

Lieferung der Daten an Portale



Die virtuelle Zusammenführung von Informationen aus verschiedenen institutionellen Quellen in sammlungsübergreifenden Portalen bietet einen enormen Gewinn für die wissenschaftliche Forschung, die damit in die Lage versetzt wird, größere Datenmengen in semantischen Strukturen auszuwerten. Zum Beispiel können Objekte, die aus demselben Entstehungskontext stammen, heute aber auf verschiedene Standorte verteilt sind, aufgrund der Verwendung kontrollierter Datenwerte nachträglich in Suchergebnissen zusammengeführt werden. Der noch bessere Weg ist es jedoch, die Objektdaten bereits so an die Portale zu liefern, dass sie automatisch mit den verwandten Objekten aus anderen Institutionen in übersichtlichen Verknüpfungen dargestellt werden können. Man denke etwa an die einzelnen Teile eines mehrflügeligen Altarretabels, die sich heute in verschiedenen Museen befinden. Würde jede aufbewahrende Sammlung den Bezug ihrer Einzelteile zum Gesamtwerk in strukturierter Form angeben, so könnte der Altar in einem übergeordneten Portal in seinem Gesamtzusammenhang anschaulich gemacht werden. Dies gelingt jedoch nur, wenn die Daten formal standardisiert (LIDO XML) und in strukturell aufeinander abgestimmter Form geliefert werden. Deshalb liegen den üblichen Portalanwendungen konzeptuelle Referenzmodelle (CIDOC CRM) zugrunde, in denen Regeln dafür festgelegt werden, welche Informationseinheiten (Entitäten) anhand welcher Logik zueinander in Beziehung gesetzt werden dürfen. Werden mehrere solcher Aussagen miteinander verknüpft, so bildet sich ein semantisches Netz, welches Informationen logisch miteinander verbindet.

Daten über Sammlungsobjekte sind Forschungsdaten und sollten hohen wissenschaftlichen Standards genügen. Dazu zählt auch, dass sie gemäß der FAIR Prinzipien (Siehe Glossar → FAIR) zur Verfügung gestellt werden, um optimal von anderen Forscher*innen nachgenutzt werden zu können. Jedem publizierten Datensatz sollte ein global eindeutiger, persistenter Identifier zugewiesen werden, mit dem auf diesen Datensatz aus anderen Kontexten heraus verwiesen werden kann. In den Metadaten muss vermerkt werden, wer den Datensatz erstellt hat und wann er veröffentlicht wurde. Die Daten sollten langfristig zitierbar bleiben, unter möglichst offenen Lizenzen weitergegeben werden und leicht mit Daten aus anderen Quellen vernetzbar sein, indem nach Möglichkeit mit Verweisen auf Normdateien gearbeitet wird, die in Form von URIs (Siehe Glossar → URI) angegeben sind. Ihre formale Struktur und das verwendete Vokabular sollten so aufbereitet sein, dass die Daten-

sätze sowohl von Menschen (auch aus fremden Disziplinen) als auch von Computern gelesen und ausgewertet werden können.

▮ Praxistipps

- * Wenn die Metadaten zu einem Kunstwerk online veröffentlicht werden, ist darauf zu achten, dass ein eindeutiger Identifier für das Werk genannt wird. Das kann eine Werkverzeichnisnummer, ein Werknormdatensatz oder ein Digital Object Identifier (DOI) sein. Damit kann anschließend in anderen Erschließungskontexten auf diese Ressource eindeutig verwiesen werden. So können zum Beispiel Autor*innen digitaler Werkverzeichnisse auf Sammlungsdaten Bezug nehmen und diese um ihre Forschungsergebnisse erweitern. Der Datensatz selbst und die Abbildung sollten durch persistente Adressen (URI) zugänglich und damit dauerhaft zitierfähig gemacht werden.
- * Wenn rechtlich möglich, sollten beschreibende Metadaten und auch digitale Abbildungen mit offenen Lizenzen an Portale weitergegeben werden, um optimal von der Wissenschaft genutzt werden zu können (Open Access).
- * Es ist durchaus empfehlenswert, dieselben Daten an verschiedene Portale zu liefern, um ihre Sichtbarkeit zu erhöhen, dabei sollte die Lieferung derselben Inhalte an dasselbe Portal über unterschiedliche Mittler (= Aggregatoren) unbedingt vermieden werden.
- * Liegen die Daten in korrekter inhaltlicher und formaler Struktur vor, so sollten Wege für die regelmäßige Aktualisierung der Daten in externen Portalen gefunden werden, damit überall im World Wide Web derselbe Informationsstand zu den eigenen Objekten zugänglich ist. Eine automatische Datenaktualisierung gelingt am besten auf der Basis einer OAI PMH Schnittstelle = **O**pen **A**rchive **I**nitiative **P**rotocol for **M**etadata **H**arvesting. Im Rahmen einer solchen technischen Schnittstelle wird automatisch überprüft, innerhalb welchen Zeitraums welche Datensätze eine Änderung erfahren haben, ergänzt oder gelöscht worden sind. So kann eine externe Partnerinstitution ausschließlich die innerhalb eines festgelegten Zeitraums veränderten Daten übernehmen und weiternutzen. Es handelt sich beim OAI-PMH Standard also um ein Protokoll, das den Austausch von Daten erleichtert.

» Literaturhinweise und Links

<https://www.openarchives.org/pmh/>

Angelina Kraft, Die FAIR Data Prinzipien für Forschungsdaten, Leibniz-Informationszentrum Technik und Naturwissenschaften Universitätsbibliothek: <https://blogs.tib.eu/wp/tib/2017/09/12/die-fair-data-prinzipien-fuer-forschungsdaten/>