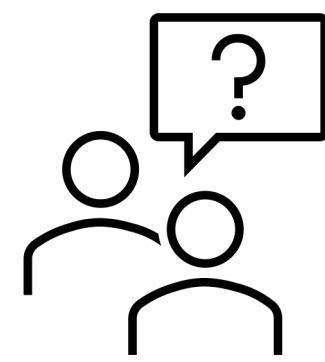


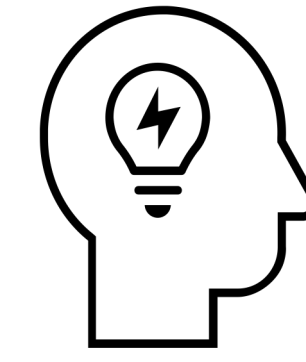
## Urheberrechtlich geschützte Texte nachnutzen: Der XSample-Workflow

### Problemstellung



- Das deutsche Urheberrecht
  - schränkt die Weitergabe urheberrechtlich geschützter Texte ein,
  - beeinflusst die Auswahl der Untersuchungsgegenstände und
  - führt so zu Verzerrungen in der Forschungslandschaft.
- Lösungsansatz von Schöch et al. (2020), „abgeleitete Textformate“, führt zu Verfremdung der Textgestalt und ist nicht für alle Nachnutzungsfälle sinnvoll.

### Lösung



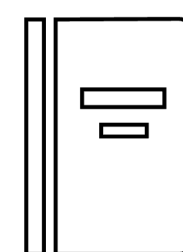
- Nutzung von §60c UrhG: erlaubt Weitergabe von Auszügen im Umfang von max. 15%
- Auszugskonzept: Interaktive Auswahl von Auszügen
- Vorteile:
  - Bewahrung der ursprünglichen Textgestalt
  - Zuschnitt der „hilfreichsten 15%“ erlaubt Anpassung der Auswahl an Forschungsinteressen
- Test des Konzepts anhand zweier Use Cases aus Linguistik und Literaturwissenschaft

### Gesetzeslage

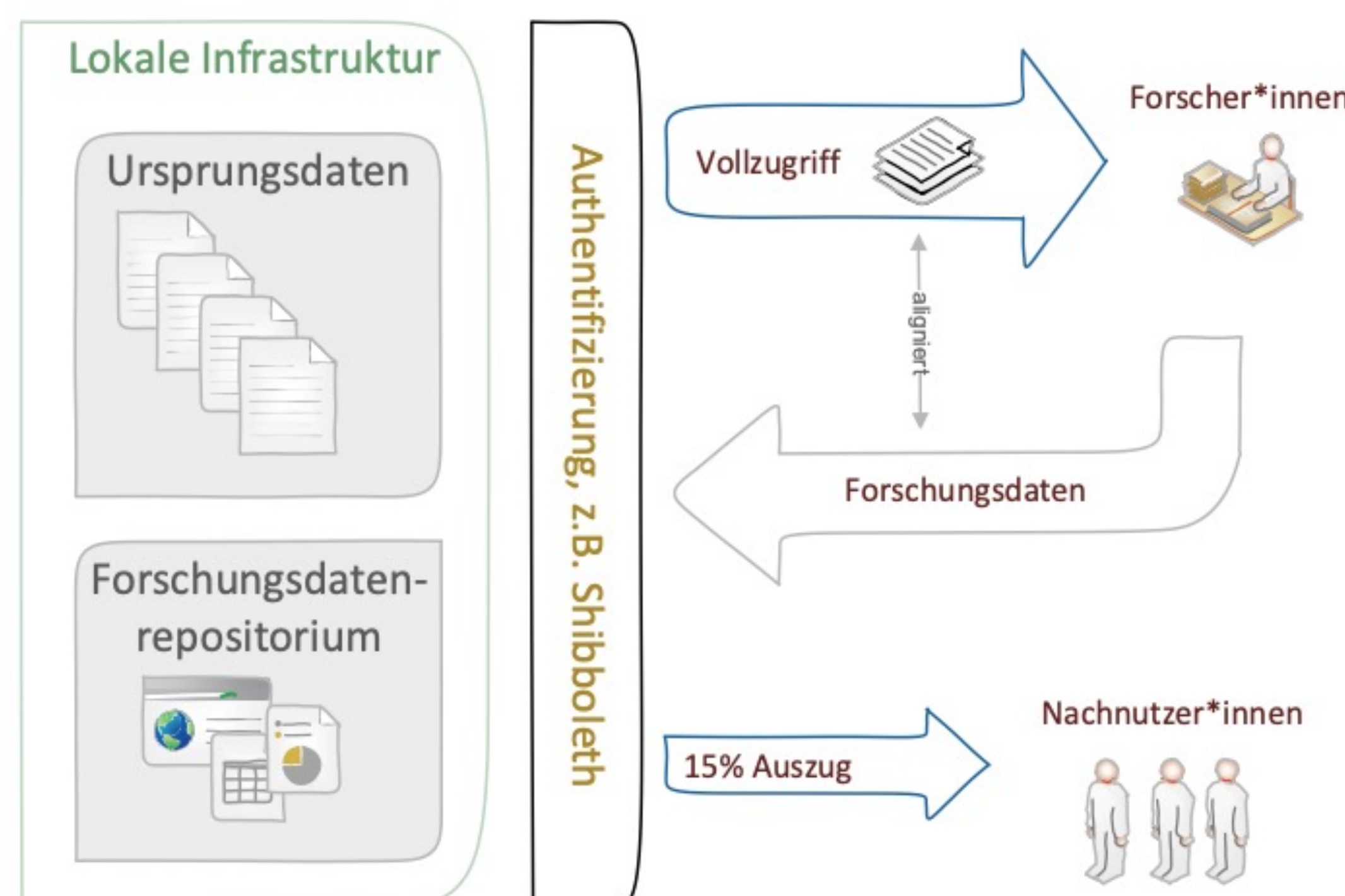


- §60c UrhG: Erlaubnisnorm für Zwecke der nicht-kommerziellen wissenschaftlichen Forschung
- Abs. 1 Nr. 1 erlaubt die Weitergabe von bis zu 15% von Werken an abgegrenzte Personengruppen
- §60d UrhG: Erlaubnisnorm zur (Nach-)Nutzung und Aufbewahrung von TDM-Korpora (Text und Data Mining)
- Aufbewahrungsdauer muss der Forschung angemessen sein (i.d.R. 10 Jahre); dauerhafte Speicherung durch Kulturerbe-Einrichtungen zulässig
- Kombination zweier Erlaubnisnormen ist möglich

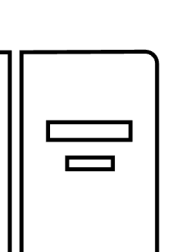
### Use Case I



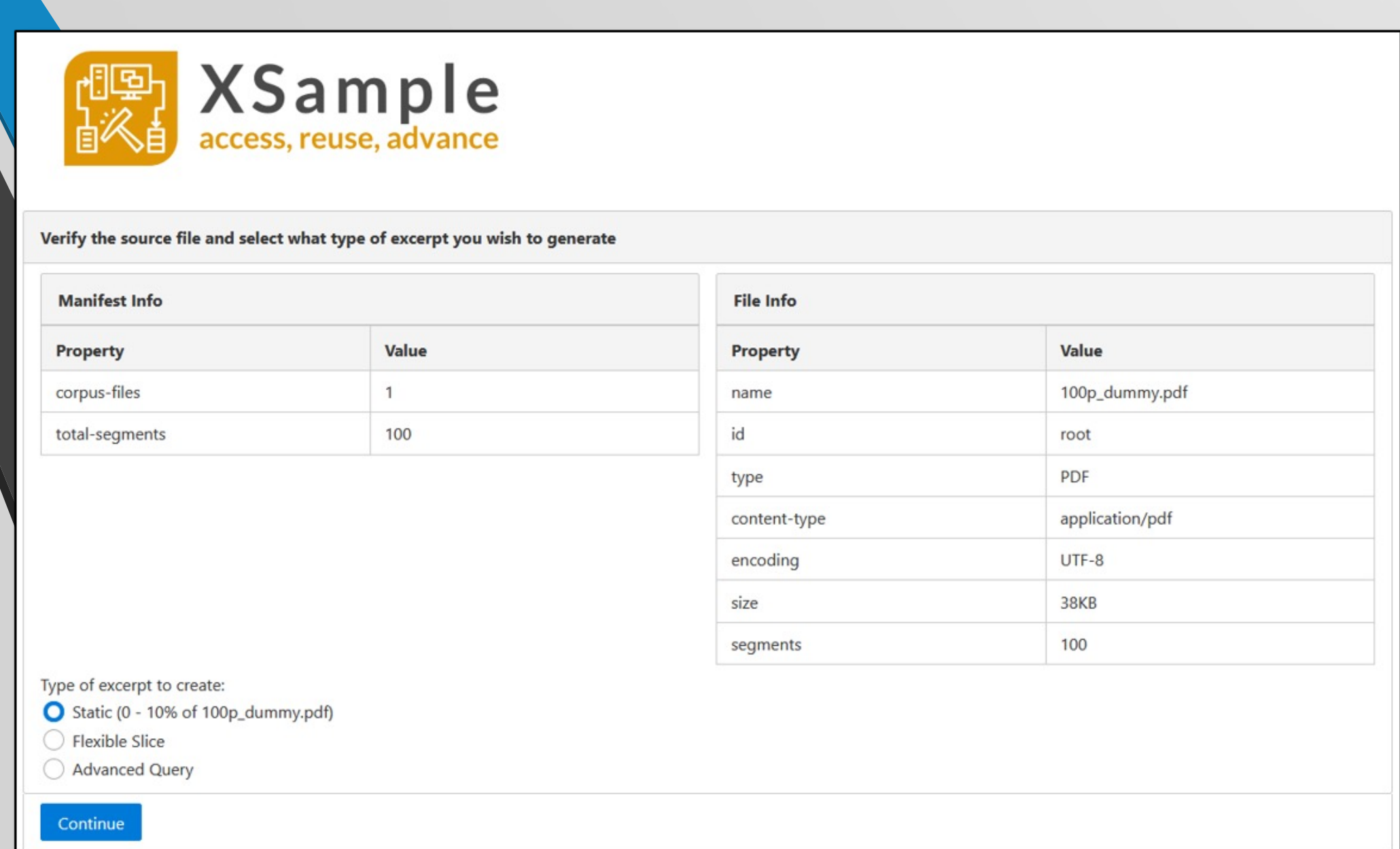
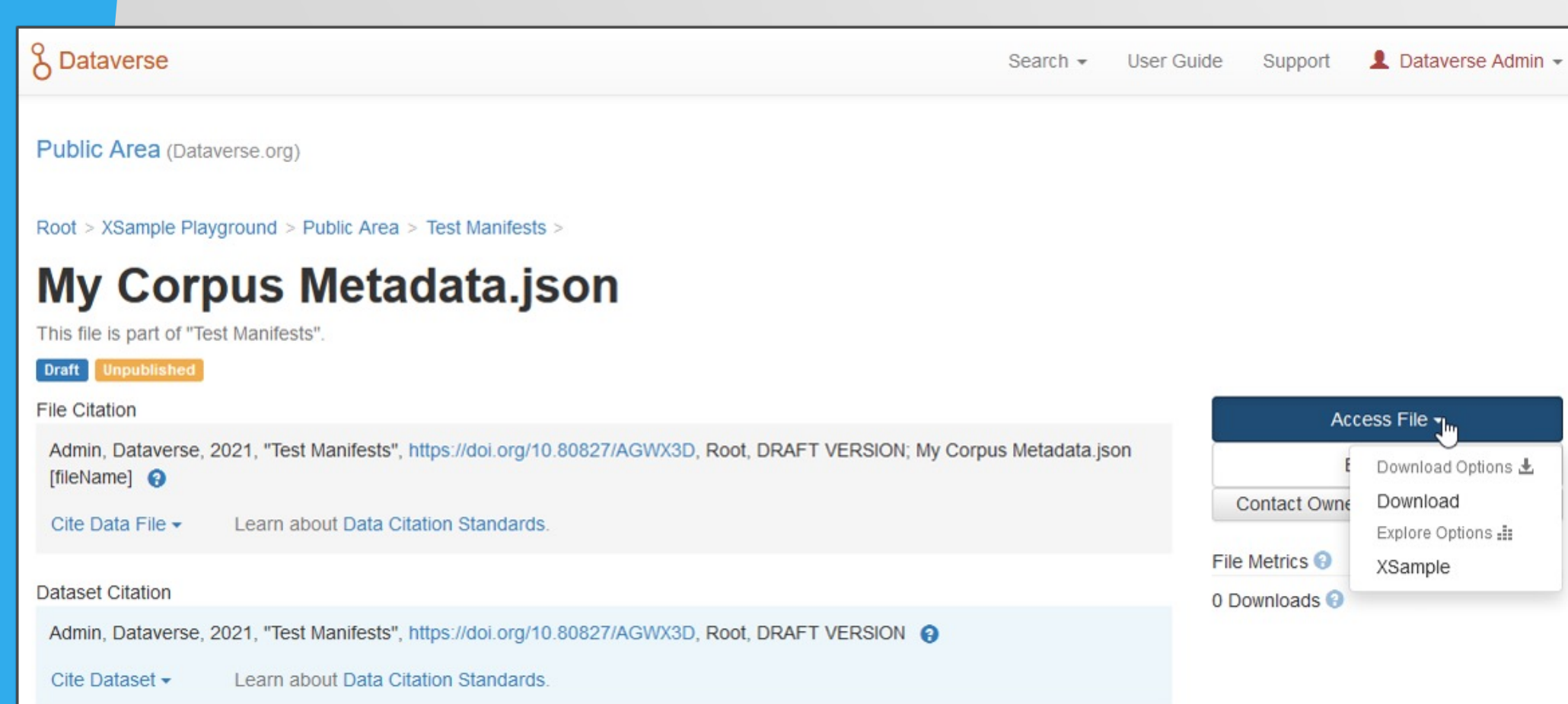
- Linguistischer Anwendungsfall: „Wissenschaftssprache“
- Forschungsfrage: Worin bestehen die linguistischen Spezifika der Wissenschaftssprache von Linguistik, Literaturwissenschaft und Philosophie?
- Methode: Datengeleitete linguistische Analyse von Beiträgen in Fachzeitschriften
- Relevanz der Auszüge: Voraussetzung, um die im Rahmen der Analyse ermittelten distinkten Merkmale tatsächlich verstehen zu können



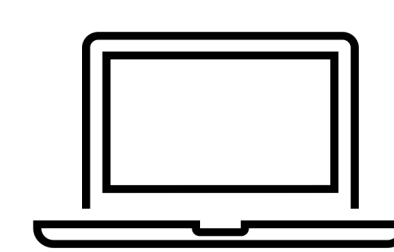
### Use Case II



- Literaturwissenschaftlicher Anwendungsfall: „Unzuverlässiges Erzählen“
- Forschungsfrage: Wie können wir feststellen, an welchen Stellen Erzählinstanzen inkorrekte Äußerungen über die fiktive Welt eines fiktionalen Textes tätigen?
- Methode: automatische und manuelle Annotation unterschiedlich komplexer Textphänomene in Erzählungen
- Relevanz der Auszüge: Nachvollzug der Operationalisierung, Einblick in erste Ergebnisse (z.B. Korrelation relevanter Phänomene)



### Implementierung



- Prototypische Implementierung im Rahmen des Projekts
- Umsetzung als eigenständiger Webserver
- Integration in Dataverse UI
- JSON Metadaten als Einstiegspunkt im Dataverse (suchbar, mit DOI)
- Die Nutzer\*innen werden authentifiziert, um die rechtlichen Beschränkungen einzuhalten.
- Die Nutzer\*innen greifen auf eine Weboberfläche für die individuelle Auszugskonfiguration zu.
- Der XSample-Server bietet mehrere Auszugsvarianten an:
  - von Autor\*innen statisch vordefiniert
  - verschiebbarer Ausschnitt
  - anfragebasiert (z.B. auf Basis von Annotationen) mit Korpusanfrage-schnittstelle

