

Fotis Jannidis

# **eScience / Grid: Anforderungen und Potential aus Sicht der Geisteswissenschaften**

Dini Tagung 2007



# The ARCHway Project

Architecture for Research in Computing for the Humanities  
through collaborative research, teaching, and learning

Principal Investigators

*Kevin Kiernan, Department of English*

*Jerzy W. Jaromczyk, Department of Computer Science*

*Alexander Dekhtyar, Department of Computer Science*

University of Kentucky

---

## **Publications, Presentations, and related Theses, Dissertations, and Student Projects**

---

The ARCHway Project was funded by the [Information Technology Research](#) program of the [National Science Foundation](#). The project set four high-level objectives: (1) to develop a universal software platform for creation and maintenance of Image-based Electronic Editions (IBEE); (2) to develop methods, algorithms, and data structures for management of concurrent, document-centric, XML data; (3) to develop methods, algorithms, and data structures for management of image-based XML encoding; and (4) to establish a teaching and learning infrastructure fostering interdisciplinary collaboration for graduate and



# The Perseus Digital Library

Gregory Crane, Editor-in-Chief, Tufts University

Search

[Configure Display](#) - [Help Library](#) - [Tools](#) - [Publications](#) - [Collaborations](#) - [Support Perseus](#)

[text-only home page](#)

## Perseus Contents

[Classics](#)  
Greek, Latin,  
Archaeology  
[Papyri](#)  
Duke Data Bank

[English](#)  
[Renaissance](#)  
Shakespeare,  
Marlowe, ...

[London](#)  
Bolles Collection

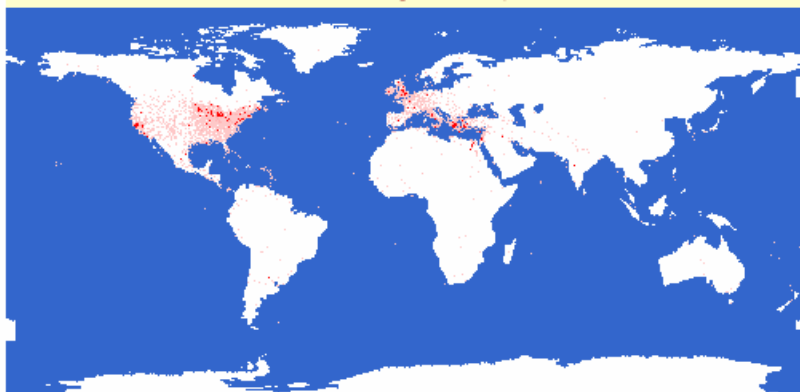
[California](#)  
[Upper Midwest](#)  
[Chesapeake](#)  
Library of  
Congress

[Tufts History](#)  
Since 1852

[Boyle Papers](#)  
History of  
Science

Home site:  
[Somerville, MA](#)  
Mirror sites:  
[Berlin, Germany](#)  
[Chicago, IL](#)

## Perseus Digital Library



A [graph of the places and dates](#) mentioned in this collection

## About Perseus

Perseus is an evolving digital library, engineering interactions through time, space, and language. Our primary goal is to bring a wide range of source materials to as large an audience as possible. We anticipate that greater accessibility to the sources for the study of the humanities will strengthen the quality of questions, lead to new avenues of research, and connect more people through the connection of ideas.

## Announcements

- **Hardware problems!**: A key server in the Perseus site has experienced intermittent hardware problems. We are in the process of upgrading this hardware and creating a more stable hardware environment. We apologize for the delays and hope to have the problem resolved in a few days.
- **[Perseus 4.0](#) released -- a new implementation of the Perseus Digital Library.**

Perseus 4.0, a new Java-based version of the Perseus Digital Library, is available for testing. It contains a faster, more manageable back-end and a more modern look and feel. Many features of Perseus are now available as XML services -- for example, developers can extract [well-formed XML fragments](#) of primary sources with full TEI-conformant markup in order to create their own front ends. [Read more...](#)

- What did an educated Roman know about the world? We've added Pliny the Elder's encyclopedic [Natural History](#), ([search](#)), a remarkable snapshot of the state of Geography, Ethnography, Astronomy, Biology, and Geology in the early Roman empire.
- New reference works: Smith's [Dictionary of Greek and Roman Antiquities](#)

## Welcome

## Welcome to TAPoR

TAPoR is a gateway to tools for sophisticated analysis and retrieval, along with representative texts for experimentation.

## Features of TAPoR

- Manage electronic texts
- Experiment with online text tools
- Learn about digital textuality

[Sign Up for an Account Here](#)

## Want to Learn More?

Take a tour:

- [About TAPoR Project](#)
- [TAPoR Tour](#)
- [TAPoR Tutorial](#)
- [WikiTADA](#)
- [What is text analysis?](#)
- [Text Analysis Recipes](#)

Browse the tutorial:

- [What's the point?](#)
- [Logging-in to the TAPoR Portal](#)
- New:** [Watch a Training Video](#)

## Try It

**1 Choose**  
a text to analyze

- Universal Declaration of Human Rights**
- Moby Dick (Chapter 1)**  
by Herman Melville
- Hyperliste (Des vilains)**  
French Medieval Poetry
- Your Web Page**  
Enter the URL below

**2 Choose**  
a tool to run

- View**  
the text
- List Words**  
in the text
- Explore with HyperPo**  
an interactive concordance
- Find a Word**  
or pattern of words

**3 Run Tool on Text**

## TAPoR Tools

Browse or search for tools that you can use. [More Info](#)



- [Browse by type](#)
- [Browse by Source Format](#)
- [New And Popular](#)
- [All A-Z](#)

## TAPoR Texts

Browse or search for text that you can use. [More Info](#)



- [Browse by Tags](#)
- [Browse by File Format](#)
- [Browse All](#)

## Credits

## TAPoR News

## General

[Call for International Network of Digital Humanities Centres](#)  
Are you involved in a digital humanities centre? If so you should consider responding to a [Call for Participation in an International Network of Digital Humanities Centers](#).

[Announcing TAPoR Version 1.0](#)

The TAPoR portal was upgraded on April 20th to Version 1.0 after three major public testing releases.

Try the portal out at [portal.tapor.ca](http://portal.tapor.ca) or check out the new [Training Videos](#) for an overview.

[Call for Papers for the second Chicago Colloquium on Digital Humanities and Computer Science](#)

The Illinois Institute of Technology, Northwestern University, and the University of Chicago are jointly sponsoring the second Chicago Colloquium on Digital Humanities and Computer Science to be held on October 21-22, 2007 at Northwestern University.

### Progress Status

- Retrieving URL...
- Text is being prepared....
- Text is being parsed....



### Report on the Text

- Text Source: [http://tapor2.mcmaster.ca/portal/coplets/frontpage/tryit/texts/moby\\_dick.xml](http://tapor2.mcmaster.ca/portal/coplets/frontpage/tryit/texts/moby_dick.xml)
- Text Title: Untitled
- Language of text: Other / Unspecified
- Total words (tokens): 2255
- Unique words (types): 859
- Highest word frequency: 124
- Average word frequency: 2.63
- Standard Deviation of word frequencies: 7.25
- Average word length: 4.30
- Standard Deviation of word lengths: 2.37
- Number of sentences: 102
- Average words per sentence: 22.1
- Number of paragraphs: 16
- Average words per paragraph: 140.9


Continue

breezes, which a few hours previous were out of sight of there.

Clicking Action: [Go to Options](#)  
 jump to next occurrence

Circumambulate the city of a dreamy Sabbath afternoon. Go from Corlears Hook to Coenties Slip, and from thence, by Whitehall, northward. What do you see? - Posted like silent sentinels all around the town, stand thousands upon thousands of mortal men fixed in ocean reveries. Some leaning against the spiles; some seated upon the pier-heads; some looking over the bulwarks of ships from China; some high aloft in the rigging, as if striving to get a still better seaward peep. But these are all landsmen; of week days pent up in lath and plaster - tied to counters, nailed to benches, clinched to desks. How then is this? Are the green fields gone? What do they here?

But look! here come more crowds, pacing straight for the water, and seemingly bound for a dive. Strange! Nothing will content them but the extremest limit of the land; loitering under the shady lee of yonder warehouses will not suffice. No. They must get just as nigh the water as they possibly can without falling. And there they stand - miles of them - leagues. Inlanders all, they come from lanes and alleys, streets avenues - north, east, south, and west. Yet here they all unite. Tell me

Options  
 Retrieve Text Segement (1000 elements)  
 • Next Segment  
 Teleporter?  
  
 Advanced Options Show >>

Clicking Action: [Go to Options](#)  
 cooccurrences

all	23	1
you	23	1
this	17	7.54
for	16	7.10
there	16	7.10
but	15	6.65
my	14	6.21
sea	13	5.76
with	13	5.76
go	12	5.32
on	12	5.32
they	12	5.32
from	11	4.88
not	11	4.88
total?	846	375.17
average?	32.54	14.43

View results in [XML?](#) [CSV?](#) or [text?](#)

Options  
[show most frequently repeating words](#)  
[show most frequent repeating series of words](#)  
[show longest repeating series of words](#)  
 Advanced Options Show >>


Information?

Clicking Action: [Go to Options](#)  
 jump to first occurrence

Corpus Information [add corpus?](#)

1) Text Source: [http://tapor2.mcmaster.ca/portal/coplets/frontpage/tryit/texts/moby\\_dick.xml](http://tapor2.mcmaster.ca/portal/coplets/frontpage/tryit/texts/moby_dick.xml) "Untitled"


	raw?	relative?	context?	relative?	expected?
to	53	23.50	12	2.82	100
a	69	30.60	11	3.67	91.67
as	26	11.53	9	1.38	75
i	43	19.07	9	2.29	75
sea	13	5.76	7	0.69	58.33
passenger	4	1.77	4	0.21	33.33
sailor	5	2.22	3	0.27	25
the	124	54.99	3	6.60	25
always	2	0.89	2	0.11	16.67




# hyperNietzsche


(Version 0.6)


[What is HyperNietzsche?](#)


 [Submit a new contribution](#)


[Search in HyperNietzsche](#)


[Manuscripts](#) 


[transcriptions](#) 

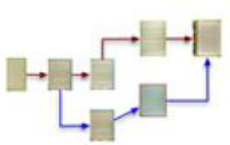
[Works](#) 

[Editions](#) 

[Authors](#) 

[Essays](#) 

[Paths](#) 

[Rhizomes](#) 

- Deutsch
- English
- Español
- Français
- Italiano
- Português

list of all the HyperNietzsche's editions, transcriptions, essays

Personal Folder New Folder wtkn

 Notebooks
  Copybooks
  Dossiers and loose sheets
  Printer's copy
  Proof sheets
  Private publications

## Notebooks

## Related Contributions



§N-I-3 description

### Facsimile

### transcriptions



Notebook (6,5 × 9cm) bound in yellow wood. 64 mainly filled pages. Pencil, mainly latin handwriting; *Memorabilia*; Aphorisms, mainly for *Human, All Too Human*, Location: Weimar, Goethe- und Schiller-Archiv, 71/178

### Facsimile

### transcriptions

73 Facsimiles

71 Transcriptions



Octavo notebook (10,5 × 17,5cm). 50 pages. Pencil. German handwriting. Notes for *The Wanderer and his Shadow*, July 1879. Location: Weimar, Goethe- und Schiller-Archiv, 71/186.

### Facsimile

### transcriptions

264 Facsimiles

151 Transcriptions



Octavo notebook (10,5 × 17,5cm). 50 pages. Pencil. Mainly German handwriting. Notes for *The Wanderer and his Shadow*, June-July 1879. Location: Weimar, Goethe- und Schiller-Archiv, 71/187.

### Facsimile

### transcriptions

106 Facsimiles

106 Transcriptions



Octavo notebook (10,5 × 17,5cm). 50 pages. Pencil. German handwriting. Notes for *The Wanderer and his Shadow*. July-August 1879. Location: Weimar, Goethe- und Schiller-Archiv, 71/188.

### Facsimile

### transcriptions

115 Facsimiles

132 Transcriptions





N IV 4,49[5]



Mp XIV 2,65v[5]



D 12,5r[1]



K 6,7[1]

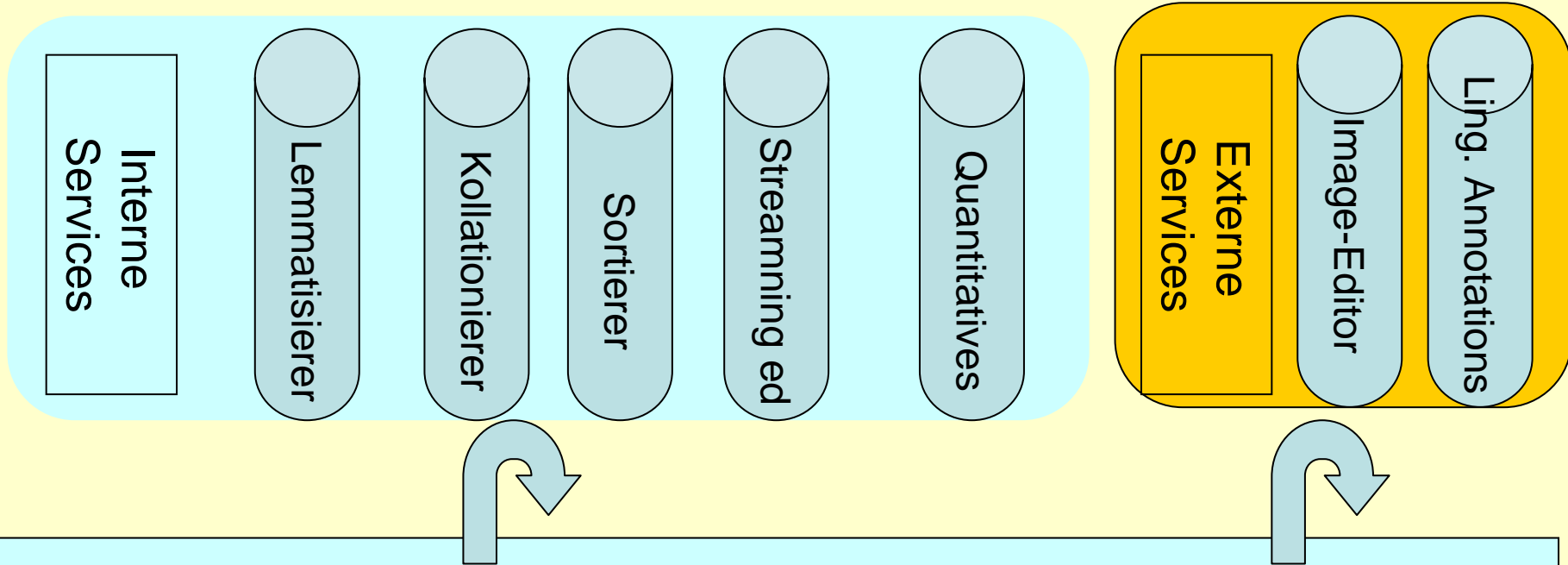


WS-1



Name	Surname	Title	Language	Status
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text" value="published"/>
<input checked="" type="checkbox"/> Essays	<input type="checkbox"/> Editing	<input type="checkbox"/> Bibliographical Cards	<input type="checkbox"/> transcriptions	<input type="checkbox"/> Facsimile
<input type="checkbox"/> Editions	<input type="checkbox"/> Path	<input type="checkbox"/> Commentary	<input type="checkbox"/> \$review	<input type="checkbox"/> \$tran

Siglum	Name	Surname	Title	Type/Subty
mmontinari-11	Mazzino	Montinari	Nietzsche mit Goethe in Italien	essay
mmontinari-13	Mazzino	Montinari	Nietzsche in Cosmopolis	essay
mmontinari-15	Mazzino	Montinari	Nietzsche e la décadence	essay
mmontinari-17	Mazzino	Montinari	Nietzsche - Hillebrand	essay
mmontinari-19	Mazzino	Montinari	Die spröde Art, Nietzsche zu lesen: die Niederschrift von "Also sprach Zarathustra" am Beispiel des Kapitels "Auf den glückseligen Inseln"	essay
mmontinari-3	Mazzino	Montinari	Nietzsche mit Goethe in Italien	essay
mmontinari-5	Mazzino	Montinari	Nietzsche in Cosmopolis	essay
mmontinari-7	Mazzino	Montinari	Nietzsche e la décadence	essay
mmontinari-9	Mazzino	Montinari	Nietzsche - Hillebrand	essay
pdiorio-1	Paolo	D'Iorio	Genèse, parodie et modernité dans "Ainsi parlait Zarathoustra"	essay
rmullerbuck-1	Renate	Müller-Buck	"... Zeit, daß ich noch einmal als Franzose zur Welt komme", Nietzsche auf der Suche nach französischen Lesern	essay
sbarbera-1	Sandro	Barbera	Eine Quelle der frühen Schopenhauer-Kritik Nietzsches: Rudolf Hayms Aufsatz "Arthur Schopenhauer"	essay
vvivarelli-1	Vivetta	Vivarelli	"Aggiungo una punta di comicità alle cose più serie": aspetti delle ultime lettere di Nietzsche	essay
gabel-1	Günter	Abel	Wahrheit und Interpretation	essay
mbrusotti-3	Marco	Brusotti	"Elemente der Rache", Sprache und Selbsterkenntnis bei Nietzsche im Aphorismus 33 von "Der Wanderer und sein Schatten"	essay
mmontinari-21	Mazzino	Montinari	Zum Verhältnis Lektüre-Nachlaß-Werk bei Nietzsche	essay
oponton-1	Olivier	Ponton	"Mitfreude", Le projet nietzschéen d'une "éthique de l'amitié" dans "Choses humaines, trop humaines"	essay
pjmvantongeren-1	Paul J. M.	van Tongeren	Der freie Geist und die Demokratie	essay
sbarbera-3	Sandro	Barbera	Il 'buon europeo' dell'Archivio Nietzsche. Momenti del culto di Nietzsche all'inizio del Novecento	essay
ahartmanncavalcanti-1	Anna	Hartmann Cavalcanti	"Fragmento 12(1): a crítica do jovem Nietzsche à metafísica da vontade de Schopenhauer"	essay



## Ressourcen

Katalogdaten  
 Goethe: Werther  
 Schiller: Wallenst  
 ....

Faksimiles

Volltext – Struktur

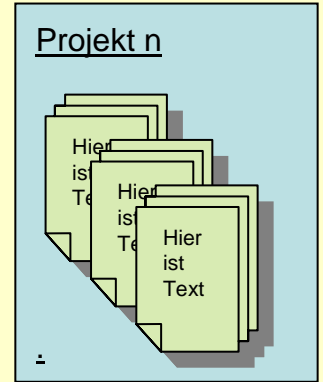
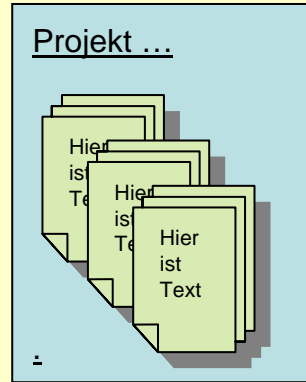
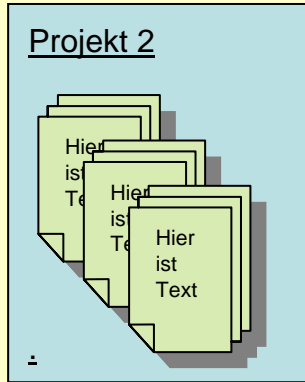
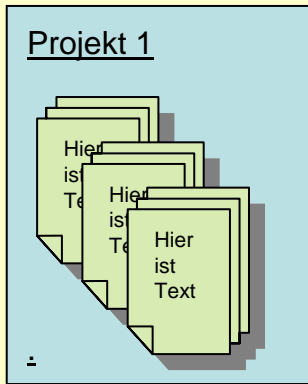
Volltext  
 -Lemmatisiert  
 -Morpho-syntakt.  
 -Buchgeschichtl.  
 -Named Entities  
 -Narratologie  
 -Thematische Ausz.  
 -- ...

Hilfsmittel  
 -Wörterbücher  
 -Biograph. DB  
 -Lexika / Enzykl.  
 -...

Externer Film: textgridlab.avi



Suche in allen Texten.  
Wieviel Semantik kann man bewahren?

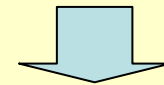
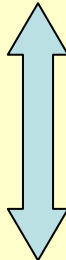
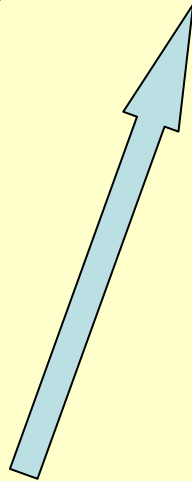
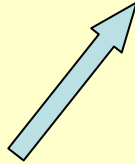
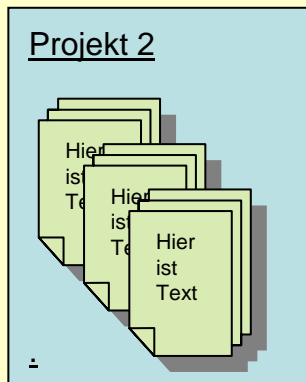
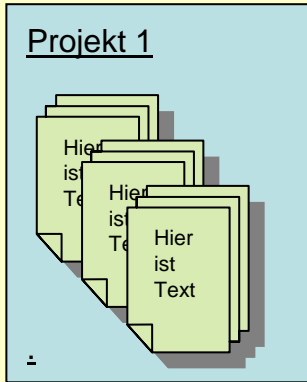


1. Hole den 1. Teil der Fassung des Werthers von 1774

2. Hole den 1. Teil der Fassung des Werthers von 1786

3. Vergleiche die Fassungen

4. Erzeuge einen Apparat in einer neuen Edition



# Anforderungen / Probleme

- Langfristige Archivierung der Daten (Volltexte, nicht nur Images)
- Langfristige Archivierung der Metadaten
- Langfristige Gültigkeit der Adressen

# Daten sind Services

- Viele wichtige Services sind definierte Schnittstellen zu Daten.
- Bsp. Wörterbuch:
  - Strukturierte Suche in einem Wörterbuch (z.B. Suche alle Einträge, die umgangssprachliche Bedeutungen definieren)
  - Strukturierte Suche über mehrere Wörterbücher hinweg (z.B. alle Einträge zum Lemma ‚Frieden‘)



# Zentrale Datenressourcen

- Bibliographische Metadaten
- Wörterbücher
- Enzyklopädien
- Personennamen (Verzeichnisse, Erl.)
- Ortsnamen (Verzeichnisse, Erl.)
- Karten
- Umfassende Volltextkorpora

# Services

- Services, die zentrale fachspezifische Ressourcen erschliessen
- Internationaler Verbund vertrauenswürdiger Register solcher Services
- Schnittstellen für Services und Register
- Fachspezifische Anreicherung vorhandener Services
- Kooperation zwischen Service-Entwicklern und langfristigen Service-Providern

# Services – Daten: Baseline Markup

- Textsortenspezifisches Markup (TEI, METS)
- 3 Funktionen
  - Textsortenspezifische häufige Suche
  - Präsentation der wichtigsten Differenzen
  - Interoperabilität via Services

# Netzeffekte

- Netzeffekte ermöglichen: Ressourcen zur Weiterverwendung anbieten (Image, Volltext, Strukturerschliessung, Semantisches Markup usw.)
- Kooperatives Arbeiten an zentralen Datenbanken, Bsp. Personendatenbank als Wissenschaftswiki

# Linking und Extraktion

- Auf Textebene
- Auf Fragmentebene
- Differenzierung und Abstraktion von
  - Werk
  - Ausgabe
  - Digitalisat

Canonical Text Services Protocol

# Organisatorisches

- Internationale Offenheit / Anschliessbarkeit (Standards + Kooperation)
- Niedrige Einstiegskosten für die Weiterentwicklung von Daten / Services (Open Source)
- Juristische Rahmenbedingungen: (Copyright, Freigabe der Rechte am Text)
- Nachhaltigkeit durch transparente Finanzierung in den *Humanities*

**Vielen Dank!**