Der Aufbau von digitalen Forschungsinfrastrukturen für die Geistes- und Kulturwissenschaften in Österreich

Karlheinz Mörth

Institute for Corpus Linguistics and Text Technology Austrian Centre for Digital Humanities Austrian Academy of Sciences

karlheinz.moerth@oeaw.ac.at

















Institut für Corpuslinguistik und Texttechnologie





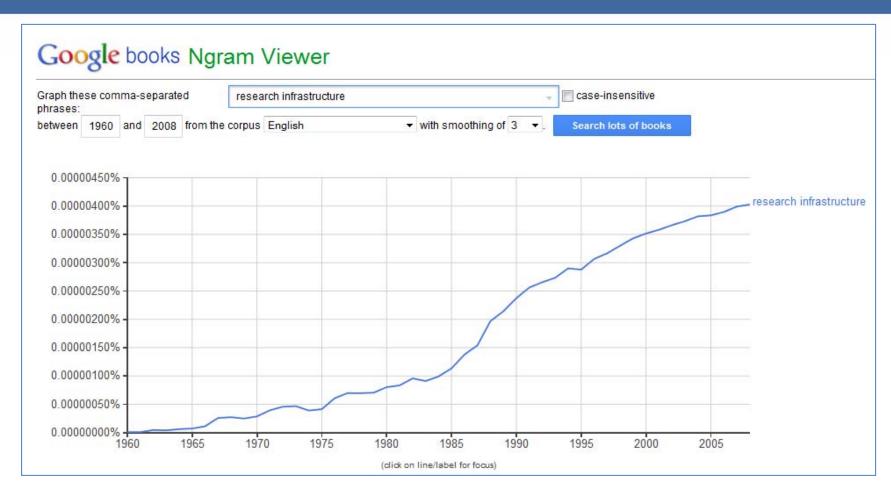
Topics



ESFRI + DH + Austria
Digital Humanities Austria
Ongoing activities and projects

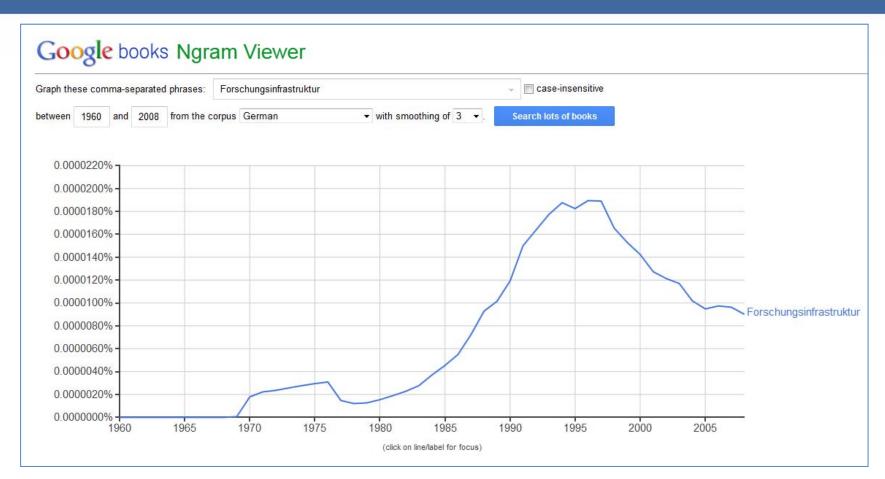
Background ,Word watching'





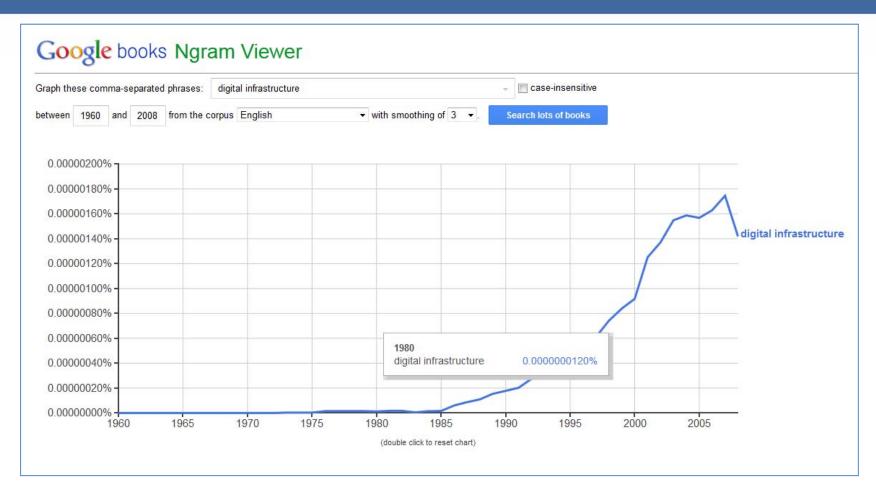
Background ,Word watching





Background ,Word watching'





Background ,Word watching'



Google books Ngram Viewer				
Graph these comma-separated phrases: Digitale Infrastrukturen		→ case-insensitive		
between 1960 and 2008 from the c	orpus German ▼ with	▼ with smoothing of 3 ▼. Search lots of books		
		No valid ngrams to plot!		
		Ngrams not found: Digitale Infrastrukturen		

Background ,Word watching





user: icltt corpus: amc

Concordance Word List

?

Save View options KWIC Sentence

Sort Left Right Node References Shuffle Sample Filter

Doc IDs Collocations ConcDesc

Frequency

Node forms

Query Forschungsinfrastrukturen 129 (0.0 per million)

age 1 of 7 Go Next | Last

DIEWIR 200810011230290034 spezifische Programm " <g/> Kapazität <g/> " - Forschungsinfrastrukturen genehmigte Projekte " <g/> Verwaltungskapazität DIEWIR 200605031949110041 Im Bereich " <q/> Kapazitäten <q/> " sollen Forschungsinfrastrukturen und Klein-und Mittelbetriebe mit 4,2 Milliarden DIEWIR 200612061651000016 Im Bereich " <q/> Kapazitäten <q/> " sollen Forschungsinfrastrukturen sowie Klein-und Mittelbetriebe mit insgesamt DIEWIR 200307011556160018 Wissenschaftler aus Drittstaaten zu erhöhen <g/> , Forschungsinfrastrukturen <g/> ; Maßnahmen zur Förderung der optimalen DIEWIR 200307011556160018 zur Förderung der optimalen Nutzung der Forschungsinfrastrukturen und zur Unterstützung der Bestimmung <e/>
<e/separatrukturen und zur Unterstützung der Bestimmung <e/separatrukturen und zur Unterstützung <e/separatruk ECHO 201002261847530065 ihrerseits den nachhaltigen Auf- und Ausbau von Forschungsinfrastrukturen innerhalb des Förderungsgebietes bewirken FORMATDB_201108191325090040 zwar eine durchaus akzeptable Anzahl an Forschungsinfrastrukturen auf <g/>
<g/>
. Im Bereich der Großgeräte <g/>
<g/>
, FORMATDB 201108191325090040 für die Anschaffung und den Betrieb von Forschungsinfrastrukturen gebildet <g/> . Dadurch kann nicht immer FORMATDB_201108191325090040 größere Einrichtungen generieren können «g/». Forschungsinfrastrukturen werden von wissenschaftlichen Organisationen FORMATDB 201108191325090040 sinnvoll ist - auch grenzüberschreitenden Forschungsinfrastrukturen <g/> , International geht der Trend in Richtung Ergebnis <g/> : Es gibt in Österreich 1.024 Forschungsinfrastrukturen <g/> · , davon 279 Großgeräte oder Großanlagen FORMATDB_201108191325090041 FORMATDB 201108191325090041 309 externe nationale und internationale Forschungsinfrastrukturen sowie 364 bereits bestehende Planungen <q/><q/>/> FORMATDB_201108191325090041 <q/>, Mehr als die Hälfte aller erhobenen Forschungsinfrastrukturen sind den Hochschulen zuzuordnen <q/><q/>). Die KLEINE 20101111035503020160013 Zeit erhebliche Mittel für den Aufbau von Forschungsinfrastrukturen akquirieren konnten <q/> , Die nächsten Ziele KLEINE_20101111035503020160024 Zeit erhebliche Mittel für den Aufbau von Forschungsinfrastrukturen akquirieren konnten <g/>
«g/». Die nächsten Ziele KURIER 201009201832270014 einer Studie in diesem Zeitraum rein für Forschungsinfrastrukturen einen Bedarf von rund 660 Millionen Euro PRESSE 201004181916400086 technologisch rasch veränderlichen Zeit müssen auch Forschungsinfrastrukturen neu bewertet und angepasst werden <q/><q/>>. PRESSE 200701261852060117 sogenannte " <g/> Kapazitäten <g/> " zu (<g/> 1 <g/>) Forschungsinfrastrukturen <g/> , (<g/> 2 <g/>) Forschung zu Gunsten von

Mittel werden insbesondere für den Ausbau der Forschungsinfrastrukturen verwendet <g/> " <g/> , so Sausgruber <g/> .

Mittel werden insbesondere für den Ausbau der Forschungsinfrastrukturen verwendet <g/> " <g/> , so Sausgruber <g/> .

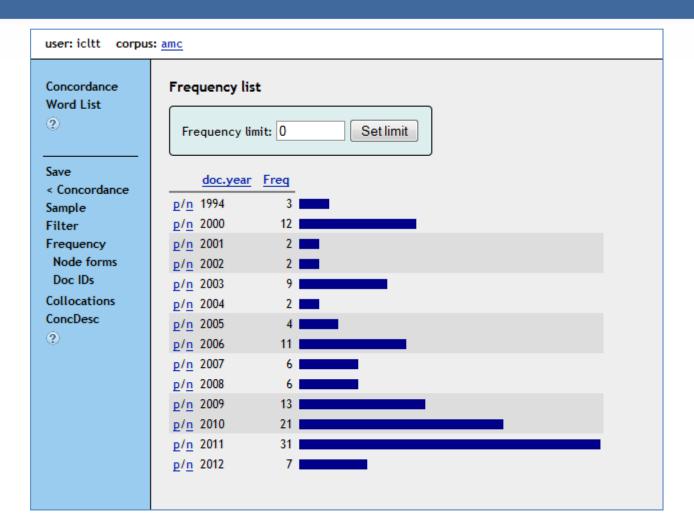
Page 1 of 7 Go Next | Last

VN_200304180746180296

VN 200304180954090296

Background ,Word watching'







INFRASTRUCTURES







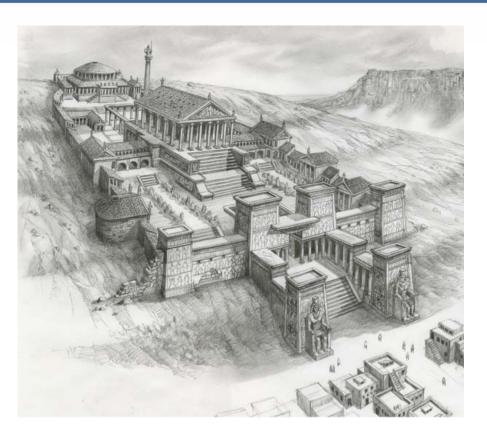


DHA

RESEARCH INFRASTRUCTURES

Archive Museen Gallerien

Bibliotheken



Μουσεῖον τῆς Ἀλεξανδρείας



DIGITAL RESEARCH INFRASTRUCTURES

European Strategy Forum on Research Infrastructures

- was founded in 2002.
- brings together national delegates and a representative of the Commission, working together to pool resources to provide Europe with the most up-to-date research infrastructures.
- is a strategic instrument to develop the scientific integration of Europe and to strengthen its international outreach.
- identifies potential new pan-European RI via a Roadmap (2006, 2010).



DIGITAL RESEARCH INFRASTRUCTURES



European Strategy Forum on Research Infrastructures Roadmap 2010

- Social Sciences and Humanities
- Environmental Sciences
- Energy
- Biological and Medical Sciences
- Materials and Analytical Facilities
- Physical Sciences and Engineering



DIGITAL RESEARCH INFRASTRUCTURES

European Strategy Forum on Research Infrastructures Roadmap 2010

- Social Sciences and Humanities
 - CLARIN (Common Language Resources and Technology Infrastructure)
 - DARIAH (Digital Research Infrastructure for the Arts and Humanities)

Digital research infrastructures



AUSTRIA (A BIT OF HISTORY)

2003 ESFRI report

2008 BMWF mandate

2009 Survey on Austrian language resources

2011 Submission of CLARIN-ERIC

29 Feb. 2012 CLARIN-ERIC (Construction Phase)

2012-15 Austrian commitment for CLARIN-ERIC

2013 Submission of DARIAH-ERIC

2014 DARIAH-ERIC

2014-16 Austrian commitment for DARIAH

2014-16 Austrian Centre for Digital Humanities (ACDH)

Digital research infrastructures AUSTRIA



Austria early on represented in both infrastructures

CLARIN + DARIAH as "sibling projects"
Spin-off projects, e.g. since 2012 in DASISH

GOALS



CLARIN aims at providing easy and sustainable access for scholars in the humanities and social sciences to digital language data and advanced tools to discover, annotate, analyse or combine this data, independent of its location or format.

DARIAH aims to enhance and support digitally-enabled research and teaching across the humanities and arts. It will develop, maintain and operate an infrastructure in support of ICT-based research practices.

GOALS



CLARIN aims at providing easy and sustainable access for scholars in the humanities and social sciences to digital language data and advanced tools to discover, annotate, analyse or combine this data, independent of its location or format.

DARIAH aims to enhance and support digitally-enabled research and teaching across the **humanities and arts**. It will develop, maintain and operate an infrastructure in support of ICT-based research practices.

GOALS



DARIAH aims to enhance and support digitally-enabled research and teaching across the humanities and arts. It will develop, maintain and operate an infrastructure in support of ICT-based research practices.

GOALS



CLARIN aims at providing easy and sustainable access for scholars in the humanities and social sciences to digital language data and advanced **tools** to discover, annotate, analyse or combine this data, independent of its location or format.

The DARIAH-EU infrastructure will be a connected network of **tools**, information, people and methodologies for investigating, exploring and supporting research across the broad spectrum of the Digital Humanities.

GOALS



CLARIN is establishing a networked federation of European data repositories, service centres and centres of expertise.

The DARIAH-EU infrastructure will be a connected network of tools, information, people and methodologies for investigating, exploring and supporting research across the broad spectrum of the Digital Humanities.

CLARIN is working on the interoperability of tools and data across the network, in order to allow researchers to combine distributed and heterogeneous data collections and to perform complex operations on these.

DARIAH will preserve, provide access to and disseminate research that stems from these collaborations and ensure that best practices, methodological and technical standards are followed.

GOALS



CLARIN is establishing a networked federation of European data repositories, service centres and centres of expertise.

The DARIAH-EU infrastructure will be a connected network of tools, information, people and methodologies for investigating, exploring and supporting research across the broad spectrum of the Digital Humanities.

CLARIN is working on the interoperability of tools and data across the network, in order to allow researchers to combine distributed and heterogeneous data collections and to perform complex operations on these.

DARIAH will preserve, provide access to and disseminate research that stems from these collaborations and ensure that best practices, methodological and technical standards are followed.

GOALS



CLARIN is establishing a networked federation of European data repositories, service centres and centres of expertise.

The DARIAH-EU infrastructure will be a connected network of tools, information, people and methodologies for investigating, exploring and supporting research across the broad spectrum of the Digital Humanities.

interoperability of tools and data across the network, in order to allow researchers to combine distributed and heterogeneous data collections and to perform complex operations on these.

DARIAH will preserve, provide access to and disseminate research that stems from these collaborations and ensure that best practices, methodological and technical standards are followed.

STRUCTURAL SIMILARITIES



"Under construction"

Stable two-tier governance structure

General Assembly

Board of Directors

National Coordinators Committee

Scientific Board

CLARIN Office

Thematic committees

SCCTC (,Centres Committee')
Technical centres

"Under construction"

More complex governance structure

General Assembly

Board of Directors

National Coordinators Forum

Scientific Board

DARIAH-EU Coordination Office

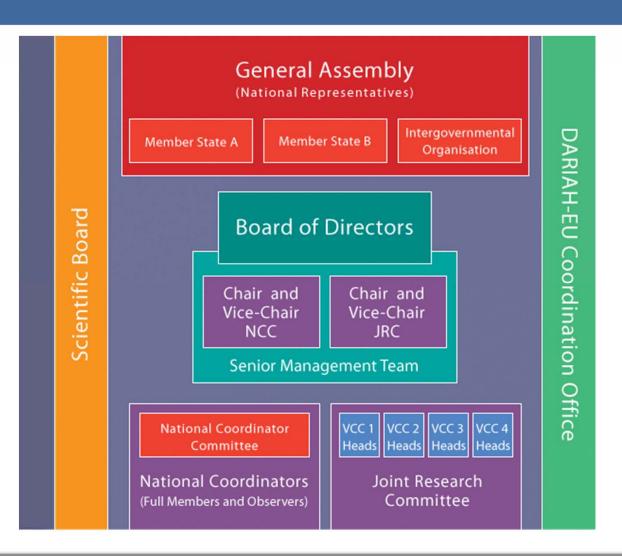
Task forces / Cross-VCC activities

Joint Research Committee
VCCs

DARIAH

GOVERNANCE





CENTRES



Network of centres

Certification (Centre Assessment)

Three types

A (infrastructure)

B (LRT data/services)

C (metadata)

Currently 11 certified centres (most of them DE, +1x NL, DK, CZ)

Coordinated by SCCTC (Standing Committee for CLARIN Technical Centres)

CENTRES



Getting ready for centre status:

- Centre of Estonian Language Resources (CELR)
- clarin.dk
- Austrian Centre for Digital Humanities
- DANS (Data Archiving and Networked Services)
- Huygens Instituut
- INL (Instituut voor Nederlandse Lexicologie)
- LINDAT-Clarin
- MI (Meertens Instituut)

INFRASTRUCTURE



Federated Identity

AAI, Single-Sign-on

Persistent Identifier

CMD - Component Metadata

Federated Content Search (FCS)

Propagation of standards (CSC)

SERVICES



Search – Deposit – Process – Visualize

Repositories (Depositing Services)

Virtual Language Observatory (~ 600.000 records)

WebLicht

CLARIN-ERIC Services

VLO - Virtual Language Observatory



search

<< < 1 2 3 4 5 6 7 8 9 10 > >>

help













Explore the world of language resources and technology from different perspectives

VLO Home >> Faceted Browser Resources | found an error?

No facets values selected

COLLECTION

Meertens collection: Liederenbank (243129) UBU Clarin Set (69966) Nederlands Instituut voor Beeld en Geluid Academia collectie (46156) TLA: childes (29266) TLA: DoBeS archive (25079)
TLA: Language and Cognition (23634) TLA: talkbank (14593) Bavarian Archive for Speech Signals (BAS) (14416) TLA: Acquisition (14363) Institut für Deutsche Sprache, CLARIN-D Zentrum, Mannheim (14320)

RESOURCECLASS

text (71428) televisie (39307) Text (32261) Sound (9995) http://purl.org/dc/dcmitype/PhysicalObject (9173)
Data Provider (4402) bioscoop (2956) Televisie (1695) niet van toepassing (1185) Image (763) more...

CONTINENT

Europe (71551) North-America (21727) Asia (18807) South-America (8335) Africa (5258) Oceania (5221) Australia (2287) Middle-America (2253) North America (1159) Europa (9) more...

COUNTRY

Germany (25763) United States (20591) Netherlands (19918) Japan (6953) United Kingdom (6643) Papua New Guinea (5963) Turkey (4707)
Belgium (3972) France (3891) Indonesia (3466)

LANGUAGE

Dutch (148326) English (72980) German (39043) French (9230) Spanish; Castilian (8798) Danish (8447) Japanese (6760) Turkish (4988) nl NL (3000) Chinese (1980)

MODALITY

speech (84617) spoken (16949) speech; gestures (7611) signs (3346) act-out (2925) verbal and non-verbal interaction (2082) written (1183) signs; speech (997) gestures (988) speech/gestures (728)

GENRE

discourse (82754) info:eu-repo/semantics/article (34945) kinderlied (19218) primary text (12813) narrative (11471) info:eu-repo/semantics/masterthesis (11178) language description (11131) liefdeslied (8381) stimuli (8203) conversation (7826) more...

SUBJECT

answer the question: 'what did you do today?' (3903) what did you do today? (3903) repeat text, answer questions (3613) jura (dk5-34) (2306) appointment scheduling (2232) to agree about one or more appointments for business meetings (2232) politiek (2154) contact data (1957)

Showing 1 to 30 of 605434

" DE STIEFMOEDER " die haar 18 jarigen STIEFZOON 10 jaar heeft opgesloten op de 3e verdieping Over-Amstelstraat 10 te AMSTERDAM " DE STIEFMOEDER " die haar 18 jarigen STIEFZOON 10 jaar heeft opgesloten op de 3e verdieping Over-Amstelstraat 10 te " De professionaliteit van medisch specialisten onder de loep" Social Welfare work " in Brits West-Indi

"... En nu die kale Fransen!" : Woerden in de patriottentijd en de Bataafs-Franse tijd

"... To improve Our Knowledge of Successful Library Architecture" - Elmar Mittler and the LIBER Architecture Group Seminars

"... of op de bagijnekamer op de toren"

... of een rouwbord"?

"...Een canael ende deurganck..." : een onderzoek sluisrekeningen van Gouda is naar de binnenvaart door Holland in de lange Gouden Eeuw (1572-1700)

...op het afnemen en vervoeren naar de galge van Van Nieuwenhuisen die op 17 Hekendorp..." : poging tot reconstructie van een merkwaardig sodomieproces in 1732

"A Change is Gonna Come" Stax Records en de zwarte vrijheidsstrijd "A Poem in Our Eyes": Community Writing in

American Literature

'A man must be himself' - The political style of Charles de Gaulle, Richard Nixon and Tony Blair

"A nation stays alive when its culture stays alive" In oorlogen en andere

- Het belang van culturele noodhulp geïllustreerd
aan de hand van drie projecten van Cultural
aardbevingen, stormen Emergency Response
"ACH, MIJN LIEVE AUGUSTIJN."

"ACH, MIJN LIEVE AUGUSTIJN." "ACH, MIJN LIEVE AUGUSTIJN." "Aanloopstraten in binnensteden De invloed van fysieke en organisatorische factoren op de

kwaliteit van aanloopstraten als

"According to security sources" : what remains ofjokes in Israel about the "Voice the Israeli media

"Aen 't ende van den lande" : enkele opmerkingen

Artikel over een onderzoek naar de gevangenenplaatsen in Rhenen in de 18e eeuw

veel te zeggen over de sociale

De neschiedenis van Adrianus oktober 1732 gehangen werd aan de galg van..

Political style is more than attractive 'packaging' of a politician and his/her message; it is the...

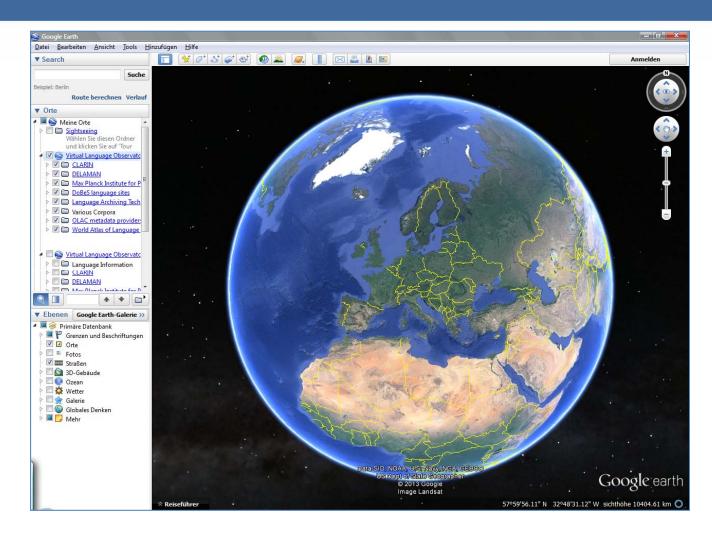
aardbevingen, stormen en andere natuurrampen wordt...

In the 1960s there were many of the UAR (United Arab Republic)..

CLARIN-ERIC Services



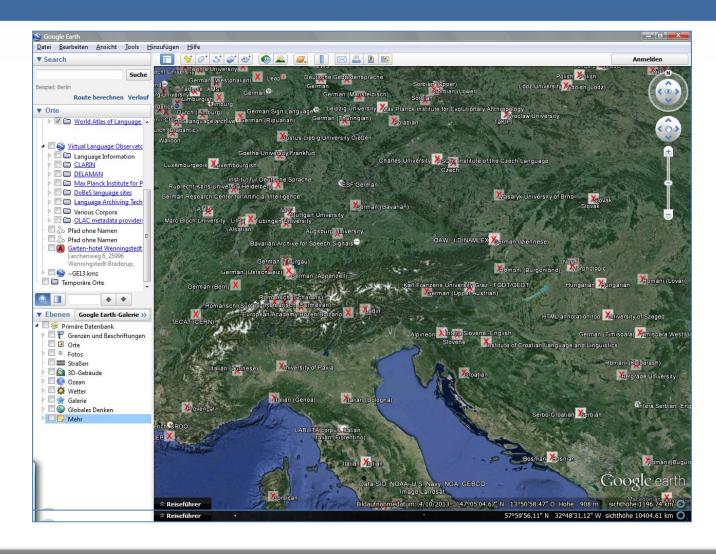




CLARIN-ERIC Services







DARIAH-EU

DA

... IS SPECIAL THROUGH

- ... working for a wide range of disciplines and topics
- ... allowing a high degree of diversity of methodological approaches
- ... fostering inclusiveness
- ... strengthening the European Research Area (ERA)
- ... supporting transdisciplinarity and integration
- ... advocating decentralised approaches
- ... creating transverse relations to other RI initiatives
- ... advocating open humanities (open data, open content, open access and open methods and sharing)
- ... basically hybrid (digitally-focused / analogue-aware)

Members



Austria

Bulgaria

Czech Republic

Denmark

Germany

Estonia

the Netherlands

Poland

the Dutch Language Union

Austria

Denmark

France

Germany

Greece

Ireland

Italy

Luxemburg

the Netherlands

Serbia

Slovenia

Members



Austria

Bulgaria

Czech Republic

Denmark

Germany

Estonia

the Netherlands

Poland

the Dutch Language Union

Austria

Denmark

France

Germany

Greece

Ireland

Italy

Luxemburg

the Netherlands

Serbia

Slovenia

DH Centres



Centre for Digital Humanities (Brock University)

Digital Humanities Centre (Nottingham)

Centre for Digital Humanities (Ryerson University)

UCL Centre for Digital Humanities

Department of Digital Humanities (DDH)

(formerly Centre for Computing in the Humanities

Ghent Center for Digital Humanities

Centre for Material Digital Culture and Digital Humanities (University of Sussex)

Trier Center for Digital Humanities

Göttingen Centre for Digital Humanities Forschungszentrums für Historische Geisteswissenschaften

ACDH FOUNDATION



Austrian Centre for Digital Humanities

August 2013: Contract with Ministry of Science and

Research





FOUNDATION



Autumn 2013: First preparations

1 January 2014: Official launch

ACDH is follow-up project of Austrian CLARIN-DARIAH activities

Financed by the Ministry of Science, Research and Economy (+ shares of participants)

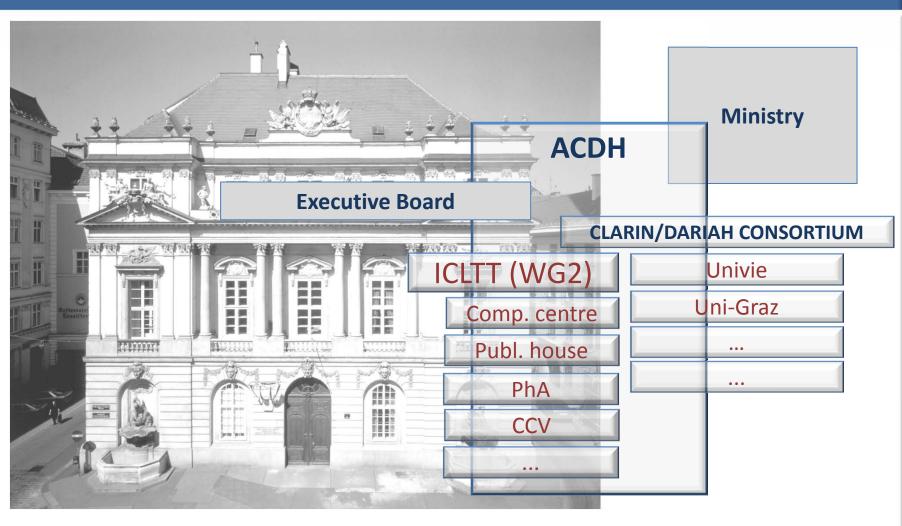
FOUNDATION





FOUNDATION

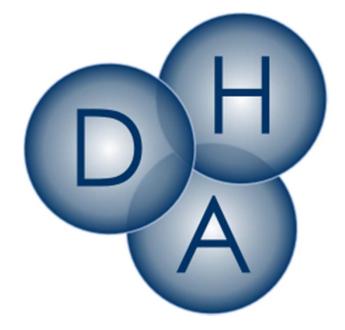




DIGITAL HUMANITIES AUSTRIA









CORE ACTIVITIES

RI4DH

Research infrastructures for the humanities Continuation of CLARIN-DARIAH-AT

go!digital

Digitisation campaign
Call for 5 "innovative" digitisation projects

dh curriculum

Establishment of DH in university curricula



Continuation of Austria's commitment to CLARIN-ERIC and DARIAH

Linking up RIs v vcc1 Setting up a Kn Working on the Coordination of

Co-head **Task 3 (Reference Software Packages)**

VCC2 (Research and Education Liaison) **Co-lead of task Training and Education Programme**



Continuation of Austria's commitment to CLARIN-ERIC and DARIAH

Linking up RIs with researchers

Setting up a Knowledge / Competence Centre

Working on the technical infrastructure

Coordination of Austrian in-kind contributions



Continuation of Austria's commitment to CLARIN-ERIC and DARIAH

Linking up RIs with researchers

Setting up a Knowledge / Competence Centre

Working on the Coordination of

Data
Tools
Best practices
Standards



Continuation of Austria's commitment to CLARIN-ERIC and DARIAH

Linking up RIs with researchers

Setting up a Knowledge / Competence Centre

Working on the technical infrastructure

Coordination of

Virtual research environments (VREs)
Depositing services (GAMS, PHAIDRA, CCV, ...)
Metadata/Content search



Continuation of Austria's commitment to CLARIN-ERIC and DARIAH

Linking up RIs with researchers

Setting up a Knowledge / Competence Centre

Working on the technical infrastructure

Coordination of Austrian in-kind contributions

Second Austrian LR Survey

Digital Humanities Austria (DHA) go!digital



Austrian Scholarly Digitisation Initiative

- 5 innovative projects
- 600.000 Euros
- Focus on "scholarly" purposes
- Focus on methods / workflows / standards
- Competitive call in April 2014
- International review of proposals



Fostering of DH in university curricula

Development of DH reference curricula

Dissemination of DH through workshops, conferences

Establishment of an expert group



KNOWLEDGE HUB

... is a platform for the dissemination of the digital paradigm in Austrian humanities studies.

The virtual network is run jointly by the Austrian Academy of Sciences, the University of Vienna and the University of Graz.

It is supported by these institutions and the Federal Ministry of Science, Research and Economy and will serve as a staging area for all kinds of DH research activities.











KNOWLEDGE HUB PRINCIPLES

Systematic use of digital infrastructures

Transdisciplinarity

Collaborative work

Participatory technologies (virtual research environments, web-based research portals)

Open Access / Open Source

Open life cycle of research, research results and research data

Digital Humanities Austria









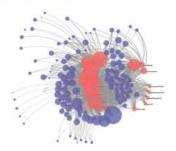
News Events RI4DH Go!Digital DH Curriculum Partners Projects Expert Bureau

Digital Humanities Austria

Digital Humanities Austria is a platform for the dissemination of the digital paradigm in Austrian humanities studies. The virtual network is run jointly by the Austrian Academy of Sciences, the University of Vienna and the University of Graz, is supported by these institutions and the Federal Ministry of Science, Research and Economy and will serve as a staging area for all kinds of DH research activities. Read more

Deutsch

SMC Browser



SMC-Browser is a web application developed in the context of CLARIN. It allows to explore the data-space of the Component Metadata Infrastructure, a framework for harmonized creation and publication of language resources descriptions (metadata). Read more

ABaC:us - Austrian Baroque Corpus



is a growing digital collection of printed German language texts dating from the Baroque era, in particular the years from 1650 to 1750. The project group which established ABaC:us focused on the creation of a thematic research collection based on the prevalence of sacred literature during the Baroque Era, Books of religious instruction and works concerning death and dying were a focal point of Baroque culture. The best-selling genre was even marked by a number of reprinted editions. Read more

Mecmua



Early Modern Ottoman Culture of Learning: Popular Learning between Poetic Ambitions and Pragmatic Concerns

In our project we intend to explore some aspects of the Early Modern Ottoman culture of learning, in particular those areas of learning used and cultivated outside the official Ottoman institutions of learning, the medreses. Read more

News

DARIAH-AT Workshop in Graz

2013-11-20 10:15

Das Zentrum für Informationsmodellierung - Austrian Centre for Digital Humanities der Universität Graz lädt zum ersten DARIAH-AT Workshop:

Digitale Geisteswissenschaften in Osterreich: Nationale Kooperationen und europäische Perspektiven Read more

Was können und wollen Digital Humanities

2013-10-25 09:00

Die Tagung organisiert von der Österreichischen Nationalbibliothek knüpft an die Veranstaltung des Ministeriums zu Forschungsinfrastrukturen in GSK inhaltlich an. Read more

Tagung zu Forschungsinfrastrukturen in GSK

2013-10-24 08:30

Das BmWF organisierte die Veranstaltung:

Herausforderungen und Chancen

User Services

Search

Process

Repositories





First Austrian node in the network of CLARIN centres

- Assessment end of March
- Language Resources Portal
- Depositing and publishing services for Austrian LR
- Knowledge Hub (under construction)
- Metadata Repository (domestic VLO)
- Semantic Mapping Component
- Vocabulary repository
- Triple store (for linked data)



CLARIN CENTRE VIENNA





CLARIN Centre Vienna Language Resources Portal





Home About DHA FAQ Impressum OAI-PMH

Login Federated Login

CLARIN Centre Vienna / Language Resources Portal

CLARIN Centre Vienna (CCV) is Austria's main connection point to the European network of <u>CLARIN Centres</u>. It is an Austrian contribution to CLARIN-ERIC and being hosted by the *Institute for Corpus Linguistics and Text Technology (ICLTT)* of the *Austrian Academy of Sciences*. It is jointly funded by the *Academy* and the *Federal Ministry of Science*, *Research and Economy*.

If you have language resources, would like to share these with the scientifc community (and/or the public) and want to make sure that the data will be around in the future, contact us. We offer archiving and online availability of the resources. If needed, we will assist you in converting data and metadata into required formats.

Contact: icltt-tech@oeaw.ac.at

Read more

Deutsch



AAC-FACKEL



The AAC digital edition of the journal »Die Fackel« offers free online access to the

vicav



The Viennese Corpus of Arabic Varieties is a diverse collection of materials concerning the

Mecmua



Early Modern Ottoman Culture of Learning: Popular Learning between Poetic Ambitions

and Dragmatic Concorns



GAMS (Geisteswissenschaftliches Asset Management System) is an asset management system for the storage and administration of digital resources produced in a scientific context.

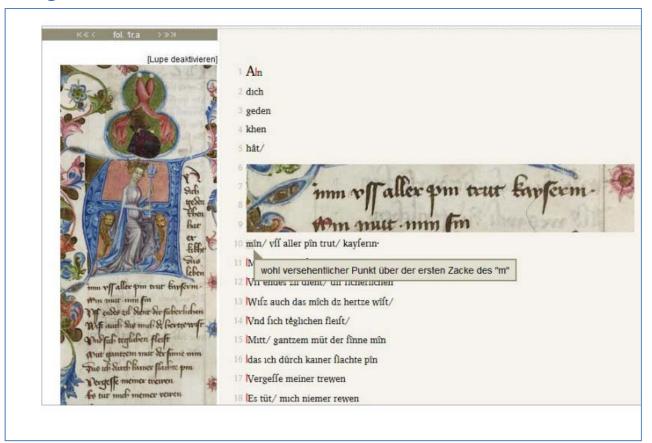


GAMS supports the electronic dissemination of research data and related metadata, with special emphasis on the arts and humanities.





Hugo Montfort Edition

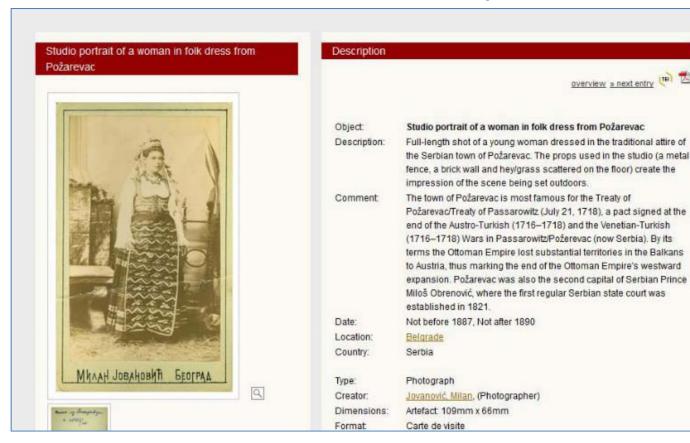








Visual Archive Southeastern Europe

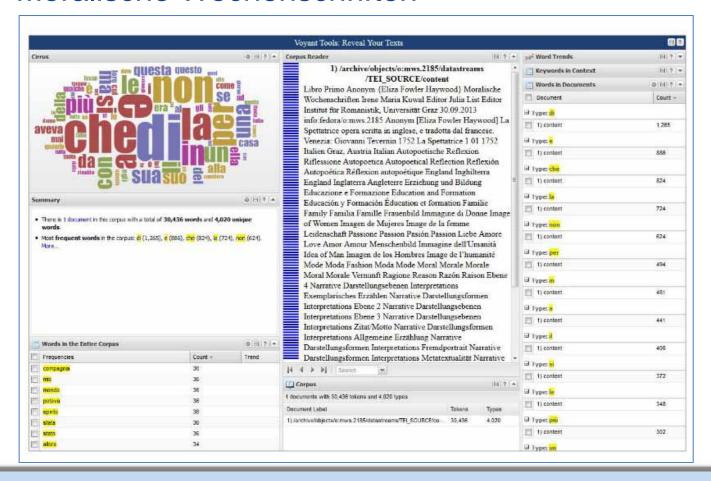








Moralische Wochenschriften







LEXICOGRAPHY INFRASTRUCTURE

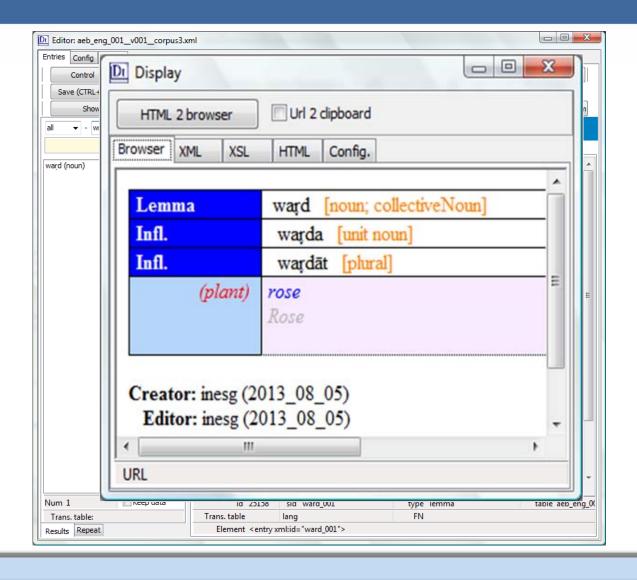
Viennese Lexicographic Editor (VLE)

- XML editor specialized for editing lexicographic data
- Generic support for any (XML) format (LMF, TEI, TBX, RDF)
- Making use of cognate technologies, (XSLT, XPath, XSD)
- Various editing modes, configurable keyboard layouts
- Optimised corpus-dictionary interface
- On-the-fly data visualisations





LEXICOGRAPHY INFRASTRUCTURE





LEXICOGRAPHY INFRASTRUCTURE

Dictionary Server

- Open and freely available software that can be readily distributed (MySQL, PHP)
- Integrated with corpus_shell (Querying works on the basis of SRU/CQL (Search/Retrieval via URL + Contextual Query Language)
- Connected to the clients through a REST-style web service
- Next steps: versioning



LEXICOGRAPHY INFRASTRUCTURE

DictGate is an infrastructure

- to allow (groups of) researchers to host lexicographic data.
- for producing and publishing lexicographic data.
- for working on smaller lexicographic projects that are in need of solutions that can be applied without much logistical and technical overhead.
- The service will be run as part of LRP.



LEXICOGRAPHY INFRASTRUCTURE

Extensible core dictionaries

- Uniform structure
- TEI, LMF, RDF
- focus on under-resourced languages
- Open Access (Creative Commons licence, CC-BY)

Projects

- VICAV (Vienna Corpus of Arabic Varieties)
- TUNICO
- Persian English Dictionary
- German Russian Dictionary
- Dictionary of Bavarian Dialects in Austria
- ...



STANDARDS

CLARIN-ERIC and DARIAH consider standards as a major concern of their activities and have institutionalised their respective work.

- Text Encoding Initiative
- Language related ISO standards
- RDF, SKOS

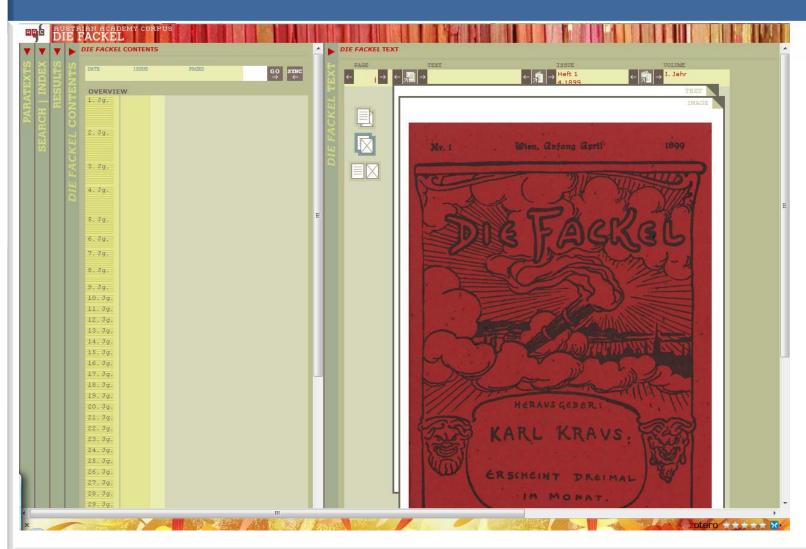


CORPUS_SHELL

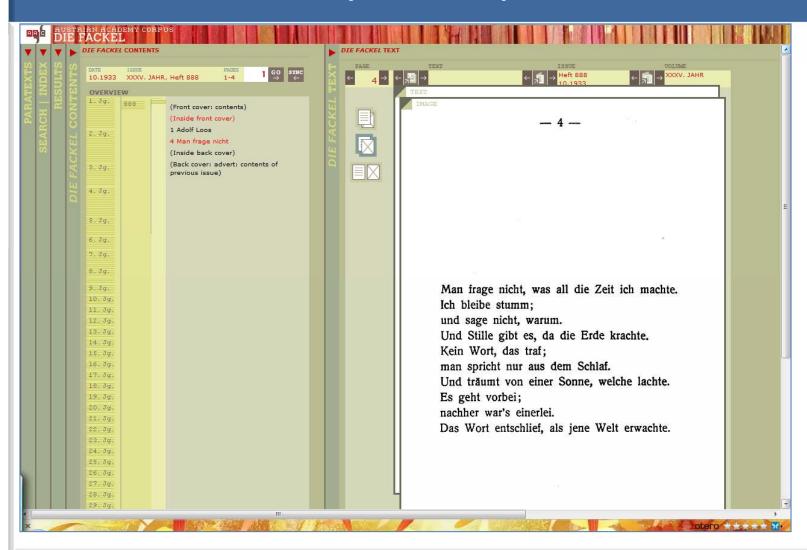
... is a modular framework for publishing a wide range of language resources designed to operate in a distributed and heterogeneous environment

- distributed setup, FCS-based
- integration with CLARIN metadata infrastructure
- (reusing) specialized Resource viewers for specific types of resources
- multiple implementations (php, perl, XQuery)

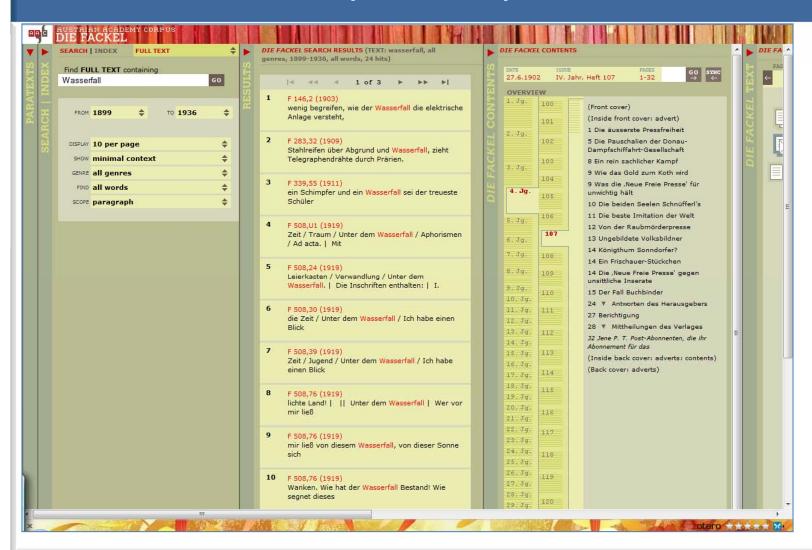




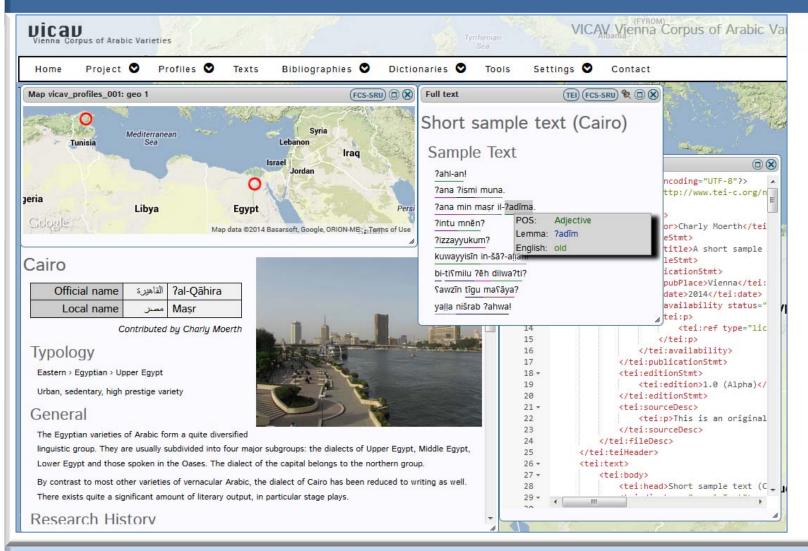




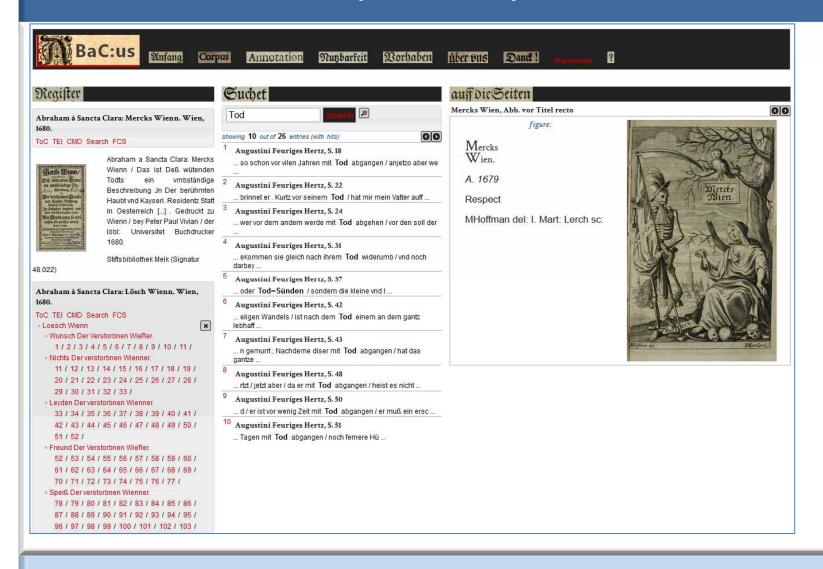




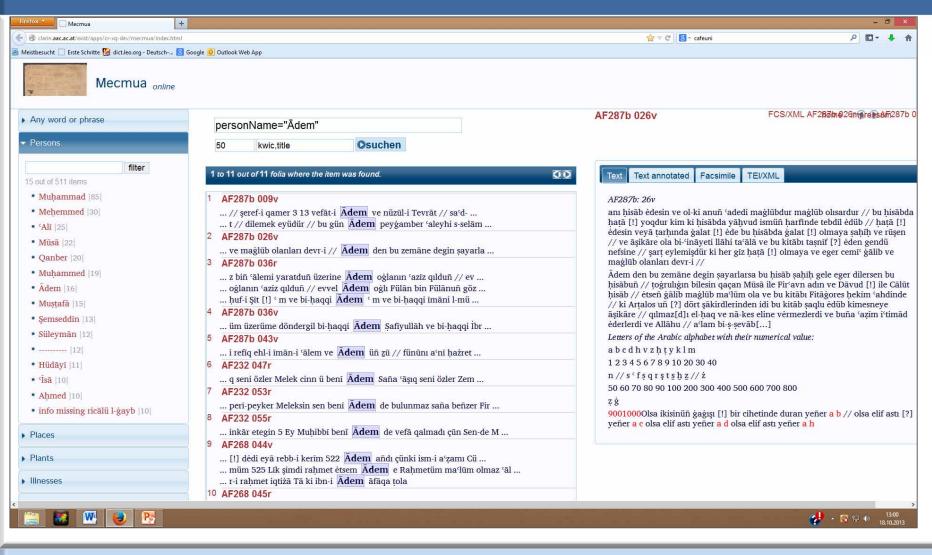




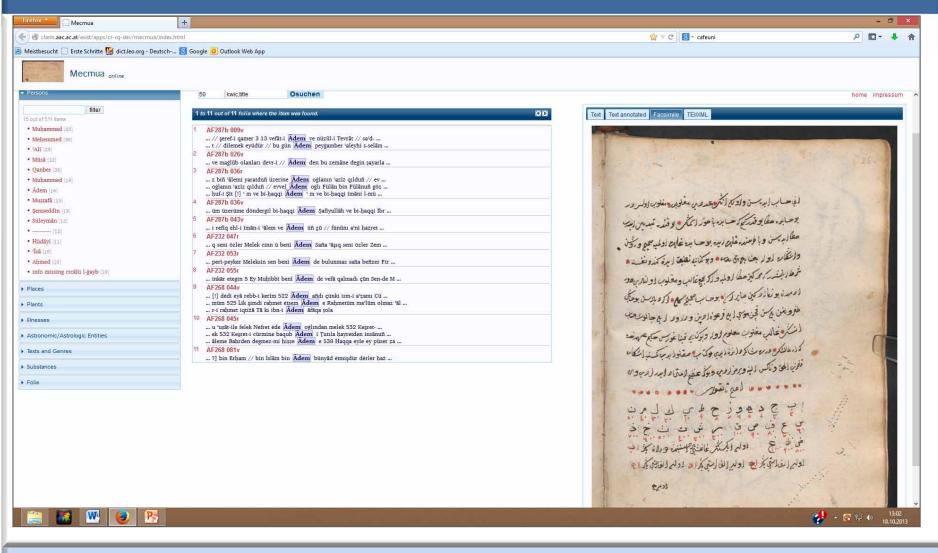














LONG-TERM STRATEGIES

Making use of synergies by cooperation between the ESFRI projects

Expanding the community

Involving more institutions, in particular ones that guarantee a certain degree of sustainability Installation of a team of experts to counsel decision makers

Der Aufbau von digitalen Forschungsinfrastrukturen für die Geistes- und Kulturwissenschaften in Österreich

Karlheinz Mörth

Institute for Corpus Linguistics and Text Technology Austrian Centre for Digital Humanities Austrian Academy of Sciences

karlheinz.moerth@oeaw.ac.at

















Institut für Corpuslinguistik und Texttechnologie



