

GeSIG

Ein Projekt im Spannungsfeld einer sich entwickelnden Infrastrukturlandschaft

Cordula Meißner / Franziska Wallner

Tagung Forschungsdaten in den Geisteswissenschaften (FORGE)

Universität Hamburg, 18.09.2015

Gliederung

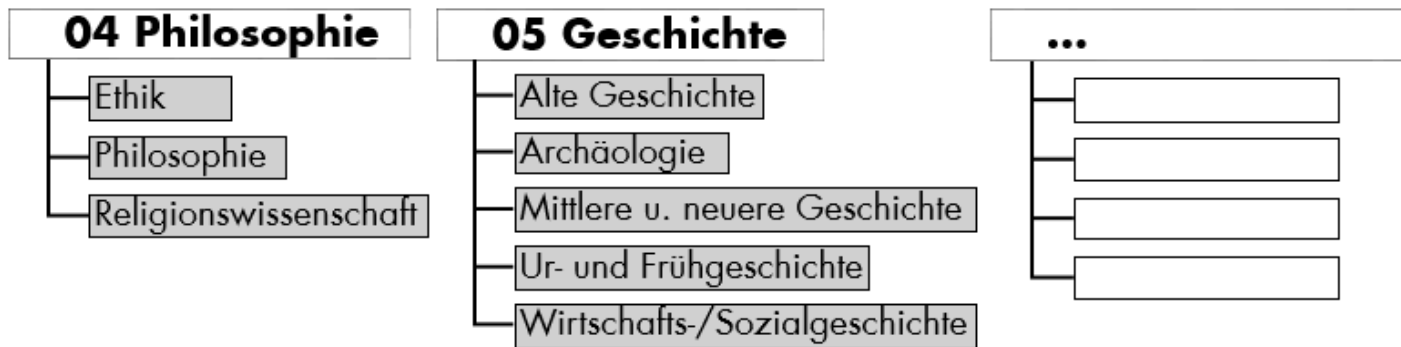
1. GeSIG

2. Ein Projekt im Spannungsfeld einer sich entwickelnden Infrastrukturlandschaft

1.1 GeSIG - Pilotphase (2015-2017)

Allgemeine Wissenschaftssprache





Begriff_NN
 Beweis_NN
 Frage_NN
 fragen_V
 fraglich_ADJ
 ...

Begriff_NN
 Münze_NN
 Frage_NN
 fragen_V
 fraglich_ADJ
 ...

...



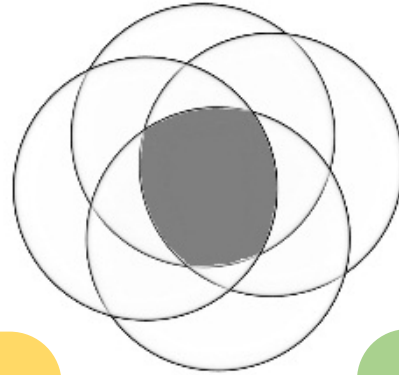
19 geisteswissenschaftliche Fachbereiche

19 Teilkorpora je mindestens 10 Dissertationen 1 Mio. Token getaggt lemmatisiert

eine Lemmaliste je Fachbereich

Schnittmenge aus 19 Fachbereichslisten

1.2 GeSIG - das lexikografische Projekt (ab 2018?)

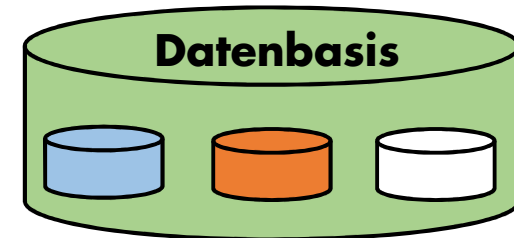
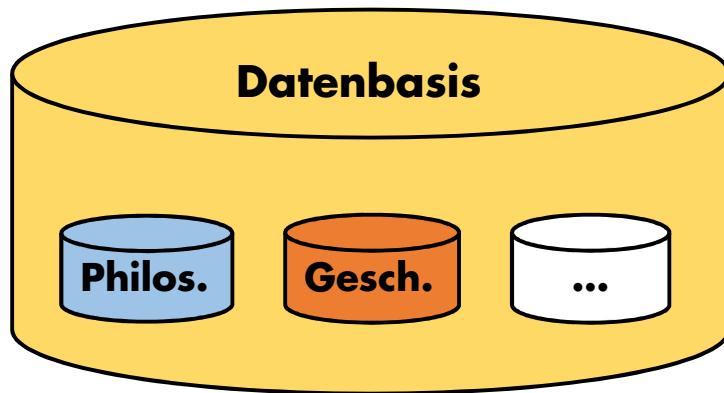


Lexikografische Ressource

Kollokationen / Kontextmuster /
Konstruktionen / Gebrauchsspezifika /
Fachspezifik u.a.

Korpuslinguistische Analyseplattform

Konkordanzsuche / N-Gramm-Analyse /
Filterung nach Fach und Textsorte u.a.



Anforderungen: - Zugänglichkeit
- Ausgewogenheit
- Größe

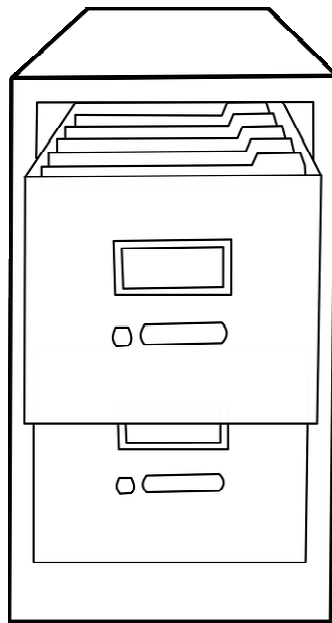
Anforderungen: - Nachnutzbarkeit
- Ausgewogenheit

Die geplante Ressource ...

- schafft ein Informationssystem für die Vermittlung der deutschen Wissenschaftssprache
- erschließt jenen Bereich, der für den Transfer zwischen Wissenschaft und Gesellschaft besonders relevant ist
- bietet den Geisteswissenschaften eine Basis
 - für die Reflexion der sprachlichen Voraussetzungen der Erkenntnisgewinnung sowie
 - für den interdisziplinären Austausch

2 GeSIG – im Spannungsfeld einer sich entwickelnden Infrastrukturlandschaft

Langfristige Zugänglichkeit der Projektergebnisse



Quelle:
https://pixabay.com/static/uploads/photo/2013/07/12/14/44/cabinet-148680_640.png - CC0

alle Teilkorpora | nutzen | Verbspur

1 2

1 - 50 von 59 Gesamttreffern

Kommunikati	Sprecher	Linker Kontext	Treffer	Rechte
PG_DE_131	LG_0306	fiktionales tage buch is kann ma n das eben gut	nutzen	ja ja ((lacht)) ich hab gar
PG_DE_131	LG_0306	sacht ähm (0.2) ah das könn wa doch gut	nutzen	das waren die wa die gleich a
PG_PL_062	MZ_0454	können wir die sistuatio äh diese information (auch/aus) benutzen (aus	nutzen	ja) hm_ hm
PG_UK_003	KW_060 6	nur dreiundfünfzig komma vier p rozent der wahlberechtigten (0.3) machten	nutzen	(0.3) von ihre 5) ein mögliche
PG_UK_016	PM_0622	stimmung in dermalgen (.) deut schland zu d deuten und	nutzen	und °h hat eii t (.) orientierte
PG_UK_014	KS_0605	(.) was sind ihre empfehlungen w ist diese mein das was	nutzen	°hh

<https://gewiss.uni-leipzig.de/index.php?id=search>

Langfristige Zugänglichkeit der Projektergebnisse

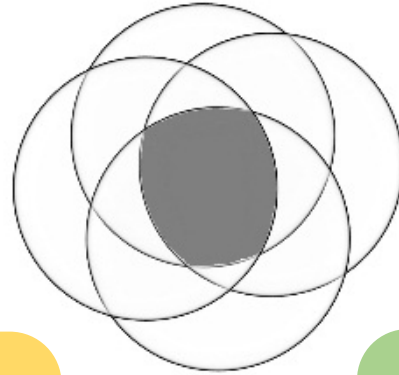
Wie kann eine langfristige Nutzung der Projektergebnisse sichergestellt werden, wenn der Zugang an ein Portal gebunden ist?

Nutzungsrechte

als Datenbasis sollen ausschließlich Open Access Publikationen genutzt werden

Hier sind zwei Arten zu unterscheiden:

1. Texte, die vom Urheber mit einer Lizenz versehen sind, die für jedermann die Nutzungsmöglichkeiten klärt (CC-Lizenz)
2. Texte, die nicht lizenziert sind und somit dem Urheberrecht unterliegen

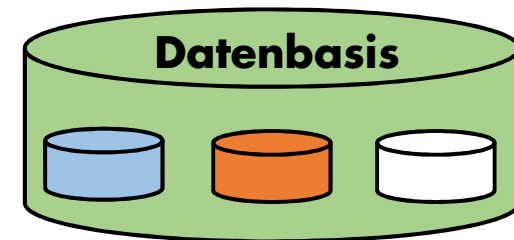
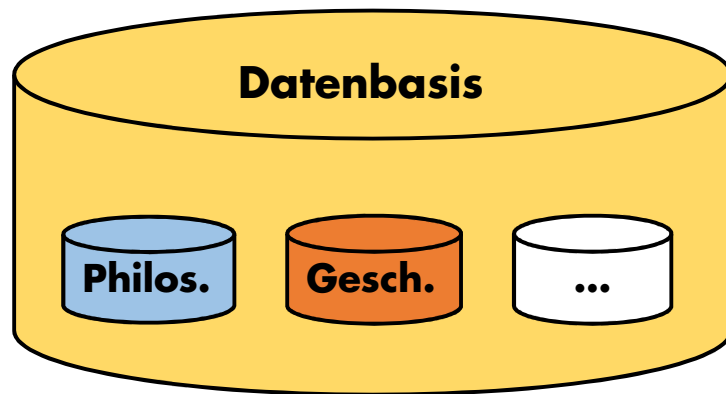


Lexikografische Ressource

Kollokationen / Kontextmuster /
Konstruktionen / Gebrauchsspezifika /
Fachspezifik u.a.

Korpuslinguistische Analyseplattform

Konkordanzsuche / N-Gramm-Analyse /
Filterung nach Fach und Textsorte u.a.

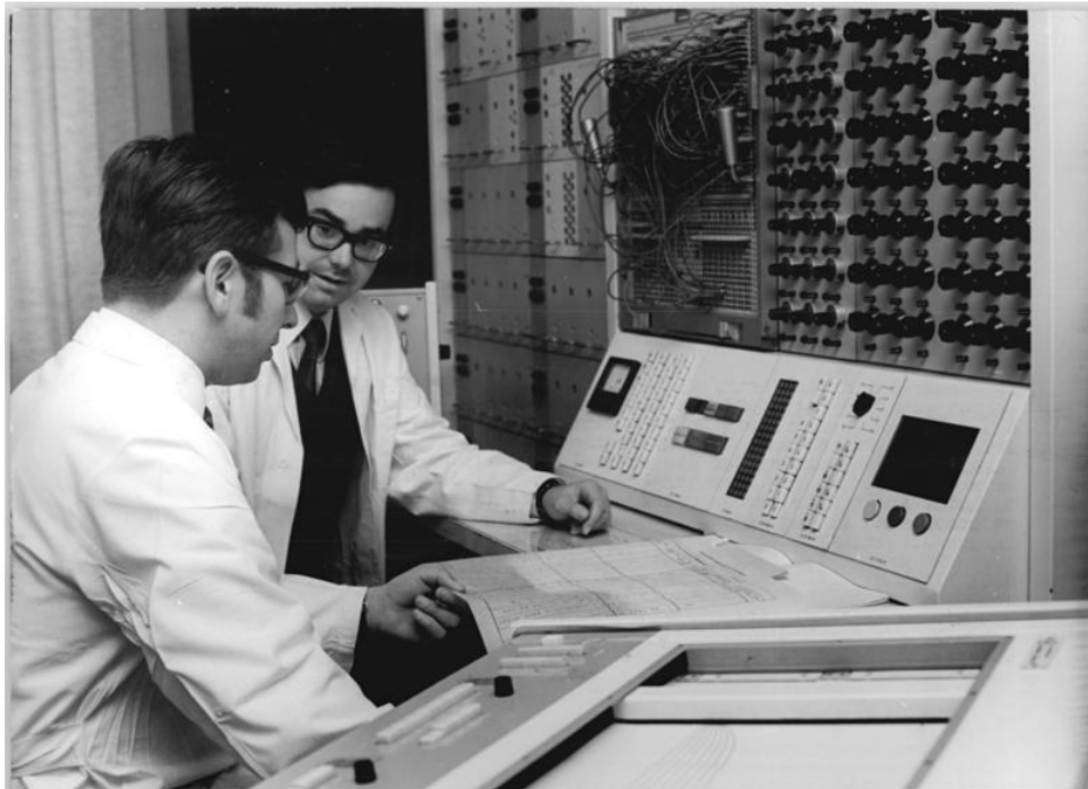


Anforderung: CC-Lizenz, die eine
Veröffentlichung im
Korpus ermöglicht

Nutzungsrechte

Welche Nachnutzungsmöglichkeiten bestehen für die Datenbasis der lexikografischen Ressource?

Zugänglichkeit sichernde Einrichtungen



Bundesarchiv, Bild 183-N0813-0307
Foto: Link, Hubert | 13. August 1974

Quelle: Bundesarchiv, Bild 183-N0813-0307 / CC-BY-SA 3.0

Zugänglichkeit sichernde Einrichtungen

Wie sollte ein Projekt unter den aktuellen Bedingungen bei der Wahl eines Datenzentrums vorgehen?

Vielen Dank für die Aufmerksamkeit!

Diskussion

- Wie kann eine langfristige Nutzung der Projektergebnisse sichergestellt werden, wenn der Zugang an ein Portal gebunden ist?
- Welche Nachnutzungsmöglichkeiten bestehen für die Datenbasis der lexikografischen Ressource (unlizenzierte Texte)?
- Wie sollte ein Projekt unter den aktuellen Bedingungen bei der Wahl eines Datenzentrums vorgehen?