

## Forschungsdatenmanagement: Übersicht Erfüllbarkeit der Anforderungen des Leistungskatalogs

erstellt am: 11.09.2018	von: Beate Rajski (TUHH) Kai Wörner (UHH)
geändert am: 11.09.2018	von: Beate Rajski (TUHH)
Version: 1.0	
Status: Abgenommen	

### Übersicht:

<b>Programm:</b>	Hamburg Open Science
<b>Projekt:</b>	Forschungsdatenmanagement
<b>Projektlaufzeit:</b>	2018
<b>M4:</b>	Konzeption und Umsetzung: 2. Iteration
<b>Z3:</b>	Übersicht Erfüllbarkeit der Anforderungen des Leistungskatalogs
<b>Auftraggeber:</b>	Prof. Dr.-Ing. Timm-Giel, Vizepräsident Forschung TUHH
<b>Projektleitung:</b>	Beate Rajski, TUHH

DOI: <https://doi.org/10.5281/zenodo.2386948>



# Übersicht Erfüllbarkeit der Anforderungen des Leistungskatalogs

## Inhalt

Einleitung.....	2
Kurzbeschreibung der Softwarelösungen .....	3
DSpace-CRIS.....	3
Invenio / Zenodo .....	3
Anforderungen nach den FAIR-Data-Prinzipien .....	3
Auffindbarkeit .....	4
Zugänglichkeit .....	7
Interoperabilität .....	10
Nachnutzbarkeit .....	11
Anhang 1: Technical Specifications DSpace .....	13

## Einleitung

Die Anforderungen an ein Forschungsdatenrepositorium (FDR) im Rahmen von Hamburg Open Science sind in der Leistungsbeschreibung vom 31.05.18 festgelegt worden.

Parallel dazu wurden zwei Prototypen als Softwarealternativen für ein institutionsspezifisches „Long Tail“<sup>1</sup>-Forschungsdatenrepositorium aufgebaut:

- An der TUHH ein Repository auf Basis von DSpace-CRIS Version 5.8 (FDR TUHH)
- An der UHH ein Repository auf Basis von Zenodo/Invenio (FDR UHH)

In diesem Dokument wird beschrieben, wie die Anforderungen aus der Leistungsbeschreibung von den beiden Systemen abgebildet werden können.

Dieses Dokument soll den an Hamburg Open Science beteiligten Einrichtungen als Entscheidungsgrundlage für die Auswahl einer Softwarealternative dienen.

---

<sup>1</sup> „Long Tail“- Forschungsdaten weisen in der Regel ein geringes Datenvolumen auf, liegen in verschiedenen Datenformaten vor und sind in Bezug auf Archivierung und Management nur schwer standardisierbar, wodurch häufig keine domänenspezifische Lösung existiert.

## Kurzbeschreibung der Softwarelösungen

### DSpace-CRIS

DSpace ist ein Open-Source-Softwarepaket für den Betrieb von Repositorien. Ziel ist die Bereitstellung digitaler Inhalte für Endnutzer und die Bereitstellung von Werkzeugen für die Verwaltung und Erhaltung von Inhalten innerhalb der Anwendung (s.a. Anhang 1: Technical Specifications DSpace). DSpace wurde ursprünglich am MIT entwickelt und ist die am weitesten verbreitete Repository-Softwareplattform (Open Source oder proprietär), mit mehr als 2.000 Installationen weltweit. Die DSpace-Entwicklung wird über die gemeinnützige DuraSpace koordiniert. Die TU Hamburg ist dort über das deutsche DSpace-Konsortium in der DSpace-Leadership Group vertreten.

Für den Prototyp wird DSpace mit der Erweiterung DSpace-CRIS (zur Zeit Version 5.8) genutzt. Dies erweitert die eigentliche Repository Software um Komponenten eines Forschungsinformationssystems, so dass z.B. Autorenprofile, Projektinformationen und Organisationseinheiten die Forschungsdaten ergänzen können. Die Forschungsdaten werden in einer eigenen „Sammlung“ abgebildet, an die eigene Metadatendefinitionen und Bearbeitungsworkflows gekoppelt werden können. Dadurch ist eine Abgrenzung von anderen Aufgaben wie Open Access Repository oder Forschungsinformationssystem möglich.

Zurzeit wird DSpace 7 entwickelt. DSpace 7 wird eine völlig neue Oberfläche und REST-Schnittstelle erhalten. Bei lokalen Anpassungen ist daher mit einem hohen Aufwand für die Übernahme zu rechnen. Die Migration von DSpace 5.X/6.X wird voraussichtlich 2020 möglich sein.

### Invenio / Zenodo

Zenodo ist ein kostenloser Repositoriendienst, der am CERN in Genf betrieben und entwickelt wird. Er basiert auf dem Open-Source Framework Invenio, das ebenso am CERN entwickelt und koordiniert wird. Für den Prototyp des FD-Repositoriums wird die Codebasis von Zenodo verwendet und an die Bedürfnisse für den Betrieb an der Universität Hamburg angepasst. Das Team des Zentrums für nachhaltiges Forschungsdatenmanagement ist aktiv an der Entwicklung einer Invenio RDM (Research Data Management)-Version beteiligt, mit der die Zenodo-Funktionalität leichter an individuelle Bedarfe angepasst werden kann. Technische Spezifikationen und weitere Informationen finden sich unter <https://invenio.readthedocs.io/en/latest/> bzw. <https://zenodo.readthedocs.io/en/latest/index.html>.

## Anforderungen nach den FAIR-Data-Prinzipien

Die “FAIR Data Principles” formulieren Grundsätze, die nachhaltig nachnutzbare Forschungsdaten erfüllen müssen und die Forschungsdateninfrastrukturen dementsprechend im Rahmen der von ihnen angebotenen Services implementieren sollten.

Gemäß der FAIR-DATA-Prinzipien sollen Forschungsdaten auffindbar (**F**indable), zugänglich (**A**ccessible), interoperabel (**I**nteroperable) und nachnutzbar (**R**eusable) sein. Für ein Forschungsdatenrepositorium ergeben sich hieraus die folgenden Anforderungen:

## Auffindbarkeit

Forschungsdaten werden durch Metadaten beschrieben. Damit wird sichergestellt, dass die Forschungsdaten über die zugehörigen Metadaten sowohl von Menschen als auch von Computersystemen leicht auffindbar sind und die Entdeckung interessanter Datensätze unterstützt wird.

1. Für jede Version eines (Meta)Datensatzes wird ein global eindeutiger und dauerhaft persistenter Identifier vergeben.

- Für veröffentlichte (Meta)Daten wird ein Digital Object Identifier (DOI) über DataCite, die Registrierungsagentur für Forschungsdaten, registriert.

FDR TUHH	FDR UHH
Folgende persistente Identifier werden bei Veröffentlichung eines Datensatzes vergeben: <ul style="list-style-type: none"><li>• DOI über DataCite</li><li>• Handle als DSpace-Standard</li><li>• URN für die Deutsche Nationalbibliothek (DNB).</li></ul>	Bei der Veröffentlichung eines Datensatzes wird eine DOI über DataCite vergeben.

- Der persistente Identifier kann bereits vor der Veröffentlichung des Datensatzes reserviert werden.

FDR TUHH	FDR UHH
Bei der Eingabe der Metadaten werden die persistenten Identifier bereits reserviert und für die weitere Verwendung angezeigt.	Bei der Eingabe der Metadaten kann die DOI durch Anklicken einer entsprechenden Schaltfläche reserviert und angezeigt werden.

2. Daten werden mit umfangreichen Metadaten beschrieben. Das DataCite Metadata Schema<sup>2</sup> dient als Referenz für die Erfassung.

- Folgende Felder sind verpflichtend: Identifier, Creator, Title, Publisher, PublicationYear, ResourceType

FDR TUHH	FDR UHH
Die Felder sind vorhanden und als Pflichtfelder deklariert. Zu Autoren ist die Ergänzung der authentifizierten ORCID iD möglich.	Die Felder sind vorhanden und als Pflichtfelder deklariert. Zu Autoren ist die Ergänzung der einer ORCID iD möglich.

- Folgende Felder sind optional: Subject, Contributor, Date, Language, AlternateIdentifier, RelatedIdentifier, Size, Format, Version, Rights, Description, GeoLocation, FundingReference

---

<sup>2</sup> DataCite Metadata Schema 4.1: <https://doi.org/10.5438/0014>

FDR TUHH	FDR UHH
Die Felder sind entsprechend dem DataCite-Metadaten-Schema abgebildet. Ausnahme: GeoLocation	Die Felder sind entsprechend dem DataCite-Metadaten-Schema abgebildet; GeoLocation wird im ersten Quartal 2019 ergänzt.

- Optionale Felder können als Pflichtfelder deklariert werden.

FDR TUHH	FDR UHH
Jedes Feld kann als verpflichtend oder nicht verpflichtend deklariert werden.	Jedes Feld kann als verpflichtend oder nicht verpflichtend deklariert werden.

- Das Metadaten-Schema ist erweiterbar.

FDR TUHH	FDR UHH
Das Metadaten-Schema ist beliebig erweiterbar. Vorhandene Datensätze können über sogenannte Curation Tasks an veränderte Metadaten-Schemata angepasst werden.	Das Metadaten-Schema ist beliebig erweiterbar.

### 3. Metadaten werden in einem durchsuchbaren Verzeichnis registriert oder indiziert.

- Für die Eingabe der Metadaten stehen intuitiv bedienbare Webformulare zur Verfügung.

FDR TUHH	FDR UHH
Die Eingabeformulare sind frei konfigurierbar und werden in Deutsch und Englisch angeboten. Weitere Sprachen sind möglich. Es stehen verschiedene Feldtypen für die Konfiguration zur Verfügung. Vorbelegungen und Auswahllisten können hinterlegt werden. Zu jedem Feld kann ein Hinweis angezeigt werden. Bei der Speicherung werden über fehlenden Pflichtfeldern Hinweistexte angezeigt. Eine Hilfedatei steht zur Verfügung. Für den Freischaltungsworkflow ist es möglich, das Formular um weitere Metadatenfelder zu ergänzen. Außerdem gibt es eine Administrations-sicht, in der jedes definierte Metadatum geändert, ergänzt oder gelöscht werden kann.	Für die Eingabe der Metadaten stehen intuitiv bedienbare und umfangreich getestete Webformulare zur Verfügung. Die Formulare stehen auf Englisch zur Verfügung, können aber über ein i18n-Modul in beliebig viele weitere Sprachen übersetzt werden.

- Alternativ können Metadaten über festgelegte Dateiformate importiert werden.

FDR TUHH	FDR UHH
<p>Die Formulare können aus Dateien mit Metadaten z.B. einer BibTeX-Datei vorbelegt werden. Über das Protokoll SWORD ist die automatisierte Eingabe von Metadaten und Dateien möglich, die in den Veröffentlichungsworkflow eingespeist werden.</p>	<p>Metadaten können über verschiedene Schnittstellen (OAI, SWORD V3) und über eine API importiert werden. Perspektivisch ist geplant, Daten über diese Schnittstellen automatisiert in das Repositorium zu übernehmen.</p>

- Die Metadaten werden strukturiert und maschinenlesbar gespeichert und sind dadurch indexierbar.

FDR TUHH	FDR UHH
<p>Die Metadaten werden strukturiert abgelegt und gespeichert und können entsprechend differenziert gesucht werden. Die Verwendung für Facetten ist möglich. Die Daten können in verschiedenen Formaten für unterschiedliche Harvester / Suchmaschinen bereitgestellt werden. Bei Bedarf können sie z.B. über Curation Tasks an verschiedene Ausgabeanforderungen angepasst werden.</p>	<p>Die Metadaten werden strukturiert abgelegt und gespeichert und können entsprechend differenziert gesucht werden. Für das Browsing werden bestimmte Facetten automatisch erzeugt. Die Daten werden für Harvester / Suchmaschinen bereitgestellt.</p>

- Sie werden über Schnittstellen für weitere Suchmaschinenindizes bereitgestellt.

FDR TUHH	FDR UHH
<p>Die Metadaten werden bei der DOI-Registrierung an DataCite übergeben und dort indiziert. Über die OAI-PMH-Schnittstelle stehen die Daten außerdem allen interessierten Datenbank Anbietern zum Harvesting zur Verfügung. Das Repositorium ist OpenAIRE 3.0 (OA, funding) kompatibel. Zur Suche im Linked-Data-Datenbestand kann ein SPARQL-Interface angeboten werden. Die (Meta-)Daten werden in den Solr-Index des lokalen Katalogs (tub.find) integriert und auch darüber durchsuchbar gemacht. Für die DNB wird X-Metadiss-Plus bereitgestellt. Die Landing-Pages werden schema.org Metadaten für die Indizierung durch Suchmaschinen wie Google Data enthalten.</p>	<p>Die Metadaten werden bei der DOI-Registrierung an DataCite übergeben und dort indiziert. Über die OAI-PMH-Schnittstelle stehen die Daten außerdem allen interessierten Datenbank Anbietern zum Harvesting zur Verfügung. Das Repositorium wird über re3data bei OpenAIRE gemeldet, um EU-geförderte Projekte automatisch an OpenAIRE zu melden. Die Landing-Pages enthalten schema.org Metadaten für die Indizierung durch Suchmaschinen wie Google Scholar.</p>

4. Metadaten enthalten klar und eindeutig den Identifier, der die Daten referenziert.

FDR TUHH	FDR UHH
Alle persistenten Identifier sind in den Metadaten enthalten. Die Forschungsdaten selbst werden über einen eindeutigen URL referenziert.	Die DOI verweist auf die Metadaten, die Referenzen auf die eigentlichen Daten ist in den Metadaten enthalten, um bei Bedarf eine Löschung der Daten ohne Verlust der DOI-Referenz zu ermöglichen.

## Zugänglichkeit

Daten und Metadaten sollten langzeitarchiviert und so verfügbar gemacht werden, dass sie mit Hilfe von Standard-Kommunikationsprotokollen leicht von Maschinen und Menschen abgerufen und heruntergeladen oder lokal genutzt werden können.

1. (Meta)Daten sind über ihren Identifier mithilfe eines standardisierten Kommunikationsprotokolls auffindbar.

- Es werden ausschließlich offene Protokolle wie http(s), ftp oder smtp genutzt.

FDR TUHH	FDR UHH
Der Zugriff auf Daten und Metadaten erfolgt über https. Die Kommunikation zum Zugriff auf geschützte Daten erfolgt über Mail (smtp).	Der Zugriff auf Daten und Metadaten erfolgt über https. Die Kommunikation zum Zugriff auf geschützte Daten erfolgt über Mail (smtp).

- Es wird sichergestellt, dass der Identifier auf die Landing Page des Forschungsdatums auflöst.

FDR TUHH	FDR UHH
Bei der Registrierung von DOIs bei DataCite und URNs bei der DNB werden die URLs zu den Metadaten mit dem Identifier verknüpft. Bei Änderungen der URLs werden diese an die Registrierungsagenturen weitergegeben.	Bei der Registrierung von DOIs bei DataCite und URNs bei der DNB werden die URLs zu den Metadaten mit dem Identifier verknüpft. Bei Änderungen der URLs werden diese an die Registrierungsagenturen weitergegeben.

- Bei Zugriffsbeschränkungen auf Forschungsdaten werden die Zugangsregeln in den Metadaten und auf der Landing Page des Forschungsdatums hinterlegt.

FDR TUHH	FDR UHH
Embargodaten werden auf der Landingpage angezeigt. Für Anfragen zum Zugriff steht ein Webformular zur Verfügung, über das die Anfrage an den Einstellenden per Mail geschickt wird.	Embargodaten werden auf der Landingpage angezeigt. Für Anfragen zum Zugriff steht ein Webformular zur Verfügung, über das die Anfrage an den Einstellenden per Mail geschickt wird.

2. Das Repositorium bietet ein Rollen- und Rechtemanagement an, welches die Authentifizierung und Autorisierung von Benutzerinnen und Benutzern, einschließlich maschineller Zugriffe, unterstützt.

- **Authentifizierung:** Die Authentifizierung für Hochschulangehörige erfolgt identitätsgesichert über Shibboleth oder LDAP. Für Administration, maschinelle Zugriffe und externe Nutzende können lokale Accounts angelegt werden. Eine zusätzliche O-Auth-Authentifizierungsmöglichkeit über ORCID<sup>3</sup> ist optional.

FDR TUHH	FDR UHH
Das Login erfolgt für TU-Angehörige über Shibboleth, für Externe über lokale Accounts. Alternative Authentifizierungsverfahren wie LDAP, OAuth über ORCID oder Zertifikate können konfiguriert werden. Die TU Hamburg nutzt für Shibboleth die Verlässlichkeitsklasse „Advanced“. Deshalb ist die Identität der Anmeldenden bereits im Vorfeld sichergestellt.	Das Login erfolgt für UHH-Angehörige über Shibboleth. Ein externer, autorisierter Zugriff ist nicht vorgesehen, könnte aber über lokale Accounts erfolgen. Alternative Authentifizierungsverfahren können konfiguriert werden. Die UHH nutzt für Shibboleth die Verlässlichkeitsklasse „Advanced“. Deshalb ist die Identität der Anmeldenden bereits im Vorfeld sichergestellt.

- **Autorisierung:** Forschungsdaten können, wenn notwendig, mit einem Zugriffsschutz versehen werden. Die Metadaten sind immer frei zugänglich. Der Zugriff kann durch den Einstellenden auf ausgewählte, autorisierte Konten oder Gruppen beschränkt werden. Für Interessierte wird die Möglichkeit der Kontaktaufnahme mit dem Einstellenden angeboten.

FDR TUHH	FDR UHH
Der Hochladende kann die Daten mit einem Zugriffsschutz bzw. einem Embargo versehen. Über einen „Request-Button“ kann der Zugang zu den Daten per Mail angefragt werden. Der Einstellende und seine Adresse bleiben dabei nach außen hin anonym. In der Administration kann ein dezidiertes Rollen- und Rechtemanagement eingerichtet werden. Dies kann auch für den Zugriff auf Daten genutzt werden. Die Einrichtung ist auf die Administratoren beschränkt und in der aktuellen DSpace-Version relativ aufwendig.	Der Hochladende kann die Daten mit einem Zugriffsschutz bzw. einem Embargo versehen. Über einen „Request-Button“ kann der Zugang zu den Daten per Mail angefragt werden. Der Einstellende und seine Adresse bleiben dabei nach außen hin anonym. In der Administration kann ein dezidiertes Rollen- und Rechtemanagement eingerichtet werden. Dieses ist aber bislang für den Betrieb nicht vorgesehen.

---

3 ORCID: Open Researcher and Contributor ID: <https://orcid.org>



### 3. Langzeitarchivierung und Löschung

- Forschungsdaten können gelöscht werden. Die Metadaten bleiben nach der Löschung dauerhaft über ihren persistenten Identifier verfügbar.

FDR TUHH	FDR UHH
Die Einstellenden haben nach der Freigabe der Daten keine Änderungs- oder Löschrerechtigung. Dadurch wird sichergestellt, dass der Datensatz nicht gelöscht wird. Die Forschungsdaten können, wenn notwendig durch berechtigtes Personal gelöscht werden. Sollten die Metadaten gelöscht werden müssen, wird bei DataCite der entsprechende Status gesetzt. Forschungsdaten, die bei der DNB abgeliefert werden, sind dort ohne Löschmöglichkeit langzeitarchiviert.	Die Einstellenden haben nach der Freigabe der Daten keine Änderungs- oder Löschrerechtigung, worauf sie auch hingewiesen werden. Die Löschung von Daten kann nur durch die Administratoren auf Anfrage erfolgen, die Metadaten und damit die Landingpage bleiben dann erhalten. Sollten die Metadaten gelöscht werden müssen, wird bei DataCite der entsprechende Status gesetzt.

- Für die Langzeitarchivierung der Forschungsdaten können S3-kompatible Speicher angebunden werden.

FDR TUHH	FDR UHH
DSPACE kann an S3-kompatible Speicher angebunden werden. Der Proof of Concept für das Repositorium der TUHH kann erst zu einem späteren Zeitpunkt mit dem LZA des RRZ erfolgen.	Invenio kann durch ein entsprechendes Modul an S3-kompatible Speicher angebunden werden. Entsprechende Tests für das FDR der UHH haben stattgefunden.

- Für die Langzeitarchivierung von Forschungsdaten zu Dissertationen einschließlich der beschreibenden Metadaten ist die Ablieferung als OAIS-kompatibles Datenpaket an die Deutsche Nationalbibliothek<sup>4</sup> optional möglich.

FDR TUHH	FDR UHH
Das Projekt eDissPlus ist noch bei den Projektpartnern in der Erprobungsphase. Es ist im Repositorium der TUHH aber grundsätzlich möglich, Forschungsdaten mit TUHH-Dissertationen zu verknüpfen und entsprechende Sets für das Harvesting durch die DNB bereitzustellen.	Bislang ist keine Anbindung an eDissPlus vorgesehen. Es ist grundsätzlich möglich, Forschungsdaten im Repositorium mit Dissertationen der UHH zu verknüpfen und für das Harvesting bereitzustellen.

<sup>4</sup> s.a. eDissPLUS: Archivierung und Veröffentlichung von elektronischen Dissertationen mit Forschungsdaten <http://gepris.dfg.de/gepris/projekt/280718174>

## Interoperabilität

Die Daten sollten so vorliegen, dass sie ausgetauscht, interpretiert und (semi-)automatisiert mit anderen Datensätzen von Menschen und Computersystemen kombiniert werden können.

1. Metadaten nutzen eine formale, zugängliche, gemeinsam genutzte und breit anwendbare Sprache

- Für die Metadaten werden die disziplinübergreifenden Standards von DataCite und OpenAIRE<sup>5</sup> zugrunde gelegt.

FDR TUHH	FDR UHH
Die Metadaten schemata von DataCite (V4.1) und OpenAIRE (V3.0) sind implementiert und die Metadaten werden in diesen Formaten für die Nachnutzung bereitgestellt.	Die Metadaten schemata von DataCite (V4.0, 4.1 ab Q1/2019) und OpenAIRE (V3.0) sind implementiert und die Metadaten werden in diesen Formaten für die Nachnutzung bereitgestellt.

- Die Erweiterung des Metadaten schemas um disziplinspezifische Metadaten ist optional.

FDR TUHH	FDR UHH
Das Repositorium der TU Hamburg hat einen institutionellen und keinen fachlichen Fokus. Eine Einbindung fachspezifischen Vokabulars ist deshalb zum jetzigen Zeitpunkt nicht vorgesehen. Das Metadaten schema kann bei Bedarf aber dahingehend erweitert werden.	Das Repositorium der UHH hat einen institutionellen und keinen fachlichen Fokus. Ob zu einem späteren Zeitpunkt Communities mit erweiterten Metadaten (für Fakultäten und Forschungsbereiche) eingerichtet werden, hängt vom entsprechenden Bedarf ab.

- Die Metadaten lassen sich in verschiedenen Formaten wie z.B. JSON LD abrufen.

FDR TUHH	FDR UHH
Eine Ausgabe der Metadaten ist derzeit nur über XML über OAI-PMH möglich. Mit der Linked Data-Schnittstelle von DSpace können weitere Formate wie Turtle oder N-Triples bereitgestellt werden. Dies ist derzeit aber noch nicht im Einsatz. Auf der Landing Page wird ein BibTeX-Export angeboten.	Die Metadaten lassen sich auf der Landingpage in den Formaten BibTeX, CSL, DataCite, Dublin Core, JSON, JSON-LD, und MARCXML abrufen sowie direkt nach Mendeley exportieren.

<sup>5</sup> OpenAIRE Guidelines for Data Archives: <https://guidelines.openaire.eu/en/latest/data/index.html>

- Das Metadatenschema ermöglicht die Referenzierung von anderen Metadatensätzen über entsprechende Metadatenfelder und die Eingabe des zugehörigen Persistenten Identifiers.

FDR TUHH	FDR UHH
<p>Die Referenzierung ist über das DataCite-Feld RelatedIdentifier mit den Attributen relatedIdentifierType und relationType möglich. Für die beiden Attribute wird kontrolliertes Vokabular genutzt. Beziehungen zwischen Datensätzen erfolgen über persistente Identifier, in der Regel ist dies der DOI. Datensätze können im Repository versioniert werden. Die Verknüpfung über die Metadatenfelder für relationType mit „IsNewVersionOf“ und „IsPreviousVersionOf“ erfolgt dabei automatisch. Weitere Beziehungen können manuell über die Eingabemaske eingetragen werden.</p>	<p>Die Referenzierung ist über das DataCite-Feld RelatedIdentifier mit den Attributen relatedIdentifierType und relationType möglich. Für die beiden Attribute wird kontrolliertes Vokabular genutzt. Beziehungen zwischen Datensätzen erfolgen über persistente Identifier, in der Regel ist dies der DOI. Datensätze können im Repository versioniert werden. Die Verknüpfung über die Metadatenfelder für relationType mit „IsNewVersionOf“ und „IsPreviousVersionOf“ erfolgt dabei automatisch. Weitere Beziehungen können manuell über die Eingabemaske eingetragen werden.</p>

## Nachnutzbarkeit

Eine gute Beschreibung von Daten und Metadaten sorgt dafür, dass die Daten für zukünftige Forschung wiederverwendet werden können und mit anderen, kompatiblen Datenquellen vergleichbar sind. Ein ordnungsgemäßes Zitieren der Daten muss ermöglicht werden, und die Bedingungen, unter denen die Daten nachgenutzt werden können, sollten für Maschinen und Menschen verständlich dargestellt sein.

- Die (Meta-)Daten enthalten, wenn notwendig, eine präzise Beschreibung der Daten und ihrer Entstehung. Der Kontext für Nachnutzung und Reproduzierbarkeit kann in den (Meta-)Daten beschrieben werden.

FDR TUHH	FDR UHH
<p>In den Metadaten Beschreibung sowie bei den Forschungsdaten selbst kann der Kontext für Nachnutzung- und Reproduzierbarkeit beschrieben werden. Eine Kontrolle, ob diese Beschreibung vorhanden bzw. notwendig ist, kann im Freischaltungsworkflow erfolgen. Als institutionelles und damit nicht disziplin- oder projektspezifisches Repository wird die Eingabe dieser Daten nur nach dem aktuellen DataCite-Schema unterstützt. Eine maschinelle Auswertung wie sie bei vielen gleichartigen Daten möglich wäre, ist deshalb nicht gegeben.</p>	<p>Im Metadatenfeld „Beschreibung“ sowie bei den Forschungsdaten selbst kann der Kontext für Nachnutzung- und Reproduzierbarkeit beschrieben werden. Eine Kontrolle, ob diese Beschreibung vorhanden bzw. notwendig ist, erfolgt zunächst nicht. Als institutionelles und damit nicht disziplin- oder projektspezifisches Repository wird die Eingabe dieser Daten nur nach dem aktuellen DataCite-Schema unterstützt. Eine maschinelle Auswertung wie sie bei vielen gleichartigen Daten möglich wäre, ist deshalb nicht gegeben.</p>

2. Die Metadaten enthalten eine eindeutige, für Menschen und Maschinen lesbare Angabe einer Lizenz für die Datennutzung.

FDR TUHH	FDR UHH
Creative Commons Lizenzen können ausgewählt werden und werden in den Metadaten maschinenlesbar mit einem Link auf die Lizenz gespeichert. Die Lizenzauswahl kann erweitert werden, wenn sich im Betrieb weitere Bedarfe ergeben.	Creative Commons Lizenzen können ausgewählt werden und werden in den Metadaten maschinenlesbar mit einem Link auf die Lizenz gespeichert. Die Lizenzauswahl kann erweitert werden, wenn sich im Betrieb weitere Bedarfe ergeben.

3. Die Beziehung zu verwendeten Datensätzen kann abgebildet werden.

FDR TUHH	FDR UHH
Die Referenzierung ist über das DataCite-Feld RelatedIdentifier mit den Attributen relatedIdentifierType und relationType möglich. Für die beiden Attribute wird kontrolliertes Vokabular genutzt. Beziehungen zwischen Datensätzen erfolgen über persistente Identifier, in der Regel ist dies der DOI. Datensätze können im Repository versioniert werden. Die Verknüpfung über die Metadatenfelder für relationType mit „IsNewVersionOf“ und „IsPreviousVersionOf“ erfolgt dabei automatisch. Weitere Beziehungen können manuell über die Eingabemaske eingetragen werden.	Die Referenzierung ist über das DataCite-Feld RelatedIdentifier mit den Attributen relatedIdentifierType und relationType möglich. Für die beiden Attribute wird kontrolliertes Vokabular genutzt. Beziehungen zwischen Datensätzen erfolgen über persistente Identifier, in der Regel ist dies der DOI. Datensätze können im Repository versioniert werden. Die Verknüpfung über die Metadatenfelder für relationType mit „IsNewVersionOf“ und „IsPreviousVersionOf“ erfolgt dabei automatisch. Weitere Beziehungen können manuell über die Eingabemaske eingetragen werden.

## Anhang 1: Technical Specifications DSpace

CC-BY 4.0 DSpace <https://duraspace.org/dspace/resources/technical-specifications/>

**Description:** DSpace is an out-of-the-box open source software package for creating repositories focused on delivering digital content to end users and providing a full set of tools for managing and preserving content within the application. DSpace is the most widely used repository software platform (open source or proprietary), with more than 2,000 installations worldwide representing a continuously growing and active user community.

**History:** DSpace was originally developed by MIT Libraries and Hewlett-Packard (HP) Labs. Since its initial open source release in 2002, the platform has been guided by a global community of committers, developers, repository managers, and other stakeholders who contribute to project governance. DSpace became a DuraSpace project in 2009 when the Fedora Commons and DSpace organizations merged to form DuraSpace.

**Cost:** Open source software, no charge. DSpace is distributed under the terms of the BSD open source license.

**Use case highlight:** The most common use of the DSpace software is by academic and research libraries as an open access institutional repository for managing and disseminating scholarly output. There are also many organizations using the software to host and manage subject based, dataset, or media-based repositories. See examples here: <https://duraspace.org/dspace/features/use-cases> .

<b>Architectural Overview</b>	DSpace is a set of cooperating Java web applications and utility programs that maintain an asset store and an associated metadata store. The web applications provide interfaces for administration, deposit, ingest, search, and access. The asset store is maintained on a file system or similar storage system. The metadata, including access and configuration information, is stored in a relational database.
<b>Service Providers</b>	DSpace has a global, active network of registered service providers who provided commercial support, hosting, training or site customization. <a href="https://duraspace.org/dspace/resources/service-providers/">https://duraspace.org/dspace/resources/service-providers/</a>
<b>Technical Aspects</b>	<b>Operating System:</b> Written in Java, tested under Linux, Windows, and Mac OSX  <b>License:</b> BSD  <b>Release version:</b> 6.0, <a href="https://duraspace.org/dspace/download/">https://duraspace.org/dspace/download/</a>  <b>Documentation:</b> <a href="https://wiki.duraspace.org/display/DSDOC/">https://wiki.duraspace.org/display/DSDOC/</a>

**Other prerequisite software:** Java 7 or 8, Apache Maven, Apache Ant, Relational Database - PostgreSQL or Oracle, Servlet 3.0 container (Tomcat 7+ or similar).

## Key Features

**Application Architecture:** DSpace is a full stack web application, consisting of a database, storage manager and front end web interface. The architecture includes a specific data model with configurable metadata schemas, workflows and browse/search functionality.

**Application Architecture:** DSpace is a full stack web application, consisting of a database, storage manager and front end web interface. The architecture includes a specific data model with configurable metadata schemas, workflows and browse/search functionality.

**Modern, RESTful Web UI (*coming soon*):** DSpace 7.0 will feature a completely rewritten web user interface based on the Angular 2 javascript platform.

**Built-in workflows:** Originally designed for libraries, the embedded DSpace data model and approval workflows are familiar to librarians and archivists.

**Built-in search engine:** DSpace comes packaged with Apache Solr, an open source enterprise search platform that enables filtered (faceted) searching and browsing of all objects. The full text of common file formats is searchable, along with all metadata fields. Browse by interfaces are also configurable.

**Unlimited File types:** DSpace can store any type of file. In addition, it auto-recognizes files of most common formats (e.g., DOC, PDF, XLS, PPT, JPEG, MPEG, TIFF).

**Metadata:** By default, DSpace uses a Qualified Dublin Core (QDC) based metadata schema. Institutions can extend that base schema or add custom QDC-like schemas. DSpace can import or export metadata from other major metadata schemas such as MARC or MODS.

**Tools/plugins:** DSpace comes with a suite of tools (batch ingest, batch export, batch metadata editing, etc.) and plugins for translating content into DSpace objects. Additionally, commercial plugins are available through service providers.

**Security:** DSpace provides its own built-in authentication / authorization system, but can also integrate with existing authentication systems such as LDAP or Shibboleth.

**Permissions:** DSpace allows you to control read/write permissions site-wide, per community, per collection, per item and per file. You may also delegate administrative permissions per community or per collection.

**Disaster Recovery:** DSpace allows you to export all of your system content as AIP (Archival Information Packages) backup files. These AIPs can be used to restore your entire site, or restore individual communities, collections or items.

**OAI-PMH / SWORD (v1 and v2) / OpenAIRE /**

**Driver:** DSpace complies with standard protocols and best practices for access, ingest, and export.

**REST:** DSpace provides RESTful APIs in accordance with modern web standards.

**Configurable Database:** Organizations can choose either PostgreSQL or Oracle for the database in which DSpace manages items and metadata.

**Configurable File Storage:** Files in DSpace can be stored either using a local filesystem (default) or a cloud-based solution, such as Amazon S3.

**Data Integrity:** On upload, DSpace calculates and stores a checksum for each file. Optionally, you may ask DSpace to verify those checksums to validate file integrity.

**Languages:** DSpace is available in over 20 languages.

**Software Development**

Duraspace hosts a group of leading-edge open source projects characterized by thriving, consensus-driven, developer communities, including governance and membership, that strive to produce high quality products that insure sophisticated access and management of durable digital information.