

Sammlungsstrategie

Präambel

ARCHE bietet eine stabile Langzeitarchivierung digitaler Forschungsdaten für die österreichische geisteswissenschaftliche Forschungsgemeinschaft. Dieses Dokument beschreibt die Grundsätze der Akquise, Kuratierung und Verwaltung von Ressourcen in ARCHE und erläutert neben der Sammlungsstrategie auch die Art der zulässigen Inhalte.

Sammlungsbereich

Als Teil der Infrastruktur von CLARIAH-AT, ist ARCHE primär als digitaler Dienst zur Aufbewahrung von Daten aus den Geisteswissenschaften in Österreich intendiert. Daher sind Daten aus allen Fachbereichen der Geisteswissenschaften gleichermaßen willkommen. Dazu gehören moderne Philologien, klassische Philologien, Linguistik, Literaturwissenschaft, Geschichte, Jura, Philosophie, Archäologie, vergleichende Religionswissenschaft, Ethik und Kunstgeschichte. ARCHEs Vorgänger, das CLARIN Centre Vienna / Language Resources Portal (<https://clarin.oeaw.ac.at>) war auf digitale Sprachressourcen spezialisiert, während ARCHE nun Dienste für einen breiteren Fachbereich anbietet.

Der Dienst wurde so gestaltet, dass eine breite Auswahl digitaler geisteswissenschaftlicher Forschungsdaten unterstützt werden kann. Wir akzeptieren (annotierte) digitale Texte, lexikalische Ressourcen, tabellarische Daten, Datenbanken, Bilder, Sammlungen mit GIS-, 3D- oder CAD-Daten, multimediale Dateien (Audio und/oder Video), Webseiten und soziale Medien etc. Darüber hinaus wird auch Software (Anwendungen, Quellcode etc.) angenommen.

ARCHE ist vor allem für Ressourcen mit Österreich-Bezug gedacht: Daten, die in Österreich erstellt oder erfasst wurden bzw. solche, die in historischer oder geografischer Hinsicht für österreichische Forschende von Interesse sind. Jedoch werden Ressourcen ohne direkten Bezug zu Österreich nicht kategorisch ausgeschlossen.

ARCHE zielt sowohl auf die Erhaltung primärer Forschungsdaten, als auch auf abgeleitete und prozessierte Daten in unterschiedlichen Versionen ab. Herkömmliche Publikationen zu bewahren, ist nicht der Schwerpunkt von ARCHE, da es dafür zahlreiche Alternativen gibt.

ARCHE ist nicht nur als Repository für vollständige und abgeschlossene Datensammlungen von beendeten Projekten gedacht, sondern begrüßt ausdrücklich auch Daten aus laufenden Projekten in Form von Zwischenständen, die weitere Überarbeitungen erwarten, wie beispielsweise Transkriptionen historischer Texte, die schrittweise mit zusätzlichen Informationen oder Annotationsebenen angereichert werden. Nichtsdestotrotz werden alle gespeicherten Daten als unveränderlich betrachtet, das heißt, Daten mit dem gleichen persistenten Identifikator bleiben unverändert.

ARCHE wird von Fall zu Fall über die Annahme einer Einreichung entscheiden und gegebenenfalls auf besser geeignete Archive oder Repositorien verweisen.

Evaluierung

ARCHE erkennt an, dass es für Datengebende und -nutzende von Vorteil ist, wenn eingereichte Daten ein effektives und rigoroses Peer-Review-Verfahren durchlaufen. Die Daten werden durch Kurator:innen nach den unten genannten Kriterien evaluiert, wobei auch deren Handhabbarkeit und eventuell notwendige Maßnahmen zu deren Erhaltung und Dissemination berücksichtigt werden. In diesem Zusammenhang ist beabsichtigt, geeignete Kurator:innen für die jeweilige Fachdisziplin, aus der die Daten stammen, einzusetzen.

ARCHE ist sich seiner Verantwortung bewusst, eng mit den Datengebenden zu kooperieren, um die Verfügbarkeit der erforderlichen Metadaten zu gewährleisten.

Die folgende Liste enthält allgemeine Auswahlkriterien zur Beurteilung von eingereichten Daten. Diese Kriterien dienen als allgemeine Richtlinie; die Nichterfüllung eines Kriteriums muss nicht zwingend zum Ausschluss von Daten führen. In jedem Fall wird die endgültige Entscheidung zur Annahme oder Ablehnung der eingereichten Daten gemeinsam mit dem:der Datengebenden gefällt.

- Risiko: Sind die Daten in Gefahr, zu verschwinden oder auf andere Weise unzugänglich zu werden?
- Grad der Verfügbarkeit: Können die Daten bereits anderswo gefunden werden?
- Nachfrage: In welchem Ausmaß werden die Daten verwendet?
- Geografischer oder historischer Bezug: Haben die Daten eine Beziehung zu Österreich?
- Kuratierungsaufwand: Wie viel Arbeit ist notwendig, um die Daten zur Archivierung vorzubereiten?
- Stabilität der Daten: Handelt es sich bei den Daten um einen Zwischenstand oder eine finale Version?

2

Speicherdauer

Es ist beabsichtigt, alle in ARCHE aufgenommenen Ressourcen permanent zu speichern. Auf Ersuchen von Datengebenden können Daten entfernt werden, wobei ARCHE jedoch mindestens einen Hinweis auf das Vorhandensein der Daten bis zum Löschdatum beibehalten wird. Daher bleiben Metadaten erhalten und werden mit dem zusätzlichen Hinweis darauf versehen, dass die Dateien selbst entfernt wurden. Der zugewiesene PID bleibt erhalten und verweist auf eine sogenannte Tombstone-Seite, welche die Metadaten enthält.

Quellen

<https://digital.library.yorku.ca/tags/digital-preservation-policy>

<https://archaeologydataservice.ac.uk/about/policies/collections-policy>

<http://datenportal.ianus-fdz.de/resources/documents/Sammlungsstrategie.pdf>

<https://spotdocs.scholarsportal.info/display/OAIS/Collection+Policy>