

## Informationen zum Datenrepositorium HeiDATA Dataverse Network

Das "Dataverse Network™ Project" ist eine webbasierte Software zur Zitation, Analyse und Veröffentlichung von Forschungsdaten. Sie ist Open Source und wird seit 2006 vom Institute for Quantitative Social Science (IQSS) der Harvard Universität entwickelt.

Im Kompetenzzentrum Forschungsdaten der Universität Heidelberg wird mit HeiDATA ein Langzeit-Repository für Forschungsdaten mit Hilfe dieser Software aufgebaut. Im Zuge der Zusammenarbeit mit der Universität Heidelberg und dem begleitenden Aufbau der Virtuellen Bibliothek Altertumswissenschaften "Propylaeum" besteht für Autoren von Digital Classics Online die Möglichkeit, das Datenrepositorium für begleitendes oder zusätzliches Material (Rohdaten) zu benutzen.

Im Zuge der Open-Access-Politik des Journals und aus Gründen der vorgesehenen Langzeitarchivierung werden die Autoren gebeten, die Daten in möglichst offenen, nicht proprietären Formaten nach den BSI-Richtlinien zur Langzeitarchivierung zur Verfügung zu stellen.

Laut dem IT-Grundschutzkatalog (M 4.170) des Bundesamts für Sicherheit in der Informationstechnik (BSI) müssen Dateiformate, die bei der elektronischen Langzeitarchivierung eingesetzt werden, folgende Kriterien erfüllen:

- Das Datenformat sollte möglichst langfristige Relevanz haben.
- Die Dokumentstruktur sollte eindeutig interpretiert werden können.
- Der Dokumentinhalt sollte elektronisch weiterverarbeitet werden können.
- Gesetzliche Vorschriften müssen berücksichtigt werden.
- Grammatik und Semantik des Datenformates sollten ausführlich dokumentiert sein, sodass eine spätere Migration zu anderen Formaten problemlos möglich ist.
- Merkmale des Originaldokuments (elektronisch oder in Papierform) sollen später eindeutig nachweisbar sein, auch wenn das Originaldokument nicht mehr vorhanden ist.

Diese Voraussetzungen erfüllen folgende Dateitypen und Dokumentenbeschreibungssprachen:

A. Strukturformate: PDF/A, SGML, XML. Bei der Archivierung in SGML oder XML sind jedoch unbedingt auch die Semantikspezifikation (DTD - Document Type Definition) und ggf. auch die Layout-Informationen, in XSL oder als zusätzliches TIFF-Dokument, zu archivieren.

B. Bildformate: TIFF und JPEG. Für eine reversionssichere Archivierung wird empfohlen, bei der Auswahl der Kompressionsstufe eine verlustfreie Kompression zu wählen.

C. Audio- und Video-Formate: Bei der digitalen Verarbeitung von Audio- und Videodaten entstehen schon bei zeitlich kurzen Aufzeichnungen sehr große Datenmengen, die Eignung verlustbehafteter Kompressionsverfahren für die Archivierung von Video- und Tonmaterial ist deshalb anwendungsspezifisch zu prüfen.

Darüber hinaus empfehlen wir für zweidimensionale Datentabellen UTF-8-kodierte Plain Text Dokumente als Delimiter-separated value, also CSV (Comma-separated values) oder TSV (Tab-Separated Values).