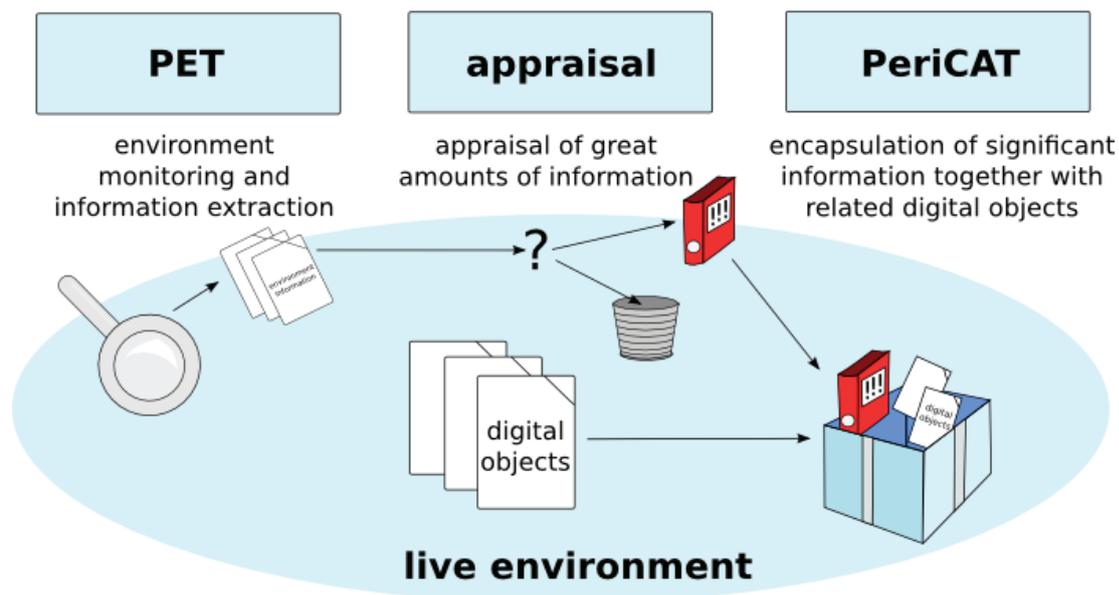
Aktivitäten zur Langzeitarchivierung die schon während der Entstehung digitaler Objekte ansetzen

- Leichtgewichtig
- Integriert in den Workflow
- Volle Kontrolle des Benutzers
  - Private und sensible Daten schützen
- Nicht den gewohnten Workflow beeinträchtigen

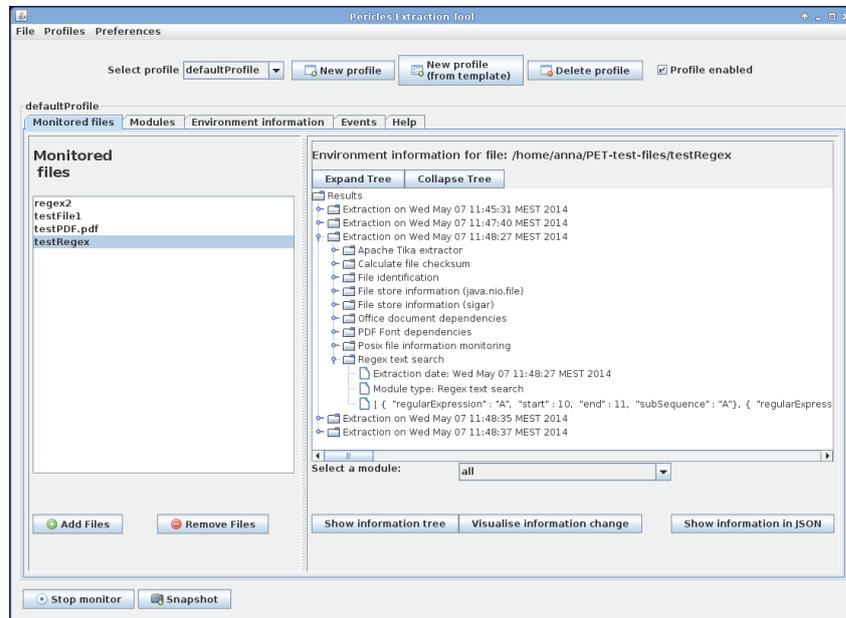
# Sheer Curation Tools Entwicklerperspektive

- KISS - "Keep it simple, stupid"
  - Nachvollziehbarkeit (Open Source, Usability)
  - einfach konfigurierbar
- Geringer Ressourcenverbrauch
- Remote Features vermeiden
- Läuft kontinuierlich im Hintergrund
  - Interfaces, um das Tool besser in den Workflow zu integrieren
- Verschlüsselung

## PERICLES sheer curation



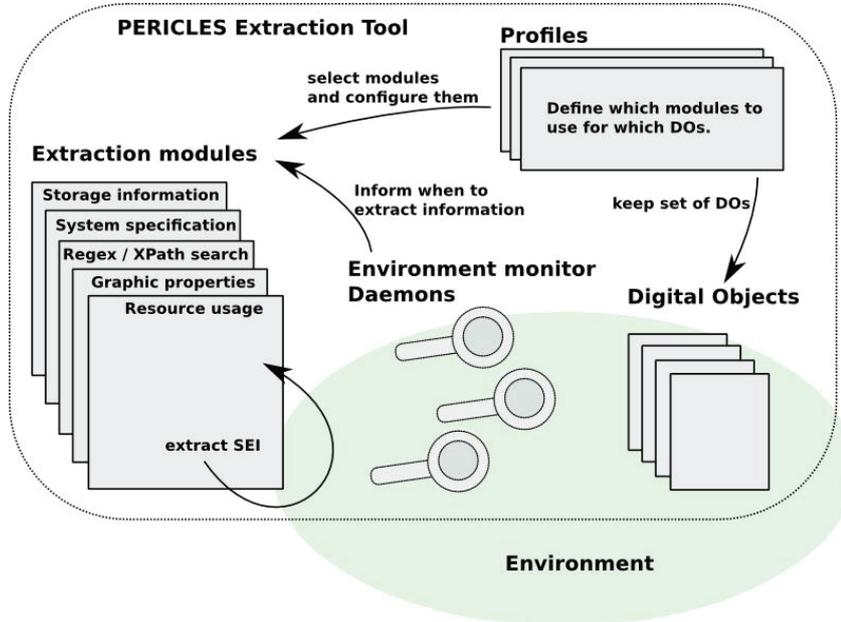
# Das PERICLES Extraction Tool



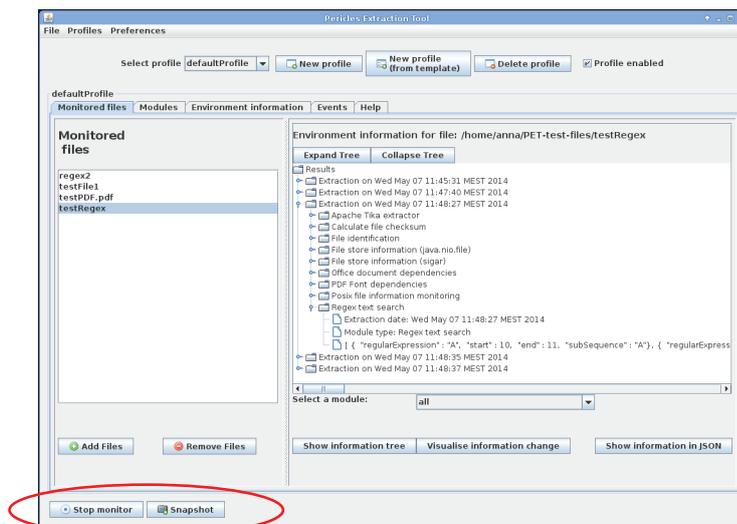
# Das PERICLES Extraction Tool

1. Sammelt Umgebungsinformationen
2. Beobachtet die Umgebung und reagiert auf Events
3. Modular aufgebaut → Der Benutzer definiert was gesammelt werden soll
4. Läuft kontinuierlich im Hintergrund
5. Informationen über wichtige Dateien
6. Kontrolle liegt beim Benutzer

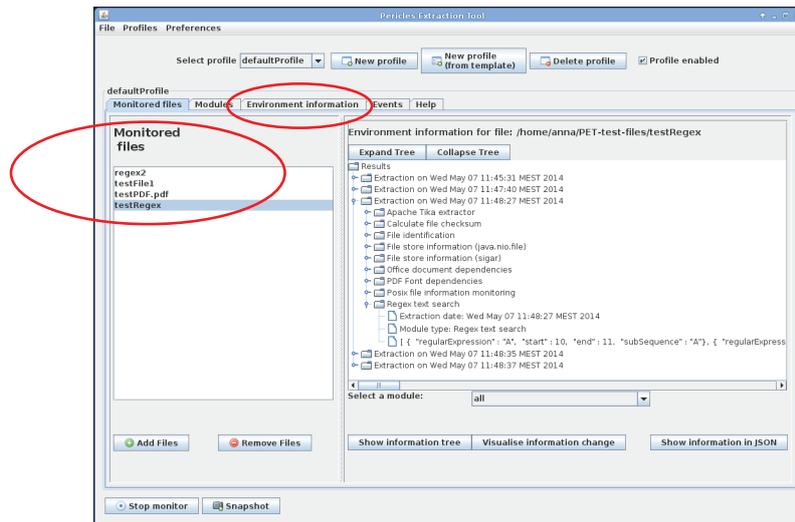
# PET Architektur



# Schnappschuss oder kontinuierliche Extraktion



# Informationen bezogen auf eine Datei, oder allgemeingültige Umgebungsinformationen

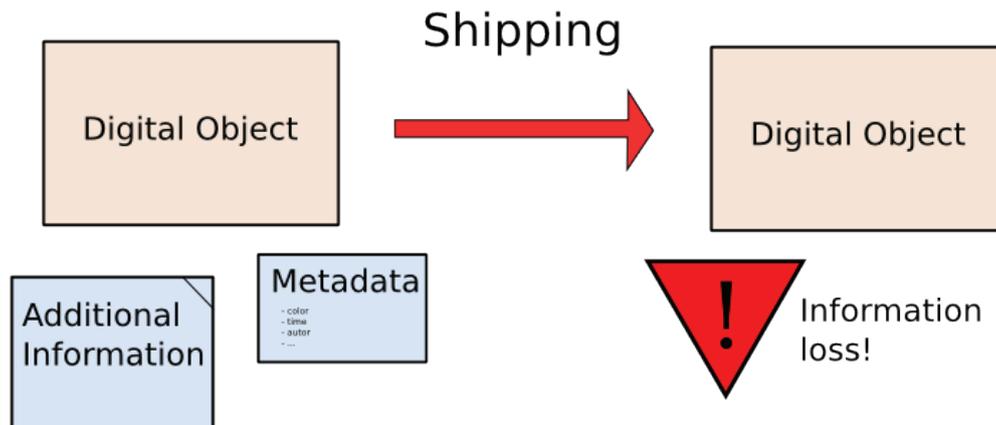


## DEMO

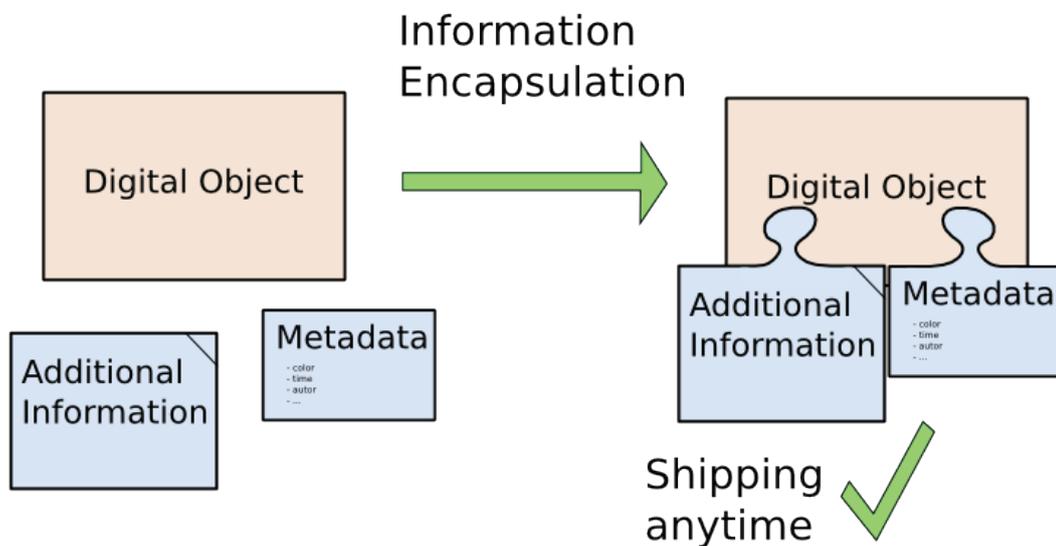
The screenshot shows the GitHub interface for the repository 'pericles-project / pet'. At the top, there is a search bar and navigation links for 'Explore', 'Features', 'Enterprise', and 'Pricing'. The repository name is displayed with 'pericles-project' in blue and 'pet' in black. Below the name, there are statistics: 'Watch 14', 'Star 13', and 'Fork 0'. The main content area is titled 'Examples' and includes a sub-header 'PET examples'. The text describes the experiments and lists three steps: 1. The PET tool is installed, configured and started on the machine where the digital objects are used. 2. The user interacts with the system while PET collects environment information in the background. 3. The environment information, digital object events and changes are stored for future use and analysis. On the right side, there is a 'Pages' sidebar with a list of links: Home, Building from source, Development of Extraction Modules and Monitoring Daemons, Eclipse installation, Examples, PET Development and Software Architecture, Quick start guide, Scenarios, and TODO list, further developments. At the bottom of the repository page, there is a 'Clone this wiki locally' button with the URL 'https://github.com/'.

Die Informationen sind gesammelt.  
Und nun?

## Problem: Informationen oder Links zu ihnen gehen verloren



## Lösung: Informationsverkapselung



Which information encapsulation technique to use?

PACKAGING

EMBEDDING

SECURITY  
AUTHENTICITY  
COMPRESSION  
CAPACITY  
RESTORABILITY  
ACCESSIBILITY  
SPREADING

INVISIBILITY  
RISK OF DATA LOSS  
PROCESSABILITY  
ENCRYPTION  
COMPLEXITY  
INTEGRITY  
STANDARDISATION

**PeriCAT**

nestor  [www.langzeitarchivierung.de](http://www.langzeitarchivierung.de)

nestor PERICLES School 2016  
Vortrag IV: PERICLES Sheer Curation Tools  
Anna Eggers – Folie 27

 **pericles**  
FP7 Digital Preservation

## Informationsverkapselung

Informationen verpacken oder einbetten?

**Verpacken:** Alle Informationen in ein Paket

**Einbetten:** Nutzlast-Informationen werden in eine Trägerinformation eingebettet

## Einfache Archivierungsformate



<https://www.flickr.com/photos/stoolhog/8153406149> (CC BY-SA 2.0)

## Pakete mit Metadateiendateien



<https://www.flickr.com/photos/quinnanya/7206706162> (CC BY-SA 2.0)

## Strukturiertes Verpacken



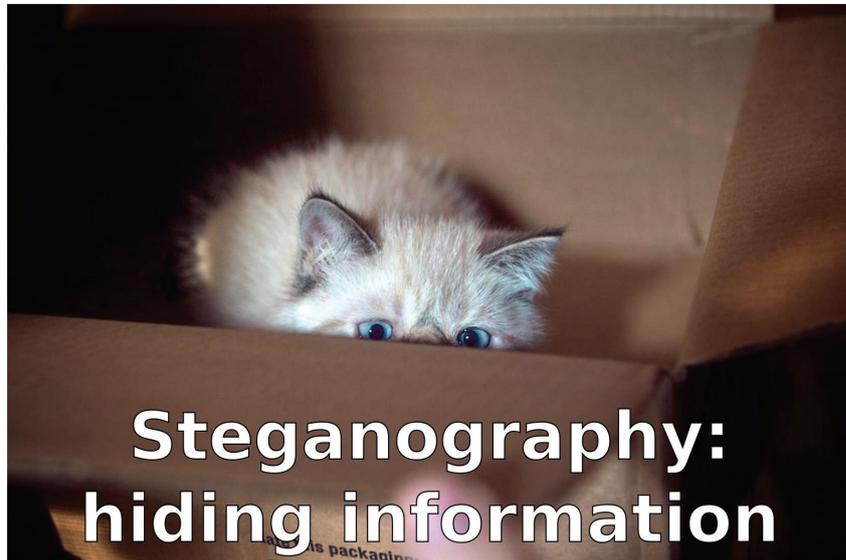
<https://www.flickr.com/photos/oskay/2716368532> (CC BY 2.0)

## Digitale Wasserzeichen



<https://www.flickr.com/photos/oskay/2716368532> (CC BY 2.0)

# Steganographie



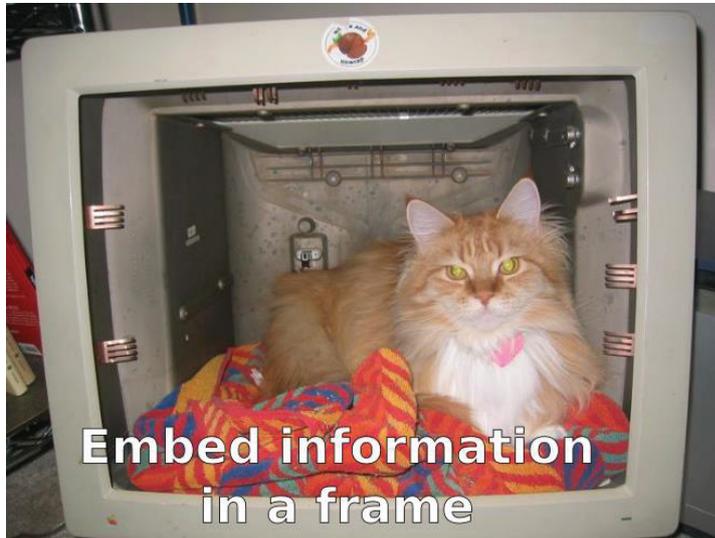
<https://www.flickr.com/photos/70006548@N04/6359363281> (CC BY 2.0)

# Eigenschaften von Dateien



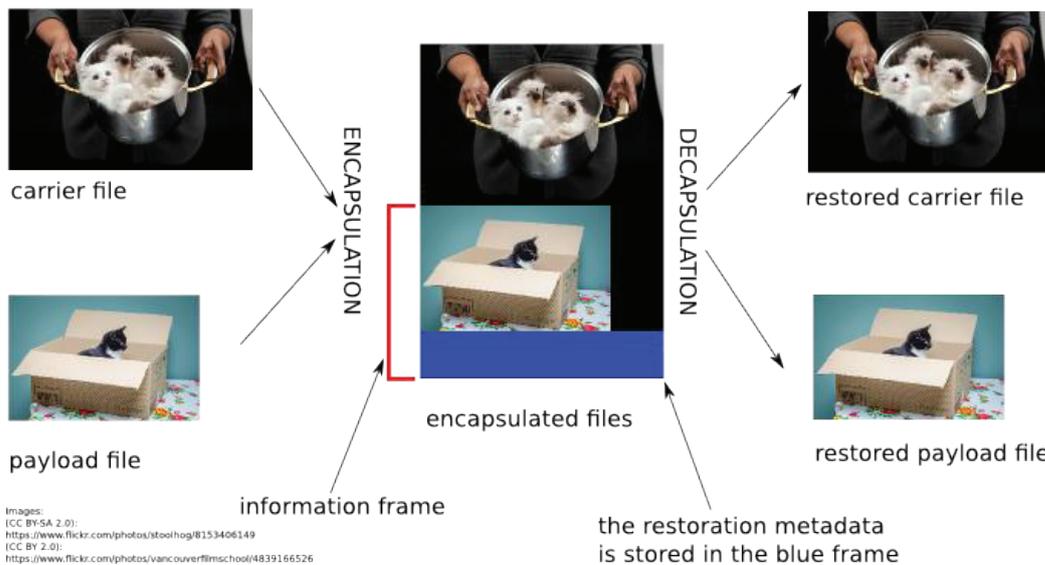
<https://www.flickr.com/photos/vancouverfilmschool/4839166526> (CC BY 2.0)

# Informationsrahmen



<https://www.flickr.com/photos/oskay/156242299> (CC BY 2.0)

## Image information frame encapsulation



Pericat - PERICLES Content Aggregation Tool

Information Encapsulation | Information Decapsulation | Scenario

### Information Encapsulation

#### Datasets

**Carrier file:**  
/home/anna/carrierPicture.png  
Add / change carrier file

**Payload files:**  
/home/anna/payloadText  
Add payload file | Remove payload file | Encapsulate

**Datasets:**  
new Dataset  
Create dataset  
Delete dataset  
Change name

#### Encapsulation Algorithms

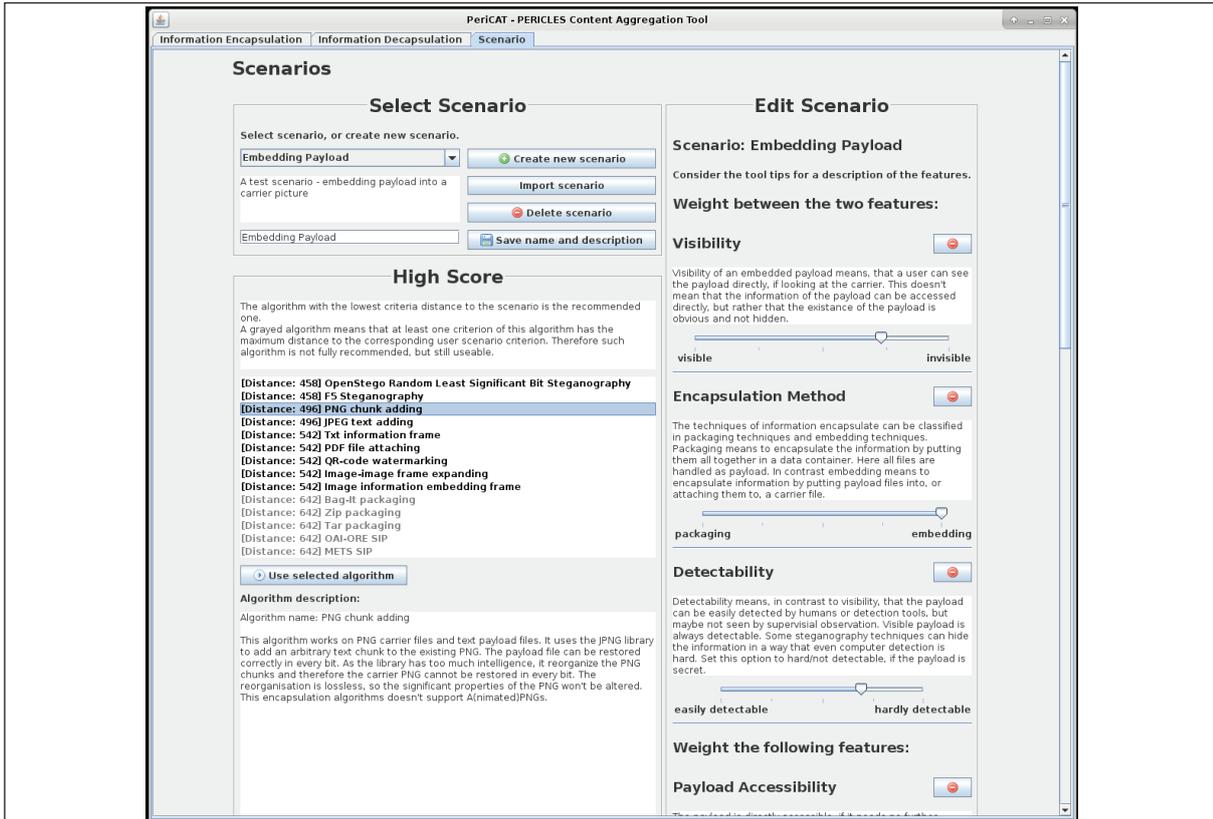
Note: Red algorithms can't be applied on the given data.

- [Distance: 550] Txt information frame
- [Distance: 550] OpenStego Random Least Significant Bit Steganography
- [Distance: 550] F5 Steganography
- [Distance: 550] Bag-It packaging
- [Distance: 550] PNG chunk adding
- [Distance: 550] JPEG text adding
- [Distance: 550] PDF file attaching
- [Distance: 550] Zip packaging
- [Distance: 550] Tar packaging
- [Distance: 550] OAI-ORE SIP
- [Distance: 550] METS SIP
- [Distance: 550] QR-code watermarking
- [Distance: 550] Image-image frame expanding
- [Distance: 550] Image information embedding frame

nestor [www.langzeitarchivierung.de](http://www.langzeitarchivierung.de) | nestor PERICLES School 2016 | Vortrag IV: PERICLES Sheer Curation Tools | Anna Eggers – Folie 37 | pericles FP7 Digital Preservation

## Den passenden Algorithmus für ein Szenario finden

- Algorithmen aus verschiedensten Bereichen  
– Unterscheidbar anhand ihrer Eigenschaften
  - Das Szenario hat Anforderungen an den Algorithmus
- Abbildung der Algorithmen-eigenschaften auf die Szenarioanforderungen und eine Distanzberechnung.



DEMO

## Abschließend

- Welche Informationen sind für die Zukunft wichtig?
- Wie sieht "mein" Szenario eigentlich aus und welche Einflüsse wirken auf die Objekte?
- Welche anderen Anwendungen sind denkbar?

Tools können nur eine Hilfe sein um diese Fragen zu beantworten → Planung

## Fragen?

Pericles Projektseite:

<http://pericles-project.eu>

Beide Tools sind auf GitHub:

<https://github.com/pericles-project>



## nestor School 2016: Übung 1

Der Leiter einer Universität hat VertreterInnen verschiedener Gruppen damit beauftragt, bei einer gemeinsamen Sitzung zu überlegen,

- ob sich die Universität im Bereich Forschungsdaten engagieren soll,
- wieso oder wieso nicht Aktivitäten in diesem Bereich verfolgt werden sollen und
- für welche Forschungsdaten.

Welche Aktivitäten im Einzelnen verfolgt werden sollen, würde in einem separaten Projekt entwickelt werden, wenn in der Sitzung eine Einigung erzielt wird (und ist deshalb nicht Teil der Übung).

Überlegen Sie sich in Ihren jeweiligen Gruppen die Argumente, die für Sie relevant sind. Danach werden wir in einem kurzen Rollenspiel die Sitzung durchspielen, d.h. wir brauchen aus jeder Gruppe eine/n Freiwillige/n, der/die dann die Situation spielt. Nachdem wir in dem Rollenspiel ein paar der Argumente ausprobiert haben, machen wir noch eine Runde, in der jede Gruppe ihre Sichtweise insgesamt vorstellen kann.

1. Überlegen Sie zuerst: Was sind Ihre eigenen Interessen?
2. Gehen Sie verschiedene Arten von Forschungsdaten durch: Was wollen Sie mit dieser Art von Forschungsdaten erreichen oder nicht?
3. Was sind ihre Argumente dafür?
4. Überlegen Sie sich, was wohl andere Gruppen wollen. Wenn das Ihrem Interesse widerspricht, was sind ihre Argumente dagegen?

Die verschiedenen Gruppen:

- WissenschaftlerInnen
- Bibliothek
- Förderer
- Verwaltung/Forschungsabteilung

## Das Dateninterview

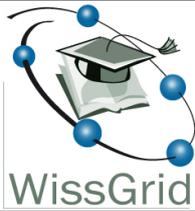
---

Szenario: Eine/ein Embedded Data Manager arbeitet als Projektmitarbeiter vollständig in einem interdisziplinären Forschungsprojekt (30 ForscherInnen) und ist für das Datenmanagement im Projekt verantwortlich. Jede/r ForscherIn hat ein eigenständiges Forschungsthema. Um mehr Informationen über die anfallenden Daten zu erfahren und etwaige Maßnahmen zu identifizieren, hat der/die Embedded Data Manager einen Termin mit einer/m ForscherIn vereinbart. Die Datenerhebung ist noch nicht abgeschlossen, hat jedoch bereits begonnen. Eine nähere Beschäftigung mit Datenmanagement hat bisher noch nicht stattgefunden.

Lässt sich die Wissensgrid-Liste als Basis für ein solches Gespräch nutzen und lässt über ein Interview mit ForscherInnen das passende Forschungsdatenmanagement entwickeln?

Ein Vorschlag zur Vorgehensweise:

1. Streichen Sie alle Fragen, die der Embedded Data Manager allein durch den Projektkontext, in dem er sich befindet, bereits weiß oder klären könnte.
2. Streichen Sie alle Fragen, die im Verantwortungsbereich des Embedded Data Manager liegen könnten und damit nicht durch den Wissenschaftler beantwortet werden müssten.
3. Die Interviewzeit ist begrenzt: Wählen Sie ca. 20 Fragen aus, die Sie den Wissenschaftler fragen möchten und begründen Sie die Auswahl und die Reihenfolge.
4. Wenn Zeit bleibt: Stellen Sie eine Interview-Situation nach.



---

# WissGrid

## Deliverable 2.3

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

---

**Arbeitspaket 2: Blaupausen und Beratung,  
Arbeitspaket 3: Langzeitarchivierung**

# Checkliste zum Forschungsdaten-Management<sup>1</sup>

**Version 0.6:  
Entwurfsversion zur öffentlichen Kommentierung**

---

<sup>1</sup> This work is created by the WissGrid project. The project is funded by the German Federal Ministry of Education and Research (BMBF).

**Kurzfassung:**

1. Planung und Erstellung
  - 1.1 Sind alle Rahmendaten des Projekts dokumentiert (Name, Ziele, Finanzgeber, Laufzeit, Partner, Leiter)?
  - 1.2 Wie können bereits existierende Daten integriert/nachgenutzt werden?
  - 1.3 Welche Bedeutung haben die Daten für das Vorhabensziel?
  - 1.4 Wie lassen sich die verwendeten/erzeugten Daten charakterisieren (Datenarten, Formate, Reproduzierbarkeit)?
  - 1.5 Wie werden die Daten erfasst/erstellt?
  - 1.6 Wie groß ist die Datenmenge/Produktionsrate?
  - 1.7 Wie erfolgt die Qualitätssicherung?
2. Auswahl und Aufbewahrungsdauer
  - 2.1 Wieso müssen welche Daten aufbewahrt werden?
  - 2.2 Wann, durch wen und womit erfolgt die Datenauswahl?
  - 2.3 Wie lange müssen die Daten aufbewahrt werden?
  - 2.4 Was geschieht bei Ablauf der Aufbewahrungsdauer?
3. Ingest: Einspeisen und Verantwortungsübernahme
  - 3.1 Wann und wie werden die Daten übergeben/übertragen?
  - 3.2 Wann und von wem werden welche Metadaten erfasst?
  - 3.3 Wie werden Daten und Metadaten validiert?
  - 3.4 Wie wird mit sensiblen Daten umgegangen?
  - 3.5 Sind Rechte und Pflichten von Datenproduzent und -archiv geklärt (Protokollierung, Fehlerbehandlung)?
4. Speicherung und Infrastruktur
  - 4.1 Wer ist während des Projekts und danach für die Speicherung der Daten verantwortlich?
  - 4.2 Mit welchen Technologien und an welchen Orten werden die Daten gespeichert?
  - 4.3 Werden regelmäßig Sicherheitskopien erstellt und überprüft?
  - 4.4 Wie hoch ist die erwartete Datenmenge?
  - 4.5 Gibt es besondere Infrastruktur-Anforderungen für Datentransfer, -zugriff und -nutzung? (Netzwerk-Bandbreite, Hardware etc.)
5. Erhaltungsmaßnahmen
  - 5.1 Sind die eingesetzten Technologien sowie Abhängigkeiten, Nutzungszielgruppe und -anforderungen dokumentiert?

- 
- 5.2 Wird regelmäßig überprüft, ob sich die Anforderungen, verfügbaren Technologien oder Abhängigkeiten verändert haben?
  - 5.3 Wird die Aufbewahrungswürdigkeit regelmäßig überprüft?
  - 5.4 Gibt es eine Nachfolgeregelung bei einem Wechsel der aufbewahrenden Institution?
6. Zugriff und Nutzung
- 6.1 Können die Daten auch von Anderen innerhalb oder außerhalb des Projekts genutzt werden?
  - 6.2 Gibt es Verpflichtungen, Daten freizugeben oder nicht freizugeben?
  - 6.3 Welche Einrichtungen/Gruppen werden an den Daten interessiert sein?
  - 6.4 Wie werden Veröffentlichung, Suchbarkeit und Zugriff realisiert?
  - 6.5 Zu welchem Zweck und mit welcher Software werden die Daten voraussichtlich genutzt?
  - 6.6 Wie werden ggf. Erstnutzungsrecht, Nutzungsbeschränkungen oder Lizenzbedingungen durchgesetzt?
  - 6.7 Spielt Interoperabilität eine Rolle?
7. Management, Organisation und Policies
- 7.1 In welcher Art von Repository werden die Daten aufbewahrt?
  - 7.2 Ist der Bezugsrahmen national oder international?
  - 7.3 Welche Institutionen sind für das Datenmanagement zuständig oder daran beteiligt?
  - 7.4 Haben alle Beteiligten eingewilligt und sind deren Beiträge definiert?
  - 7.5 Ist der Workflow des Datenmanagements beschrieben?
  - 7.6 Sind alle Ressourcen beschrieben und abgeschätzt?
  - 7.7 Welche Anforderungen/Auflagen/Policies müssen berücksichtigt/umgesetzt werden? Von wem?
  - 7.8 Wie wird die Einhaltung der Planung überprüft?
8. Kosten
- 8.1 Wie hoch sind die Kosten und der Personalaufwand für das Datenmanagement während des Projektes/nach dem Projekt?
  - 8.2 Wer übernimmt die Kosten?
  - 8.3 Stehen alle Verantwortlichen und Beteiligten hinter den Plänen zum Datenmanagement?
9. Rechtliche Aspekte von Forschungsdaten
- 9.1 Sind die Daten aufgrund des Datenschutzes oder aus anderen Gründen sensibel?
  - 9.2 Werden fremde Daten oder Software verwendet, welche dem Urheberrecht, dem Patentrecht o.ä. unterliegen?

- 
- 9.3 Unterliegen eigene Daten oder Software dem Urheberrecht/Patentrecht und sind die Nutzungsbedingungen und Rechte geklärt?
  - 9.4 Werden alle notwendigen Maßnahmen zum Datenmanagement eingeräumt?
  - 9.5 Sind evtl. auslaufende Schutzfristen zu berücksichtigen?
10. Metadaten
- 10.1 Wozu sollen die Metadaten dienen oder benutzt werden?
  - 10.2 Was soll durch Metadaten beschrieben werden?
  - 10.3 Welche Semantik und Syntax wird verwendet?
  - 10.4 Inwieweit können die Metadaten automatisch erstellt werden?
  - 10.5 Welche Voraussetzungen bestehen für die Verarbeitung und das Verständnis der Metadaten?
11. Identifikatoren und Informationsobjekte
- 11.1 In welchem Verhältnis stehen die Informationsobjekte zueinander?
  - 11.2 Für welche Informationsobjekte werden dauerhafte Identifikatoren benötigt?
  - 11.3 Was für Identifikatoren werden benutzt?
  - 11.4 Wie werden die Identifikatoren nachgewiesen?
  - 11.5 Wer wird die Aktualisierung und Pflege von Identifikatoren vornehmen?

## Integrationsmodelle

---

Wie tief ein/e Embedded Data Manager in ein Forschungsprojekt eingebettet sein sollte und kann – darüber gibt es keine klare Vorstellung. Das gilt ebenso für die Frage, was eigentlich die Tiefe des „Eingebettetseins“ ausmacht und welchen Bedingungen die „Einbettung“ unterliegt. Eine Annäherung könnten lebenspraktische Szenarien bieten.

Das *Five-Level Model on Embedded Librarianship for Health Sciences Librarians* beschreibt ein Stufenmodell, in dem BibliothekarInnen bestimmte Rollen/Funktionen, Aufgaben/Dienstleistungen und Kenntnisse/Fähigkeiten zugeordnet werden. Die Stufe bewertet die Integrationstiefe bzw. Grad des "Eingebettetseins".

### **Lassen sich ähnliche Szenarien für "Embedded Data Manager" beschreiben?**

Es muss dabei kein Stufenmodell entstehen, jedoch sollte der Grad der Integration definiert oder beschrieben sein. Ein Vorschlag für Elemente eines Szenarios:

- die beteiligten Institutionen/Organisationen, je für Embedded Data Manager und ForscherInnen
- Rolle/Funktion der/des Embedded Data Manager
- Aufgaben/Tätigkeiten der/des Embedded Data Manager in Bezug auf das Datenmanagement
- Aufgaben/Tätigkeiten/Eigenschaften/Umstände in Bezug auf die Integration der/des Embedded Data Manager in die Forschung (z.B. Teilnahme an Projekttreffen)
- notwendige Ausbildung/Kenntnisse/Fähigkeiten

## Five-Level Model on Embedded Librarianship for Health Sciences Librarians

Lin Wu & Misa Mi: Sustaining librarian vitality: embedded librarianship model for health sciences libraries. *Medical reference services quarterly*, 32(3), 2013, 257-265

zit.n. / Übersetzung:

Marina Betker: Die „embedded library“ als Strategiekonzept für Gesundheits-und Medizinbibliotheken, Berlin 2014

Stufe	Rolle des Bibliothekars	Dienstleistungen	Wissen und empfohlene / wünschenswerte Fähigkeiten
1	Der Bibliothekar nimmt die Rolle des Dienstleisters, Erwerbers und Bestandsentwicklers für eine spezifische Zielgruppe ein (und nicht für die gesamte Hochschule).	Der Bibliothekar kümmert sich z.B. um den Bestandsaufbau für einen bestimmten Fachbereich, Kurs oder ein Forschungsprojekt. Die Bibliotheksdienstleistungen sind physisch in der Bibliothek verortet.	Der Bibliothekar kennt die Inhalte des Kurses. Er erkennt außerdem den Literaturbedarf des Fachbereichs bezogen auf Forschung und Lehre und den Lernbedarf der Studierenden.
2	Der Bibliothekar nimmt die Rolle des Gastdozenten und Bestandsentwicklers für eine spezifische Zielgruppe ein.	Der Bibliothekar unterstützt den Fachbereich in Forschung und Lehre. Er entwickelt Einzelschulungen und führt diese bei Bedarf innerhalb eines Modules/Kurses durch. Er erstellt eine webbasierte Linksammlung von elektronischen Bibliotheksressourcen.	Der Bibliothekar kennt die Inhalte des Kurses und kann den Literaturbedarf des Fachbereichs und der Studierenden voraussehen. Er verfügt über Grundkenntnisse von Lernprinzipien und didaktischen Methoden.
3	Der Bibliothekar nimmt die Rolle eines Teammitgliedes oder Co-Dozenten ein, ist persönlicher Ansprechpartner für eine spezifische Zielgruppe.	Der Bibliothekar führt im Curriculum verankerte Schulungen durch. Er vereinbart Termine für individuelle Einzelschulungen. Er stellt einen gemeinsamen Arbeitsraum für Forschungsprojekte zur Verfügung. Er pflegt ein institutionelles Repositorium.	Der Bibliothekar verfügt über Kenntnisse im Bereich Open Access, Wissenschaftskommunikation und Forschungsprozesse. Er kennt Lerntheorien und Möglichkeiten der Unterrichtsgestaltung. Er verfügt über technologisches Know How, um Informationen zu organisieren und zu verwalten.

4	<p>Der Bibliothekar nimmt die Rolle des Lehrers, Kursentwicklers, Informationsspezialisten, klinisch medizinischer Bibliothekar, Patientenbibliothekar ein und ist als Bibliothekar Mitglied der Fakultät.</p>	<p>Der Bibliothekar führt benotete Lehrveranstaltungen durch. Sein Arbeitsplatz befindet sich teilweise oder ganz in der Fakultät und in der Nähe der Mitarbeiter und Studierenden. Im klinischen Kontext begleitet er Ärzte bei der Morgenvisite und bietet spezialisierte und personalisierte Services an. Er stellt gesundheitsbezogene Literatur für die Patienten bereit.</p>	<p>Der Bibliothekar kennt Lerntheorien und Möglichkeiten der Unterrichtsgestaltung. Er besitzt kulturelle- und Forschungskompetenz. Er versteht translationale Forschungsprozesse. Er verfügt über technologisches Know How, um Informationen zu organisieren und zu verwalten. Er besitzt Kenntnisse im Projektmanagement. Er hat ein zusätzliches Fachgebiet studiert oder eine zusätzliches Zertifikat.</p>
5	<p>Der Bibliothekar nimmt die Rolle eines Fakultätsmitglieds, eines Co-Autors, Forschungsmitarbeiter oder -partner ein oder ist an der Entwicklung des Curriculums beteiligt.</p>	<p>Der Bibliothekar ist an wichtigen Entscheidungsprozessen oder Ausschüssen der Institution beteiligt. Sein Arbeitsplatz befindet sich in der Nähe der Fakultät bei Mitarbeitern und Studierenden. Er entwickelt selbstständig Teile des Lehrplans. Er wirkt als Co-Autor bei systematischen Reviews mit oder unterstützt ein Forscherteam bei der Erstellung von Forschungsanträgen oder der Durchführung von Projekten.</p>	<p>Der Bibliothekar kennt Lerntheorien und die Möglichkeiten der Unterrichtsgestaltung. Er verfügt über kulturelle - und Forschungskompetenz. Er versteht translationale Forschungsprozesse. Er besitzt Kenntnisse im Projektmanagement. Er kennt den gesundheitsbezogenen Informationsbedarf von Patienten. Er verfügt über inhaltliches Fachwissen in einer Bezugsdisziplin der Gesundheitswissenschaften, führt einen Doktorgrad oder einen zusätzlichen Mastertitel in Bildung/Didaktik oder einer Bezugsdisziplin. Mit einer speziellen Fachkompetenz auf einem speziellen Gebiet leistet er Beiträge zu den strategischen Zielen der eigenen Institution.</p>

# Anzeige

---

Neben der Schwierigkeit, ForschungsdatenmanagerInnen zu finden, besteht die Schwierigkeit, sie zu suchen.

Szenario: Sie suchen für einen Sonderforschungsbereich eine/n "Embedded Data Manager" im Rahmen des INF-Projektes. Gehen Sie von der heutigen Perspektive aus, Sie können keine geeigneten Personen backen, sondern müssen sich an dem orientieren, was der Markt aus Ihrer Sicht hergibt.

Eine Ausschreibung sollte Angaben zu Ausbildung, Kenntnisse, Fähigkeiten, Tätigkeitsbeschreibung - diese können bewusst umfangreich oder spärlich, weit oder eng ausfallen, sollten aber begründet werden.

Anregungen (ohne Vorbildfunktion):

SFBs:

- SFB 1182: Entstehen und Funktionieren von Metaorganismen
- SFB 1187: Medien der Kooperation

Ausschreibungen:

- SUB Göttingen, 2012
- University of Sidney, 2015
- IZ/GIGA Hamburg, 2016

## Forschungsinfrastrukturaufbau" (SUB Göttingen)

---

Georg-August-Universität Göttingen, Göttingen, 01.03.2012  
Bewerbungsschluss: 31.01.2012

Die Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB) ist eine der grössten wissenschaftlichen Bibliotheken Deutschlands und engagiert sich seit Jahren in der Forschung und Entwicklung von digitalen Forschungsinfrastrukturen, virtuellen Forschungsumgebungen und Forschungsdaten-Management. Die Abteilung Forschung und Entwicklung (FE) der SUB sucht für drei Projekte in diesem Bereich wissenschaftliche Verstärkung. Sie koordiniert im Bereich geisteswissenschaftlicher Infrastrukturen das BMBF- geförderte Projekt DARIAH-DE, koordiniert den wissenschaftlichen Aufbau von Forschungsinfrastrukturen für ein Verbundprojekt im Bereich Digital Humanities am Wissenschaftsstandort Göttingen und organisiert im Sonderforschungsbereich "Ökologische und sozioökonomische Funktionen tropischer Tieflandregenwald-Transformationssysteme (Sumatra, Indonesien)" das Forschungsdaten-Management. Für diese Projekte besetzen wir drei Stellen für eine/einen

wissenschaftliche Mitarbeiterin/wissenschaftlichen Mitarbeiter - Entgeltgruppe 13 TV-L -

mit 39,8 bzw. 19,9 Wochenstunden zur Koordination des Aufbaus dieser Forschungsinfrastrukturen für eine Dauer von bis zu vier Jahren.

Zu den Aufgaben gehören:

Projektstelle 1: DARIAH-DE (19,9 WStd. für zwei Jahre)

[...ENTFERNT...]

Projektstelle 2: Verbundprojekt Digital Humanities (19,9 WStd. für drei Jahre)

[...ENTFERNT...]

Projektstelle 3: SFB Sumatra/Indonesien (39,8 WStd. für vier Jahre)

- Unterstützung bei der Koordination des wissenschaftlich-technischen Infrastrukturaufbaus
- wissenschaftliche Moderation der Diskussion zwischen Wissenschaftlern, Entwicklern und IT-Experten sowie die Abstimmung mit verwandten Projekten und Aktivitäten
- inhaltliche und technische Konzeption von Forschungsdaten-Management

Vorausgesetzt werden:

- Ein wissenschaftlicher Hochschulabschluss im Bereich Informatik oder Geisteswissenschaften mit ausgewiesenen technologischen Kenntnissen. Ein nachgewiesenes Verständnis digitaler geisteswissenschaftlicher Forschung ist von Vorteil.
- Fundierte Kenntnisse von verteilten Architekturkonzepten und den in den Aufgabenbeschreibungen benannten Technologien und Informationsinfrastrukturen.
- Nachgewiesene Erfahrung in der Projektarbeit in großen, dezentralen und heterogenen Teams.
- Gute organisatorische Fähigkeiten, gutes Zeitmanagement.
- Hohe Kommunikationskompetenz, sehr gute Teamfähigkeit und Englischkenntnisse.

Eine Bewerbung auf mehrere der Stelleanteile ist möglich. Die SUB Göttingen bietet Ihnen die Gelegenheit eng eingebettet in laufende Aktivitäten der Abteilung Forschung und Entwicklung am Aufbau nationaler und internationaler Forschungsinfrastrukturen und -dienste mitzuwirken. Die Aufgaben werden im lokalen Team (knapp 10 KollegInnen) in Abstimmung mit diversen verwandten Projekten der Abteilung (ca. 30 KollegInnen) und nationalen und internationalen Partnern geleistet.

[...ENTFERNT...]

URL <http://hsozkult.geschichte.hu-berlin.de/chancen/type=stellen&id=6870>

Das GIGA German Institute of Global and Area Studies / Leibniz-Institut für Globale und Regionale Studien ist ein unabhängiges, sozialwissenschaftliches Forschungsinstitut mit Sitz in Hamburg. Es analysiert politische, ökonomische und soziale Entwicklungen in Afrika, Asien, Lateinamerika und Nahost. Dabei forscht es aus vergleichender Perspektive zu internationalen Beziehungen, Entwicklung und Globalisierung, Gewalt und Sicherheit sowie zu politischen Systemen. Das GIGA bietet zum nächstmöglichen Zeitpunkt eine Position als

## **Wissenschaftliche/r Mitarbeiter/in im Informationszentrum.**

Es handelt sich um eine halbe Stelle, zunächst befristet auf drei Jahre; Entgeltgruppe 13 TV-AVH/TVöD.

Das GIGA Informationszentrum (IZ) ist die Spezialbibliothek für Area- und Comparative Area Studies. Kernaufgabe des IZ ist die Versorgung der WissenschaftlerInnen des GIGA mit forschungsrelevanten Informationen und damit zusammenhängenden Dienstleistungen. Das IZ unterstützt die WissenschaftlerInnen des GIGA in allen Phasen des Forschungsprozesses und ist bestrebt, Dienstleistungen u.a. in den Bereichen Forschungsdatenmanagement, Open Access, Science 2.0 auszubauen.

Zu Ihren Aufgaben gehören:

- Entwicklung einer Strategie zum Management quantitativer und qualitativer Forschungsdaten im engen Dialog mit den GIGA WissenschaftlerInnen, Umsetzung der Strategie zu einer an Forschungsthemen und Forschungsprozessen orientierten Dienstleistung
- Konzeption und Durchführung von Beratungs- und Schulungsangeboten, bspw. zu Software der quantitativen Datenanalyse, Datenbanken, Zotero
- Konzeption, Aufbau und Ausbau von Dienstleistungen des Informationszentrums sowie deren Betreuung
- Kooperation mit externen Institutionen

Gewünschtes Qualifikationsprofil:

- Abgeschlossenes Hochschulstudium (Master/Magister) in einem sozialwissenschaftlichen Fach, vorzugsweise Politikwissenschaften oder Wirtschaftswissenschaften
- Nachgewiesene Erfahrungen in der Arbeit mit gängigen sozialwissenschaftlichen Datenbanken (Mikro/Makro)
- Erfahrungen in der Arbeit mit Software zur quantitativen Datenanalyse (z.B. Stata, R)
- Kenntnisse im Forschungsdatenmanagement und entsprechender Werkzeuge wünschenswert
- Hochschulstudium (Master/Magister) im Bibliotheks- und Informationswissenschaftlichen Bereich von Vorteil
- Interesse an aktuellen Entwicklungen im Bereich von Forschungs- und Informationsinfrastrukturen
- Fähigkeit zum selbständigen konzeptionellen Arbeiten und Innovationsfreude
- Ausgesprochen dienstleistungsorientiertes Aufgabenverständnis, ausgeprägte Team- und Kommunikationsfähigkeit sowie Kooperationsbereitschaft
- Verhandlungssichere Deutsch- und Englischkenntnisse in Wort und Schrift

Liegt kein Abschluss im Bibliotheks- und Informationswissenschaftlichen Bereich vor, fördert das GIGA bei Interesse der Stelleninhaberin/des Stelleninhabers ein entsprechendes berufsbegleitendes Studium.

Die Vereinbarkeit von Beruf und Familie ist ein wichtiges Anliegen des Instituts. Das GIGA fördert die Gleichstellung von Frauen und Männern und bemüht sich aktiv um Bewerbungen von Frauen. In Bereichen, in denen Frauen unterrepräsentiert sind, werden sie bei gleicher Qualifikation vorrangig berücksichtigt.

Bitte richten Sie Ihre Bewerbung (**Ref.-Nr. GIGA-16-03**) mit aussagekräftigen Unterlagen (Anschreiben; Lebenslauf; Zeugnisse sowie das unter <https://giga.hamburg/de/stellen-und-praktika> herunterzuladende Bewerbungsformular) an:

**Urte Schneider - IZ Sekretariat, GIGA German Institute of Global and Area Studies, Neuer Jungfernstieg 21, 20354 Hamburg, Germany E-Mail: [jobs-iz@giga.hamburg](mailto:jobs-iz@giga.hamburg)**  
(Bewerbungen per E-Mail sind ausdrücklich erwünscht).

Das Auswahlverfahren beginnt am 7. März 2016.

Für weitere Informationen besuchen Sie bitte die Webseiten des GIGA ([www.giga.hamburg](http://www.giga.hamburg)) oder kontaktieren Sie Jan Lüth (Leiter IZ, Tel.: +49 (0)40-42825-565) oder Sonja Bartsch (Wissenschaftsmanagement, Tel.: +49 (0)40-42825-592).

Das GIGA ist ausgezeichnet.  
Hamburger Familiensiegel



[HOME](#)[MAIN MENU](#)[SEARCH](#)[CONTACT](#)[MEMBERS PORTAL](#)

Australian Library and  
Information Association

# RESEARCH DATA MANAGER

## OVERVIEW

### RESEARCH DATA MANAGER

### UNIVERSITY LIBRARY

### REFERENCE NO. 783/0414A

- **A truly diverse and interesting role allowing you to develop and lead processes**
- **Opportunity to think and act strategically and foster solid internal and external relationships**
- **Full-time continuing, remuneration package: \$109K-\$123K p.a. (which includes leave loading and up to 17% super)**

The University of Sydney is Australia's first university and has an outstanding global reputation for academic and research excellence. It employs over 7300 permanent staff, supporting over 50,000 students.

The University Library's modernisation agenda is focused on delivering high quality, client-centred services, information resources and learning spaces to support world-class research and research-informed teaching. As such, we are seeking to appoint innovative, energetic and collaborative leaders like you, to join our new organisational structure.

The Access Services division is responsible for the acquisition, organisation, discovery, access and management of library information resources, so you can be assured of plenty of variety in the role. As Research Data Manager, you will be tasked with providing advice, client-focused leadership and expertise within the University Library and wider University on the management, curation and access of research data. As you foster strong working relationships with the faculties and research portfolio, you will have the opportunity to develop and implement a wide range of data management services and set protocols and guidelines to ensure the University meets compliance on data use.

This is a role that you can truly make your own, where your contribution can make a huge difference. As such, you will see yourself as someone who is analytical and comfortable with process who possesses the ability to engage, collaborate and consult with ease. There is a strong people management component to the position, so your thirst to mentor and build up individuals into cohesive and effective teams will certainly be quenched.

This is a prime opportunity to truly effect change for the better, where your vision and experience will ultimately serve to enhance the University's research profile on a global scale. We very much look forward to

receiving your application.

**Remuneration package:** \$109,610 p.a.-\$123,615 p.a. (which includes a base salary of \$92,621p.a. - \$104,456 p.a., leave loading and up to 17% employer's contribution to superannuation).

All applications must be submitted via the University of Sydney careers website. Visit [sydney.edu.au/recruitment](http://sydney.edu.au/recruitment/) (<http://sydney.edu.au/recruitment/>) and search by the reference number 783/0414A for more information and to apply.

General enquiries can be directed to Abdul Macauley, Lead Recruitment Consultant, at [abdul.macauley@sydney.edu.au](mailto:abdul.macauley@sydney.edu.au) (<mailto:abdul.macauley@sydney.edu.au>) or on +61 2 8627 1238. Specific enquiries about the role can be directed to Anne Bell, University Librarian, at [anne.bell@sydney.edu.au](mailto:anne.bell@sydney.edu.au) (<mailto:anne.bell@sydney.edu.au>) or on +61 2 9351 2990.

**CLOSING DATE:** 5pm, 8 March 2015

The University is an equal opportunity employer committed to equity, diversity and social inclusion. Applications from equity target groups, including women and people with disabilities are encouraged. As the University of Sydney has established a scheme to increase the number of Aboriginal and Torres Strait Islander staff employed across the institution, applications from people of Aboriginal and Torres Strait Islander descent are also encouraged.

© The University of Sydney

## DETAILS

**Job region** [NSW \(/jobreg/nsw\)](#)

**Employer** The University Of Sydney

**Type** Full time

**Salary** \$109K-\$123K p.a. (which includes leave loading and up to 17% super)

**Contact**

Abdul Macauley, Lead Recruitment Consultant, at [abdul.macauley@sydney.edu.au](mailto:abdul.macauley@sydney.edu.au) or on +61 2 8627 1238

**Closing date** Sunday 08 March 2015

## SFB 1182: Entstehen und Funktionieren von Metaorganismen

**Fachliche Zuordnung** Biologie  
Medizin  
**Förderung** Förderung seit 2016

### Projektbeschreibung

Keine Zusammenfassung vorhanden

**DFG-Verfahren** Sonderforschungsbereiche  
Laufende Projekte

- [A01 - Molekulare Basis und Evolutionsdynamik von C. elegans-Mikrobiota Interaktionen](#) (Teilprojektleiter [Dierking, Katja](#) ; [Leippe, Matthias](#) ; [Schulenburg, Hinrich](#) ; [Tholey, Andreas](#) )
- [A02 - Wirt-Mikroben Ko-Evolution im Säugerdarm](#) (Teilprojektleiter [Baines, John](#) ; [Franke, Ph.D., Andre](#) )
- [A03 - Von der Domestikation abhängige Evolution des Weizen-Metaorganismus](#) (Teilprojektleiter [Dagan, Tal](#) ; [Holtgrewe-Stukenbrock, Ph.D., Eva](#) )
- [A04 - Mathematische Modellierung der Interaktionen in evolvierenden Metaorganismen](#) (Teilprojektleiter [Bosch, Thomas C. G.](#) ; [Fraune, Sebastian](#) ; [Schulenburg, Hinrich](#) ; [Traulsen, Arne](#) )
- [B01 - Wirt-Bakterien Interaktionen in den basalen Tierstämmen Porifera und Cnidaria](#) (Teilprojektleiter [Fraune, Sebastian](#) ; [Hentschel-Humeida, Ute](#) )
- [B02 - Mikroben-Wirt Interaktionen zu Beginn der vielzelligen Stammesentwicklung](#) (Teilprojektleiter [Reusch, Ph.D., Thorsten](#) ; [Schmitz-Streit, Ruth Anne](#) )
- [C01 - Einen denkbar einfachen Vielzeller verstehen: welchen Einfluss haben symbiotische Mikroben und Algen auf Entwicklungsprozesse in Hydra und wie passen sie sich an den Wirtsorganismus an?](#) (Teilprojektleiter [Bosch, Thomas C. G.](#) ; [Hülter, Nils](#) )
- [C02 - Der Einfluss von Fehlernährung auf Epithel-Mikroben Interaktionen im Darmtrakt von Fliegen und Mäusen.](#) (Teilprojektleiter [Roeder, Thomas](#) ; [Rosenstiel, Ph.D., Philip Caspar](#) )
- [C03 - Der Lebenszyklus des Tiefsee Metaorganismus Bathymodiolus: Populationsdynamik der Symbionten und Besiedelung des Kiemenepithels des Wirtes](#) (Teilprojektleiter [Dagan, Tal](#) ; [Dubilier, Nicole](#) )
- [Z01 - Administration und Koordination](#) (Teilprojektleiter [Bosch, Thomas C. G.](#) )
- [Z02 - Mikroorganismen: Kultur und Einzelzellanalyse](#) (Teilprojektleiter [Hentschel-Humeida, Ute](#) ; [Schmitz-Streit, Ruth Anne](#) )
- [Z03 - Einsatz von Sequenzierungsmethoden der nächsten Generation zum Studium von Metaorganismen](#) (Teilprojektleiter [Baines, John](#) ; [Häsler, Robert](#) ; [Rosenstiel, Ph.D., Philip Caspar](#) )
- [INF - Data management and integrated analyses](#) (Teilprojektleiter [Dutheil, Ph.D., Julien](#) ; [Franke, Ph.D., Andre](#) ; [Lorenz, Sören](#) ; [Marten, Holger](#) )

**Antragstellende Institution:** [Christian-Albrechts-Universität zu Kiel](#)  
24098 Kiel

**Beteiligte Institution:** [GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel](#)

Wischhofstraße 1-3  
24148 Kiel

[Max-Planck-Institut für Evolutionsbiologie](#)  
August-Thienemann-Straße 2  
24306 Plön

[Max-Planck-Institut für Marine Mikrobiologie](#)  
Celsiusstraße 1  
28359 Bremen

**Beteiligte Hochschule:** [Universität Hamburg](#)  
Mittelweg 177  
20148 Hamburg

**Sprecher:** [Professor Dr. Thomas C. G. Bosch](#)  
Christian-Albrechts-Universität zu Kiel  
Mathematisch-Naturwissenschaftliche Fakultät  
Zoologisches Institut  
Am Botanischen Garten 1-9  
24118 Kiel  
Telefon: +49 431 8804169  
Telefax: +49 431 8804747  
E-Mail: [tbosch@zoologie.uni-kiel.de](mailto:tbosch@zoologie.uni-kiel.de)

**fachlicher  
DFG-Ansprechpartner:** [Christoph Limbach](#)

---

GEPRIS ist ein Projekt der Deutschen Forschungsgemeinschaft.

Sie erreichen GEPRIS unter <http://www.dfg.de/gepris>

(c) 1999 - 2016 Deutsche Forschungsgemeinschaft (<http://www.dfg.de>)

## SFB 1187: Medien der Kooperation

<b>Fachliche Zuordnung</b>	Sozial- und Verhaltenswissenschaften Geisteswissenschaften Elektrotechnik, Informatik und Systemtechnik
<b>Förderung</b>	Förderung seit 2016

### Projektbeschreibung

Keine Zusammenfassung vorhanden

**DFG-Verfahren** Sonderforschungsbereiche  
Laufende Projekte

- [A01 - Digital vernetzte Medien zwischen Spezialisierung und Universalisierung](#) (Teilprojektleiter [Schüttpelz, Erhard](#) )
- [A02 - Standardisierungskultur der Telekommunikation im Spannungsfeld der digitalen und neoliberalen 'Doppelrevolution' seit den 1980er Jahren](#) (Teilprojektleiter [Henrich-Franke, Christian](#) )
- [A03 - Navigation in Online-/Offline-Räumen](#) (Teilprojektleiter [Gerlitz, Carolin](#) ; [Thielmann, Tristan](#) )
- [A04 - Normale Betriebsausfälle. Struktur und Wandel von Infrastrukturen im öffentlichen Dienst](#) (Teilprojektleiter [Pothast, Jörg](#) )
- [A05 - Kooperative Herstellung von Nutzerautonomie im Kontext von Open Source und der alternden Gesellschaft](#) (Teilprojektleiter [Müller, Claudia](#) ; [Stevens, Gunnar](#) )
- [A06 - Visuell integrierte klinische Kooperation](#) (Teilprojektleiter [Kolb, Andreas](#) ; [Schubert, Cornelius](#) )
- [B01 - Literarische Öffentlichkeiten im deutschsprachigen 18. Jahrhundert: Medienpraktiken von Patronage und Freundschaft](#) (Teilprojektleiter [Ghanbari, Nacim](#) ; [Stanitzek, Georg](#) )
- [B03 - "Going Public" in medienkooperativen Engagementformen](#) (Teilprojektleiter [Baringhorst, Sigrid](#) )
- [B04 - Digitale Öffentlichkeiten und gesellschaftliche Transformation im Maghreb](#) (Teilprojektleiter [Rohde, Markus](#) ; [Wulf, Volker](#) ; [Zillinger, Martin](#) )
- [B05 - Frühe Kindheit und Smartphone. Familiäre Interaktionsordnung, Lernprozesse und Kooperation.](#) (Teilprojektleiter [Wiesemann, Jutta](#) )
- [B06 - Un-/Erbetene Beobachtung: Die Überwachungsgesellschaft und das soziale Feld der Medien](#) (Teilprojektleiter [Ludwig-Mayerhofer, Wolfgang](#) )
- [B07 - Medienpraktiken und Urheberrecht - Soziale und juristische Rahmenbedingungen kooperativen und derivativen Werkschaffens im Netz](#) (Teilprojektleiter [Hoffmann, Dagmar](#) ; [Klass, Nadine](#) )
- [P01 - Wissenschaftliche Medien der Praxistheorie: Harold Garfinkel und Ludwig Wittgenstein](#) (Teilprojektleiter [Rawls, Anne](#) ; [Schüttpelz, Erhard](#) ; [Thielmann, Tristan](#) )
- [Ö - Öffentliche Infrastrukturen: Sichtbarkeit und Unsichtbarkeit](#) (Teilprojektleiter [Henrich-Franke, Christian](#) ; [Neutsch, Cornelius](#) )
- [INF - Infrastrukturkonzepte für die Beforschung kooperativer Medien](#) (Teilprojektleiter [Pipek, Volkmar](#) ; [Roller, Sabine](#) )

■ **Z - Zentrales Verwaltungsprojekt des SFB (Teilprojektleiter Schüttpelz, Erhard )**

**Antragstellende Institution:** [Universität Siegen](#)  
Herrngarten 3  
57072 Siegen

**Beteiligte Hochschule:** [Universiteit van Amsterdam](#)  
Kruislaan 409  
1098 Amsterdam  
Niederlande

[Universität zu Köln](#)  
Albertus-Magnus-Platz  
50931 Köln

**Sprecher:** [Professor Dr. Erhard Schüttpelz](#)  
Universität Siegen  
Fakultät I  
Medienwissenschaftliches Seminar

**fachlicher DFG-Ansprechpartner:** [Claudia Althaus](#)

---

GEPRIS ist ein Projekt der Deutschen Forschungsgemeinschaft.  
Sie erreichen GEPRIS unter <http://www.dfg.de/gepris>  
(c) 1999 - 2016 Deutsche Forschungsgemeinschaft (<http://www.dfg.de>)

# Förderung

---

Derzeit scheinen große Verbundprojekte die beste Möglichkeit für den Einsatz von Embedded Data Manager zu sein, weil der Bedarf und Finanzierungsmöglichkeit gleichsam gegeben sind. Die Organisationsform als Projekt ermöglicht eine direkte Einbindung, die Größe des Projektes rechtfertigt die Einbindung mindestens einer ganzen Person.

Allerdings wird in den Empfehlungen der DFG noch wenig auf Formen der Integration von Forschungsdatenmanagement und der Betreuung durch Personen (statt technischer Systeme) eingegangen.

Zwei Fragen als Diskussionsansatz:

1. Wie passt ein Embedded Data Manager argumentativ in das INF-Programm und wie wäre er zu rechtfertigen, gegenüber a) den Empfehlungen für INF-Projekte, b) der Forschungsdatenleitlinie und c) den Empfehlungen für die Gutachter?
2. Was könnten sinnvolle Ergänzungen zu den "Empfehlungen für INF-Projekte" sein, die ein betreutes Forschungsdatenmanagement durch Embedded Data Manager anregen und direkter empfehlen? Wäre dies überhaupt sinnvoll?

Anlagen:

- Empfehlungen für INF-Projekte:  
DFG-Vordruck 50.06, Merkblatt Sonderforschungsbereiche, III.1.3 Teilprojekt Informationsinfrastruktur
- Empfehlungen für Gutachter:  
DFG-Vordruck 60.14, Hinweise zur Begutachtung von Anträgen im Programm Sonderforschungsbereiche, I.4. Serviceprojekt Informationsinfrastruktur
- Forschungsdatenleitlinie der DFG

# DFG Merkblatt 50.06

---

## III.1.3 Teilprojekt Informationsinfrastruktur

In einem Sonderforschungsbereich dient ein Teilprojekt Informationsinfrastruktur vor allem dem systematischen Management der im Kontext des Sonderforschungsbereichs relevanten Daten. Unter Forschungsdaten werden alle Ergebnisse und Bezugsquellen des Forschungsprozesses (u.a. auch Software oder Forschungsobjekte bzw. Proben) verstanden, die im Projekt erhoben, ausgewertet und/oder entwickelt werden. Ein auf diese Daten bezogener Einsatz sowie die Erprobung oder Entwicklung neuer wissenschaftlicher Kommunikationsformen ist ebenfalls möglich. Es können die Entwicklung und Umsetzung eines Konzeptes sowie die Bereitstellung einer dafür notwendigen leistungsfähigen Informationsinfrastruktur gefördert werden. Damit soll erreicht werden, dass sich im Sonderforschungsbereich wissenschaftliche Synergien durch gemeinsame Datenplattformen und/oder Kommunikationsforen sowie eine effiziente Datennutzung einstellen. Grundsätzlich wird ein professionelles Management der Daten, die im Sonderforschungsbereich erhoben, ausgewertet und/oder entwickelt werden, erwartet. In der Regel soll daher mit den einschlägig ausgewiesenen Informationseinrichtungen am Standort zusammengearbeitet werden (z.B. Bibliotheken, Rechenzentren oder Biobanken der antragstellenden Hochschule). Der Nutzung bereits vorhandener Repositorien (siehe auch <http://www.re3data.org> oder <http://resources.dfg.de>), Werkzeuge und Techniken ist gegenüber der Entwicklung neuer Instrumente in der Regel der Vorzug einzuräumen. Auch nach Ende der Projektförderung soll ein Zugang zu den Daten oder eine Übergabe an bestehende Infrastrukturen gewährleistet sein. Grundsätzlich soll zum Zeitpunkt der Antragstellung, bei experimentellen Entwicklungen spätestens vor Projektabschluss zur Frage der Nachhaltigkeit Stellung genommen werden. Im Einzelnen kann sich ein Teilprojekt Informationsinfrastruktur auf eins oder mehrere der folgenden Ziele beziehen:

- Zugang zu Forschungsdaten
- Erschließung und Pflege von Forschungsdaten
- Nachnutzung von Forschungsdaten
- Virtuelle Forschungsumgebungen
- Aufbau von interoperablen Komponenten
- Adaption und Implementierung neuartiger Technologien und Verfahren

Finanziert werden können neben Mitteln für Personal (u.a. wissenschaftliche Mitarbeiter / Anwendungsinformatiker / technisches Personal) auch Sach- und Investitionsmittel (u.a. Server und Hardwarekomponenten, Software und Updates, Speichermedien, Lizenzen, Support und externer Service), sofern diese projektspezifisch begründet sind.

# DFG Merkblatt 60.14

---

## I.4. Serviceprojekt Informationsinfrastruktur (falls beantragt)

Wie beurteilen Sie die Qualität des Projekts und seine Einbettung in den Verbund?

- Ist das Arbeitsprogramm überzeugend?
- Gibt es besondere, neuartige Ideen zum Umgang mit Daten (alle Bezugsquellen und Ergebnisse des Forschungsprozesses), die im Sonderforschungsbereich erhoben, ausgewertet und/oder entwickelt werden? Lassen die Planungen erwarten, dass prototypische Entwicklungen entstehen?
- Wie werden fachspezifisch vorhandene Standards zu Dateninhalten und -formaten in den Arbeiten umgesetzt? Wie beurteilen Sie die formulierten Alternativen in Fällen fehlender Standards?
- Wie verhält sich das Projekt zu außerhalb der antragstellenden Hochschule bestehender relevanter Informationsinfrastruktur? Wird überzeugend begründet, wie die Arbeiten im Teilprojekt an andere Entwicklungen Anschluss finden oder sich davon absetzen und eigenständige Lösungen entwickeln wollen?
- Ist gesichert, dass Daten aus den wissenschaftlichen Teilprojekten im Teilprojekt INF angemessen inhaltlich gewürdigt werden? Wie sollen sie informationstechnisch aufbereitet werden?
- Besitzen die Teilprojektleitenden sowohl wissenschaftlich als auch informationstechnisch eine passende Expertise?
- Wie ist die Einbettung am Standort gelungen?
- Ist das Projekt mit den einschlägigen Einrichtungen zur Informationsinfrastruktur am Standort vernetzt?
- Wird die Langzeitverfügbarkeit der im Projekt zu entwickelnden Informationsinfrastruktur inhaltlich und technisch überzeugend gesichert bzw. wo dies noch nicht absehbar ist, werden überzeugende Kriterien für eine nachfolgende Beurteilung der Nachhaltigkeit beschrieben?

Falls das Projekt förderungswürdig ist:

- Sind die beantragten Mittel unter Berücksichtigung der eingebrachten Grundausstattung angemessen veranschlagt? Inwieweit gehen die Sachmittel über die zur Einhaltung der Guten Wissenschaftlichen Praxis (Umgang mit Primärdaten) notwendigen Maßnahmen hinaus?

## Leitlinien zum Umgang mit Forschungsdaten

Forschungsdaten sind eine wesentliche Grundlage für das wissenschaftliche Arbeiten. Die Vielfalt solcher Daten entspricht der Vielfalt unterschiedlicher wissenschaftlicher Disziplinen, Erkenntnisinteressen und Forschungsverfahren. Zu Forschungsdaten zählen u.a. Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Surveydaten, Objekte aus Sammlungen oder Proben, die in der wissenschaftlichen Arbeit entstehen, entwickelt oder ausgewertet werden. Methodische Testverfahren, wie Fragebögen, Software und Simulationen können ebenfalls zentrale Ergebnisse wissenschaftlicher Forschung darstellen und sollten daher ebenfalls unter den Begriff Forschungsdaten gefasst werden. Die langfristige Sicherung und Bereitstellung der Forschungsdaten leistet einen Beitrag zur Nachvollziehbarkeit und Qualität der wissenschaftlichen Arbeit und eröffnet wichtige Anschlussmöglichkeiten für die weitere Forschung. Die Allianz der Wissenschaftsorganisationen hat sich bereits mit den im Jahr 2010 verabschiedeten „Grundsätze[n] zum Umgang mit Forschungsdaten“ für die langfristige Sicherung von, den grundsätzlich offenen Zugang zu und die Berücksichtigung fachdisziplinärer Regularien im Umgang mit Forschungsdaten ausgesprochen.<sup>1</sup> Die „Leitlinien zum Umgang mit Forschungsdaten“ konkretisieren den mit den „Grundsätzen“ vorgegebenen Rahmen im Kontext der DFG-Förderregularien.

*Folgende übergeordnete Leitlinien gelten bei der Deutschen Forschungsgemeinschaft für Antragstellende:*

### 1. Projektplanung und Antragstellung

Bereits in die Planung eines Projekts sollten Überlegungen einfließen, ob und welche der aus einem Vorhaben resultierenden Forschungsdaten für andere Forschungskontexte relevant sein können und in welcher Weise diese Forschungsdaten anderen Wissenschaftlerinnen und Wissenschaftlern zur Nachnutzung zur Verfügung gestellt werden können. In einem Antrag sollten die Antragstellenden daher ausführen, welche Forschungsdaten im Verlauf eines wissenschaftlichen Forschungsvorhabens entstehen, erzeugt oder ausgewertet werden. Dabei sollten fachspezifisch angemessene Konzepte und Überlegungen für die Qualitätssicherung, für den Umgang mit und die langfristige Sicherung der Forschungsdaten zugrunde gelegt werden. Die einschlägigen Erläuterungen müssen Informationen zu Datentypen, falls vorhanden zu disziplinspezifischen Standards und zur Wahl geeigneter Repositorien enthalten, sofern diese für ein bestimmtes Fachgebiet oder bestimmte Datentypen vorhanden sind. Zusätzlich werden Angaben zu ggf. betroffenen Rechten Dritter sowie erste Planungen zum zeitlichen Rahmen der Datenveröffentlichung erbeten.

### 2. Bereitstellung

Soweit einer Veröffentlichung der Forschungsdaten aus einem DFG-geförderten Projekt Rechte Dritter (insbesondere Datenschutz, Urheberrecht) nicht entgegenstehen, sollten Forschungsdaten so zeitnah wie möglich verfügbar gemacht werden. Die Forschungsdaten sollten dabei in einer Verarbeitungsstufe (Rohdaten oder bereits weiter strukturierte Daten) zugänglich sein, die eine sinnvolle Nach- und Weiternutzung durch Dritte ermöglicht. Um dies sicherzustellen, ist darauf zu achten, dass der Zugang zu den Forschungsdaten auch dann gewährleistet bleibt, wenn im Zusammenhang mit einer Publikation Verwertungsrechte an den Forschungsdaten an Dritte, i.d.R. einen Verlag, übertragen werden müssen.

### 3. Langfristige Sicherung

Den Regeln der Guten Wissenschaftlichen Praxis folgend sollen Forschungsdaten in der eigenen Einrichtung oder in einer fachlich einschlägigen, überregionalen Infrastruktur für mindestens 10 Jahre archiviert werden.

---

<sup>1</sup> <http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaeetze.html>

*Folgende Unterstützung bietet die Deutsche Forschungsgemeinschaft zur Umsetzung der Leitlinien:*

1. Unterstützung und Beratung

Die Nachnutzung von Forschungsdaten spielt in nahezu allen Wissenschaftsbereichen eine wachsende Rolle. Gleichzeitig sind die fachspezifischen Unterschiede im Hinblick auf den Stand der Diskussionen und die zur Verfügungen stehenden Infrastrukturen erheblich. Um die Planung von Forschungsprojekten zu erleichtern, wurden Informationen und Anregungen zum Umgang mit Forschungsdaten zusammengetragen: [www.dfg.de/antragstellung/forschungsdaten/](http://www.dfg.de/antragstellung/forschungsdaten/).

Ein Verzeichnis von Forschungsdatenrepositorien findet sich unter: <http://www.re3data.org/>.

Einen Überblick über weitere, bereits existierende und qualitätsgesicherte Infrastrukturen bietet unter anderem das Informationsportal RISources: <http://risources.dfg.de/>.

2. Kosten für die Aufbereitung von Daten und für die Nutzung existierender Infrastrukturen

Projektspezifische Kosten, die im Rahmen eines wissenschaftlichen Projekts bei der Aufbereitung von Forschungsdaten für eine Anschlussnutzung bzw. für die Überführung von Forschungsdaten in existierende Infrastrukturen entstehen, können mit dem Antrag bei der DFG eingeworben werden. Ebenso können Mittel zur Finanzierung derjenigen Kosten eingeworben werden, die für die Nutzung einschlägiger Infrastrukturen entstehen. Gefördert werden können Personalkosten, projektspezifische Hard- und Software sowie Nutzungsgebühren.

3. Förderung des Auf- und Ausbaus von Informationsinfrastrukturen

Die Deutsche Forschungsgemeinschaft unterstützt mit dem Programm „Informationsinfrastrukturen für Forschungsdaten“ die Wissenschaft dabei, Strukturen für einen verbesserten Umgang mit Forschungsdaten und Forschungsdatenrepositorien auszuarbeiten und umzusetzen.

Weiterführende Informationen finden Sie unter:

[http://www.dfg.de/foerderung/programme/infrastruktur/lis/lis\\_foerderangebote/index.html](http://www.dfg.de/foerderung/programme/infrastruktur/lis/lis_foerderangebote/index.html)

*Folgenden Appell richtet die Deutsche Forschungsgemeinschaft an die wissenschaftlichen Fachgemeinschaften:*

1. Disziplinspezifische Regularien

Der Umgang mit Forschungsdaten ist in hohem Maße von den Arbeitsweisen wissenschaftlicher Disziplinen geprägt. Daher fordert die Deutsche Forschungsgemeinschaft Fächer, Fachgesellschaften und Communities dazu auf, ihren Umgang mit Forschungsdaten zu reflektieren und angemessene Regularien zur disziplinspezifischen Nutzung und ggf. offenen Bereitstellung von Forschungsdaten zu entwickeln. Beispiele für bislang unter Einbindung der DFG-Fachkollegien erarbeitete, fachspezifische Regularien im Kontext der DFG-Förderung finden sich unter [www.dfg.de/antragstellung/forschungsdaten](http://www.dfg.de/antragstellung/forschungsdaten).

2. Anerkennung der Leistung bei der Verfügbarmachung von Forschungsdaten

Das Engagement und die Bemühungen von Wissenschaftlern und Wissenschaftlerinnen um die Verfügbarmachung von Forschungsdaten, wie beispielsweise die fachspezifische Weiterentwicklung des Diskussionsprozesses oder der technischen Möglichkeiten der Sicherung, Auswertung oder Vernetzung von Forschungsdaten sollten bei der Würdigung von wissenschaftlichen Qualifikationen und Leistungen zukünftig stärker berücksichtigt werden.

**Verabschiedet durch den Senat der DFG am 30. September 2015**

## Übung III

### Auswahl und Bewertung von Forschungsdaten für die Archivierung und Nachnutzung

Es ist Karneval. Sie sind dem Getümmel entflohen und haben nun endlich Zeit, etwas Ordnung in Ihre persönlichen digitalen Dokumente zu bringen. Sie sind überrascht, was sich über die Jahre alles angesammelt hat! Unter anderem finden Sie

- Rechnungen
- Verträge
- Kontoauszüge
- E-Mails
- Musikbibliothek mit Titeln aus verschiedenen Quellen (gekaufte mp3 Alben, konvertierte CDs, digitalisierte LPs)
- Fotos
- Urlaubsvideos
- Steuererklärungen
- Scans von Zeugnissen und Urkunden
- Ihre Dissertation (Texte, Literatur, Daten)
- Backup Ihrer Handydateien
- E-Books für Ihren Kindle
- Zugangsdaten und Passwörter für verschiedenste Onlinedienste
- Kapitel aus einem Roman, an dem Sie schreiben
- eine App, die Sie selbst programmiert haben
- eine Kopie Ihrer Facebook-Daten
- ...und vieles mehr.

Sie entscheiden sich, das ganze systematisch anzugehen und beginnen damit, ein persönliches digitales Archiv anzulegen.

#### Aufgabenteil 1

Legen Sie zunächst fest, welche Zielsetzung(en) Sie mit dem Aufbau des Archivs verfolgen (z.B. Nutzbarkeit der digitalen Objekte für Sie selbst erhalten, Überlieferung an Dritte...).

Überlegen Sie sich dann für diese Zielsetzung(en) Auswahlkriterien, mit denen Sie entscheiden können, welche Ihrer digitalen Objekte Sie in Ihr persönliches Archiv übernehmen möchten. Stellen Sie sicher, dass die formulierten Kriterien die nachfolgenden Aspekte berücksichtigen:

- inhaltliche Aspekte,
- rechtliche Aspekte,
- technische Aspekte,
- zur Verfügung stehende Ressourcen.

Achten Sie darauf, die Kriterien hinreichend konkret auszuformulieren, sodass eine Bewertung der digitalen Objekte tatsächlich möglich ist.

#### Aufgabenteil 2

Öffnen Sie den Umschlag. Darin finden Sie ein beispielhaftes Objekt / eine Gruppe von Objekten. Entscheiden Sie anhand der formulierten Kriterien, ob die Übernahme der Objekte in Ihr persönliches Archiv gewünscht und möglich ist.

## **Gruppe 1**

Fotos Ihres 25-jährigen Dienstjubiläums aus verschiedenen Quellen:

- von Ihrem Smartphone
- vom Firmenserver heruntergeladen
- von Ihrer eigenen Digitalkamera

## **Gruppe 2**

Fotos Ihres 25-jährigen Dienstjubiläums aus verschiedenen Quellen:

- von Ihrem Smartphone
- vom Firmenserver heruntergeladen
- von Ihrer eigenen Digitalkamera

## **Gruppe 3**

Eine persönliche Webseite, auf der Sie sich und ihre Familie vorstellen.

## **Gruppe 4**

Eine persönliche Webseite, auf der Sie sich und ihre Familie vorstellen.

# Nestor School 2016 - Sheer Curation Übungen

## ***Beispielszenario Bildbearbeitung***

Diese Übung baut auf folgendem Szenario auf:

Es wurden mehrere aufwändige Fotos aufgenommen und auf einen Arbeitsrechner kopiert. Dort werden die Fotos von einem Mediengestalter weiterverarbeitet. Dieser erstellt zunächst eine Kopie der Originale an denen er die Veränderungen vornimmt. Er macht Farbanpassungen und fügt jedem Bild das Logo seines Unternehmens hinzu. Während der Arbeit hat er ein Dokument geöffnet, in dem er seine Arbeitsschritte grob dokumentiert und festhält, welche Veränderungen an welchem Bild vorgenommen wurden.

Die fertigen Bilder sollen hinterher auf eine Website geladen werden. Gleichzeitig sollen sowohl die Originale, als auch die bearbeiteten Bilder an ein Archiv übermittelt werden.

## **PERICLES Sheer Curation Tools**

Arbeiten Sie folgende Anleitungen durch und nehmen Sie dabei thematisch Bezug auf das Beispielszenario.

### ***Aufgabe 1: Informationen sammeln mit dem PERICLES Extraction Tool:***

<https://github.com/pericles-project/pet/wiki/Quick-start-guide---german-translation>

Kopieren Sie die extrahierten Informationen in einen Ordner und diskutieren Sie die Ergebnisse.

### ***Aufgabe 2: Informationen verkapseln mit dem PERICLES Content Aggregation Tool:***

<https://github.com/pericles-project/PeriCAT/wiki/Quick-Start-Guide---german-translation>

Kopieren Sie die verkapselten Informationen sowie das erstellte Szenario in einen Ordner und diskutieren Sie die Ergebnisse.

Auf der Rückseite befinden sich Zusatzaufgaben zu den verschiedenen behandelten Themen. Wählen Sie eine davon aus, wenn noch genügend Zeit über ist.

# Zusatzaufgaben

## Zusatzaufgabe A: Signifikante Umgebungsinformationen

- A) Überlegen Sie sich ein eigenes Szenario, in dem digitale Objekte entstehen und/oder verändert werden.
- B) In welcher Anwendung werden diese Objekte in dem Szenario primär verwendet?
- C) Gibt es weitere sekundäre Anwendungen in denen die Objekte sofort verwendet werden könnten?
- D) Welche Verwendung der Objekte sind in ferner Zukunft denkbar?
- E) Welche Entitäten aus der Umgebung der Objekte haben Einfluss auf die signifikanten Eigenschaften der Objekte?
- F) Welche signifikanten Umgebungsinformationen würden Sie sammeln?

Markieren Sie auf dieser Liste:

- [\*]** Informationen, die nur für bestimmte Objekte gültig sind
- [x]** Informationen, die für alle Objekte in der Umgebung gültig sind
- [o]** Informationen, die nur einmalig gesammelt werden müssen
- [#]** Informationen, die Zeit-/Eventabhängig sind

## Zusatzaufgabe B: Sheer Curation

- A) Was sollte ein ideales Sheer-Curation-Tool in ihrem Szenario leisten?

## Zusatzaufgabe C: PET

- A) Starten Sie das PERICLES Extraction Tool
- B) Extrahieren Sie Informationen, die für das Szenario wichtig sein könnten. Untersuchen Sie dabei mindestens ein beliebiges Beispielobjekt (ohne zwingend das ganze Szenario durchzuspielen).
- C) Schauen Sie sich die extrahierten Informationen an:
  - a) Innerhalb des Tools
  - b) In den JSON Dateien
- D) Löschen Sie alle Informationen aus den JSON Dateien, die ihnen unwichtig erscheinen, oder die zu sensibel sind um sie zu versenden

## Zusatzaufgabe C: PeriCAT

- A) Starten Sie PeriCAT
- B) Füllen Sie den PeriCAT Fragebogen aus, sodass er zu ihrem Szenario passt.
- C) Welcher Algorithmus wird empfohlen?
- D) Speichern Sie das Szenario ab.
- E) Verkapseln Sie die extrahierten JSON Dateien von PET zusammen mit dem untersuten Objekt mit dem empfohlenen Algorithmus
- F) Entkapseln Sie die verkapselten Informationen wieder
- G) Schauen Sie sich die Ergebnisse von Verkapselung und Entkapselung an. Was fällt ihnen auf?