



## Kontakt

**Universität Stuttgart**  
 Informations- und Kommunikationszentrum (IZUS)  
 Universitätsbibliothek  
 Holzgartenstraße 16  
 D-70174 Stuttgart

**Ansprechpartnerin**  
 Dipl. Ing. Sibylle Hermann

T 0711 685-82502  
 sibylle.hermann@ub.uni-stuttgart.de

**Herausgeber**  
 Universität Stuttgart

**Gestaltung**  
 dallidallidesign

**Fotografie**  
 Wirestock | Dreamstime.com

**Stand**  
 April 2022

**Finanziert durch**



**Projektpartner**



Universität Stuttgart



**XSample**  
 access, reuse, advance



# Handreichung zur Nutzbarmachung von Korpusauszügen für Text- und Data-Mining-Forschungen

Der hier erläuterte Ablauf geht von der Prämisse aus, dass ausschließlich textbasierte Ressourcen mit Methoden des Text- und Data-Mining (TDM) beforscht werden und in Gestalt größerer Korpora vorliegen. Besondere Einschränkungen gelten hingegen für Computerprogramme oder Datenbanken. Der Ablauf orientiert sich an urheberrechtlichen Einschränkungen, die von Forschungsinfrastruktureinrichtungen, die die betreffenden Korpora bereitstellen möchten, zu prüfen sind. Konkret werden die vorgestellten Arbeitsschritte auf § 60d sowie auf § 60c des Urheberrechtsgesetzes (UrhG) gestützt.

Die **Nachnutzung erfolgt mit dem XSample-Tool** (<https://github.com/ICARUS-tooling/xsample-server>) und dem **Repositorium-System Dataverse** (<https://dataverse.org>)

**Ausführlicheres entnehmen Sie bitte der XSample-Website:** <https://bw-bigdiwa.bib.uni-mannheim.de/projekte/xsample-stuttgart/>

## Kurzüberblick: Urheberrechtliche Rahmenbedingungen

- Einbeziehung der Forschungsinfrastruktureinrichtungen: Forschungsinfrastruktureinrichtungen, die die Korpora aufbewahren und gegebenenfalls bereitstellen, sollten von Beginn an in die Projekt-konzeptionen integriert werden.
- Inhalt der Korpora: Die Korpora dürfen nicht die unveränderten Ursprungsdaten enthalten, es sei denn, die betreffende Einrichtung hat hierzu entsprechende Lizenzen erworben.
- Aufbewahrungsdauer: Forschende, die die Korpora bereitstellen möchten, sollten eine der Forschung angemessenen Aufbewahrungsdauer für die Korpora vorschlagen. Bei einer kürzeren oder längeren Dauer als zehn Jahren sollte es gesondert begründet werden.

- **Nachnutzung:** Die Korpusauszüge dürfen nur zu nicht-kommerziellen wissenschaftlichen Zwecken an bestimmt abgegrenzte Personenkreise zugänglich gemacht werden, das heißt auf individuelle Anfrage.
- **Umfang der Korpusauszüge:** Bereitgestellte Korpusauszüge dürfen maximal 15 Prozent ganzer Werke betragen. Kurze Werke wie zum Beispiel Aufsätze aus wissenschaftlichen Zeitschriften (maximal 25 Seiten) können vollständig herausgegeben werden.

## 1 Was müssen Forschende beachten, die ihre TDM-Korpora zur Verfügung stellen möchten?

- **Workflow**
  - Die Forschungsdaten (das heißt die Ursprungsdaten und die Korpora) sollten bereits während der Forschungsarbeiten auf den Servern der Infrastruktur gespeichert und auf diesen beforscht werden.
  - Während der Forschungsarbeiten muss der Zugriff auf die Forschenden der jeweiligen Forschungsgruppen beschränkt werden.
- **Archivierung**
  - Archiviert werden dürfen nur die Vervielfältigungen für das TDM, das heißt die Korpora, zum Beispiel in den Formaten CoNLL und TEI sowie gegebenenfalls TXT (§ 60d Abs. 5 UrhG).
  - Die Ursprungsdaten (zum Beispiel PDFs und EPUBs) müssen gelöscht werden, wenn die Forschung abgeschlossen ist (§ 60d Abs. 5 UrhG). Sie dürfen nur enthalten sein, wenn diese Nutzungen lizenziert sind, dazu zählen auch Open Access publizierte Werke.
  - Wenn die Korpora kürzer oder länger als zehn Jahre aufbewahrt werden sollen, muss das plausibel begründet werden.

## 2 Was müssen Forschende beachten, die fremde TDM-Korpora nachnutzen möchten?

- Nachnutzende benötigen einen authentifizierten DaRUS- beziehungsweise Dataverse-Zugang. Dabei müssen nicht-kommerzielle, wissenschaftliche Zwecke verifiziert werden.
- Über die Suchmaske kann das betreffende Korpus aufgefunden werden.
- Um zum XSample-Auszugs-Tool zu gelangen, muss eine individuelle Anfrage gestellt werden.
- Wenn der Zugriff gewährt wurde, kann ein Auszug interessengerecht erstellt und heruntergeladen werden.
- Für die Weiternutzung der Korpusauszüge gelten die Vorgaben des Urheberrechts (insbesondere §§ 60c, 60d UrhG).

## 3 Was müssen Einrichtungen beachten, die für die Nachnutzung von TDM-Korpora Infrastrukturen bereitstellen möchten?

- Verfügbar gemacht werden dürfen nur die Korpora. Ursprungsdaten dürfen nur dann verfügbar gemacht werden, wenn diese Nutzung lizenziert wurde.
- Die Korpora müssen in Dataverse dergestalt eingestellt werden, dass auf sie, auch mittels des XSample-Tools, nur nach individueller Anfrage zugegriffen werden kann.
- Die standardmäßige Aufbewahrungsfrist für TDM-Korpora beträgt zehn Jahre, sie kann mit entsprechender Begründung aber kürzer oder länger bemessen sein. Nach Ablauf sind die Korpora zu löschen.