

Beschreibung des Online-Dienstes RADAR

Ziele

RADAR (Research Data Repository) ist ein disziplinübergreifendes Forschungsdatenrepositorium für die Archivierung und Publikation von Forschungsdaten abgeschlossener wissenschaftlicher Studien und Projekte. Forschungsdaten versteht RADAR als digitale Daten, die je nach Fachkontext Gegenstand eines Forschungsprozesses sind, während eines solchen Prozesses entstehen oder sein Ergebnis sind. RADAR möchte das in vielen wissenschaftlichen Disziplinen noch fehlende Forschungsdatenmanagement fördern und einen wichtigen Beitrag zu einer besseren Verfügbarkeit, nachhaltigen Bewahrung und (eigenständigen) Publikationsfähigkeit von Forschungsdaten leisten.

RADAR speichert Forschungsdaten grundsätzlich in Form von paketierte Zusammenstellungen einer oder mehrerer Dateien. Diese Zusammenstellung enthält die eigentlichen Forschungsdaten sowie beschreibende Daten (Metadaten). Im Folgenden wird eine solche Zusammenstellung als „Datenpaket“ bezeichnet.

Betreiber

RADAR wird von FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur angeboten („Betreiber“). Der Dienst richtet sich primär an Hochschulen und außeruniversitäre Forschungseinrichtungen („Kunden“), die ihren Forschenden („Datengebern“) die Möglichkeit zur Datenarchivierung und -publikation anbieten wollen. In RADAR gespeicherte Daten können von Dritten („Datennutzern“) durchsucht und heruntergeladen werden, sofern die Datengeber dem zugestimmt haben. Der Dienst richtet sich zurzeit nicht an Kunden aus der Wirtschaft.

Registrierung und Benutzerprofile

RADAR ist ein Online-Dienst und kann ausschließlich über das Internet genutzt werden. Jedermann kann sich bei RADAR online registrieren und dadurch eine Nutzerkennung erhalten. Die Registrierung erfragt eine frei wählbare Benutzerkennung und eine gültige E-Mailadresse. An die E-Mailadresse verschickt RADAR eine Bestätigungsnachricht, die einen Link zum Abschluss der Registrierung enthält. Gegenüber einer anonymen Nutzung der RADAR-Plattform hat ein registrierter Benutzer zunächst keine erweiterten Rechte. Diese müssen ihm erst andere Benutzer zuweisen. Optional kann der Benutzer bei der Registrierung zusätzlich Vor- und Nachname sowie seine ORCID-ID angeben. Bis auf die E-Mailadresse und ORCID-ID werden die eingegebenen Daten gegebenenfalls auch anderen Benutzern der RADAR-Plattform angezeigt. Der Benutzer kann die eingegebenen Daten (Benutzerprofil) mit Ausnahme der Benutzerkennung jederzeit ändern.

Rollen und Rechte

Kunden können bei Abschluss eines Vertrags eine vorab von ihnen angelegte Nutzerkennung benennen, der vom Betreiber daraufhin mit administrativen Rechten für diesen Vertrag versehen wird („Administrator“). Der Kunde kann über diese Nutzerkennung für verschiedene Nutzergruppen (z.B. Forschungsgruppen, Projekte, Institute) voneinander getrennte Arbeitsbereiche einrichten. Folgend kann der Administrator weitere bei RADAR registrierte Nutzerkennungen über den Online-Dienst mit Rechten als Datengeber für einen oder mehrere dieser Arbeitsbereiche versehen

(„Kurator“). Üblicherweise wird der Kunde dabei Mitarbeiter/innen der eigenen Institution als Kurator bestimmen; bei Bedarf kann er jedoch auch institutionsfremde Personen entsprechend berechtigen. Die Verantwortung für die korrekte Rechtevergabe liegt ausschließlich beim Kunden. RADAR stellt sicher, dass Kuratoren ausschließlich in die vom Administrator bestimmten Arbeitsbereiche Forschungsdaten hochladen, bearbeiten, archivieren und gegebenenfalls publizieren können. Vom Kunden bestimmte Kuratoren können wiederum andere, bei RADAR registrierte Nutzerkennungen als Subkuratoren für ihren Arbeitsbereich berechtigen. Subkuratoren sind ebenfalls Datengeber, haben jedoch nicht die Möglichkeit, Datenpakete zu archivieren bzw. zu publizieren oder andere Nutzer als Subkuratoren zu bestimmen.

Dienstleistungen

RADAR kann über eine webbasierte Benutzungsoberfläche mit aktuellen Webbrowsern oder über eine REST-basierte Programmierschnittstelle (API) genutzt werden. Der Datengeber kann innerhalb des ihm zugewiesenen Arbeitsbereichs Datenpakete anlegen und diesem einzelne Dateien oder ZIP-Dateien mit mehreren enthaltenen Dateien zuordnen, die er über das Internet zu RADAR transferiert. Über die Benutzungsoberfläche kann er einzelne Daten hinzufügen oder löschen.

Der Datengeber kann Datenpakete mit Metadaten beschreiben. Dazu stellt der Dienst auf der RADAR-Plattform ein Formular bereit. Alternativ ist es möglich, die Metadaten offline als XML-Datei zu erstellen und dann auf die RADAR-Plattform hochzuladen. Datengeber können über die RADAR-Plattform eine Vorlage für eine entsprechende XML-Datei sowie ein XML Schema zum Validieren der Metadaten in der jeweils aktuellsten Version herunterladen. Der Administrator hat darüber hinaus die Möglichkeit, für Arbeitsbereiche Default-Werte für beschreibende Metadaten zu hinterlegen. Diese werden dann Datengebern beim Erstellen der beschreibenden Metadaten vorgeschlagen.

Ist die Zusammenstellung eines Datenpakets und dessen Beschreibung mit Metadaten abgeschlossen, kann der Kurator zwischen zwei Optionen wählen: Archivierung oder Publikation der Forschungsdaten.

Die **Archivierung** erlaubt die optionale Beschreibung des Datenpakets mit deskriptiven Metadaten. Üblicherweise werden weder Daten noch Metadaten öffentlich zugänglich gemacht, dies kann jedoch durch den Kurator eigenständig durch entsprechende Rechtevergabe über das Online-System verändert werden. Der Kurator muss die gewünschte Aufbewahrungsfrist festlegen.

Für die **Publikation** eines Datenpakets sind dessen valide Beschreibung in Form deskriptiver Metadaten sowie die Vergabe einer Lizenz durch den Datengeber erforderlich. Grundsätzlich sind Metadaten und Daten öffentlich recherchier- und zugreifbar. Für die Daten kann der Datengeber optional einen Embargozeitraum bestimmen, innerhalb dessen nur die Metadaten öffentlich recherchier- und zugreifbar sind. Nach Ablauf der Embargofrist werden auch die Daten automatisch öffentlich zugreifbar. RADAR vergibt für jedes publizierte Datenpaket einen DOI-Namen und registriert diesen bei DataCite. Über diesen DOI-Namen ist das publizierte Datenpaket persistent identifizierbar, zitierfähig und kann mit einer herkömmlichen wissenschaftlichen Publikation verknüpft werden. Gleichzeitig werden die deskriptiven Metadaten an DataCite übertragen. Der Datengeber muss sie dazu zwingend unter Creative Commons Zero (CC0) 1.0 Universell lizenzieren. Für die Übertragung erfolgt eine automatische Umwandlung der deskriptiven Metadaten vom RADAR-Format in das DataCite-Format. Weiterhin werden die deskriptiven Metadaten sowohl im RADAR- als auch automatisch umgewandelt im DublinCore-Format über einen OAI Provider öffentlich zum Harvesting angeboten.

Für publizierte und archivierte Datenpakete muss eine Lizenz aus einer vorgegebenen Liste von empfohlenen Lizenzen ausgewählt werden. Die Liste umfasst alle Creative Commons 4.0 International Lizenzen (inklusive der CC0 1.0 Universell), eine „All rights reserved“-Lizenz sowie die Möglichkeit, eine proprietäre Lizenz anzugeben. Der Betreiber behält sich vor, die Liste um weitere Lizenzen zu ergänzen.

Temporärer Speicher

Für die Zusammenstellung und Beschreibung von Datenpaketen stellt der Betreiber sogenannten temporären Speicher zur Verfügung. Die Größe des insgesamt für alle Arbeitsbereiche eines Kunden verfügbaren temporären Speichers kann der Administrator selbst über den Online-Dienst festlegen und jederzeit ändern. Die durch den Administrator festgelegte Größe des temporären Speichers fließt, unabhängig von der tatsächlichen Belegung, in die Berechnung des Preises für die Nutzung des Online-Dienstes RADAR ein. Für jeden Vertrag stellt RADAR mindestens 200 GB an temporären Speicher ohne Aufpreis zur Verfügung.

Ist der insgesamt verfügbare temporäre Speicher eines Kunden gefüllt, können solange keine weiteren Daten durch die berechtigten Datengeber mehr hinzugefügt werden, bis Dateien gelöscht, Datenpakete archiviert bzw. publiziert wurden oder der temporäre Speicher durch den Kunden vergrößert wurden. Im Fall einer Archivierung oder Publikation wandern die Datenpakete aus dem temporären Speicher in den permanenten Speicher. Diese Datenpakete belegen danach im temporären Speicher keinen Speicherplatz mehr. RADAR beschränkt die Nutzung des temporären Speichers zeitlich. Die maximale Nutzungsdauer des temporären Speichers wird dabei pro Datenpaket überprüft. Datenpakete können maximal sechs Monate im temporären Speicher belassen werden. Danach müssen sie entweder archiviert, publiziert oder gelöscht werden. Einen Monat vor Ende der Speicherfrist versendet RADAR eine E-Mail mit einem Hinweis auf das baldige Ende der Speicherfrist an den Datengeber. Im wöchentlichen Rhythmus sendet RADAR daraufhin eine Erinnerung per E-Mail mit dem Hinweis auf das nahende Ende der Speicherfrist. Wurde das Datenpaket nach sechs Monaten durch den Datengeber weder gelöscht noch archiviert bzw. publiziert, löscht RADAR das Datenpaket.

Der Betreiber stellt dem Kunden im Online-System stets aktuelle Statistiken zur Nutzung von temporärem und permanentem Speicher über alle Arbeitsbereiche eines Vertrags hinweg bereit. Anhand der Statistiken kann der Kunde die Speicherbelegung bis auf Ebene einzelner Datenpakete überprüfen. Er kann den für seinen Vertrag bereitgestellten temporären Speicherplatz erweitern oder wieder verringern. Eine Vergrößerung über 200 GB hinaus ist kostenpflichtig.

Quotas

Der Kunde kann die maximal nutzbare Größe des temporären Speichers, des für archivierte Datenpakete und des für publizierte Datenpakete genutzten permanenten Speichers festlegen („Quota“). Datengeber können Daten nur so lange auf die RADAR-Plattform hochladen (temporärer Speicher) bzw. archivieren oder publizieren (permanenten Speicher), bis die jeweilige Quota überschritten ist. Der Kunde kann jederzeit die festgelegten Quotas erhöhen, was zu Mehrkosten führen kann. Alle drei Quotas gelten pro Vertrag über alle in ihm angelegten Arbeitsbereiche hinweg.

Darüber hinaus kann der Kunde Quotas nochmals pro Arbeitsbereich festlegen. Diese Quotas dürfen nicht höher sein als die jeweilige Quota auf Vertragsebene. Die Quota auf Arbeitsbereichsebenen sind keine Zusicherungen über verfügbaren Speicherplatz, sondern nur eine Obergrenze. Daher kann

pro Vertrag die Summe der Quotahöhen über alle Arbeitsbereiche hinweg größer sein als die entsprechende Quotahöhe des Vertrags.

Haltefristen und Unveränderlichkeit von Datenpaketen

RADAR ermöglicht die dauerhafte und unverfälschte Aufbewahrung von Datenpaketen über eine definierte Zeitspanne („Aufbewahrungsfrist“). Für archivierte Datenpakete legt der Datengeber die Aufbewahrungsfrist fest. Die tatsächliche Dauer der Speicherung für archivierte Datenpakete kann kürzer sein, wenn der Kunde den Vertrag mit RADAR vor Ablauf der Aufbewahrungsfrist beendet. Für publizierte Datenpakete ist die Aufbewahrungsfrist unbeschränkt. RADAR garantiert eine tatsächliche Aufbewahrung von mindestens 25 Jahren, unabhängig von der Laufzeit des Vertrags mit dem Kunden.

Während der Aufbewahrungsfrist verändert RADAR die gespeicherten Datenpakete nicht mehr, sondern sichert ausschließlich deren physischen Erhalt („bitstream preservation“). Entsprechend garantiert RADAR weder die dauerhafte Nutzbarkeit noch die Interpretierbarkeit der in einem Datenpaket enthaltenen Daten, da diese von der Verfügbarkeit der vom Datengeber gewählten Datenformate und entsprechender Programme zu deren Interpretation abhängen.

Datenpakete im permanenten Speicher können nicht mehr verändert werden. Löschungen sind in begründeten Ausnahmefällen durch den vom Kunden als Administrator bestimmten Nutzer möglich. Zu begründeten Ausnahmefällen gehören z.B. Rechtsverstöße oder fehlerhafte Daten. Bei einer Löschung werden nur die Daten gelöscht, nicht aber die Metadaten. Diese enthalten einen Hinweis, dass die Daten gelöscht wurden.

Begutachtung von Forschungsdaten

RADAR unterstützt einen Review-Prozess vor einer Datenpublikation. Dazu kann ein Datenpaket vor der Publikation in den Status „in Begutachtung“ versetzt werden. In diesem Zustand ist das Datenpaket nicht mehr editierbar. RADAR erzeugt einen eindeutigen Link, den der Datengeber an den zuständigen Verlag bzw. die Gutachter weitergeben kann. Dieser Link ermöglicht einen Zugriff auf das noch nicht veröffentlichte Datenpaket ohne vorherige Authentifizierung. Nach Abschluss der Begutachtung kann der Datengeber den Status für das Arbeitspaket entweder wieder in den Bearbeitungsmodus überführen oder aber das Datenpaket publizieren. In beiden Fällen wird der erzeugte eindeutige Link ungültig, so dass die Gutachter nicht mehr auf das Datenpaket zugreifen können. Der Datengeber kann ein Datenpaket mehrfach hintereinander in den Status „in Begutachtung“ versetzen.

Der Betreiber führt darüber hinaus keine weitere inhaltliche Qualitätssicherung der eingestellten Forschungsdaten durch.

Technische und organisatorische Maßnahmen zur Datensicherheit

Daten im temporären Speicher befinden sich auf Magnetplatten, die über einen RAID6-Verbund gegen Datenverlust durch Ausfälle einzelner Platten abgesichert sind. Zusätzlich werden die gespeicherten Daten täglich einmal inkrementell als Sicherheitskopie (Backup) auf Band geschrieben. Die Sicherheitskopie erfolgt auf Dateiebene und wird in zwei Generationen aufbewahrt.

Die Speicherung archivierter und publizierter Datenpakete erfolgt auf Magnetbändern. RADAR speichert alle Datenpakete, die in den permanenten Speicher verschoben werden, in drei Kopien an unterschiedlichen, räumlich getrennten Standorten auf unterschiedlichen Bändern. Die Speicherung

erfolgt am Steinbuch Centre for Computing (SCC) des Karlsruher Instituts für Technologie (KIT) an zwei Standorten und am Zentrum für Informationsdienste und Hochleistungsrechnen (ZIH) der TU Dresden. Datenpakete werden vor der Speicherung mit einer Prüfsumme („checksum“) versehen, die nach jedem Kopiervorgang kontrolliert wird. Damit können Fehler während der Datenübertragung entdeckt und beseitigt werden („Ende-zu-Ende-Überprüfung“). Beim Zugriff auf ein Datenpaket wird erneut die Prüfsumme berechnet und mit dem gespeicherten Wert verglichen, um eventuelle Datenkonsistenzfehler zu identifizieren. Sollte hierbei ausnahmsweise ein Fehler erkannt werden, greift RADAR auf die zweite (bzw. dritte) Kopie des Datenpakets an einem der anderen Standorte zu. Regelmäßige Überprüfungen der Kopien auf eventuelle Bitfehler („fixity checks“) finden nicht statt. Die Betreiber der Speichersysteme halten ihre jeweiligen Speicherinfrastrukturen stets auf dem aktuellen Stand der Technik. Dies führt dazu, dass in einem Zeitraum zwischen fünf bis acht Jahren alle Daten mindestens einmal auf neue Datenträger migriert werden. Spätestens im Rahmen dieser Migration findet ein Fixity Check statt, in dessen Rahmen eventuell ausgetretene Bitfehler identifiziert und korrigiert werden.

Die für die Archivierung oder Publikation vorgesehenen Datenpakete werden vor der Übertragung in den permanenten Speicher in eine der BagIt-Spezifikation entsprechende Struktur überführt, die neben den eigentlichen Forschungsdaten in ihrer Original-Anordnung mit allen Dateien und Verzeichnissen auch technische und deskriptive Metadaten sowie ein der Spezifikation entsprechendes Manifest enthält. Die BagIt-Struktur wird in einer ZIP-Datei zusammengefasst und als Archivinformationspaket (AIS) gemäß OAIS-Standard gespeichert.

Zugriffe auf den temporären Speicher erfolgen synchron, d.h. die Auslieferung der angefragten Daten startet ohne merkliche Verzögerung. Zugriffe auf den permanenten Speicher erfolgen asynchron, d.h. es können mehrere Minuten zwischen Anforderung und Auslieferung vergehen. In Zeiten hoher Zugriffszahlen kann die Wartezeit im Stundenbereich liegen. Häufig zugegriffene Datenpakete werden, auch wenn sie bereits im permanenten Speicher gespeichert sind, regelmäßig über einen Cache schnell (d.h. synchron) ausgeliefert. Eine Zusicherung über eine schnelle Auslieferung aus dem permanenten Speicher kann nicht gegeben werden.

Preise

Der Preis der Dienstleistung setzt sich aus drei Komponenten zusammen:

- einem jährlichen Grundpreis,
- der Größe des genutzten permanenten Speichers für archivierte Datenpakete pro Jahr und belegtem Gigabyte sowie
- der Größe der in der Rechnungsperiode publizierten Datenpakete (Einmalzahlung)

Im Grundpreis inkludiert sind

- temporärer Speicherplatz von 200 Gigabyte.
- die auf Seiten des Betreibers anfallenden Kosten für den Datentransfer;
- die Anlage aller notwendigen Nutzer im RADAR-System sowie die Vergabe von Rechten.

Der Kunde kann den temporären Speicherplatz bei Bedarf kostenpflichtig erweitern. Die Vergabe von DOI-Namen für publizierte Datenpakete erfolgt über die Technische Informationsbibliothek (TIB) in Hannover als Registrierungsagentur für DOI-Namen von DataCite e.V. und ist für den Datengeber kostenfrei.

Vergabe von Zugriffsrechten und Embargos

Datenpakete, die weder archiviert noch publiziert wurden, sind nur für die Datengeber sowie die Administratoren eines Vertrags einsehbar. Für die Begutachtung gibt es eine Ausnahmeregelung für die Gutachter (siehe Abschnitt „Begutachtung von Forschungsdaten“). Der Datengeber kann im eigenen Arbeitsbereich jederzeit an weitere bei RADAR registrierte Nutzerkennungen das Recht vergeben, als Datengeber zu agieren.

Archivierte Datenpakete sind normalerweise nur für den Datengeber und den Vertragsadministrator zugreifbar. Der Datengeber kann weiteren bei RADAR registrierten Nutzerkennungen das Recht einräumen, die beschreibenden Metadaten einzusehen und die archivierten Datenpakete abzurufen. Falls gewünscht, kann er den archivierten Datensatz auch öffentlich zugänglich machen. Diese Rechtevergaben sind jederzeit durch den Datengeber zu ändern.

Publizierte Datenpakete sowie archivierte Datenpakete, bei denen der Datengeber die Zugriffsrechte so gesetzt hat, dass sie uneingeschränkt öffentlich zugänglich sind, können von allen beim Online-System angemeldeten und anonymen (nicht angemeldeten) Datennutzern abgerufen werden. Die beschreibenden Metadaten sind im Online-System durchsuchbar und werden zusätzlich über einen OAI-Provider angeboten. Weiterhin sind sie auf www.datacite.org öffentlich einsehbar. Dies trifft auch zu, wenn die eigentlichen Forschungsdaten noch unter Embargo stehen. Der Datengeber kann für die eigentlichen Forschungsdaten eine Embargofrist von bis zu 12 Monaten nach Veröffentlichung einrichten, in der lediglich die Metadaten such- und abrufbar sind, nicht jedoch die Forschungsdaten. Nach Ablauf der Embargofrist sind dann auch die Forschungsdaten allgemein abrufbar.

Archivierte Daten (sofern sie der Datengeber nicht vollständig oder teilweise öffentlich zugänglich gemacht hat) sind weder über die Suche noch über OAI auffindbar. Dritte können weder die Daten noch die Metadaten einsehen oder recherchieren.