



**Forschung – Information – Infrastruktur:
Bausteine für Open Science**

13. DINI-Jahrestagung • 24. und 25. September 2012 • Karlsruhe

Nationale Forschungsinfrastrukturen im EU Rahmen

25.09.2012

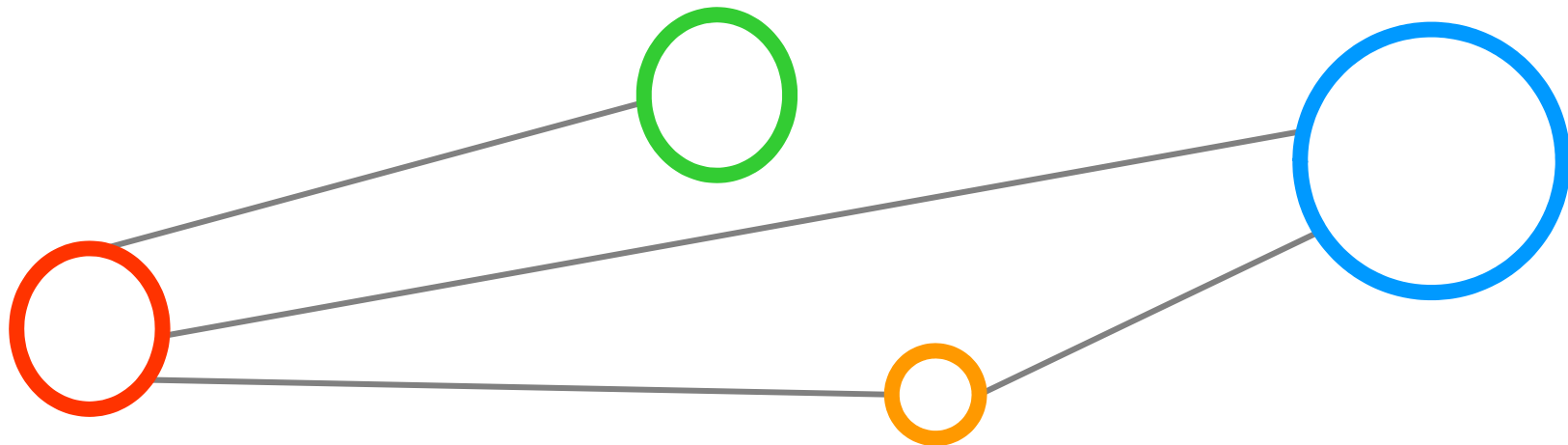
Morris Riedel

Jülich Supercomputing Centre

Abteilung Förderierte Systeme und Daten



Inhalt

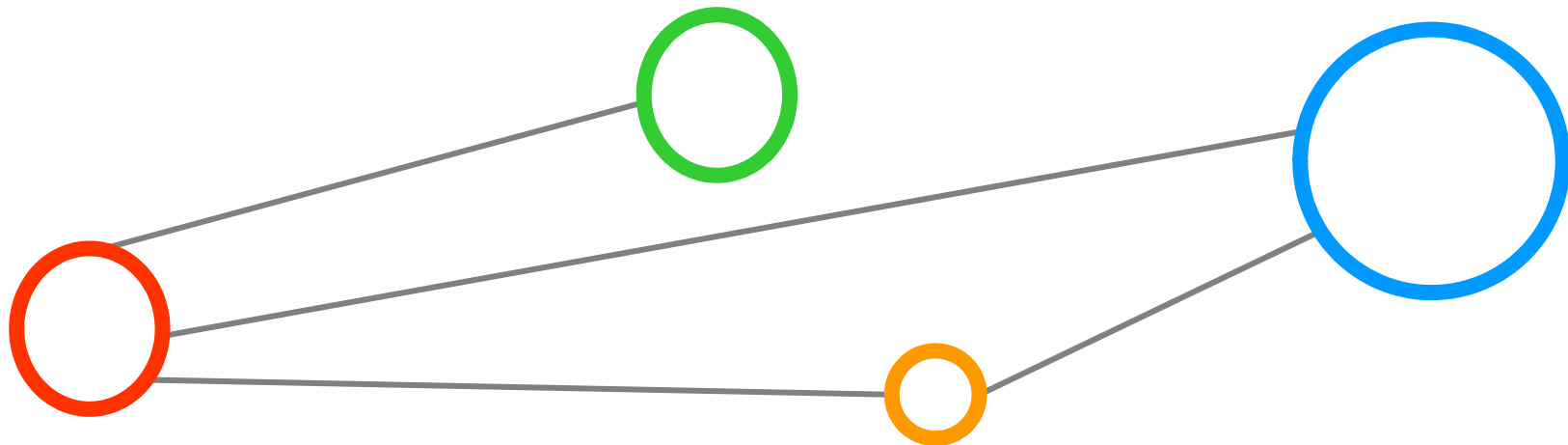


Inhalt

- Forschungsinfrastrukturen
 - Von High-Level zum Prinzip der Umsetzung
 - Komplexe Anforderungen konkreter Bausteine
 - Europäische Initiativen zur Umsetzung
- Forschungsinfrastrukturen am Beispiel
 - National und Europäisch: CLARIN und DARIAH
 - Europäisch: EUDAT
- Ausblicke
- Mögliche Erfolgsfaktoren
- Referenzen



Forschungsinfrastrukturen “High-Level”



Vorwegnahme: Begriffe für Forschungsinfrastrukturen

- ... aka Informations-Infrastruktur (u.a. dieses Event)
- ... aka Wissenschafts-Infrastruktur (u.a. auch dieses Event)
- ...aka Grid Infrastruktur (etwas älter)
- ...aka Cloud Infrastruktur (etwas neuer)
- ...aka e-Science Infrastrukturen (etwas wissenschaftlicher)
- ...aka e-Infrastructure (vor allem EU et al.)
- ...aka Research Infrastructures (vor allem EU und ESFRI et al.)
- ...aka Cyberinfrastructure (vor allem USA)



→ Die zentrale Frage bleibt doch: Wie setzen wir Anforderungen um?!

High Level Expert Group Report on Scientific Data

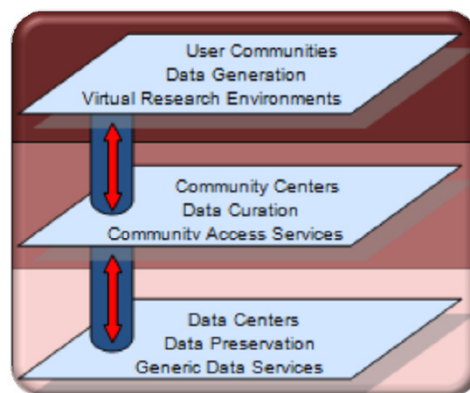
- Report beschreibt Herausforderungen bis 2030

'Researchers and practitioners from any discipline are able to find, access and process the data they need.'

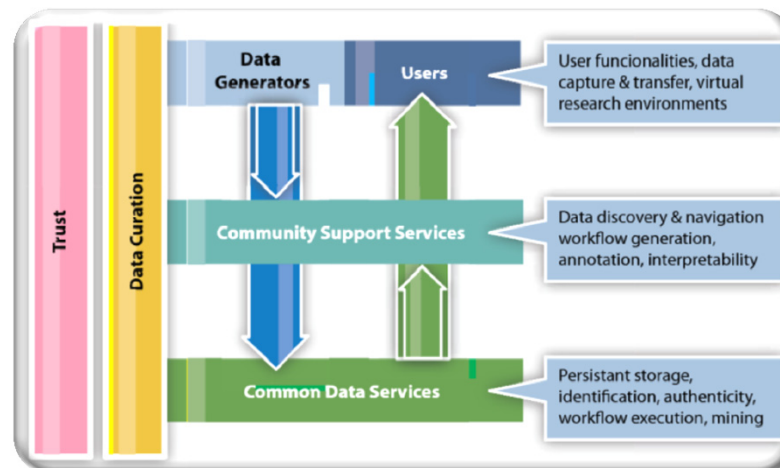
'They can be confident in their ability to use and understand data, and they can evaluate the degree to which that data can be trusted.'

Our vision is a scientific e-infrastructure that supports seamless access, use, re-use, and trust of data. In a sense, the physical and technical infrastructure becomes invisible and the data themselves become the infrastructure'

'develop an international framework for a Collaborative Data Infrastructure (CDI)'



Morris Riedel

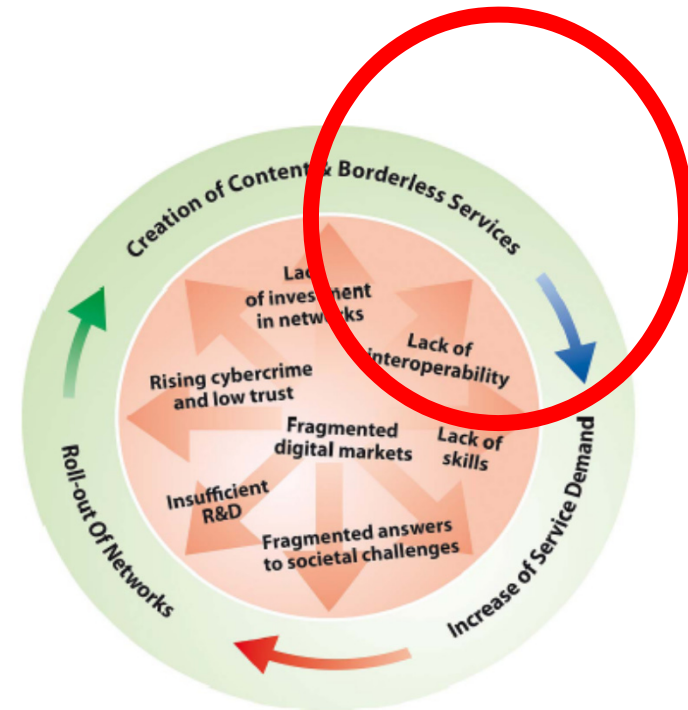


[6] Riding the Wave, 2010

Digital Agenda for Europe & Horizon 2020



OPEN ACCESS
'Every citizen cloud-enabled by 2020'





e-IRG and EEF

- e-Infrastructure Reflection Group (e-IRG)

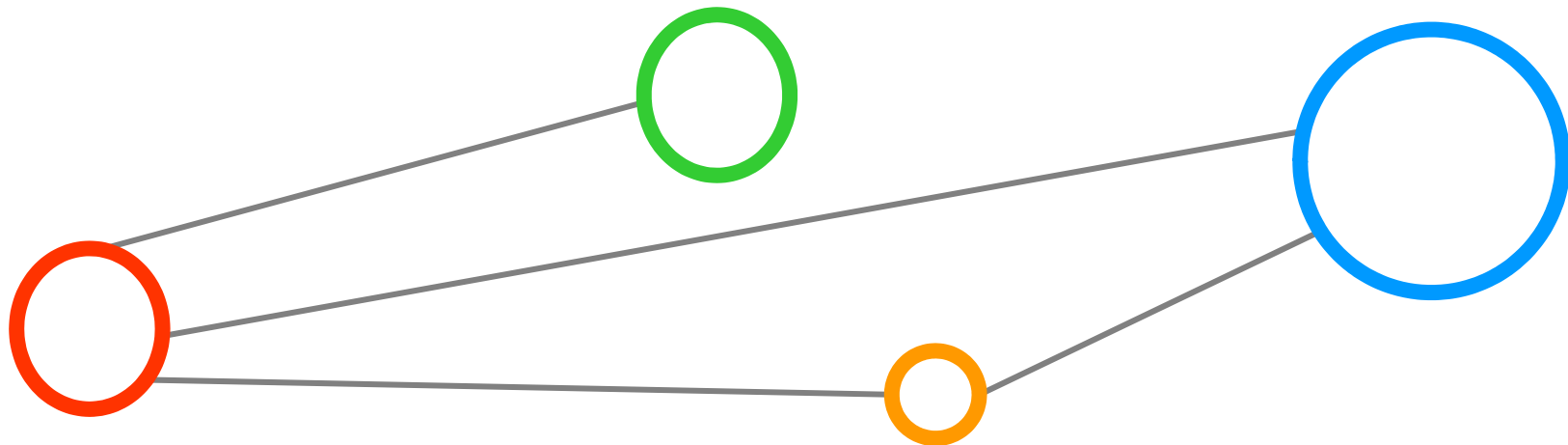
'Encourage the development of non-discipline-specific frameworks and information architectures for interoperable exchange of data...support communities for the definition of their requirements...'

'Ensure that besides hardware and services, digital objects deserve infrastructure components in their own right: ... persistent linkage of research data... Policies for long-term preservation of data, maybe focused into dedicated centers...'

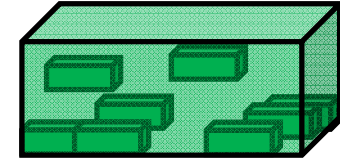
- European e-Infrastructure Forum (EEF)

'Data archiving and curation is a common need for several of the ESFRI projects'

Forschungsinfrastrukturen: Konkrete Bausteine



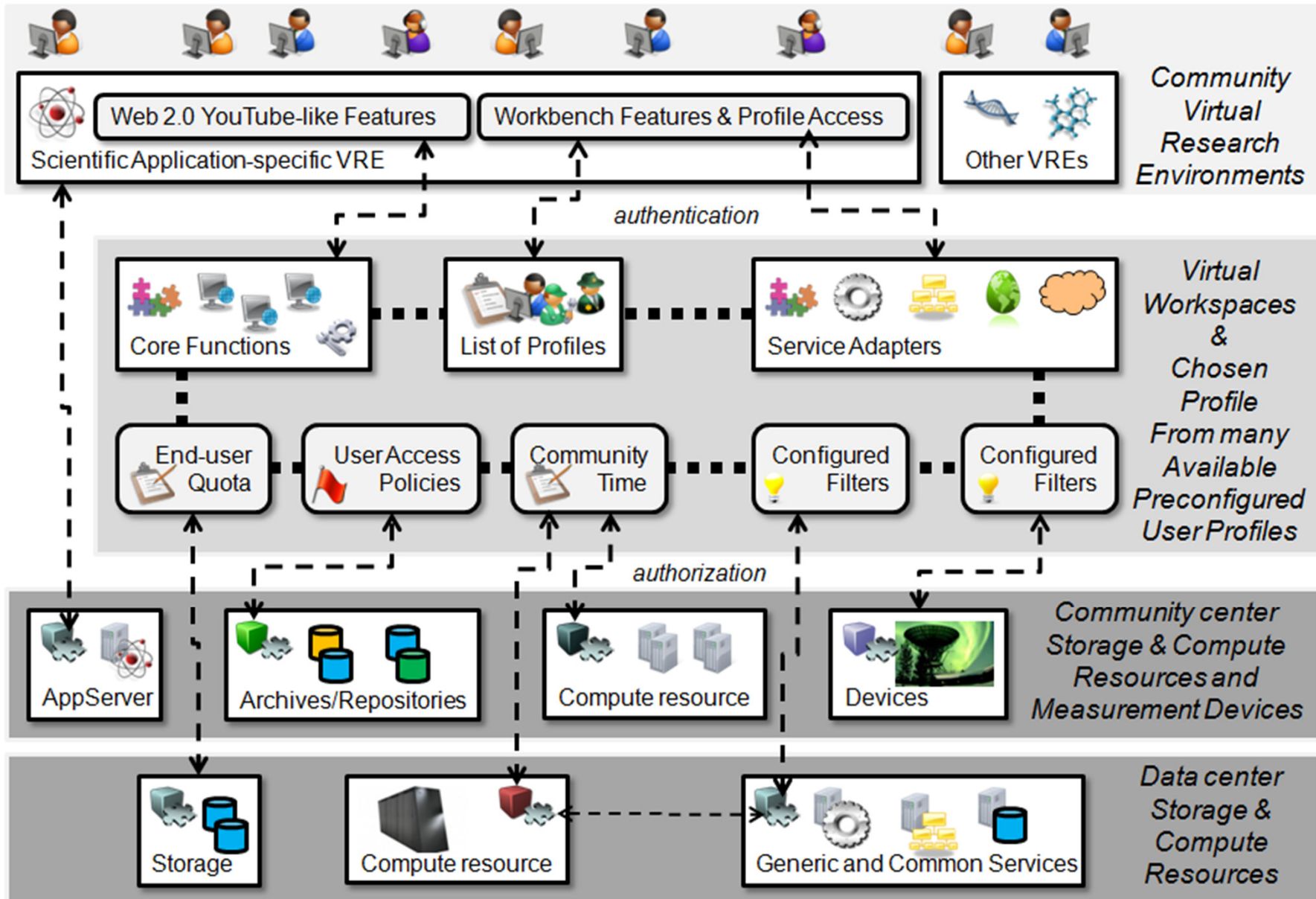
Komplexe Anforderungen von Bausteinen



- Computing
 - Prozessierung von Datenbestände mit verschiedenen Methoden
 - Dienste um von komplexen “Computing-Verfahren” zu abstrahieren
- Daten
 - Datenreplikation, Datentransfer und “einfaches Daten hochladen”
 - Persistente Referenzierung von Datenbeständen
 - Management von Metadaten
- Infrastruktur Allgemein
 - Zugriffskontrolle auf wissenschaftliche Datenbestände und Dienste
 - Vertrauensvolle Support-Strukturen und Operationspläne
 - Hosting von verschiedenen Diensten für wissenschaftliche Disziplinen
 - Evolution der Wissenschaftsinfrastruktur während der Nutzung



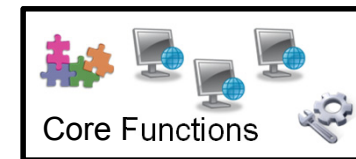
'ScienceTube' (Reference Model View)



Weitere Bausteine (Web 2.0 & Social Web abgeleitet)

- “Core Functions”

- Dynamische Web 2.0 Elements, mash-ups, Integrations-Code zwischen Funktionalitäten



- List of profiles

- Arbeitsumgebungen (“Workspaces”) sind an wissenschaftliche Profile gebunden (e.g. role dependent)



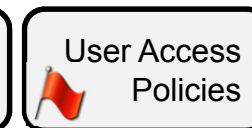
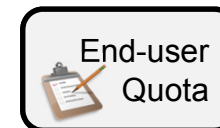
- Service Adapters

- Daten und Computing Dienste, Workflow Dienste, GreenIT Dienste, Community-spezifische Dienste



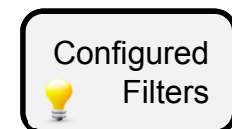
- Policies, Quota, User/Community Time

- Zugesagte Berechnungszeit für Nutzer, quotas für Speicher, Benutzer-policies



- Filter (aka vorkonfigurierte Suchen)

- Filter für vorhandene Dienste (bspw. 12 im Kontext statt 400)
 - Filter für Daten (bspw. Publikationen im letzten Jahr, Messung in 2010,...)



'ScienceTube' (User view)

Mag ich

7.557.864 Aufrufe

Alle Kommentare (6.291) [Alles anzeigen](#)

Prof. Dr. Known vor 2 Stunden
Correct. I used this great data in a recent publication supporting this claim, you can see it here: [PaperBerlinTraffic](#)

Citizen from Berlin vor 9 Stunden
I heard that you have used my data from Berlin to claim that the traffic evolution in this region is...

Warning

STORAGE: (RZG)

76% of your data in this storage is unused since 90 days – remove data?

Available Services (Filter #43):

Data-mining service

Workflow service

10:46 - CEST Meetings Today

9:00 Steering Group, see [Meeting Minutes](#)

Available Resources (Filter #443):

STORAGE: (JUELICH)

STORAGE: (RZG)

sensor ship #4455

sensor airplane #4711

HPC: (SARA)

HPC: (JUELICH)

HTC: (EGI)

STORAGE: (JUELICH)

STORAGE: (RZG)

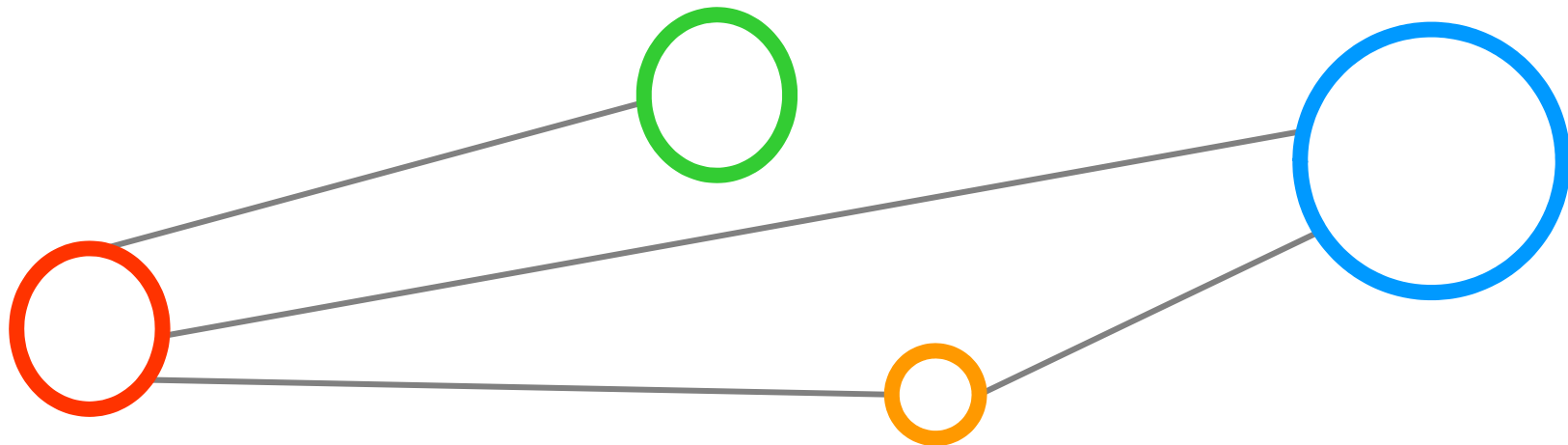
Most Viewed Today

MPI toolset usage von [DieseSchuleRock](#)

Paper about aging von [albertoson](#)

Device #46 data von [ElectronicArtsDE](#)

Forschungsinfrastrukturen: Europäische Initiativen



Multi-disziplinäre ESFRI Roadmap

- “European Strategy Forum on Research Infrastructures”
- Nutzen für Wissenschaft:
 - Bündlung von ~44 wissenschaftlichen Communities in 6 Bereichen
 - Organisation von Forschungsinfrastrukturen über Landesgrenzen
 - Identifikation und Planung von Kerndiensten für Wissenschaftler
 - Möglichkeiten zur internationalen und interdisziplinären Verknüpfung
- Idee:
 - Von Europäischen Regierungen zur Identifikation, Anschubfinanzierung und Begleitung von nachhaltigen Forschungsinfrastrukturen aufgelegt worden
 - Konzentration auf Nachhaltigkeit um Vertrauen bei Benutzer zu erlangen



Wissenschaftliche Bereiche der ESFRI



SOCIAL SCIENCES AND HUMANITIES

CLARIN > 26 DARIAH > 27



ENVIRONMENTAL SCIENCES

COPAL > 30 EISCAT_3D > 31 EMSO > 32
EPOS > 33 EURO-ARGO > 34 IAGOS > 35
ICOS > 36 LIFEWATCH > 37 SIOS > 38



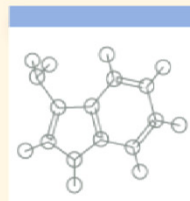
ENERGY

ECCSEL > 42 EU-SOLARIS > 43
HIPER > 44 IFMIF > 45 MYRRHA > 46
WINDSCANNER > 47



BIOLOGICAL AND MEDICAL SCIENCES

ANAEE > 50 BBMRI > 51 EATRIS > 52
ECRIN > 53 ELIXIR > 54 EMBRC > 55
ERINHA > 56 EU-OPENSREEN > 57
EUROBIOIMAGING > 58
INFRAFRONTIER > 59 INSTRUCT > 60
ISBE > 61 MIRRI > 62



MATERIALS AND ANALYTICAL FACILITIES

EMFL > 71 EuroFEL > 72 ESS > 73

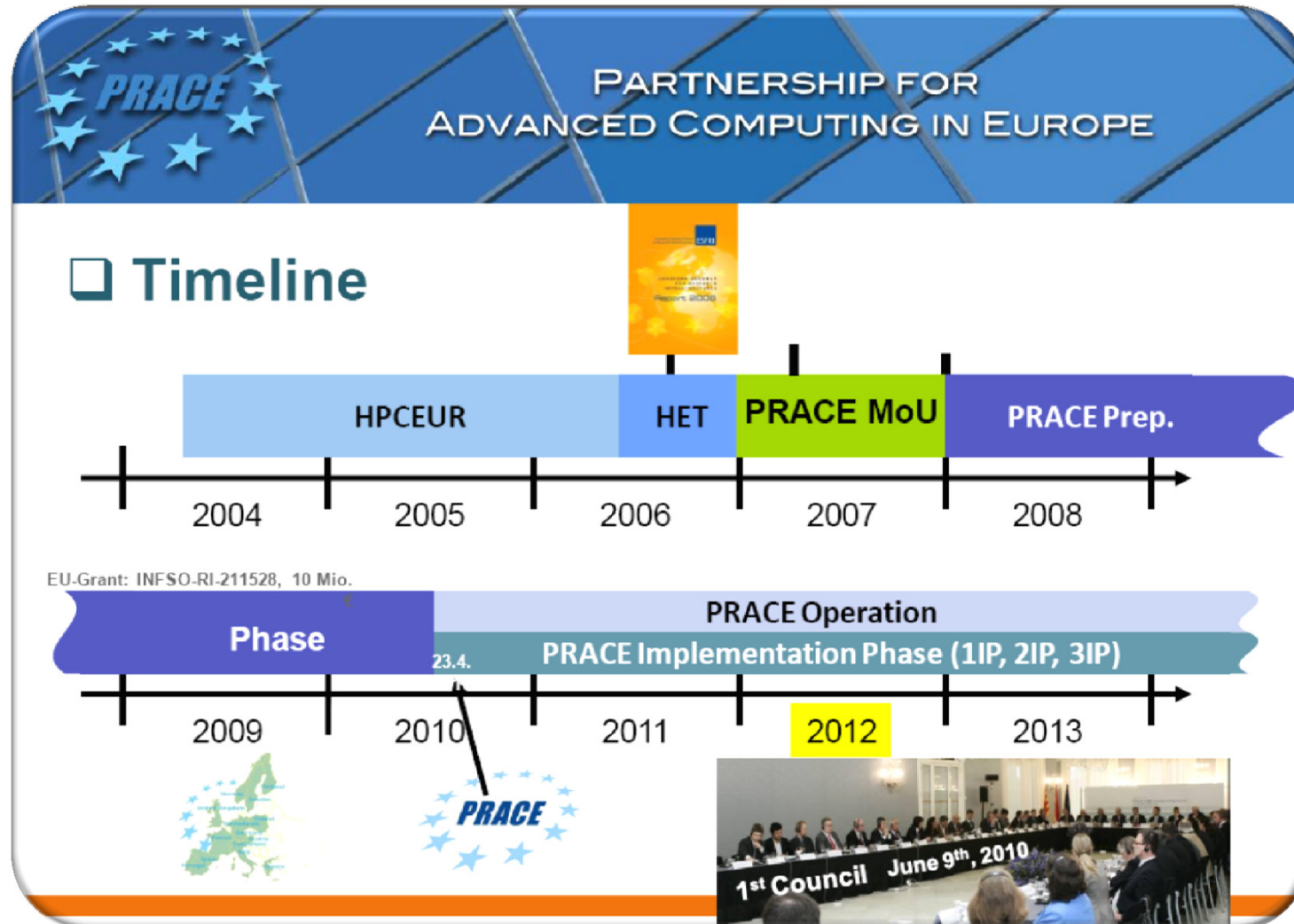


PHYSICAL SCIENCES AND ENGINEERING

CTA > 68 E-ELT > 69 ELI > 70
KM3NET > 74 SKA > 75



Computing: PRACE



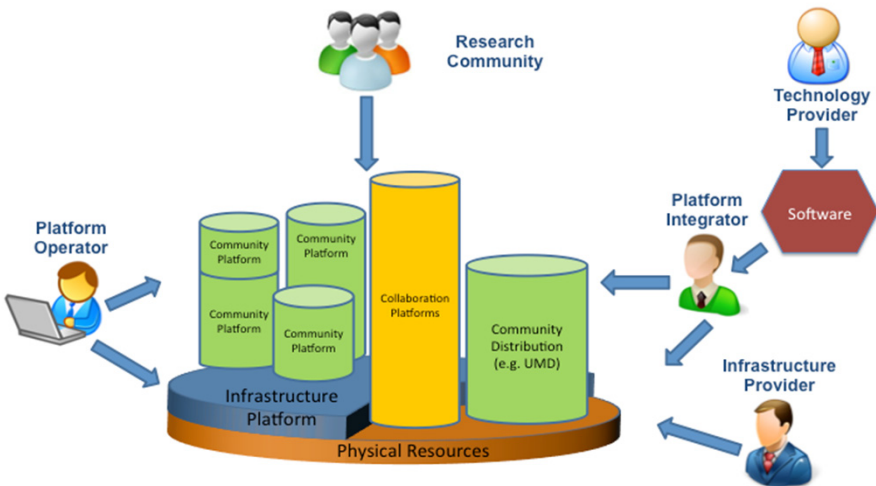
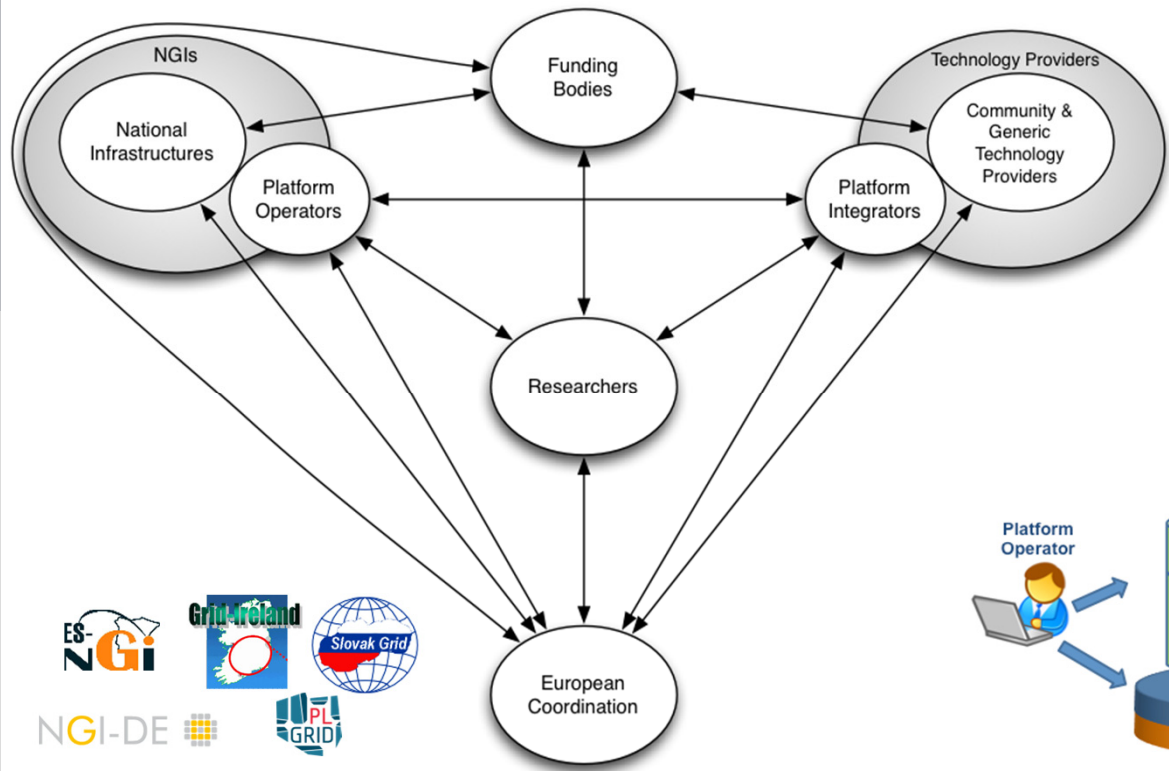
EGI wird zu EU Cloud Infrastruktur

- EC & Digital Agenda: ‘Every EU Citizen cloud-enabled in 2020’
- EGI Cloud Platform Profile Aktivitäten
 - Keine festgelegte Middleware mehr
 - Offen für alle Wissenschaften

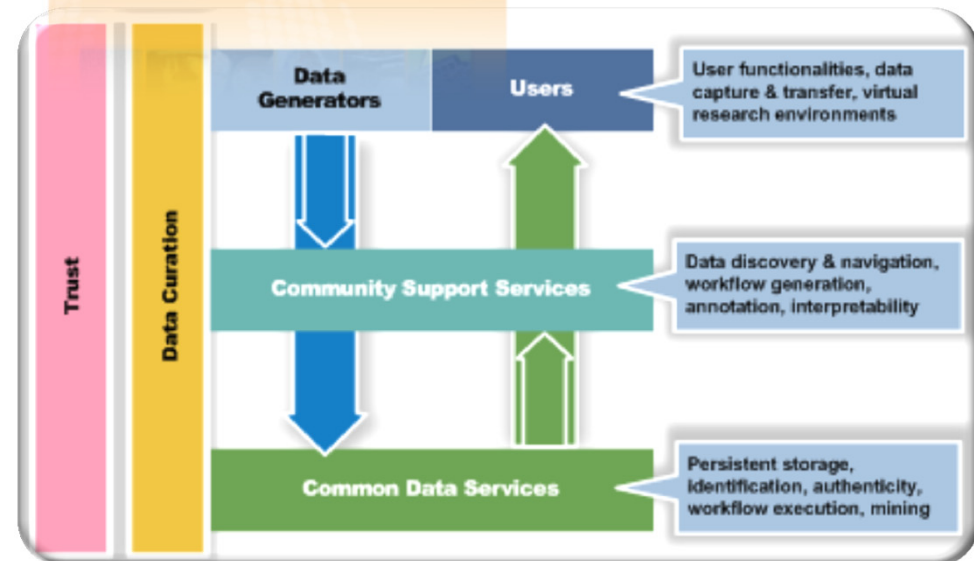
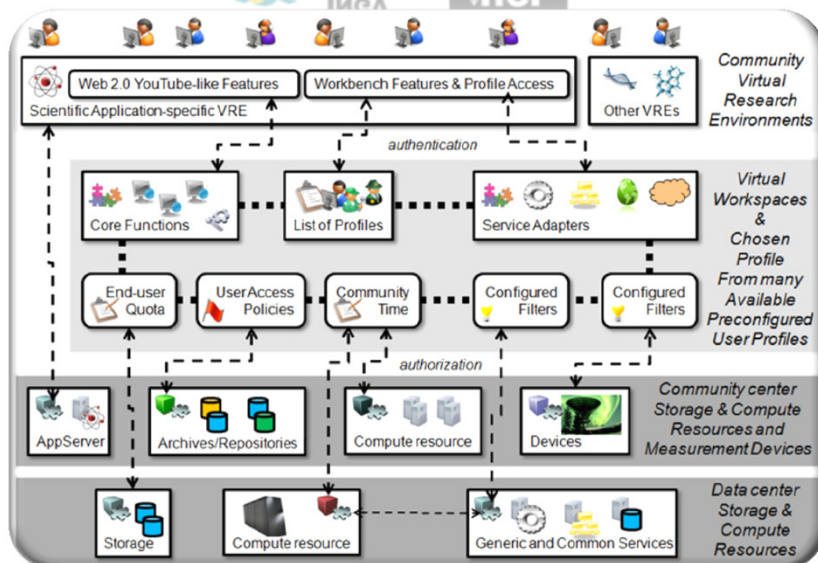
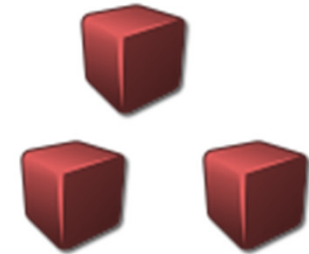
Digital Agenda
10011001010111011100001002010-2020
for Europe

[7] Digital Agenda for Europe, 2010

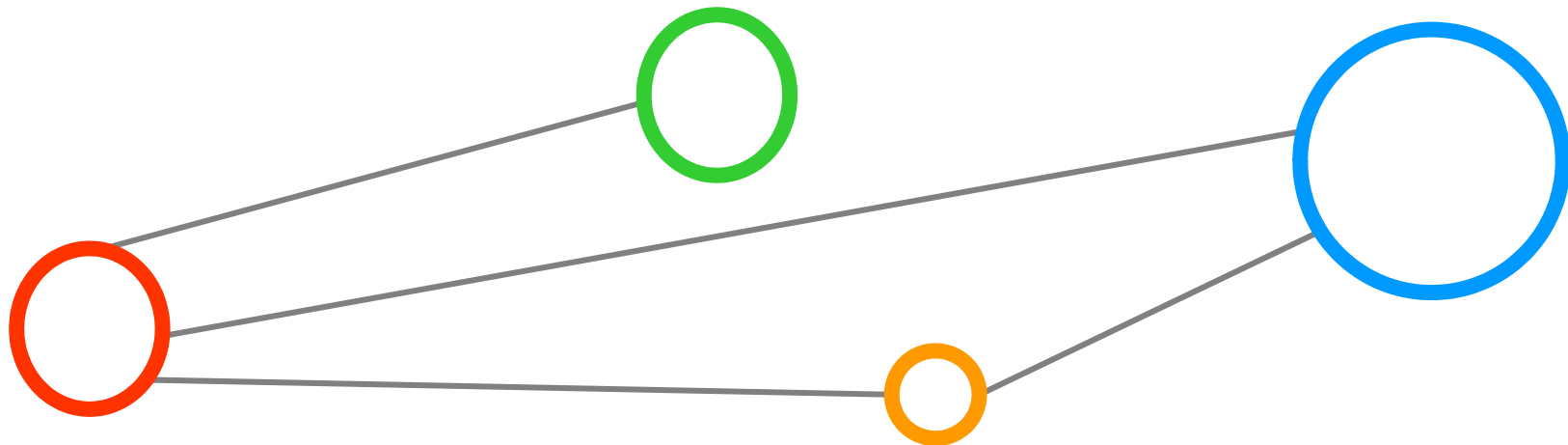
[12] Seeking new Horizons:
EGI'S Role in 2020, 2012



Daten: EUDAT



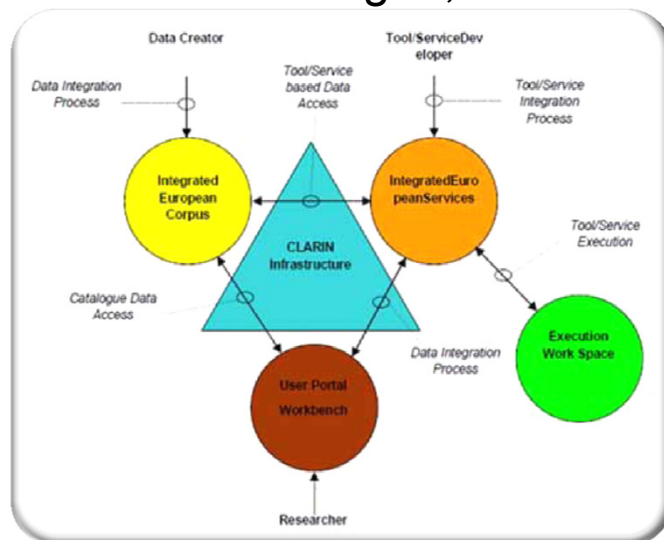
Forschungsinfrastruktur am Beispiel: CLARIN





Forschungsinfrastruktur CLARIN

- “Common Language Resources and Technology Infrastructure”
- Nutzen für die Wissenschaft:
 - Zugriff auf existierende Sprachen-Archive mit föderierter Infrastruktur
 - Bereitstellung von Sprachen-Tools und Technologien als Web Dienste
- Info zur nationalen Infrastruktur CLARIN-DE:
 - Dauer: 2011 - 2016
 - Förderung: 4,1 M€•

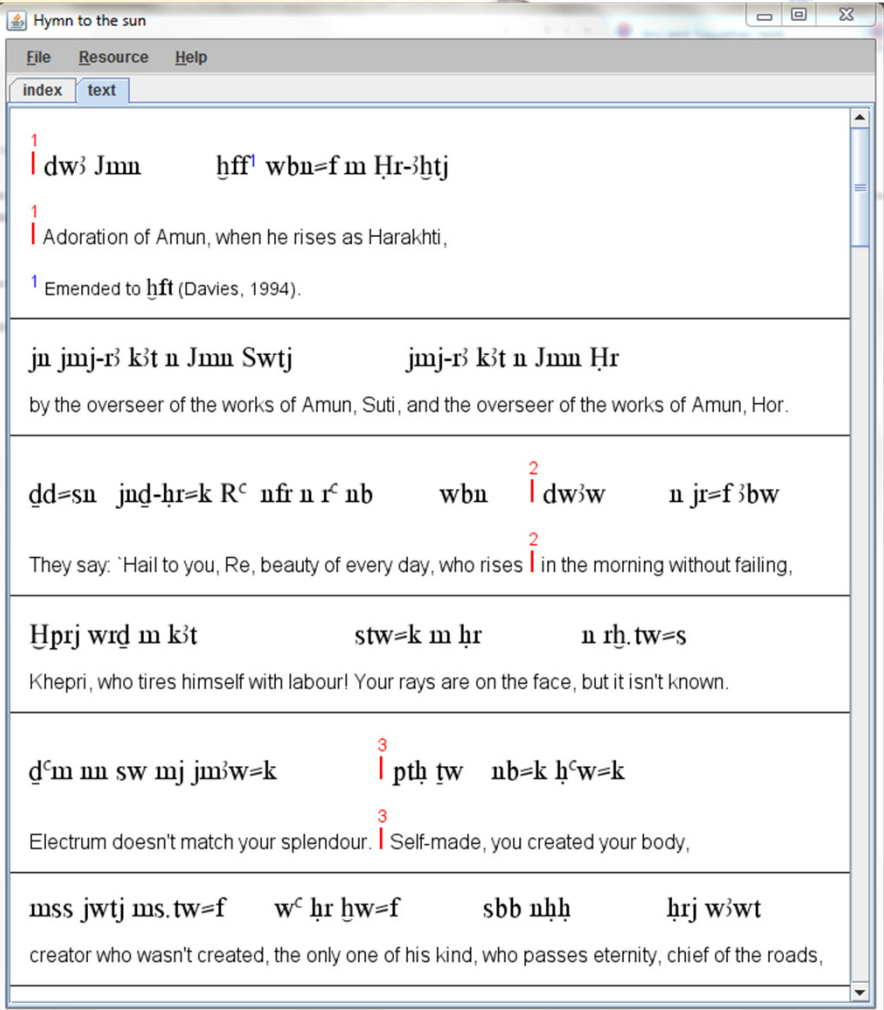


Morris Riedel

Digitalization & Archiving Services	
Natural Language Processing Services	
NLP services for Latvian by the Institute of Mathematics and Computer Science, University of Latvia	Text-to-Speech Synthesizer Morphological Analyser
NLP services for Polish by Wrocław University of Technology	The TaKPII-WS web service provides functionality for text tagging, lemmatization, segmentation, morphologic analysis and tokenization. The requests are handled in an asynchronous manner. Furthermore a web service connection to the Polish WordNet.
NLP services for Romanian and other languages by RACAI	Text Processing, Language Identification and Factored Translation
NLP Services for several languages by Leipzig University	Broad range of services, e.g. providing occurrences, neighbours and synonyms of words
NLP Services by UPF Barcelona	Statistical services, XSLT service, etc.
CLARIN-CAT-lab	Catalan portal for CLARIN technology
CLARIN-ES-lab	The Spanish virtual laboratory for CLARIN technology
NLP Services for Portuguese by the NLX group, University of Lisbon	chunker, tokenizer, tagger (WSDL, contact for WS-security credentials)

[4] CLARIN-EU, 2012

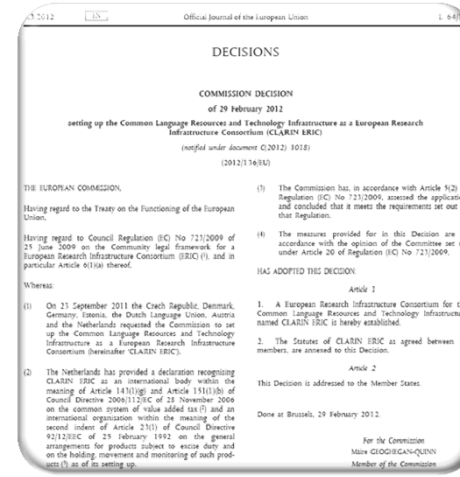
Ausgewählte Wissenschaftliche Themen



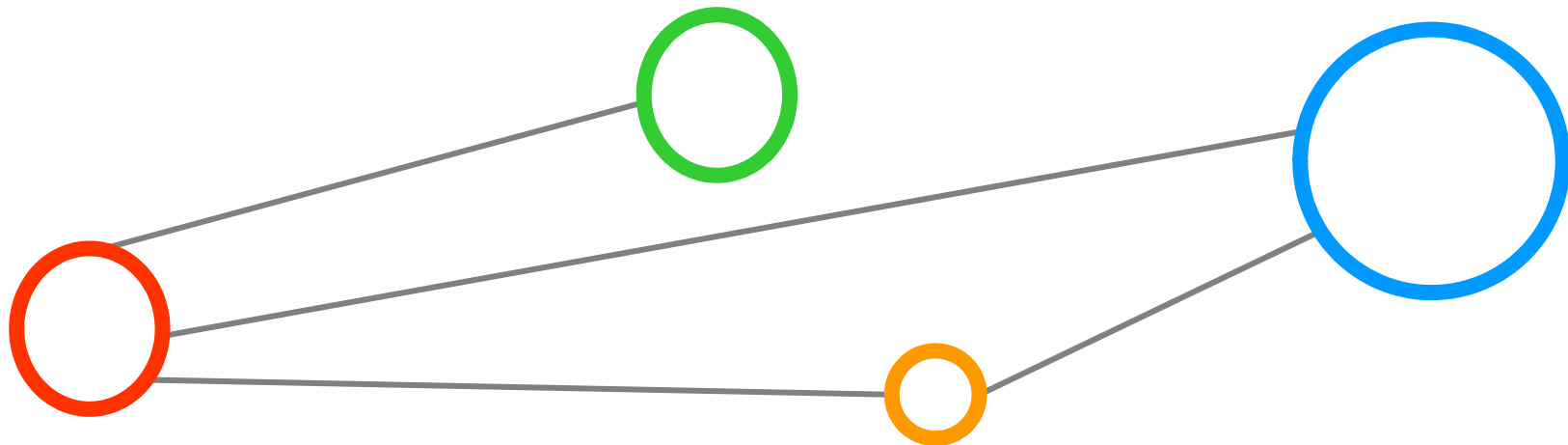
index	text
1	dwꜣ Jmn hffꜣ wbn=f m Hr-ḥtj Adoration of Amun, when he rises as Harakhti, Emended to hff (Davies, 1994).
	jn jmj-rꜣ kꜣt n Jmn Swtj jmj-rꜣ kꜣt n Jmn Hr by the overseer of the works of Amun, Suti, and the overseer of the works of Amun, Hor.
	ḏḏ=sn jnḏ-ḥr=k Rꜥ nfr n rꜥ nb wbn dwꜣw n jr=f ḥbw They say: 'Hail to you, Re, beauty of every day, who rises in the morning without failing,
	Hꜣrj wrꜥ m kꜣt stw=k m ḥr n rḥ.tw=s Khepri, who tires himself with labour! Your rays are on the face, but it isn't known.
	ḏꜥm nn sw mj jmꜣw=k pth tw nb=k ḥꜥw=k Electrum doesn't match your splendour. Self-made, you created your body,
	mss jwtj ms.tw=f wꜥ ḥr ḥw=f sbb nḥḥ ḥrj wꜣwt creator who wasn't created, the only one of his kind, who passes eternity, chief of the roads,

Zusammenspiel CLARIN-DE & CLARIN-EU

- CLARIN-DE ist der deutsche Beitrag der europaweit agierenden ESFRI Forschungsinfrastruktur CLARIN-EU
- Drei Phasen
 - eine *Vorbereitungsphase* (preparatory phase; ... - 2008),
 - eine *Aufbauphase* (construction phase; **2008-2010**) und
 - eine *Nutzungssphase* (operational phase; 2011-...).
- CLARIN-EU ist European Research Infrastructure Consortium (ERIC)
 - *Seit 29/02/2012*

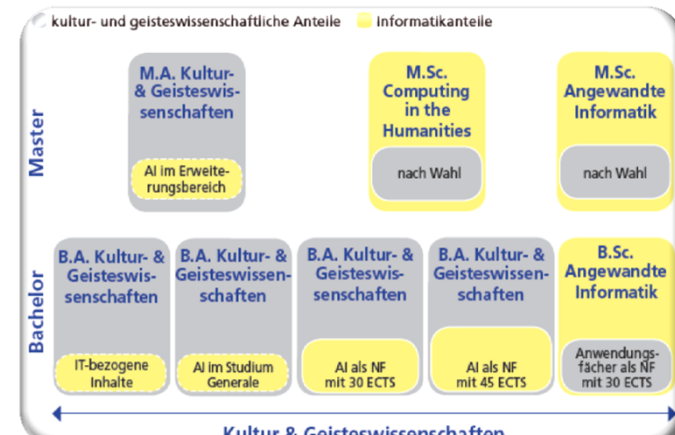


Forschungsinfrastruktur am Beispiel: DARIAH

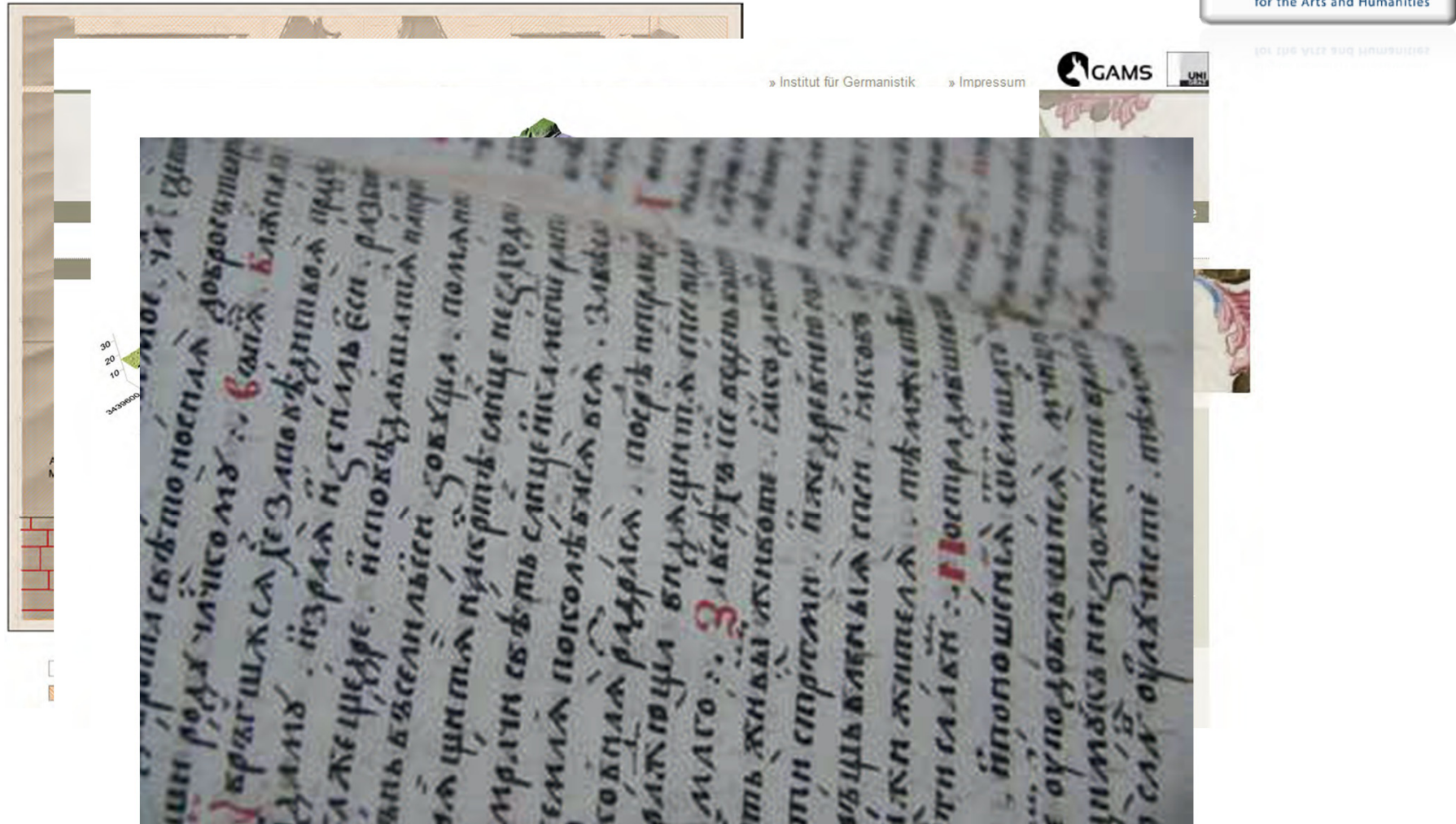


Forschungsinfrastruktur DARIAH

- “Digital Research Infrastructure for the Arts and Humanities”
- Nutzen für die Wissenschaft:
 - Bereitstellung digitalen Ressourcen für Geistes- und Kulturwissenschaftler
 - Unterstützung der “Digital Humanities” mit Methoden und Diensten
- Info zur nationalen Infrastruktur DARIAH-DE:
 - Dauer: März 2011 bis Februar 2014
 - Förderung: 5,8 M€



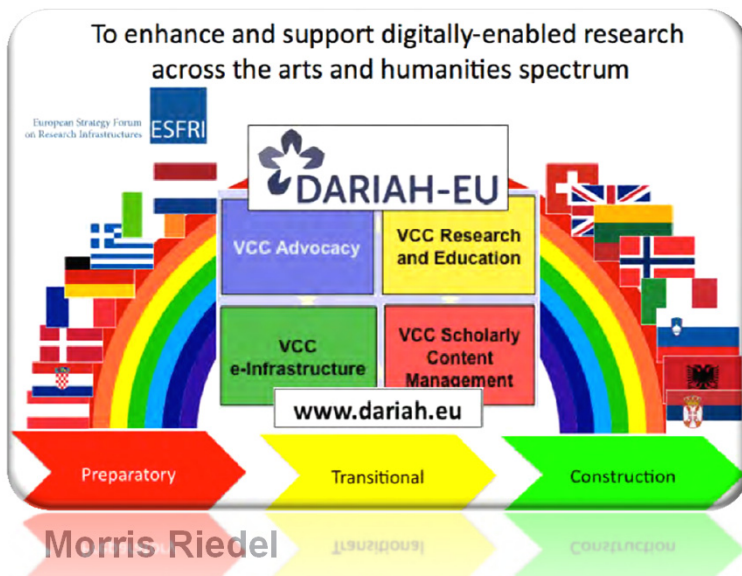
Ausgewählte Wissenschaftliche Themen



Zusammenspiel DARIAH-DE & DARIAH-EU



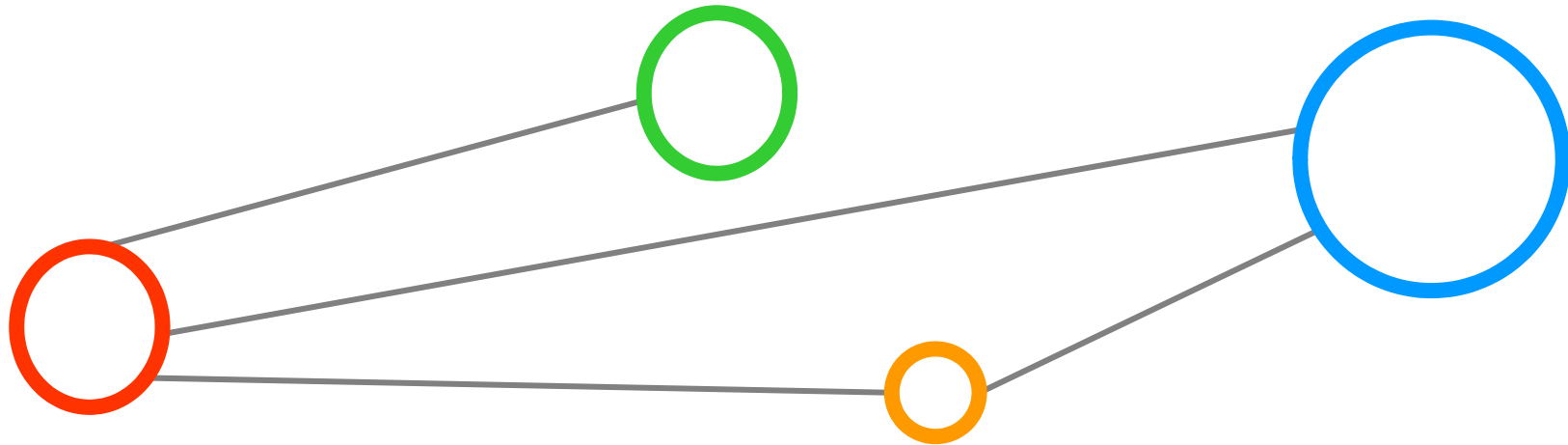
- DARIAH-DE ist der deutsche Beitrag der europaweit agierenden ESFRI Forschungsinfrastruktur DARIAH-EU
- Drei Phasen
 - eine *Vorbereitungsphase* (preparatory phase; 2008-2010),
 - eine **Aufbauphase** (construction phase; **2011-2015**) und
 - eine *Nutzungssphase* (operational phase; 2016-2020).
- DARIAH-EU plant European Research Infrastructure Consortium (ERIC)



By the end of 2011, the following 10 countries signed a Memorandum of Understanding formally stating their willingness to support the establishment of the DARIAH-ERIC:

- Austria
- Croatia
- Denmark
- France (Host Country)
- Germany (Coordinator)
- Greece
- Ireland
- The Netherlands (Coordinator)
- Slovenia
- Serbia

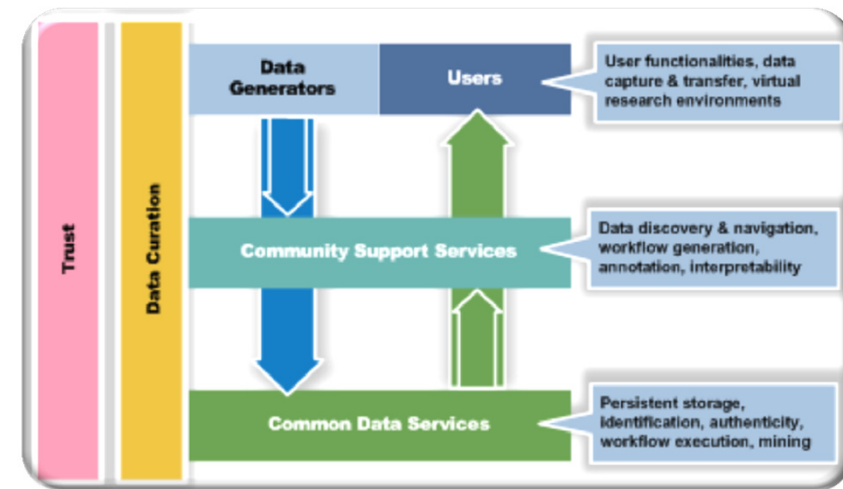
EU Forschungsinfrastruktur am Beispiel: EUDAT





European Collaborative Data Infrastructure

- Nutzen für die Wissenschaft:
 - Aufbau einer langfristig ausgerichteten förderierten Infrastruktur
 - Einfache, sichere und registrierte Policy-basierte Datenspeicherung
 - Zugriff über geografische Grenzen & wissenschaftliche Disziplinen hinweg
- Infos
 - Dauer: 10/2011 für 36 Monate
 - Förderung: 16,3 M€ (9,3M€ European Commission)





Ausgewählte Wissenschaftliche Themen

EPoS - European Plate Observing System
 CLARIN - Common Language Resource Infrastructure
 ENES - Service for Earth System Modelling
 VPH - The Virtual Physiological Human
LifeWatch - Biodiversity Data and Observatories

- Distributed data sensors
- About 200 centers in EU
- About 20 centers in EU
- Pilot project with 5 hospitals

- Distributed data sensors
- Metadata standardisation
- Interoperability reqs
- Involving most nature infrastructures
- Common reference model

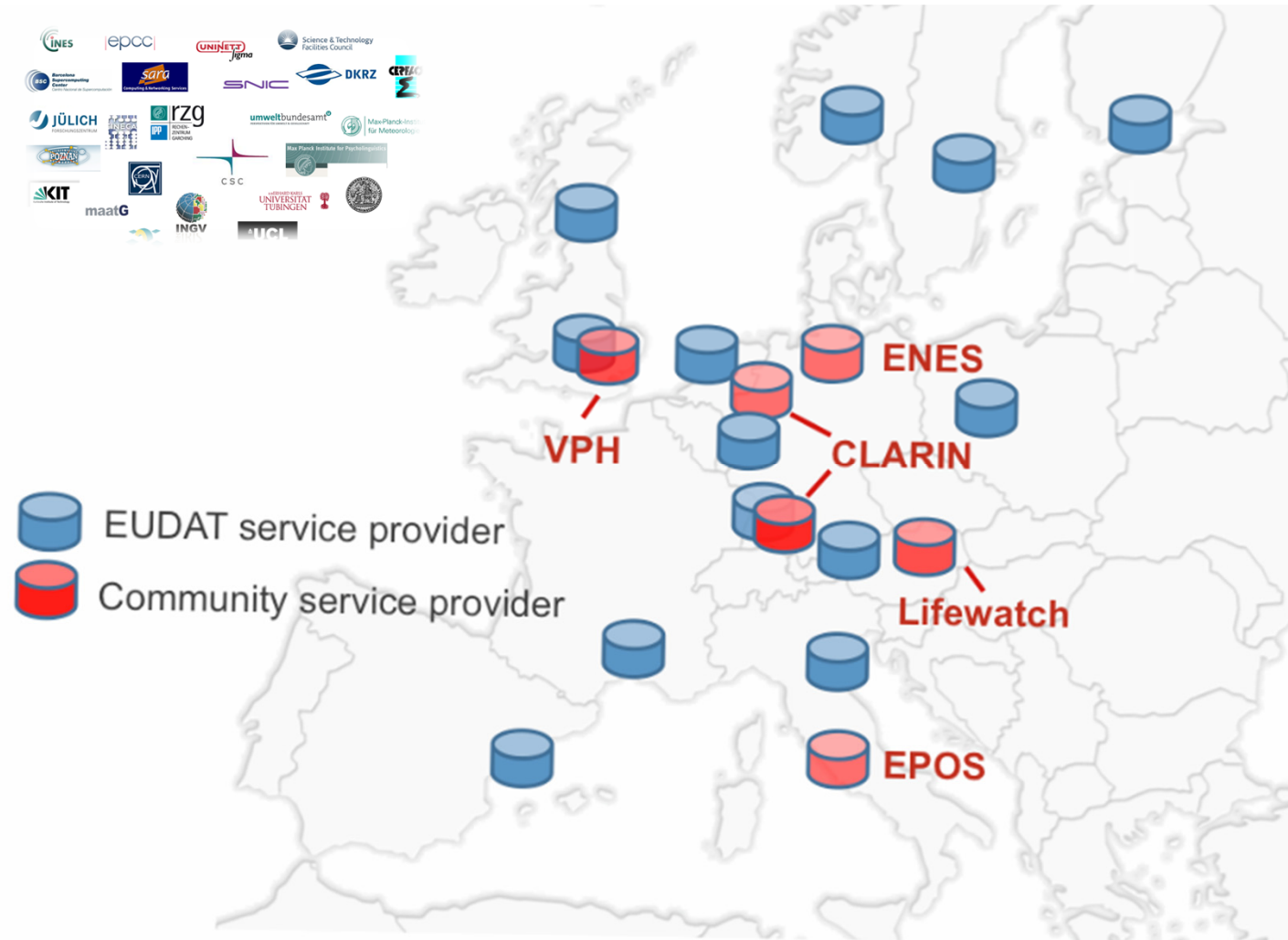


<http://envri.eu/>
<http://creative-b.eu/>

LifeWatch will construct and bring into operation the facilities, hardware, software and governance structures for all aspects of biodiversity research. Facilities for data generation and processing, data integration and interoperability. A network of observatories, virtual laboratories. A Service Center supporting scientific and policy users.

The effect of technology

Aufbau der Forschungsinfrastruktur



Dienste für Wissenschaftliche Daten



EUDAT Portal

Integrated APIs and harmonized access to EUDAT facilities



Metadata Catalog

Aggregated EUDAT metadata domain.
Data inventory

Requirement
Provide an inventory of metadata across disciplines

Function
• Join metadata domains
• Catalog indexing stored data

eudat-metadata@postit.csc.fi



AAI

Network of trust among authentication and authorization actors



Data Staging

Dynamic replication to HPC workspace for processing

Requirement
Provide a service to stage data between EUDAT infrastructure and HPC/HPC resources



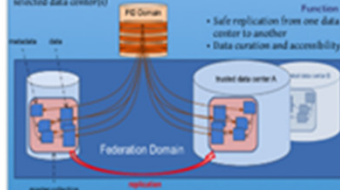
Function
Dynamically replicate subset of data stored in EUDAT to HPC workspace

eudat-datastaging@postit.csc.fi

Safe Replication

Data curation and access optimization

Requirement
Provide a service to replicate and curate data to selected data center(s)



Function
• Safe replication from one data center to another
• Data curation and accessibility

eudat-safereplication@postit.csc.fi

Simple Store

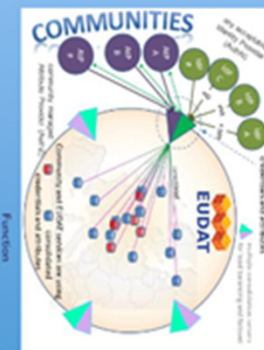
Researcher data store (simple upload, share and access)

Requirement
Provide a simple service to store user data temporarily



Function
Simple upload
Store data

eudat-simplestore@postit.csc.fi

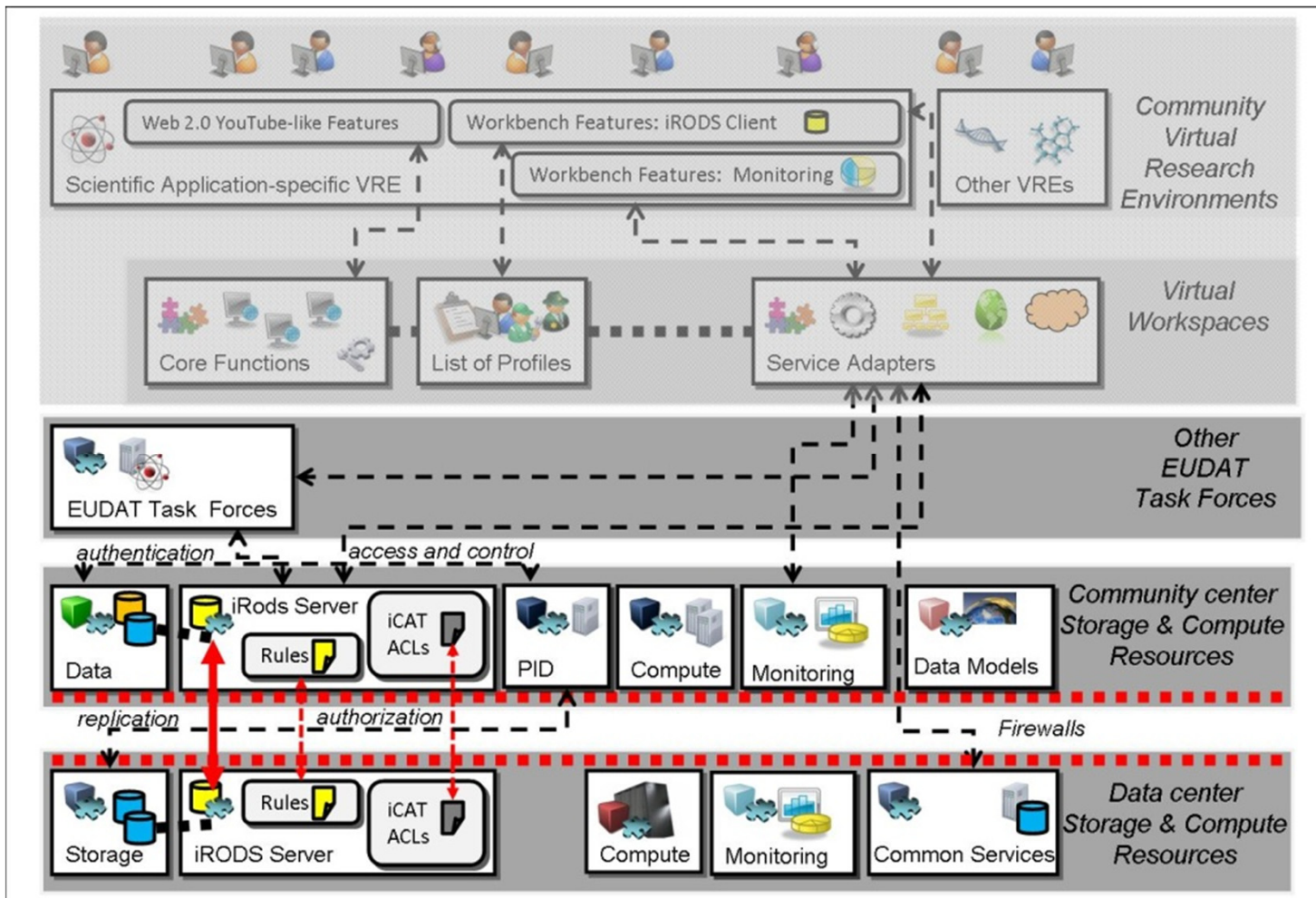


Requirement
Provide a working AAI system in a federated scenario

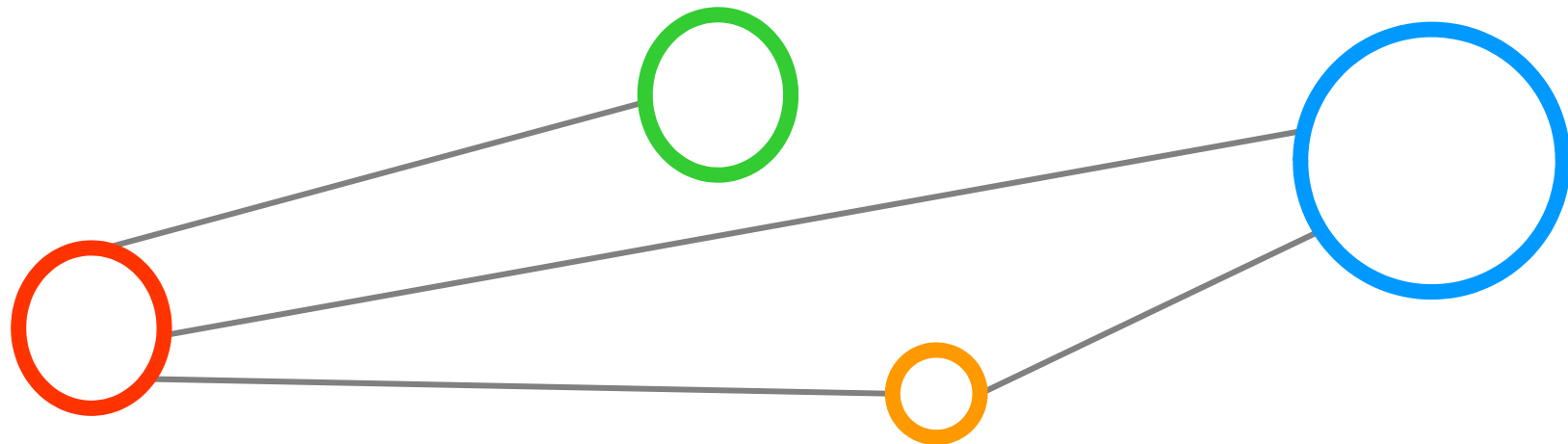
Function
• Integrate existing authentication systems
• Establish a network of trust among AAI AAI providers
• Establish a network of trust among AAI AAI providers
• Establish a network of trust among AAI AAI providers

eudat-AAI@postit.csc.fi

Technisches Beispiel: Datenreplikation

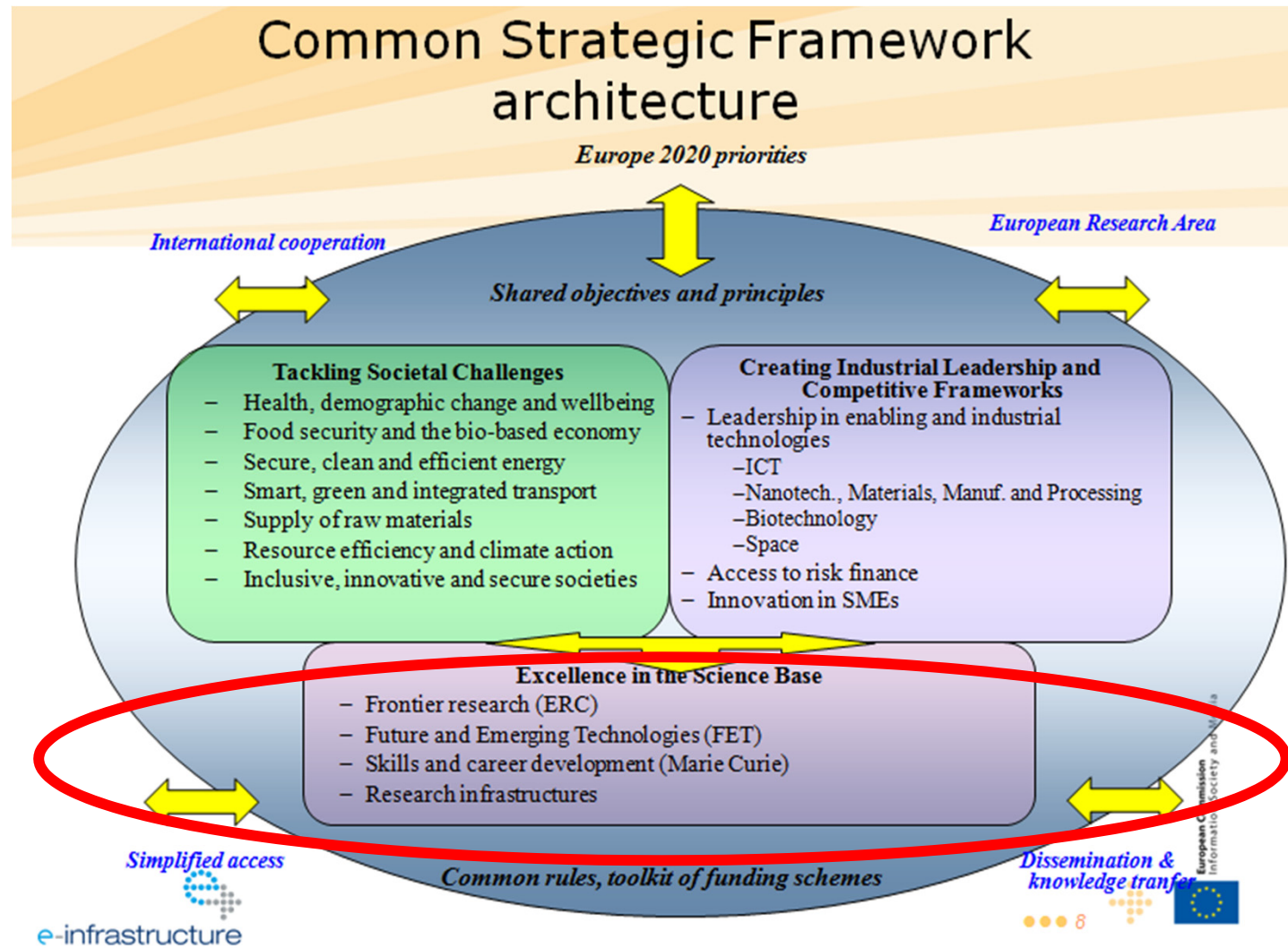


Ausblicke

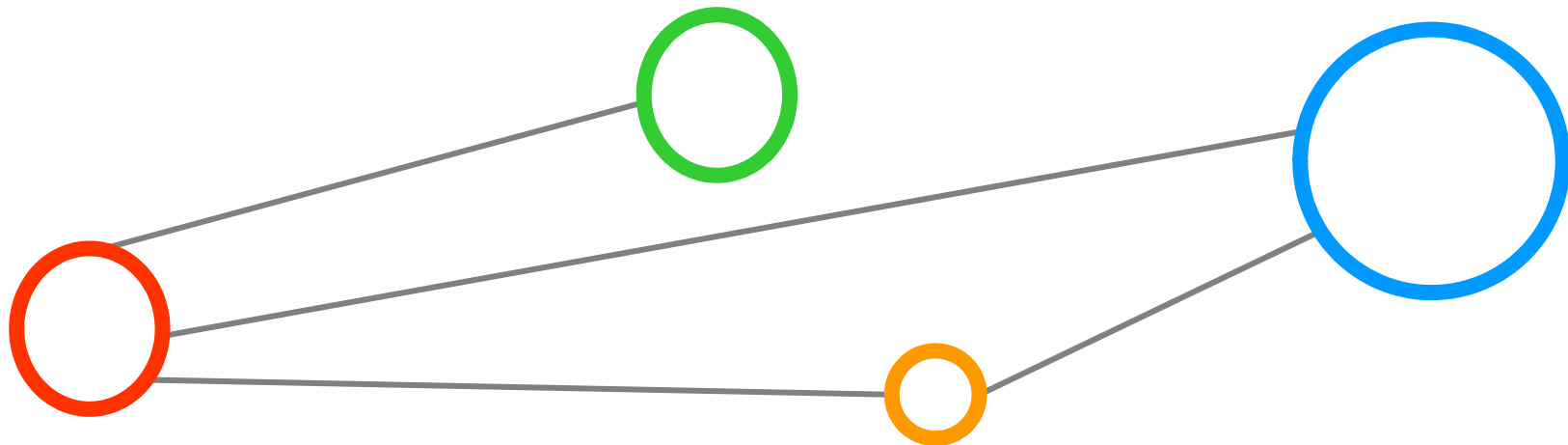




Ausblicke: EU Horizon 2020 Programm



Mögliche Erfolgsfaktoren

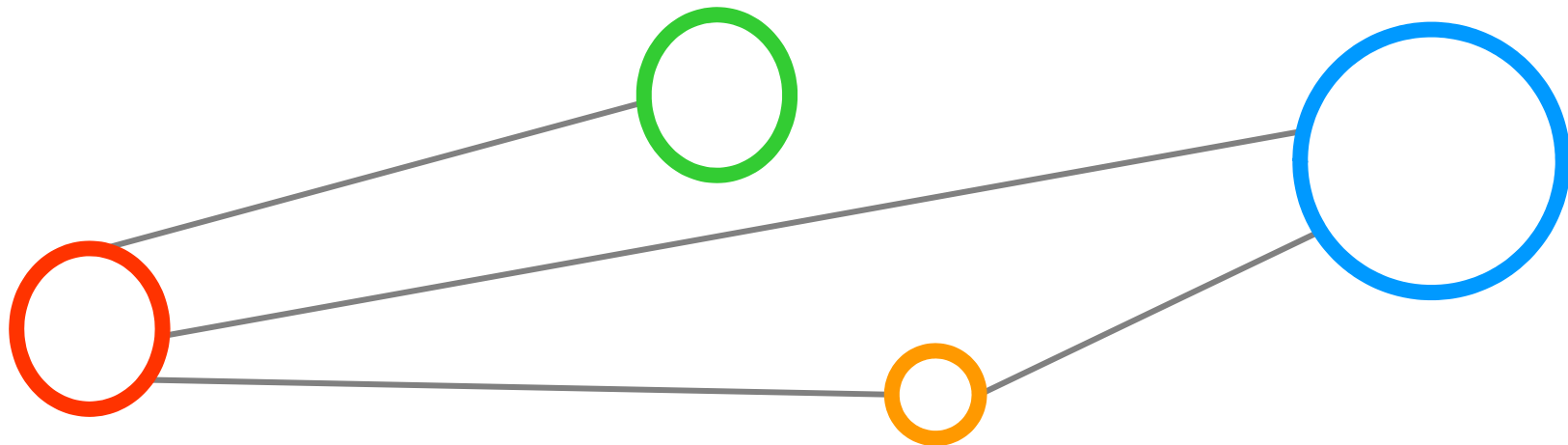


Mögliche Erfolgsfaktoren

- Bausteine von Forschungsinfrastrukturen
 - Nach dem Prinzip der Föderation aufbauen und existierendes einbinden
 - Dienste “produktisieren”, aber nicht eine dedizierte Middleware fordern
 - Modularer Aufbau von Bausteinen mit Diensten und Ressourcen
- Beitrag zur Wissenschaft im Fokus
 - Wissenschaftler im “Driving Seat” → Entscheidungen & Uptake
 - “Application Enabling” → Technicker & Wissenschaftler arbeiten zusammen
 - “Building the workforce” → Infrastrukturbenutzung in die Lehre
- Vertrauen in die Forschungsinfrastruktur
 - Langfristige Pläne, Koordinierung und Zusammenarbeit
 - Vertrauen in Nachhaltigkeit bringt “mehr Vertrauen”
 - “Bottom-up and only slightly coordinated”



Referenzen



Referenzen

[1] K. Glinos, *e-Infrastructure in EU*, 2012 (Online)

https://gridlab.man.poznan.pl/Meetings/Zakopane2012/EU_Konstantinos_Glinos.ppt

[2] DARIAH-EU, 2012 (Online)

<http://www.dariah.eu>

[3] ESFRI Roadmap, 2012 (Online)

<http://ec.europa.eu/research/esfri>

[4] CLARIN.EU, 2012 (Online)

<http://www.clarin.eu/external/>

[5] M. Riedel, P. Wittenburg, J. Reetz, M. van de Sanden, J. Rybicki, B. von St. Vieth, G. Fiameni, G. Mariani, A. Michelini, C. Cacciari, W. Elbers, D. Broeder, R. Verkerk, E. Erastova, M. Lautenschlaeger, R. Budig, H. Thielmann, P. Coveney, S. Zasada, O. Buechner, C. Manzano, M. Memon, A. Memon, D. Lecarpentier, H. Helin, K. Koski, and T. Lippert. *Reference Model Design of a Federated Data Infrastructure - Towards the Realization of a ScienceTube Vision. Journal of Internet Services and Applications: Special Issue on Data Intensive Computing*, 2012.

[6] *Riding the Wave, Report on Scientific Data*, 2010

[7] *Digital Agenda for Europe*, 2010

[8] *Horizon 2020*

[11] PRACE

<http://www.prace-ri.eu>

[12] *Seeking new Horizons: EGI'S Role in 2020*, 2012

[13] EUDAT

[14] EPOS

[15] CLARIN

[16] ENES

[17] VPH

[18] LifeWatch

Morris Riedel