Schnittstellen

OAI-PMH, GraphQL und REST-API

- Einstieg: Schnittstellen
- OAI-PMH-Schnittstelle
- Ausnahme ZIP-Archiv

Einstieg: Schnittstellen

Eine Schnittstelle legt Regeln und Protokolle fest, die den effizienten Datenaustausch zwischen den Systemen ermöglichen. Bei der Datenabfrage („Harvesting“) holen wir die Daten in regelmäßigen Intervallen ab. Eine Schnittstelle standardisiert, beschleunigt und vereinfacht den Datentransfer.

Um Ihre Daten dem Kulturpool bereitzustellen, akzeptieren wir folgende Schnittstellenstandards – empfohlen wird OAI-PMH:

OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*) wird vor allem in wissenschaftlichen und akademischen Kontexten genutzt, um Metadaten aus verschiedenen Quellen abzurufen und sie an anderer Stelle zugänglich zu machen. OAI-PMH ermöglicht es, Metadaten in einem standardisierten Format zu sammeln und zu teilen, was die Interoperabilität zwischen verschiedenen Institutionen und Repositorien fördert.

REST-API (*Representational State Transfer Application Programming Interface*) ist eine Software-Schnittstelle, die die Prinzipien des REST-architektonischen Stils für den Austausch von Daten nutzt. Sie verwendet standardisierte HTTP-Anfragen und Daten werden häufig im JSON- oder XML-Format übertragen.

GraphQL ist eine Abfragesprache für APIs sowie ein Laufzeit-System zur Ausführung dieser Abfragen durch Nutzung vorhandener Daten. Im Gegensatz zu REST-APIs, die auf dem Konzept von Ressourcen basieren, ermöglicht GraphQL genauer zu spezifizieren, welche Daten benötigt werden, wodurch überflüssige Datenübertragungen vermieden werden können.

Ein **ZIP-Archiv** kann in begründeten Ausnahmefällen zur Bereitstellung der Metadaten verwendet werden; bitte beachten Sie die hierfür geltenden Vorgaben.

Linktipps

- Handreichung Deutscher Museumsbund: Kapitel 7: Datenexport und Schnittstellen

OAI-PMH-Schnittstelle

Für die Konfiguration Ihrer OAI-PMH-Schnittstelle finden Sie als Partnerinstitution hier eine Übersicht zu den wichtigsten Fragen.

Welche Metadaten sollen übermittelt werden?

Legen Sie zuerst fest, welche Metadaten für das Harvesting über OAI-PMH zur Verfügung gestellt werden sollen. Dies kann beschreibende Metadaten wie den Titel und das Leitthema sowie technische Metadaten wie das Medienformat und die Mediennutzungsrechte beinhalten. Dabei sollen zumindest die Mindestanforderungsfelder, besser aber noch mehr Metadaten (z. B. über Ort, Zeit, Handlungsträger, Konzepte …) übermittelt werden, um das Objekt bestmöglich zu kontextualisieren und es für verschiedene Benutzerszenarien vorzubereiten.

Welches Sammlungsmanagementsystem soll ich auswählen?

Ein Sammlungsmanagementsystem (auch: Sammlungsverwaltungssystem, Collection Management Software) erlaubt es, die Daten zu den Objekten in der Institution zu organisieren, zu kontrollieren und zu verwalten. Es gibt viele verschiedene Systeme, die den OAI-PMH-Standard unterstützen. Dabei ist es wichtig, eine Software zu finden, die sowohl mit den bereits bestehenden Systemen in der Institution kompatibel ist, als auch die gewünschten Metadaten übermitteln kann.

Wie konfiguriere ich die Software?

Die Software (das Sammlungsmanagementsystem) muss so konfiguriert werden, dass die gewünschten Metadaten über die OAI-PMH-Schnittstelle abgefragt werden können. Dies kann die Konfiguration der Software für die Verwendung bestimmter Metadaten-Schemata, die Einrichtung von Zugriffskontrollen und die Konfiguration der Servereinstellungen umfassen.

Als Hilfestellung zur Implementierung bietet die Open Archives Initiative Links zu verschiedenen Programmen an, die von Mitgliedern der Community erstellt wurden.

Was soll die OAI-PMH-Schnittstelle alles können?

Das OAI-PMH-Protokoll enthält sechs Basisfunktionen: `Identify`, `ListMetadataFormats`, `ListSets`, `ListIdentifiers`, `ListRecords`, und `GetRecord`. Diese Befehle werden mit „`?verb=`“ an die URL zum Repositorium der Institution (*baseURL*) angehängt.

Ein Beispiel für eine vollständige URL: `https://meine-sammlung.tld/oaipmh/?verb=ListRecords&metadataPrefix=edm&set=kulturpool`

Identify zeigt die allgemeinen Informationen zum OAI-Repositorium an. Das sind etwa der Name des Repositoriums, der baseURL und die Version des Protokolls.

ListMetadataFormats listet die Metadatenformate, die vom Archiv für den Export unterstützt werden.

ListSets gibt hierarchische Informationen zu der Sammlung, in der alle Kataloge (Datensets) organisiert wurden.

ListIdentifiers liefert den Header (Dateikopf) aller Datensätze, die mit den angegebenen Parametern (siehe unten) übereinstimmen.

ListRecords liefert mittels der Angabe von Parametern (siehe unten) die Datensätze. Für selektives Harvesting kann die Anfrage auf Datensätze aus einem bestimmten Datenset oder einem bestimmten Zeitraum beschränkt werden, wobei die Zeit immer die koordinierte Weltzeit UTC ist.

GetRecord liefert einzelne Datensätze durch die Angabe eines Identifiers und eines Präfixes.

Mögliche Parameter

Parameter können die folgenden sein:

- `from` – Startdatum
- `until` – Enddatum
- `set` – Datenset, welches abgerufen werden soll
- `metadataPrefix` – Metadatenformat
- `resumptionToken` – Datenflusssteuerung

Es sollten auch Fehlerbehandlungsmechanismen implementiert werden, um sicherzustellen, dass der Server auch bei Fehlern oder unvollständigen Daten angemessen auf Anfragen reagieren kann. Die definierten Fehlerkennungen sind: *`badArgument`*, *`badResumptionToken`*, *`badVerb`*, *`cannotDisseminateFormat`*, *`idDoesNotExist`*, *`noRecordsMatch`*, *`noMetaDataFormats`*, und *`noSetHierarchy`*.

Die Metadaten folgen dabei dem Mindeststandard des Europeana Data Model (EDM).

Die Datendarstellung sollte einheitlich erfolgen:

- Datumsangaben im Format YYYY-MM-DD
- Datumsangaben mit Zeitstempel (in koordinierter Weltzeit) im Format YYYY-MM-DDThh:mm:ss
- Der Sprach-Tag soll einheitlich für alle Daten sein, dem ? ISO-Standard 639-1 oder 639-3 folgend.
- Mehrere Werte sollen in eigene XML Elemente aufgeteilt werden, z. B. sollte anstatt von `<dc:creator>Smith, Adam; Nahs, John</dc:creator>` Folgendes ausgegeben werden:

```
<dc:creator>Smith, Adam</dc:creator>
<dc:creator>Nahs, John</dc:creator>
```

Wie stelle ich sicher, dass alles funktioniert?

Testen Sie Ihre OAI-PMH-Implementierung, um sicherzustellen, dass sie ordnungsgemäß funktioniert und die Metadaten wie erwartet verfügbar gemacht werden. Dies kann z. B. über die von OAI-PMH auf der Website zur Verfügung gestellte Validierung geschehen.

Wie kann ich meine Daten der Öffentlichkeit zeigen?

Schließlich können Sie Ihr Repositorium bei einem OAI-PMH-Dienstanbieter oder Aggregator registrieren, damit andere die Metadaten entdecken und sammeln können.

Sie können sich auch freiwillig bei OAI-PMH registrieren, da die Registrierung bestimmte Vorteile mit sich bringt. Neben Testungen zur Konformität des Repositoriums ist dann auch die zur Verfügung gestellte Datenbank für andere Aggregatoren und OAI-PMH-Dienstanbieter öffentlich zugänglich.

Linktipps

- Liste von OAI-PMH-Tools
- OAI-PMH-Validierung und Registrierung
- Informationen über die OAI-PMH Registrierung
- Beispiel einer OAI-PMH-Dokumentation der Uni Graz
- Informationen über OAI-Harvesting bei Europeana

Ausnahme ZIP-Archiv

Wir ermöglichen in **begründeten Ausnahmefällen**, Ihre Metadaten als ZIP-Archiv von XML-Dokumenten über eine permanente URL bereitzustellen. Hierfür sind folgende Vorgaben zu erfüllen:

1. Das Archiv muss über eine **permanente Webadresse** über **HTTPS** abrufbar sein, z.B. https://partnerinstitution.tld/kulturpool_daten.zip. Diese Adresse darf sich bei einer Aktualisierung der Daten nicht verändern.

2. Die **Dateistruktur** im Archiv muss **flach** sein, alle XML-Dokumente befinden sich auf der Hauptebene (keine Ordner in Ordnern), es werden ausschliesslich Dateien mit der Endung `.xml` verarbeitet.

3. Die XML-Dateien müssen **valides XML** enthalten, siehe W3C Markup Validation Service

4. Die **Dateinamen** der XML-Dateien müssen folgendem Schema entsprechen: `[Erstellungsdatum / Letztes Änderungsdatum]__[Identifier des Datensatzes].xml` , z.B.: `2024-06-08__SM-2381-92.xml` , dabei ist das Datum im Format `YYYY-MM-DD` anzugeben und durch zwei Unterstriche `__` vom Identifier zu trennen. Sollte Ihr Identifier zu lange sein, oder Zeichen enthalten, die als Dateiname nicht zulässig sind, nehmen Sie bitte vorab Kontakt mit uns auf.

Ein Beispiel für die **Dateistruktur** kann wie folgt aussehen:

```
├── 2024-07-16__SM-1234-56.xml
├── 2024-08-22__SM-5678-91.xml
├── 2024-09-05__SM-2345-78.xml
├── 2024-10-30__SM-8765-34.xml
├── 2024-11-12__SM-3456-87.xml
├── 2024-12-25__SM-7890-12.xml
├── 2025-01-08__SM-4321-09.xml
├── 2025-02-14__SM-0987-65.xml
├── 2025-03-20__SM-6543-21.xml
└── 2025-04-28__SM-3210-43.xml
```

Der Inhalt der XML-Dateien ist bevorzugt im Format Europeana Data Model (EDM), schauen Sie sich die Beispiel-Datensätze in EDM an. Falls die XML-Dateien in einem anderen Metadaten-Format vorliegen, benötigen wir eine Mapping-Tabelle, die Sie uns im Zuge der Anbindung zukommen lassen können.

Ähnlich einer Datenschnittstelle können wir Ihre Daten in regelmäßigen Intervallen in unser System einspielen und so Aktualisierungen vornehmen.