

## Metadata Harvesting

*Stigler, Johannes; johannes.stigler@uni-graz.at*

Seit Mitte der 90er-Jahre des letzten Jahrhunderts entstanden weltweit Archive mit unterschiedlichsten wissenschaftlichen digitalen Inhalten. Daraus entwickelte sich der Wunsch nach einer anbieterübergreifenden, aber dennoch gezielten Suchmöglichkeit. In der Bibliothekswelt galt dafür lange Zeit eine sogenannte Z39.50-Schnittstelle als *State of the Art*. Bei diesem Verfahren handelte es sich um eine *Cross Database Search*, bei der in jeder Suchanfrage einer Nutzerin/eines Nutzers, die Server der einzelnen Informationsanbieter – ob verfügbar oder nicht verfügbar – angefragt wurden.

Demgegenüber hat sich in den letzten Jahre das sogenannte *Metadata Harvesting* durchgesetzt. Sammlerdienste (*Service Provider*) besuchen dabei in regelmäßigen Abständen Informationsanbieter (*Data Provider*), die über eine standardisierte Schnittstelle Metadaten (= kurze Pressemitteilungen) über die digitalen Inhalte (Publikationen, edierte Handschriften, erschlossene Artefakte u. v. m.) in ihren Forschungsdatenrepositorien bereitstellen. In diesen gesammelten Metadaten kann dann auf einem Portal des *Service Providers* anbieterübergreifend gesucht werden. Die eigentlichen Inhalte verbleiben beim *Data Provider*, der mittels eines *persistenten Identifiers* auf seine Inhalte verweist. Über diese Referenz gelangen die Suchenden dann auch zu den Vollinhalten der gefundenen Einträge.

Die *Open Archives Initiative* (OAI) hat im Jahre 2000 erstmalig das auf XML und REST basierende *OAI Protocol for Metadata Harvesting* (OAI-PMH) veröffentlicht, über das die standardisierte Kommunikation zwischen *Service* und *Data Providern* abgewickelt wird.

Unter dem Namen *OAI Object Reuse and Exchange* (OAI-ORE) wurden in der *Open Archives Initiative* ergänzend zu OAI-PMH Verfahren geschaffen, um die Binnenstruktur digitaler Objekte in Repositorien und die Verknüpfungen zwischen ihnen abzubilden. So kann die digitale Repräsentation einer Handschrift in einem Repository z. B. aus verschiedenen Versionen und Formaten (Volltext in PDF und HTML, Metadaten in RDF etc.) und auch aus verschiedenen Teilen bestehen (TEI-Transkript, Faksimiles etc.) sowie Verknüpfungen zu anderen Dokumenten besitzen (Übersetzung, Zitation, Versionierung etc.). Die Grundidee von OAI-ORE besteht darin, diese Binnenstruktur eines Dokumentes maschinenlesbar in einer *Resource Map* abzubilden und für den *Service Provider* verfügbar zu machen.

Die Verfügbarkeit von Forschungsergebnissen auf solchen Suchportalen trägt unmittelbar zu deren Sichtbarkeit bei und so kann *Metadata Harvesting* auch als wichtiges Instrument der Öffentlichkeitsarbeit in der Scientific Community verstanden werden.

Auswahl einiger für die Digitalen Geisteswissenschaften bedeutender *Service Provider*:

- <https://europeana.eu>
- <http://kulturpool.at>
- <https://correspsearch.net/>
- <https://pelagios.org/>
- <http://numismatics.org/ocre/>

### **Projekte:**

Open Archives Initiative, OAI Protocol for Metadata Harvesting, OAI Object Reuse and Exchange

### **Verweise:**

Metadaten, Persistent Identifier, Digitale Nachhaltigkeit, Bereitstellung von Forschungsdaten

### **Themen:**

Metadaten, Archivierung

### **Zitiervorschlag:**

Stigler, Johannes. 2021. Metadata Harvesting. In: KONDE Weißbuch. Hrsg. v. Helmut W. Klug unter Mitarbeit von Selina Galka und Elisabeth Steiner im HRSM Projekt "Kompetenznetzwerk Digitale Edition". URL: <https://gams.uni-graz.at/o:konde.10>