

## Eine generische Lösung für den nachhaltigen Erhalt von Daten in den Geisteswissenschaften

George Dogaru

Gesellschaft für wiss. Datenverarbeitung mbH Göttingen (GWDG)

- Das Text+-Langzeitarchiv soll 2023 an den Start gehen. Es wird im Rahmen der Measure 2 in der Text+ Task Area Infrastructure/Operations durch die GWDG, die SUB Göttingen und die DNB realisiert.
- Es adressiert den grundlegenden Bedarf nach einer Infrastruktur für den langfristigen und sicheren Erhalt von Daten mit geisteswissenschaftlichem Bezug.
- Es wird das unkomplizierte Archivieren ermöglichen, ohne Einschränkung des Datenformats.
- Es wird *bitstream preservation* bieten sowie PIDs.
- Sowohl offene als auch geschützte Archivbereiche werden möglich sein.

Das Text+-Langzeitarchiv (provisorischer Name)	
Spezialisierung	Daten mit geisteswissenschaftlichem Bezug
Modalität	Alle (keine Einschränkung)
Akzeptierte Datenformate	Alle (keine Einschränkung)
Ansprechperson(en)	George Dogaru (george.dogaru@gwdg.de)

### Langzeitarchiv vs. Repository

Anders als Repositories, die ihre Daten für die sofortige Nutzung bereithalten, dient ein Langzeitarchiv der langfristigen Aufbewahrung von Daten, die möglicherweise nicht häufig abgerufen oder aktiv genutzt werden. Eine doppelte Nutzung des Text+-Langzeitarchivs als Archiv und als Repository wäre technisch möglich.

### Langzeiterhalt vs. langfristige Verwendbarkeit

Das Text+ Langzeitarchiv garantiert zunächst *bitstream preservation*, was die Basisvoraussetzung für den langfristigen Erhalt von Daten darstellt. Für die langfristige Verwendbarkeit der Daten sind in der Regel weitere Maßnahmen nötig, wie Migration und Emulation (siehe <https://www.dnb.de/DE/Professionell/Erhalten/ErhaltungDigitaleObjekte/erhaltungDigitalObjekte.html>).

### Wann ist die Nutzung des Text+ Langzeitarchivs sinnvoll?

- Zur Sicherung wertvoller Ergebnisse, die oft aufwendig erarbeitet/gesammelt wurden
- Wenn benötigter Speicherplatz nicht dauerhaft zur Verfügung steht oder der langfristige Erhalt der Daten nicht garantiert wird
- Wenn eine uneingeschränkte Verfügbarmachung schwierig oder unerlaubt ist
  - Weil die Daten noch nicht publikationsreif sind oder nicht geeignet zur Publikation
  - Weil die Daten urheberrechtlich geschützt sind oder ihre Benutzung durch sonstige Lizenzbestimmungen geregelt ist
  - Weil die Klärung der Möglichkeiten zur Verfügbarmachung ein komplizierter Prozess ist und besondere Kompetenzen voraussetzt

Text+ could help facilitate the work of OPERAS [...] as a service that stores sensitive data [...] Most urgently needed would be a reliable central and long-term storage of the conducted data [...]. This would not only facilitate the work of all project partners as they would not need to think about how to store this data, it would also be centrally available to everyone with access rights throughout the project and after if the project's grant agreements require it.  
<https://www.text-plus.org/en/research-data/user-story-311/>  
OPERAS – open scholarly communication in the european research area for social sciences and humanities

### Die Nachhaltigkeit von Daten und Systemen ist einer der meist genannten Bedarfe in den Text+ User Stories. Das Langzeitarchiv kann eine wichtige Rolle bei der Realisierung nachhaltiger Lösungen spielen.

Getting researchers to get their valuable research data to a long-term archive is hard, so we need to lower the barrier to archive/publish research data.  
<https://www.text-plus.org/en/research-data/user-story-351/>  
Data Archiving Support

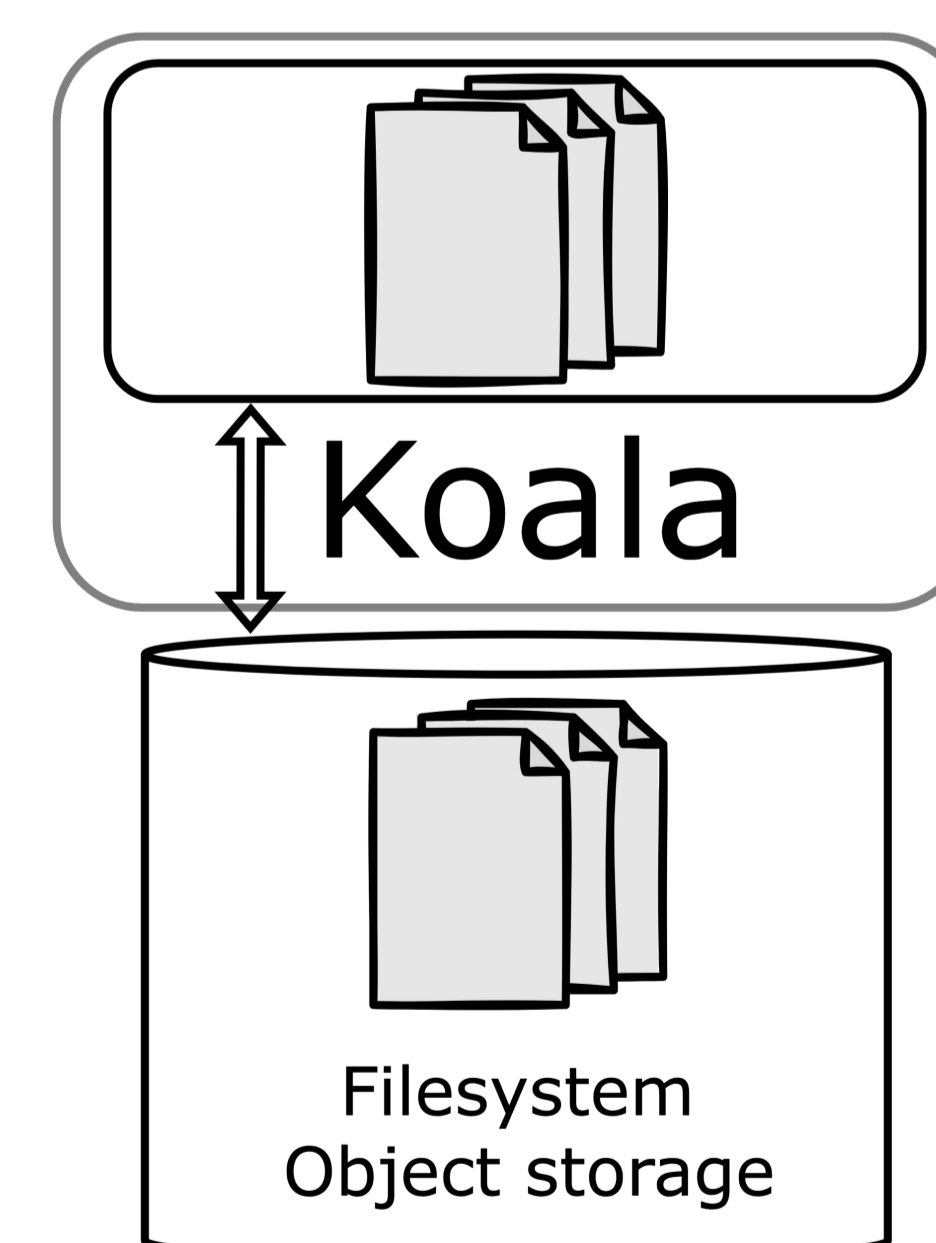
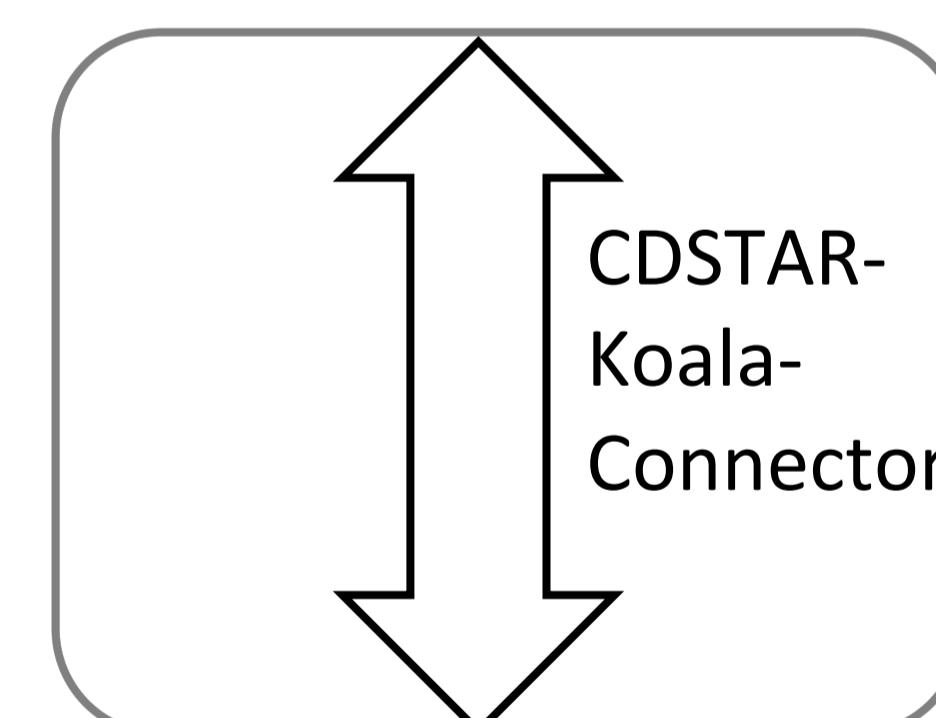
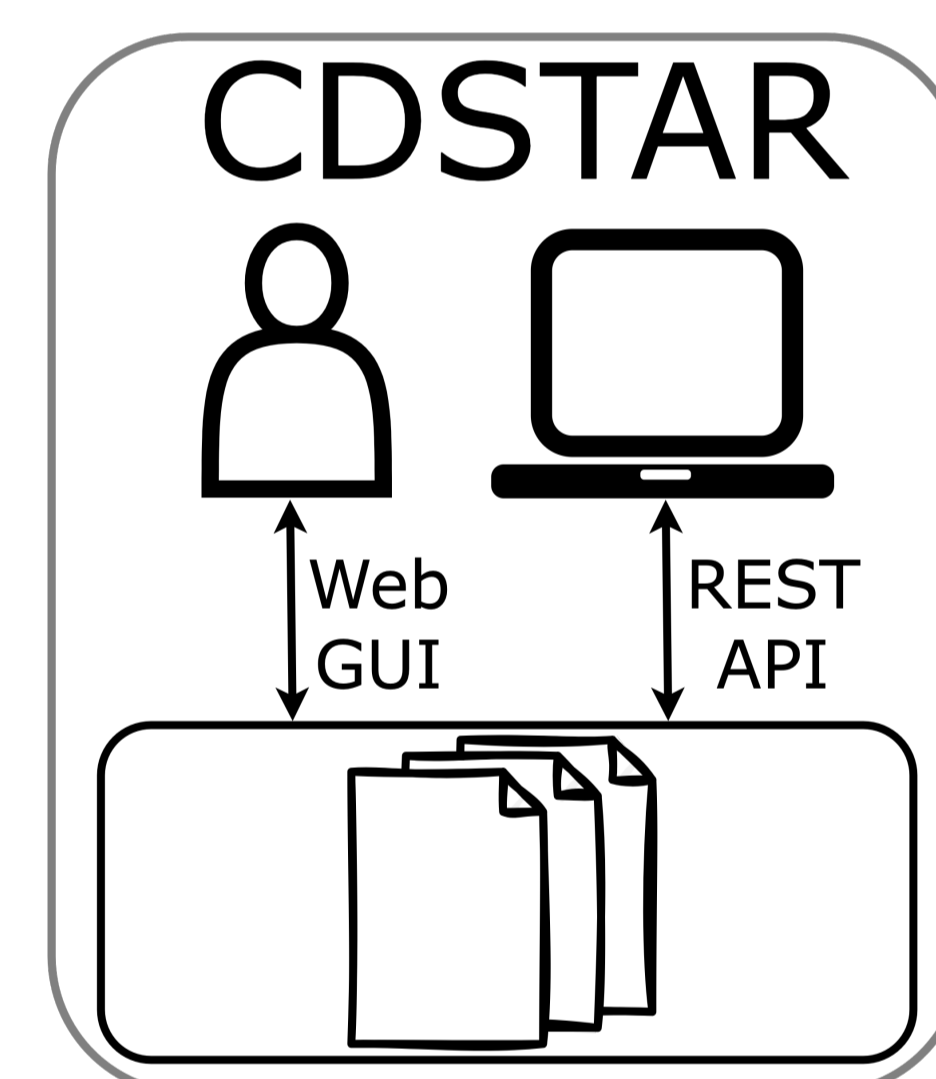
Furnishing of a central infrastructure (or networking of peripheral infrastructures via a central contact point) for the storage and the archiving of research data, with suitable user workflows.  
<https://www.text-plus.org/en/research-data/user-story-316/>  
Infrastructural Needs for Romance Research Data

Text+ could provide a repository for long term archiving of the data.  
<https://www.text-plus.org/en/research-data/user-story-334/>  
The history of the "Deutschland-Institut" in Beijing (1932–1950): an example for Sino-German academic collaboration

Creating an institution for securing and making available our research data for the long term is a matter of existential importance for the future of our discipline. The information we generate (and which should be saved for the long term) could be interconnected with databases of other German institutions, which otherwise might get lost.  
<https://www.text-plus.org/en/research-data/user-story-342/>  
Epigraphy at the BBAW

Text- and language-based data, both written and oral, has to be collected and if applicable converted into digital form, transcribed, and translated. The long-term preservation must be ensured. This is especially challenging for private websites and Facebook posts in different languages and scripts.  
<https://www.text-plus.org/en/research-data/user-story-343/>  
Knowledge Relations between Polish People's Republic and Iraq in Architecture and Planning

## Architektur, Features, Bedienung



### CDSTAR

- Repository-Software
- die Interaktion mit dem Langzeitarchiv erfolgt ausschließlich über CDSTAR
- bedienbar über eine REST-API, sowie über eine einfache eigene Web-basierte GUI mit komfortablen Möglichkeiten zum
  - Hochladen
  - Vergeben von Metadaten
  - Suchen
  - Herunterladen
- kann die Daten ins Langzeitarchiv übertragen und vom Archiv holen aber auch die Daten bereithalten, wie ein Repository
- es lassen sich sowohl offene als auch geschützte Archivbereiche realisieren

### CDSTAR-Koala-Connector

- aktuell in Entwicklung
- übersetzt zwischen dem internen CDSTAR- und dem internen Koala-Format
- übergibt die Daten von CDSTAR an Koala und umgekehrt

### Koala

- Software für Langzeitarchivierung
- sorgt für die effiziente und sichere Übertragung der Daten aus dem/in den Storage, für die dauerhafte Integrität der Daten
- die Langzeitarchivierungslösung der Deutschen Nationalbibliothek basiert auf Koala

Weiterführende Links:

[cdstar.gwdg.de](https://cdstar.gwdg.de)  
[koala-docs.gwdg.de/de](https://koala-docs.gwdg.de/de)  
<https://www.dnb.de/DE/Professionell/Erhalten/LZA-System/lza-system.html>