Perspectives: Archiving, Networking, Infrastructures

Peter Wittenburg
The Language Archive - Max Planck Institute
CLARIN European Research Infrastructure
Nijmegen, The Netherlands

Content

1. What is e-Science/e-Research?
   a) Data
   b) Operations
2. What do other communities do?
3. What does CLARIN do?
   a) Federation
   b) PIDs
   c) Metadata
   d) Web - Services
4. General Issues

Content

1. What is e-Science/e-Research?
   a) Data
   b) Operations
2. What do other communities do?
3. What does CLARIN do?
   a) Federation
   b) PIDs
   c) Metadata
   d) Web - Services
4. General Issues

What an innovation rate!

The Language Archive

1970	1990	2010
Internet (Arpanet) just started.	1 Web server at CERN.	$7*10^5$ Web servers, worldwide.
First cross-country link installed by AT&T between UCLA and BBN at 56kbps.	NSF NET backbone upgraded to T3 (44.736Mbps)	Internet speeds Mbps widespread
23 hosts	313,000 hosts	10^9 hosts
Mainframes	Desktops personal computers	Ubiquitous mobile devices
1 TB storage cost $3*10^{11}$ $	1TB disk storage costs 10M€	1TB disk storage costs 100 €
~300,000 publications	~900,000 publications	1,800,000 publications
In March 1972, Ray Tomlinson (BBN) modifies the email program he released in 1971	Email use common in academic and research institutions.	183 billion spam email messages sent daily
3 April 1973: Motorola employee Martin Cooper placed the first hand-held cell phone call to Joel Engel, head of research at AT&T's Bell Labs,	Cell phones, initially very bulky, start to appear .	6 out of 10 people worldwide have a cell phone. Mobile phones have power of 1970's mainframes
Hobby computers emerge in 1975.	Installed base of PCs worldwide is $100*10^6$	Installed base of PCs worldwide is $1{,}415*10^6$
specialized analog video devices	some started with video on computers	now digital video handling is standard

Will innovation stop?

The Language Archive

- what can we expect for YOUR decades?

- can we predict how publishing will change?
- do we know how computational methods will change?
- do we know how research will change?

- predicting the future is probably too hard
- so let's look at what happens now to be prepared

eScience & Infrastructure

The Language Archive

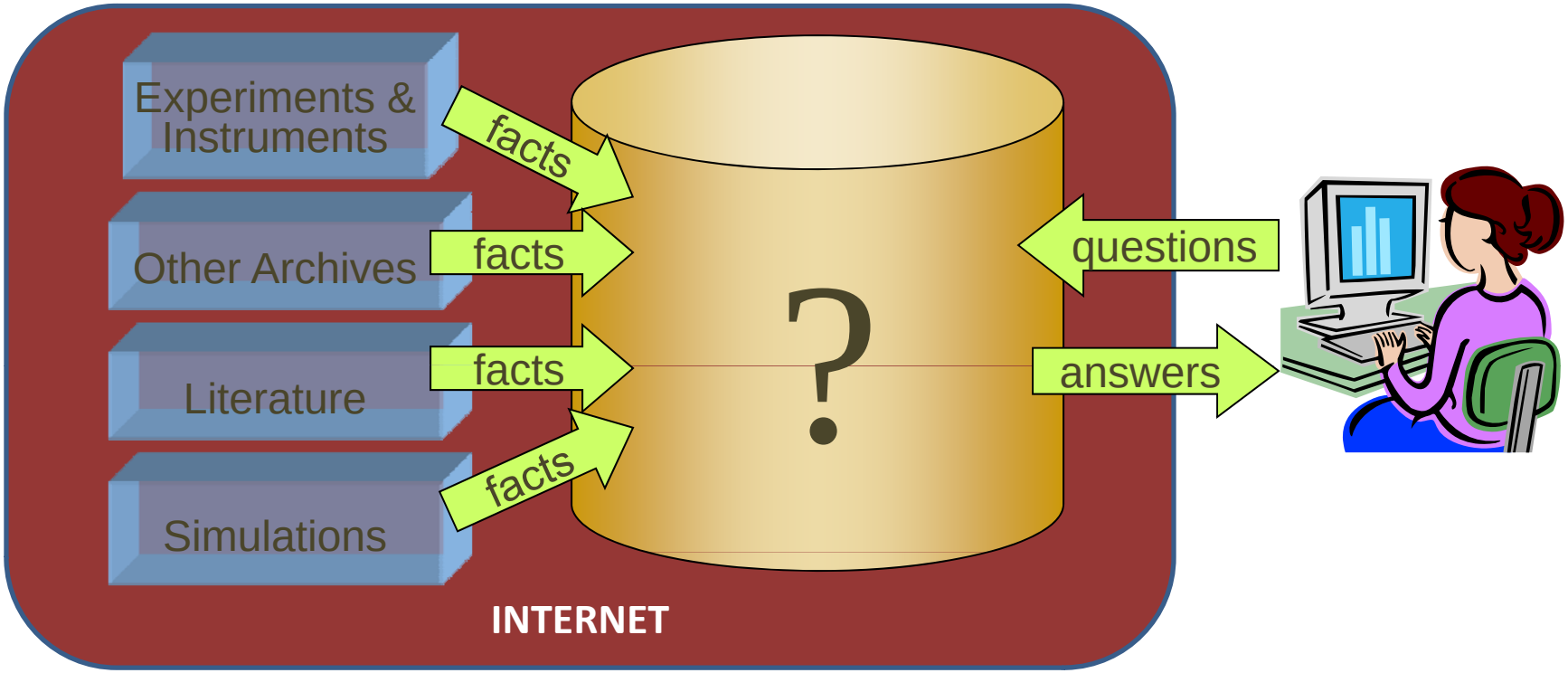
J. Taylor

"eScience is about global collaboration in key areas of science and the next generation of infrastructures that will enable it"

Need to build new types of platforms

- to allow researchers to combine existing resources easily to new ones to tackle the big challenges
- to increase the productivity of all interested researchers, since currently too much time is wasted by preparatory work
- 40 % of time of knowledge workers is spent on finding material

Dream of the eResearcher or?

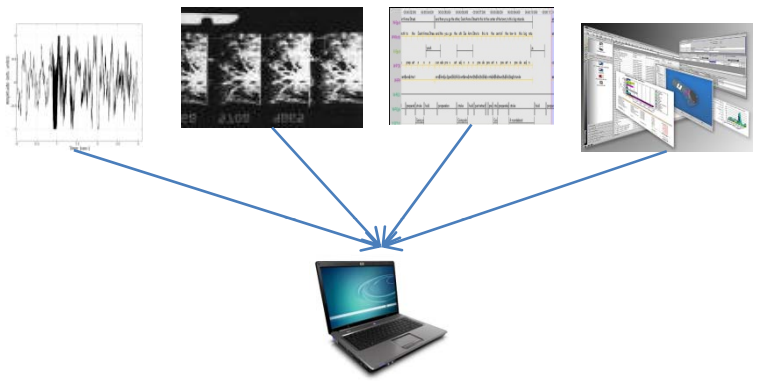


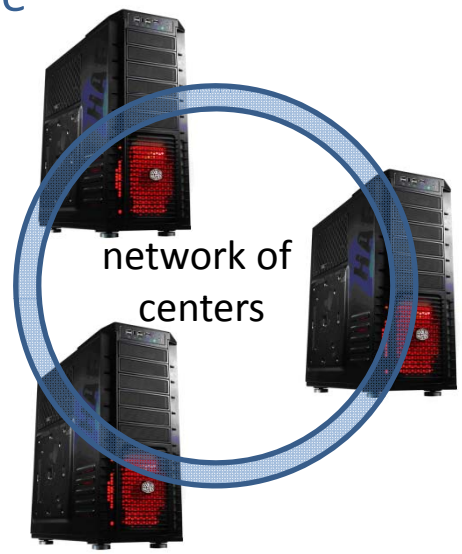
- easy data ingest
- managed petabytes+
- based on common schema(s)
- easy organization
- easy re-purposing

- easy to coexist & cooperate with other scientists and researchers?
- simple data query and visualization tools
- good support/training
- high performance

Change Download-First Behavior

down-load first vs. cyberinfrastructure





does not seem to be efficient
but has some advantages
will remain - but need another dimension

- this may facilitate working with language resources and tools
- many communities are working along same goals
  (life sciences, bioinformatics, geosciences, etc.)
- funders are changing their rules (NL, recently NSF)

e-Research paradigm

- obviously nature of research work is changing
- culture of doing research is changing
- what is it what is driving things
  - technology innovation brought us new "instruments" (digital age, smart sensors, fast networks, fast computers, huge storage capacity, Internet/Web, Semantic Web, etc.)
  - lost distance in place and time and feeling for amounts
- research work needs to be based on data also in SSH
- thus much more quantitative analysis
  - but does more data mean "more truth"?
  - but can we forget about "bright minds"?
  - let's take care - there are traps

Big challenges

- need to tackle the big questions as well
  - stability of our societies and minds
  - loosing cultural and linguistic identity
  - structural blurring and mental processing

- we can count (NLP), do minimal pattern recognition, etc
  but ...
  - can we simulate as climate people do? NO
  - can we process natural interaction? NO
  - can we operate cross-lingual? NO

How far do we want to go?

The Language Archive

- **Do you want "Collective Intelligence"**
  - If Amazon can recommend which book to offer to me based on what my friends are reading, should the cyberinfrastructure of the future recommend articles of potential interest based on what the experts in the field that I respect are reading?

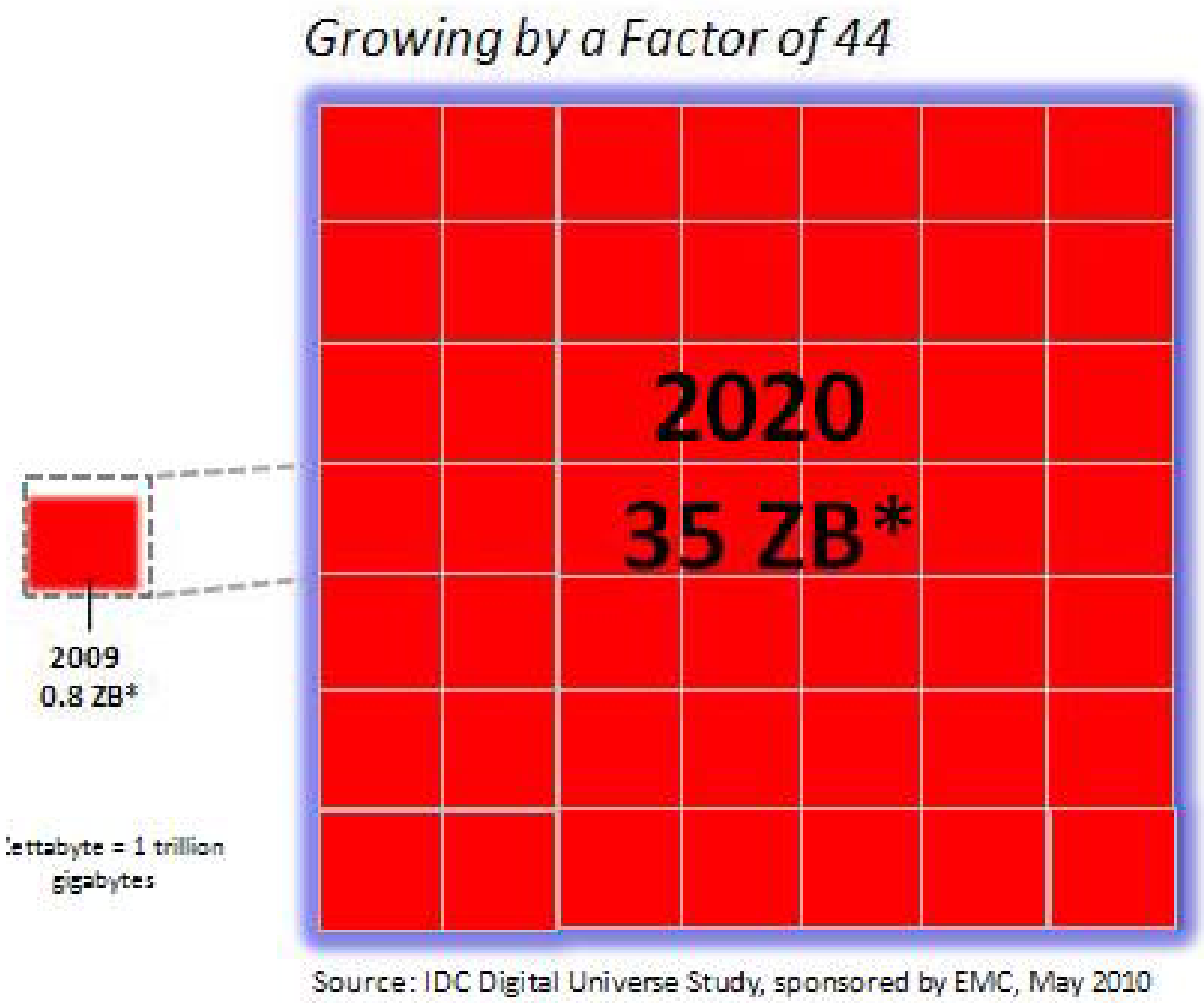
- **Semantic Computing**
  - Automatic correlation of scientific data
  - Smart composition of services and functionality

- **Leverage cloud computing to aggregate, process, analyze and visualize data**

Content

1. What is e-Science/e-Research?
   a) Data
   b) Operations
2. What do other communities do?
3. What does CLARIN do?
   a) Federation
   b) PIDs
   c) Metadata
   d) Web - Services
4. General Issues

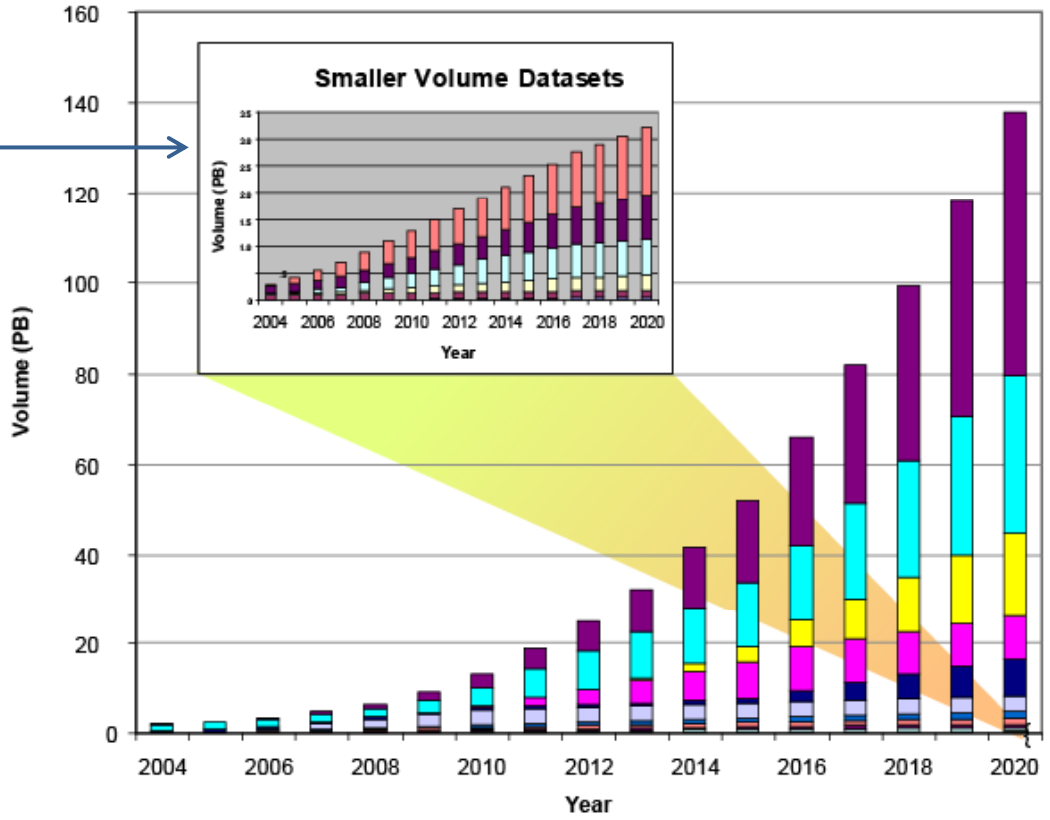
Global Data Development



US Environmental Data Archive

The Language Archive

Comprehensive Large Array-data Stewardship System (CLASS) Storage



Smaller Volume Datasets

Space Based Data

- Polar Orbiting Earth Satellites (POES)
- Defense Meteorological Satellites Program (DMSP)

Earth Based Data

- Atmosphere(Weather & Climate)
- Ocean (Weather & Climate)
- Continually Operating Reference Stations (CORS)
- Misc (Mesonets)

Large Volume Datasets

Space Based Data

- NOAA Polar-orbiting Operational Environmental Satellite System (NPOESS)
- NPOESS Preparatory Project (NPP)
- Geostationary Operational Environmental Satellites (GOES)
- NASA Earth Observing System(Moderate Resolution Spectroradiometer) (EOS MODIS)
- Meteorological Operational Satellite Program (MetOp)

Earth Based Data

- Weather Radar

Model Data

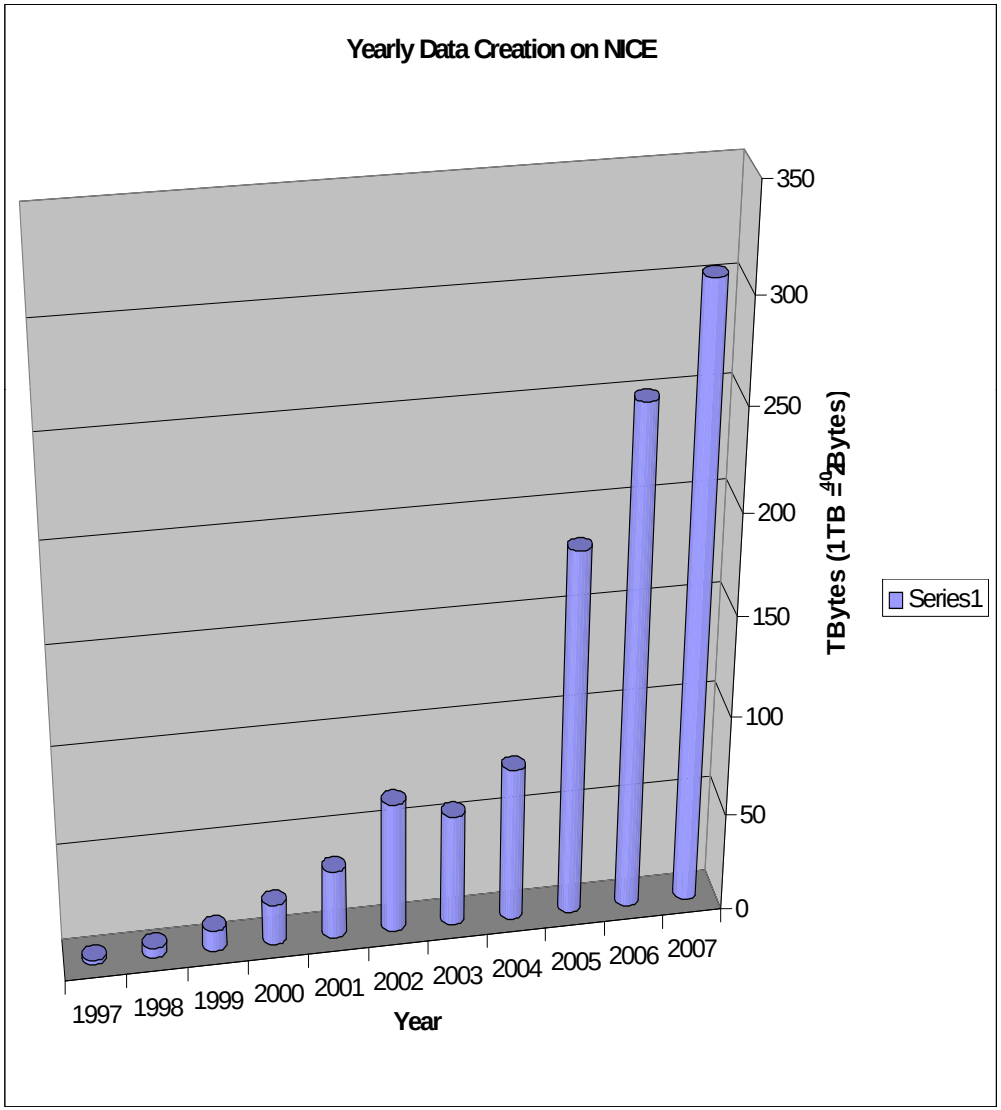
- Atmosphere & Ocean (reanalysis)

European Synchroton Radiation Facility

The Language Archive

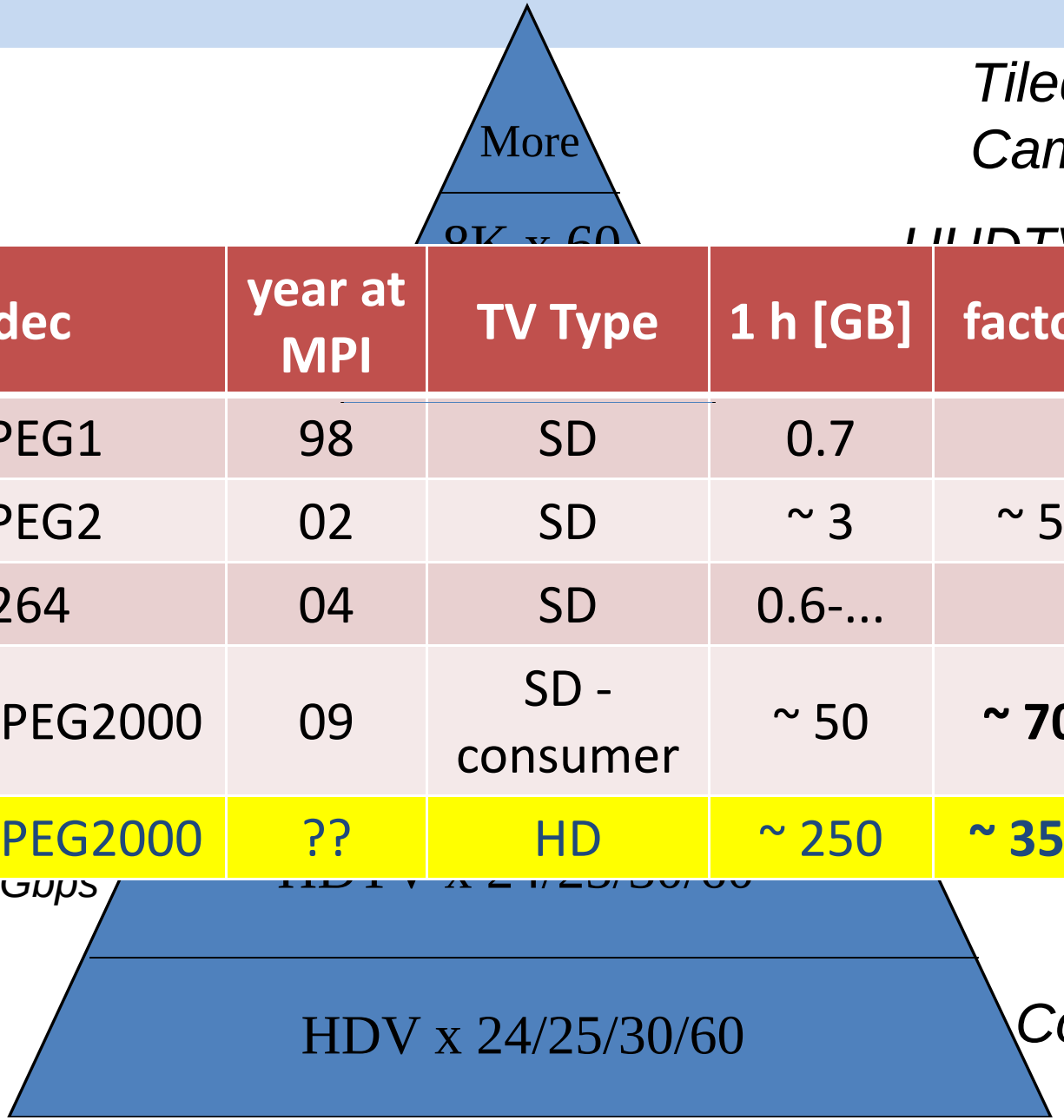
- 10 years → data volume x 300. In 2007: 300TB ~1*10^8 files
- However, doubling of the data centre infrastructure (m^2, kW, cooling)

- **green computing is an issue**



How much data do we create?

The Language Archive



Tiled Displays
Camera Arrays

1 - 24 Gbp

HUHDTV (far future)

500 Mbps

4K (future)

250 Mbs -

Quad HD)

250 Mbps

al Cinema

200 Mbps

Stereo HD

20 Mbps - 1.5 Gbps

HDTV

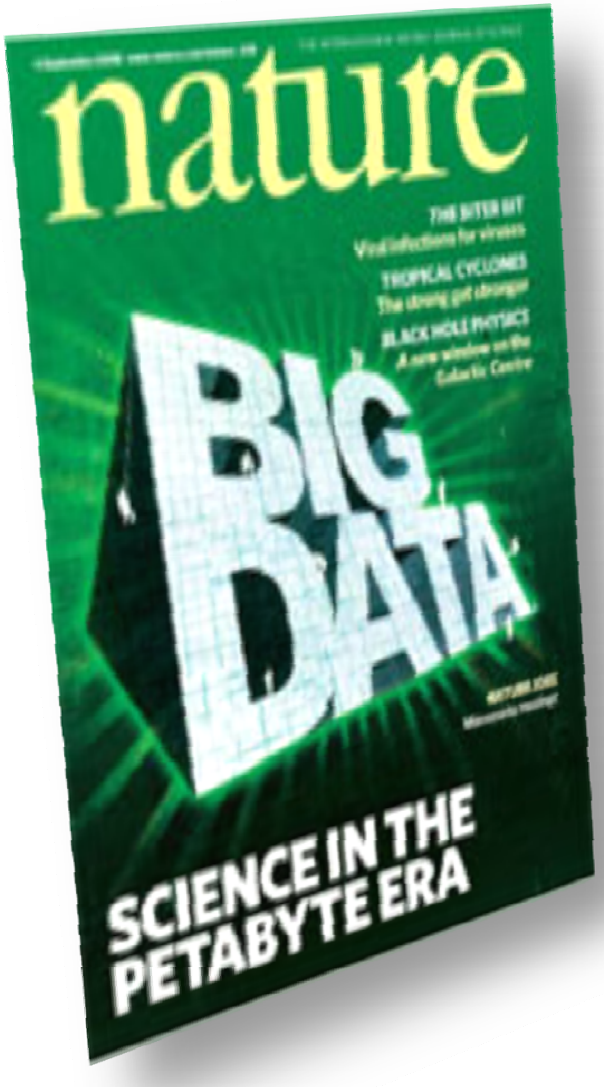
5 - 25 Mbps

Consumer HD

codec	year at MPI	TV Type	1 h [GB]	factor
MPEG1	98	SD	0.7	
MPEG2	02	SD	~ 3	~ 5
H.264	04	SD	0.6-...	
mJPEG2000	09	SD - consumer	~ 50	**~ 70**
mJPEG2000	??	HD	~ 250	**~ 350**

Data Tidal Wave

The Language Archive



Fourth Research Paradigm

The Language Archive

1. Thousand years ago – **Experimental Science**
   – Description of natural phenomena
2. Last few hundred years – **Theoretical Science**
   – Newton's Laws, Maxwell's Equations…
3. Last few decades – **Computational Science**
   – Simulation of complex phenomena
4. Today – **Data-Intensive Science**
   - Scientists overwhelmed with data sets from many different sources
     - Data captured by instruments
     - Data generated by simulations
     - Data generated by sensor networks
   - eScience is the set of tools and technologies to support data federation and collaboration
     - For analysis and data mining
     - For data visualization and exploration
     - For scholarly communication and dissemination

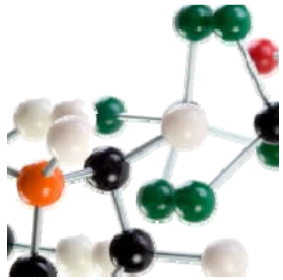
Astronomy has been one of the first disciplines to embrace data-intensive science with the Virtual Observatory (VO), enabling highly efficient access to data and analysis tools at a centralized site. The image shows the Pleiades star cluster form the Digitized Sky Survey combined with an image of the moon, synthesized within the World Wide Telescope service.

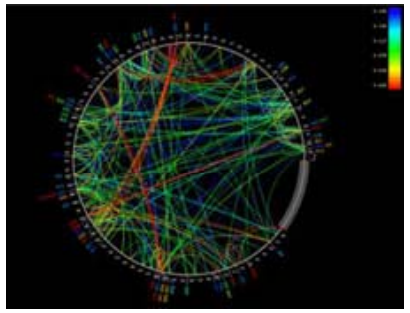
*With thanks to MS and **Jim Gray***

Concerns with Data Sharing

The Language Archive

- **Data integration / interoperability**
  - Linking together data from various sources
- **Annotation**
  - Adding comments/observations/relations to existing data
- **Provenance (and quality)**
  - 'Where did this data come from?'
- **Exporting/publishing in agreed formats**
  - To other programs, as well as people
- **Security**
  - Specifying or enforcing read/write access to your data (or *parts* of your data)





Granularity, Identity & Authenticity

The Language Archive

- identity of an object by an explicitly registered PID record with
  - checksum
  - time stamp
  - pointer to metadata
  - pointers to copies

- but what is an object in linguistics?
- which granularity is appropriate?
  - is it a whole database with all your dynamic data in it?
  - is a container appropriate requiring own application logic?
  - is it a lexical entry which is part of a large lexicon?
- these issues are not at all clear
- at MPI (and others) linguistically meaningful units
  such as a lexicon, a video, an annotation tier, etc

Nature of Research Collections

The Language Archive

- research collections are dynamic - continuous change
  (transformations, extensions, modifications/versions, relations, etc)
- collections are created with certain research purpose in mind
- but users re-combine objects in unpredicted ways
  (virtual collections crossing institutional boundaries)
- collections include a variety of resource types
- increasingly complex external relationships
  (raw data -> derived data -> annotations -> extractions ...)
- access patterns can hardly be predicted

Organization of Collections

The Language Archive

- high-quality metadata are crucial for management and access
- hq metadata allows to generate different trees based on some useful criteria
- depositors and managers need one canonical tree (for management operations and rights definitions)
- users want to create their own virtual organizations
- users want to have several types of organizations and sub-collections
- users want to express various types of relations -> graphs
- metadata are crucial for long-term interpretation
  - therefore representation in schema based format
  - various visualizations required (catalogue browsing, searching, faceted search, GIS)

Building of Collections

- build up "repository" experts or collaborate with them
- define required descriptive categories
- re-use existing ones and register new ones in open registries (ISO 12620 - ISOcat; www.isocat.org)
- map new ones to existing ones where possible
- make use of a flexible component framework where components/profiles refer to registered concepts via PIDs
- re-use or create important vocabularies
- check quality of all objects and do curation asap
- allow metadata harvesting via OAI-PMH
- provide efficient metadata organization/description tools (design of ARBIL after 10 years of experience)

Access to existing Collections -

- cater for different visualizations of metadata
  (GIS, Faceted Browser, simple/complex search, etc.
  see VLO: www.clarin.eu/vlo)
- create "community portals" - costs are high
  (nice web-pages with embedded metadata queries)
- cater for machine usage, i.e. support APIs
- provide DublinCore semantics for the occasional user
- why not social tagging, but keep it separate

Sustainability of Collections

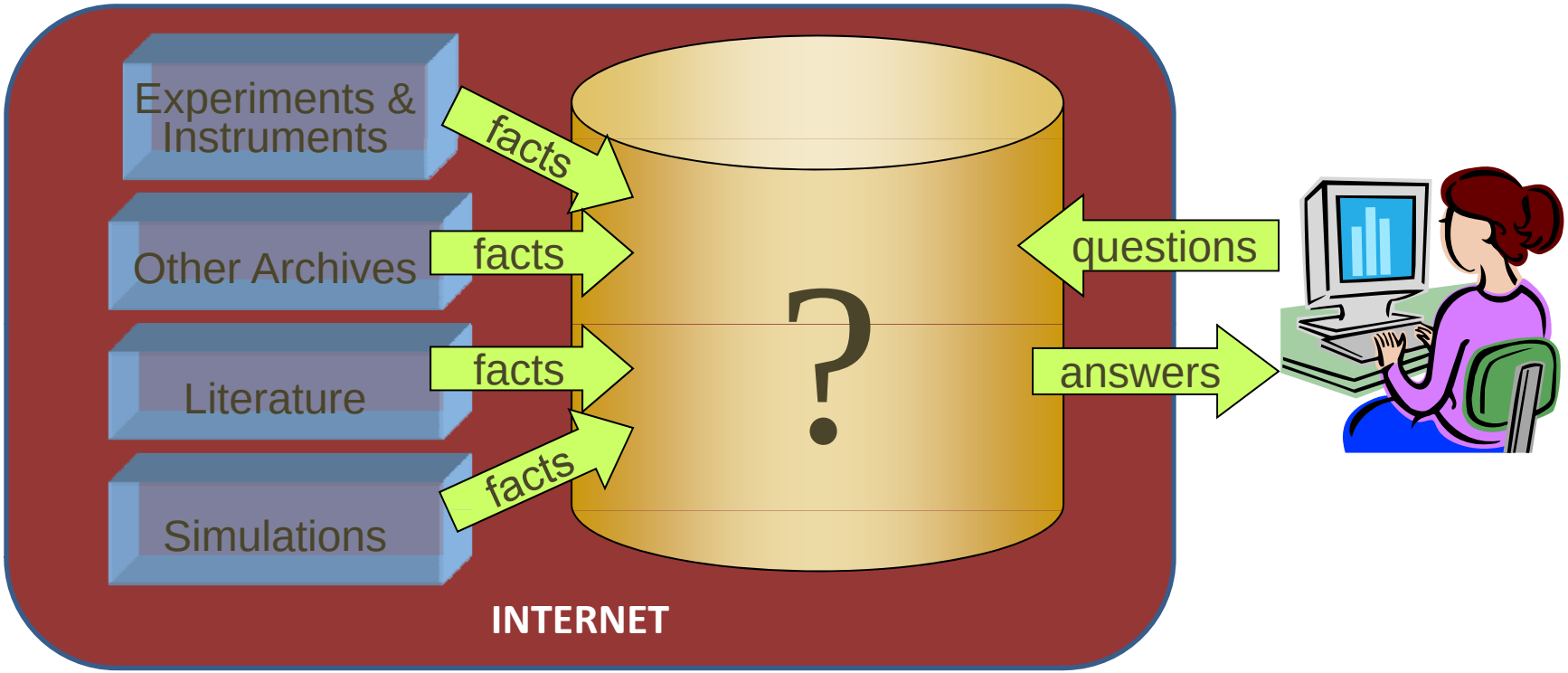
The Language Archive

- responsibility aspect
  - funding & research organizations are responsible
  - communities are responsible to enforce decisions
  - need to reserve some percentage for continuous access
  - need a process for taking decisions
    (experimental data often obsolete after N years)
  - need quality assessment procedures
    MOIMS-RAC, Data Seal of Approval

Content

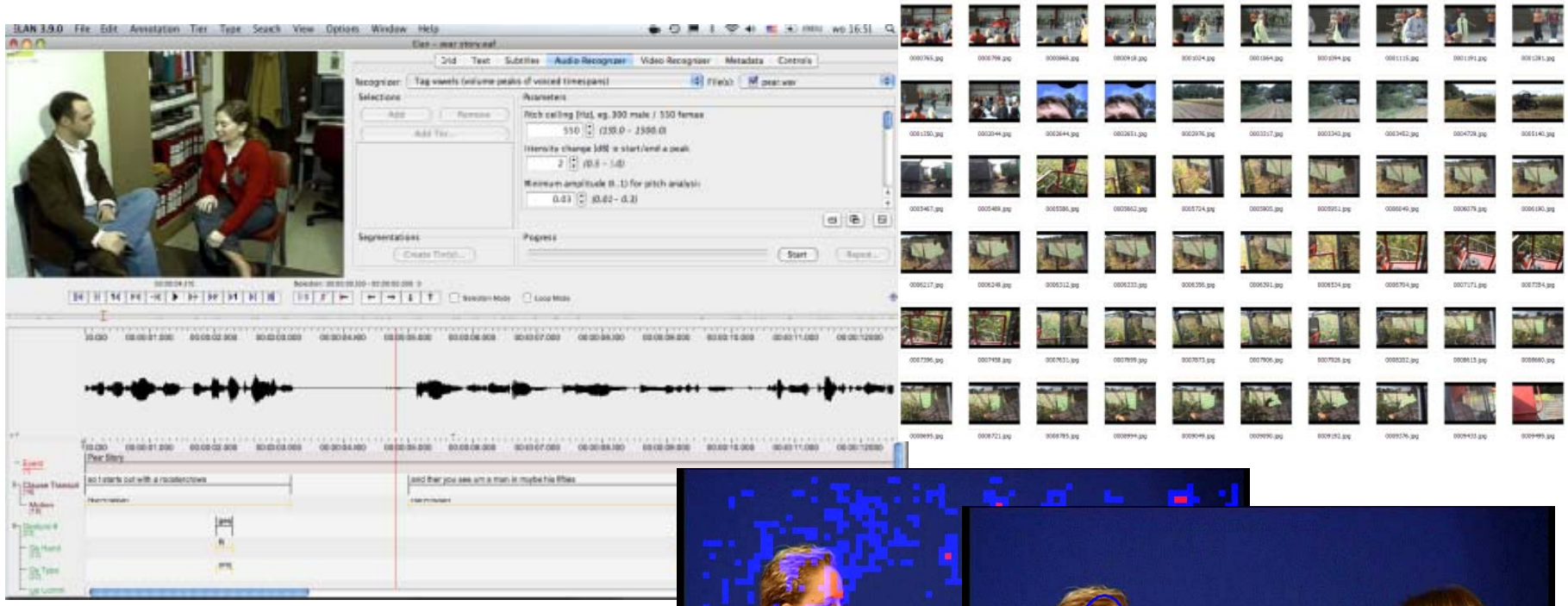
1. What is e-Science/e-Research?
   a) Data
   b) Operations
2. What do other communities do?
3. What does CLARIN do?
   a) Federation
   b) PIDs
   c) Metadata
   d) Web - Services
4. General Issues

Web-accessible Service Landscape



- like to speak about a web of services that can be started by smart web applications (these are not web-sites, but think of AMAZON etc)
- user is confronted with so-called Virtual Research Environments giving access to a bunch of related services dedicated to a certain purpose

Think of AV-Detectors



- library of AV detector services to do automatic cumulative annotation
- goal must be that you can start such services with your clip

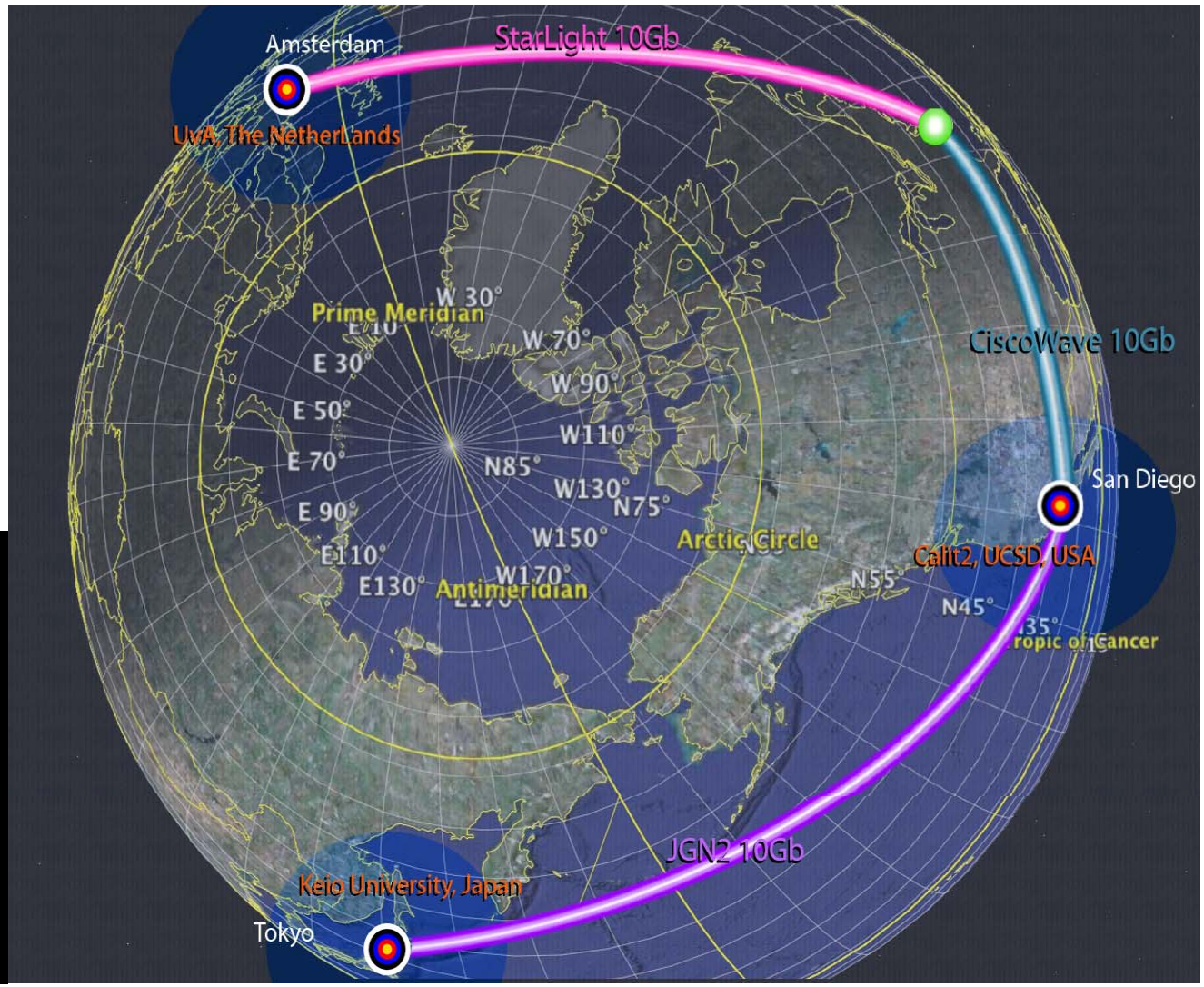
Content

1. What is e-Science/e-Research?
   a) Data
   b) Operations
2. What do other communities do?
3. What does CLARIN do?
   a) Federation
   b) PIDs
   c) Metadata
   d) Web - Services
4. General Issues

CineGrid Exchange 2008

Geographically Distributed Repositories + Fast Networks
Collaborative Film Editing

collaborative film editing with worldwide specialization based on High-quality digital media assets: 4K, 2K, HD, mono & stereo, still & motion pictures + audio



- San Diego @ UCSD/Calit2
  40 TB with 10 Gbps connectivity
- Amsterdam @ UvA
  30 TB with 10 Gbps connectivity
- Tokyo @ Keio/DMC
  6 TB with 10 Gbps connectivity

Total storage = 76 TB

CineGrid Exchange 2010

- Build multi-layer open-source framework for distributed digital media repository
- Refine middleware "rules" for robustness
- offer Collective Access

CX Node Site	Storage Type	CX Allocation
UCSD/Calit2, San Diego, USA	Sun Thumper (x4540)	66 TB
UvA, Amsterdam, Netherlands	Sun Thumper (x4540)	30 TB
UIC/EVL, Chicago, USA	RAID Array	10 TB
Keio U./DMC, Tokyo, Japan	RAID Array	8 TB
CESNET, Prague, Czech Republic	Sun Thumper (x4540)	48 TB
Ryerson U, Toronto, Canada	Sun Thumper (x4540)	57 TB
AMPAS, Los Angeles, USA	Sun Thumper (x4540)	24 TB
Total CineGrid Exchange Capacity		**243 TB**

Tier Structure in High Energy Physics (CERN)

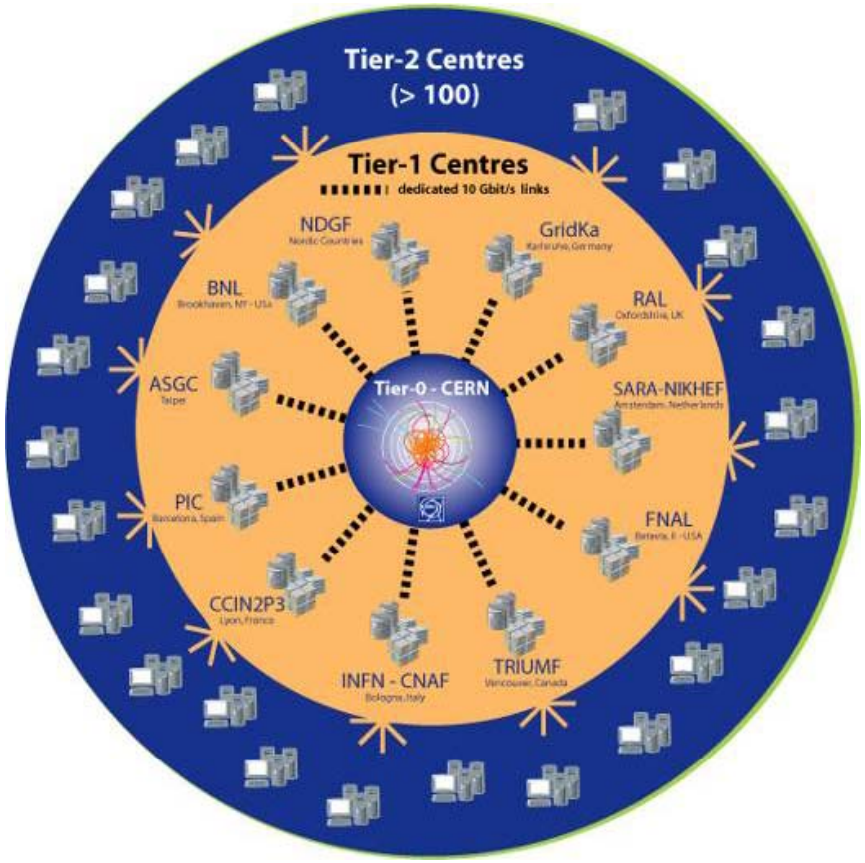
Today there are about 320 sites world-wide participating in the WLCG project.

- Tier0 center, CERN
  very large center, tape storage, first level processing
  and meta-data storage, quality service (24*7)

- Tier1 center, **11** world-wide
  large capacity, tape storage, quality service

- Tier2 center, **168** world-wide
  medium size, some large, no 24*7 service,
  no custodial storage

- Tier3 center, **140** world-wide
  very small to medium size,
  no availability guarantee,
  focus on end-user analysis activity

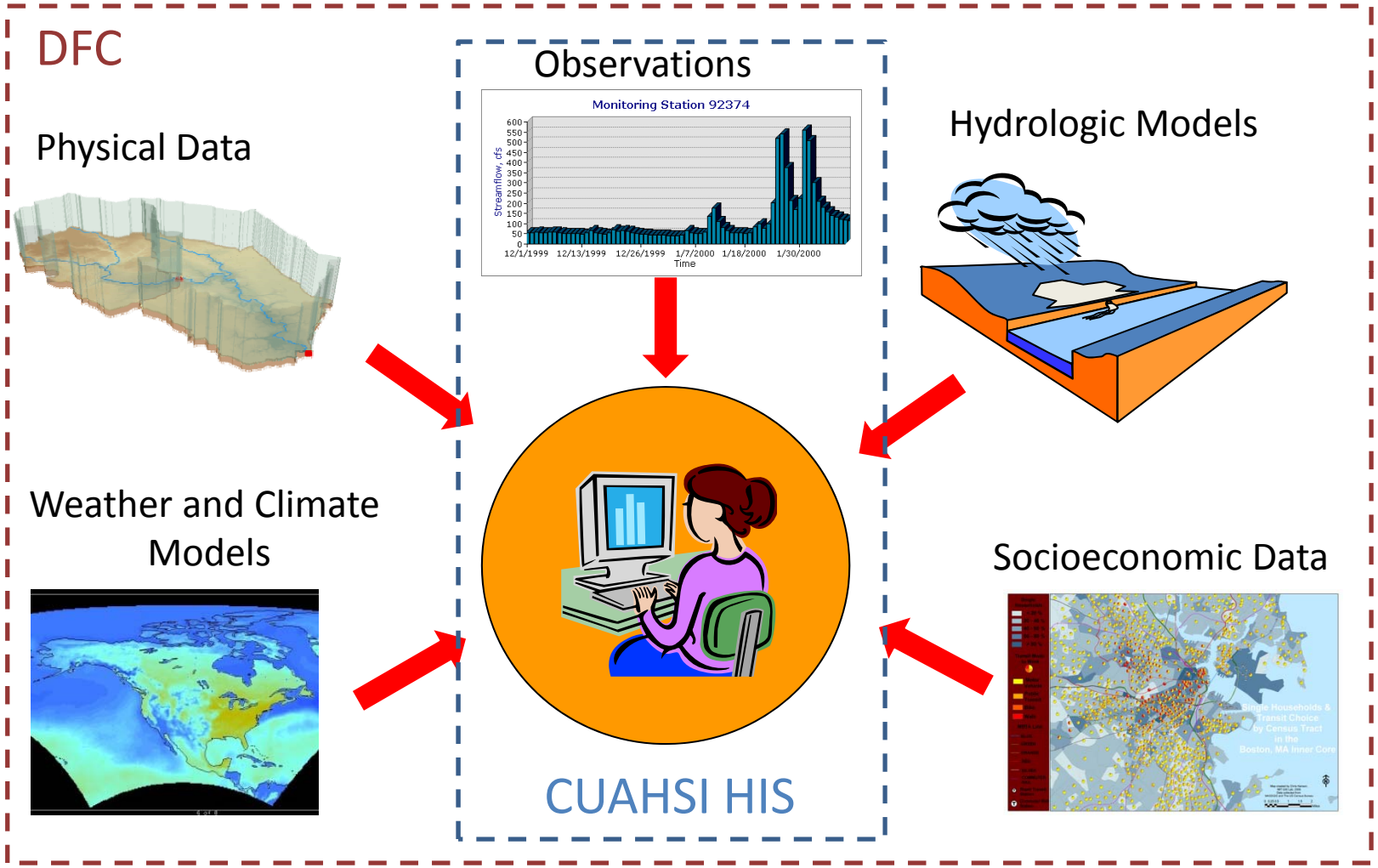


Resource installation today:
~ 110 PBytes disk space
~ 100 Pbytes tape space
~ 150000 processor cores

Data flow between Tiers:
multi GBytes/s multi PBytes/month

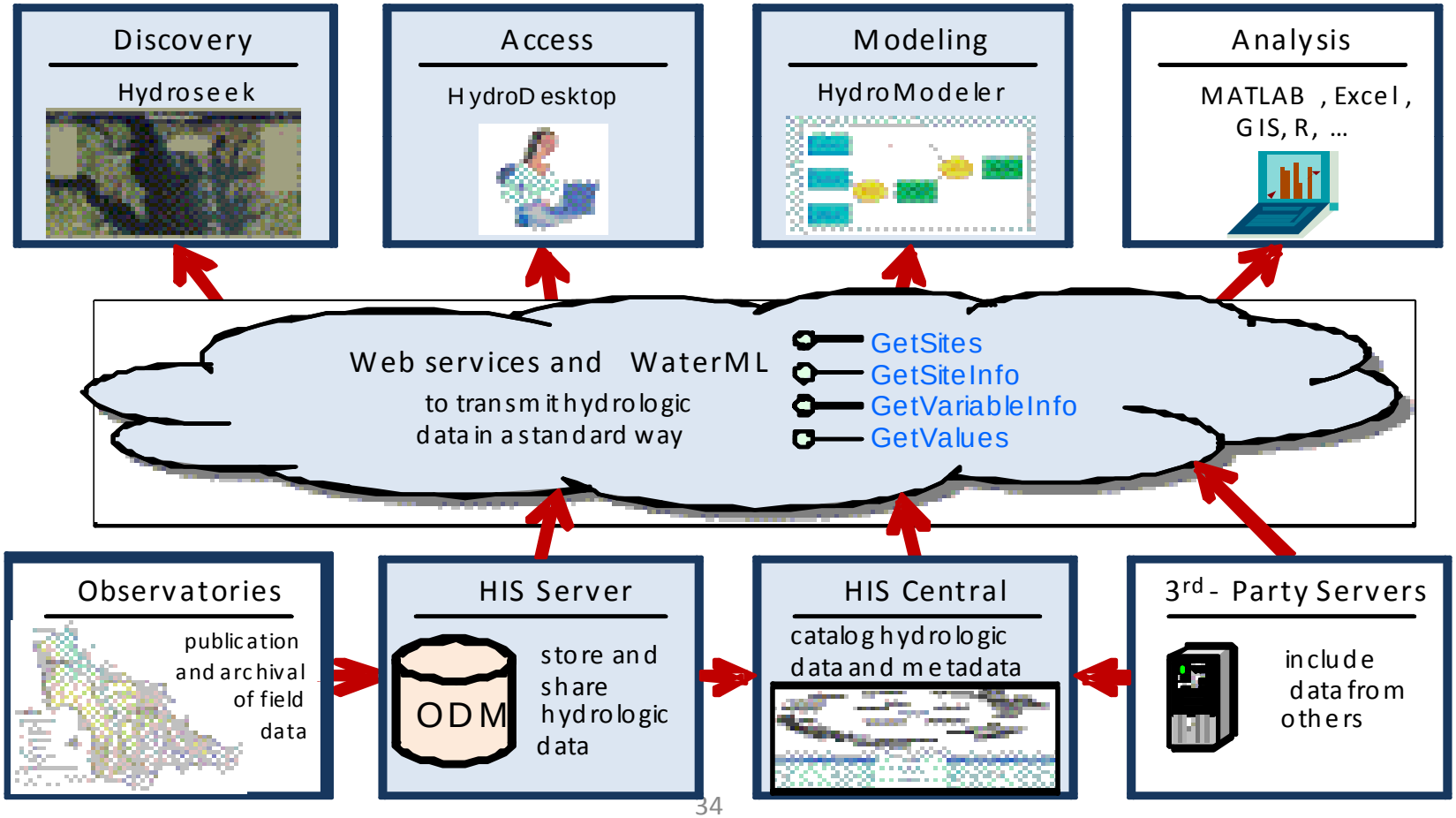
Current data sets : ~ 50 PB

Data and Model Integration Needed to Support Hydrologic Science



Hydrology Community

High Level View of HIS Service Oriented Architecture



Content

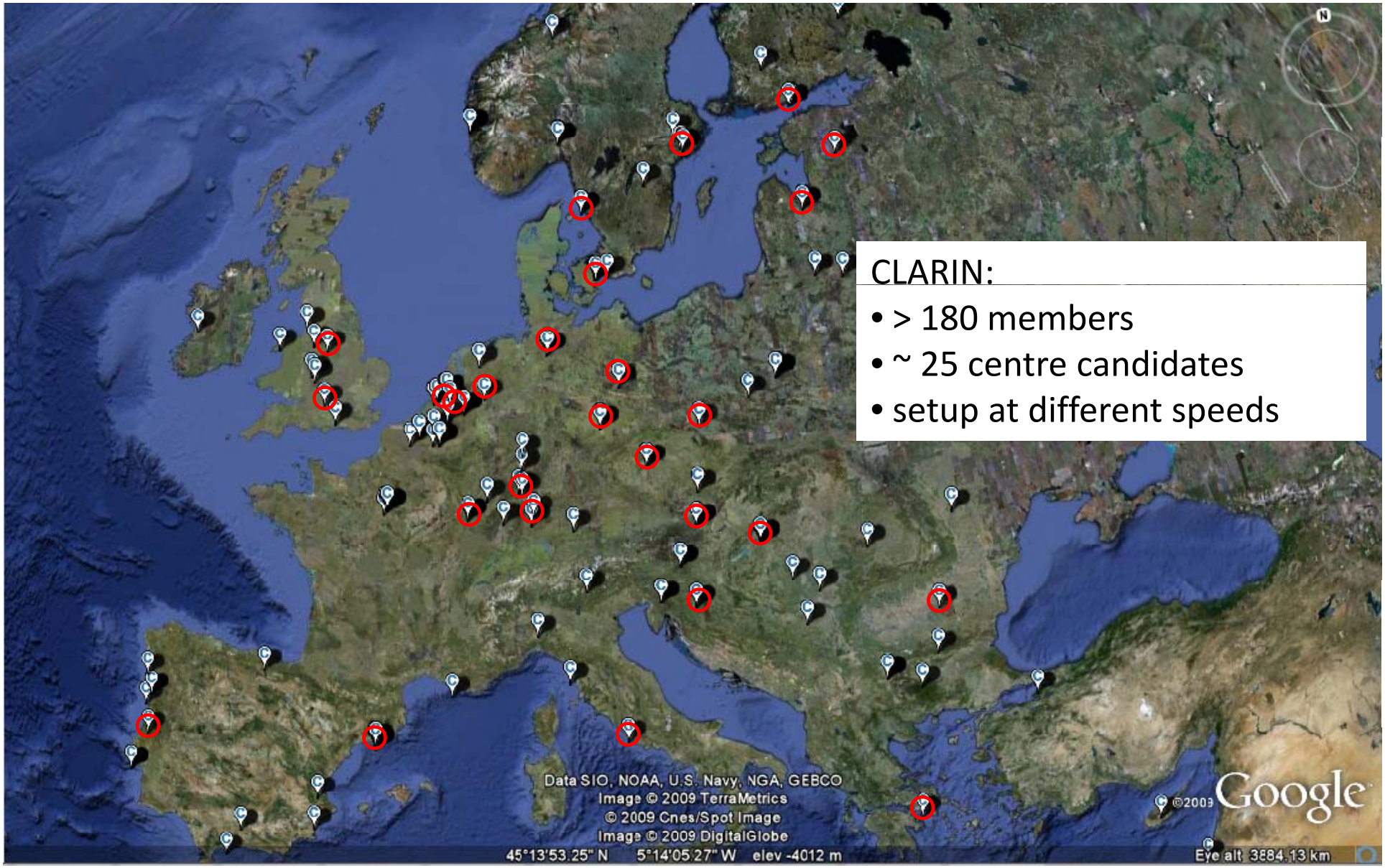
1. What is e-Science/e-Research?
   a) Data
   b) Operations
2. What do other communities do?
3. What does CLARIN do?
   a) Federation
   b) PIDs
   c) Metadata
   d) Web - Services
4. General Issues

CLARIN Goals

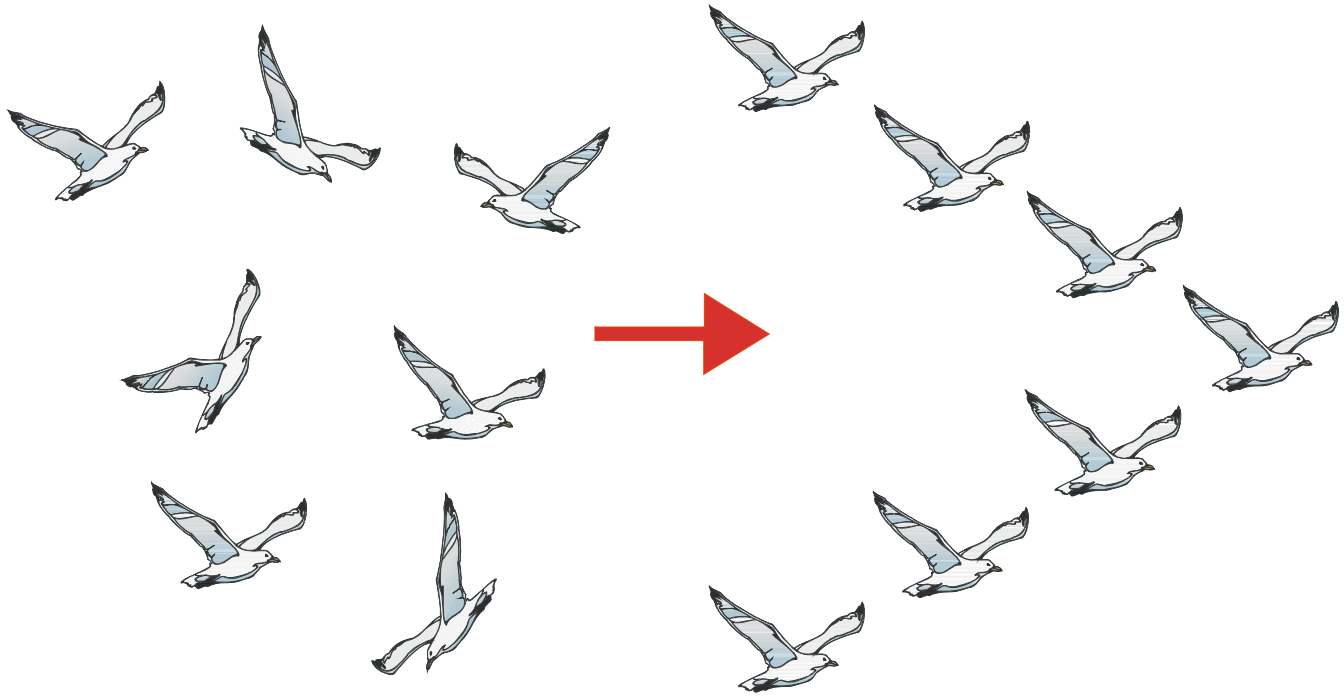
- Build an integrated and interoperable domain of Language Resources and Tools/Services
- Build a persistent infrastructure that allows researchers to use all essential components
- Integration means "get things organisationally together"
- Interoperability means "let components talk with each other"
- of course: interoperability is not easy to achieve
- both will require a change of culture

What is the state of CLARIN?

The Language Archive



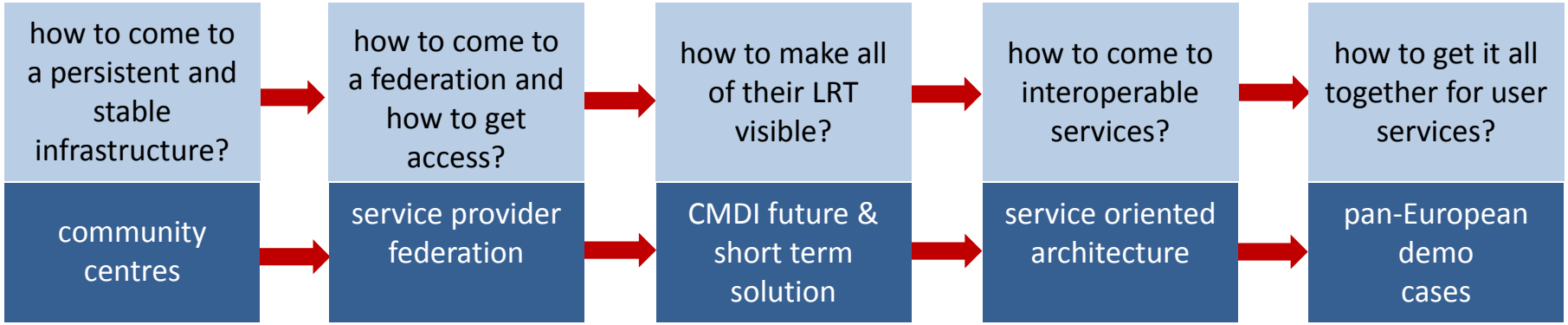
Problem to solve in CLARIN



- how to synchronize all minds in our field?
- how much synchronization is good for our field?
- how to synchronize data and tool creation?

CLARIN work dimensions

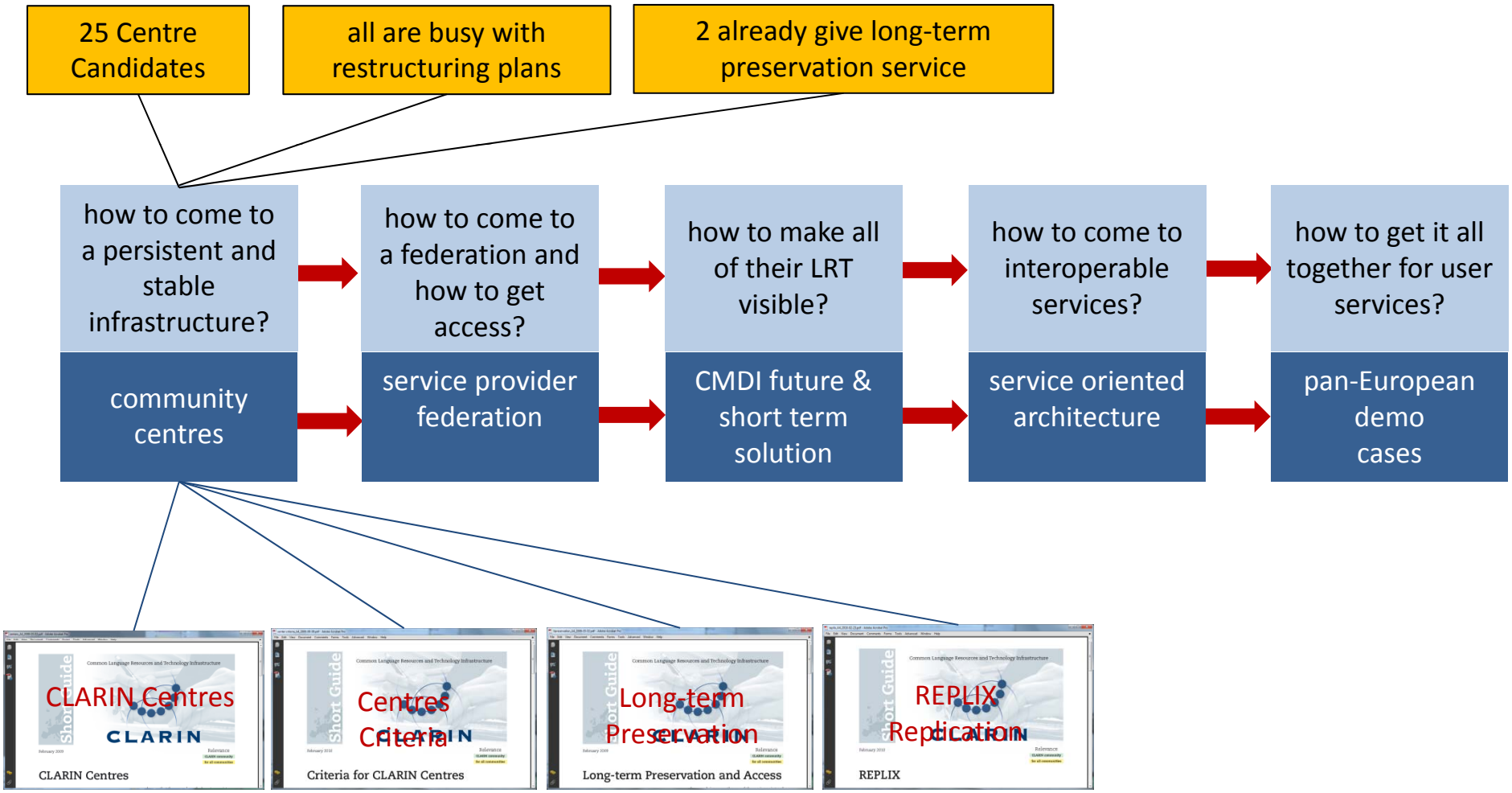
... at least from major activities in core work packages



there is more in CLARIN, but not shown here:

- link to humanities world
- education, help/support/advice, dissemination
- license and Code of Conduct harmonization
- the ERIC legal framework

Community Centres



Service Provider Federation

The Language Archive

WP2: technology

WP7: agreements, licenses,etc

setup federation technology

build initial federation

central user attribute server

setup EPIC service

how to come to a persistent and stable infrastructure?

how to come to a federation and how to get access?

how to make all of their LRT visible?

how to come to interoperable services?

how to get it all together for user services?

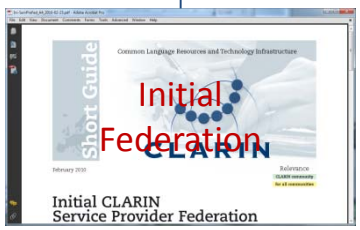
community centres

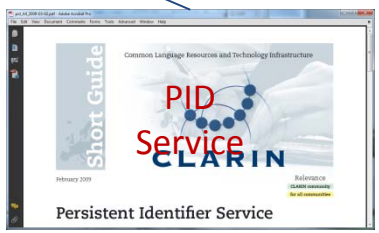
service provider federation

CMDI future & short term solution

service oriented architecture

pan-European demo cases





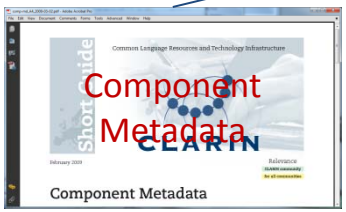
Metadata Domain

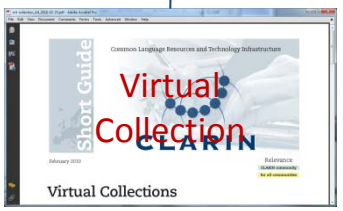
The Language Archive

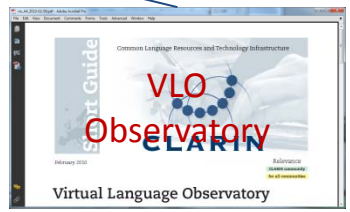
WP2: technology, standards | WP5: LRT inventory, LRT communities

CMDI Infra | ARBIL MD Editor | Category Definition | ISOcat development | LRT Inventory | setup OAI PMH machinery | Virtual Language World

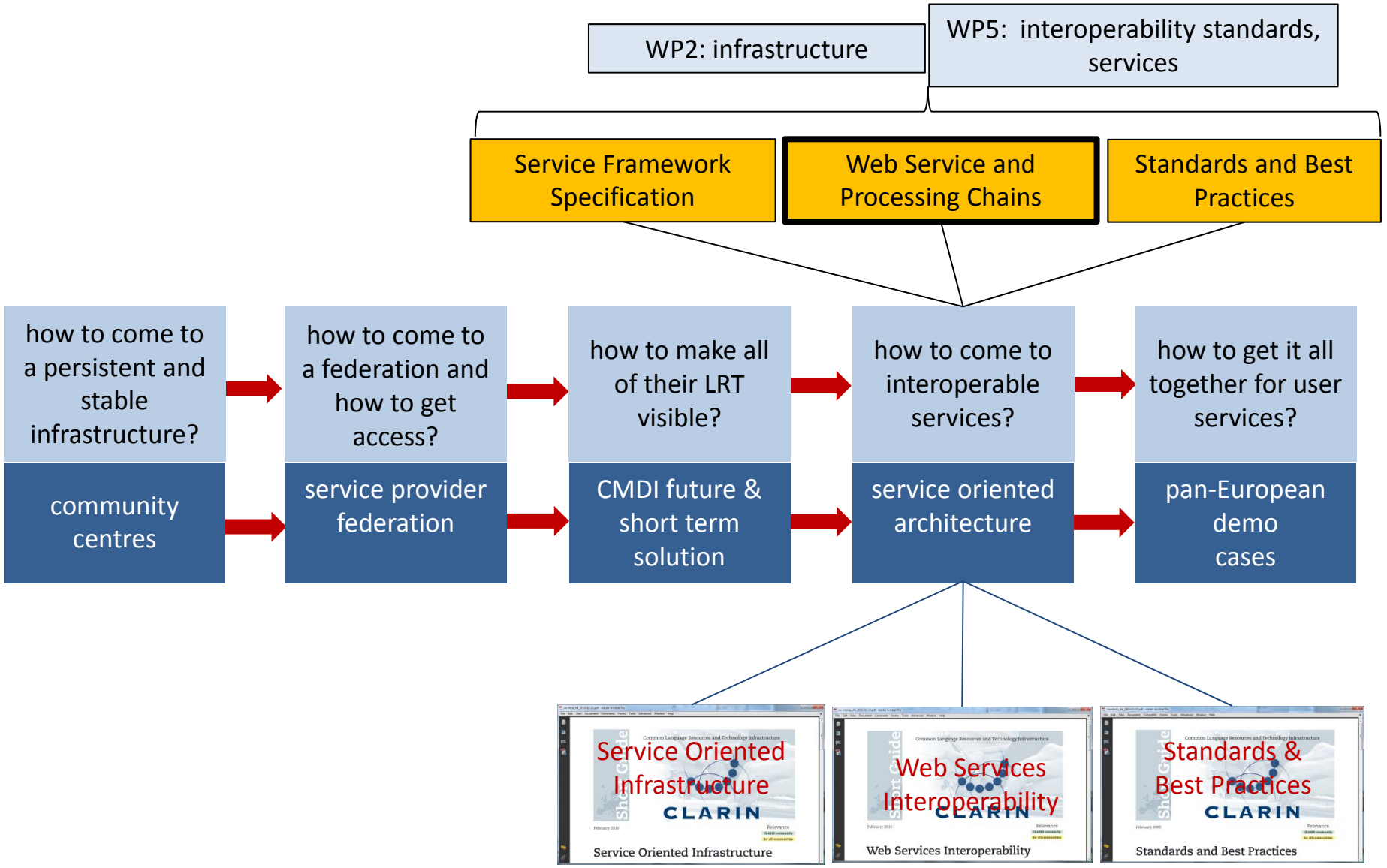
how to come to a persistent and stable infrastructure?	how to come to a federation and how to get access?	how to make all of their LRT visible?	how to come to interoperable services?	how to get it all together for user services?
community centres	service provider federation	CMDI future & short term solution	service oriented architecture	pan-European demo cases



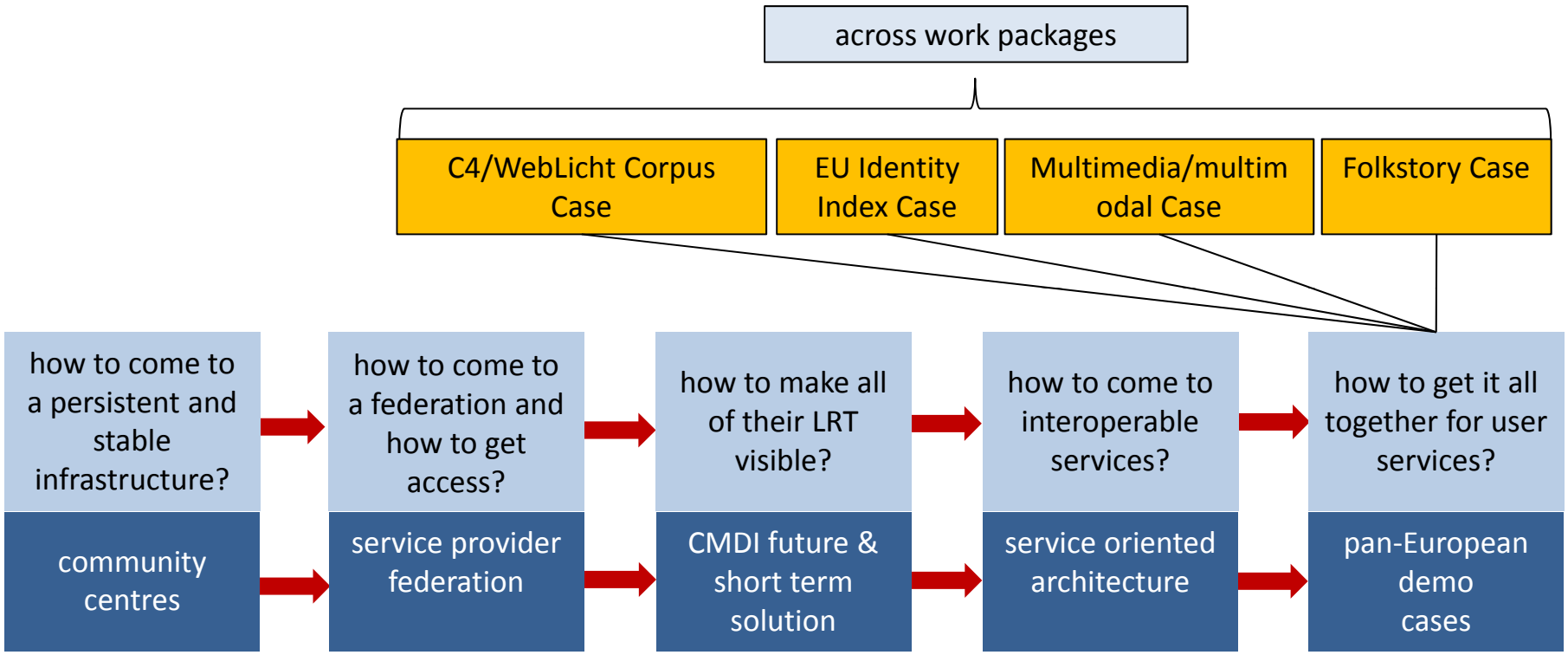




Service Oriented Architecture



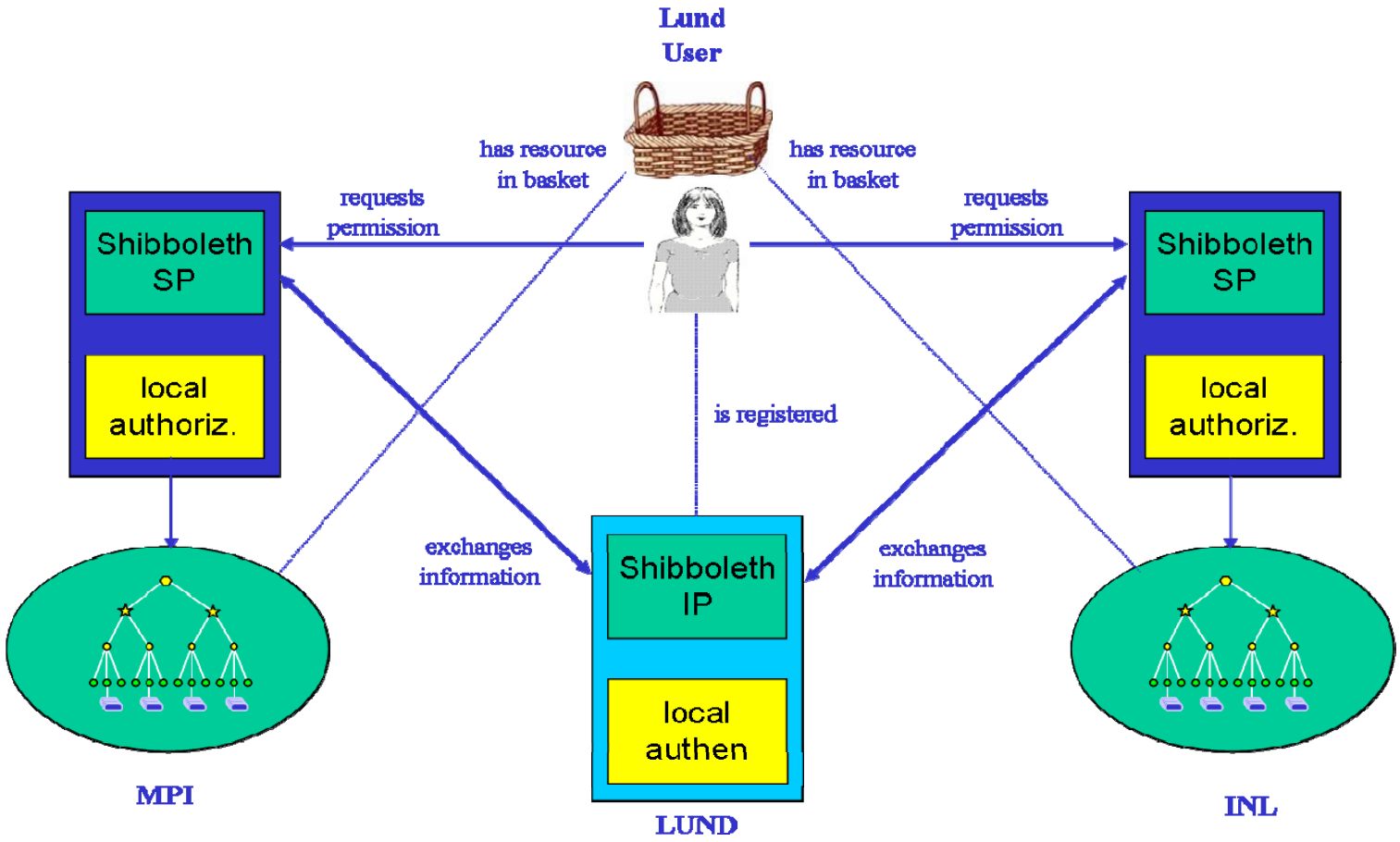
Demo Cases (just started)



Content

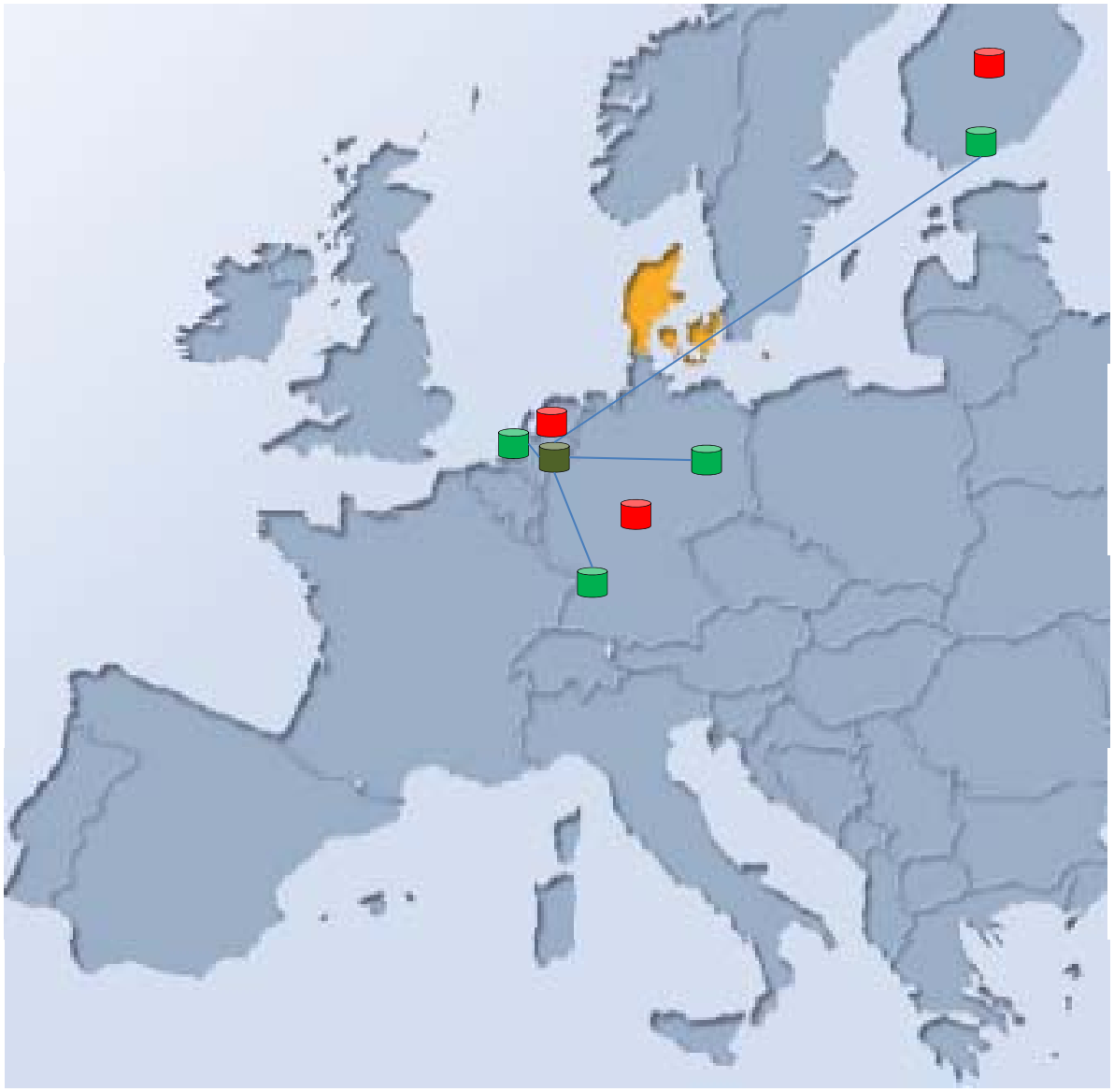
1. What is e-Science/e-Research?
   a) Data
   b) Operations
2. What do other communities do?
3. What does CLARIN do?
   a) Federation
   b) PIDs
   c) Metadata
   d) Web - Services
4. General Issues

Federation goal



- core idea: let the user operate only with his home ID
- example: build and access a virtual collection

Federation



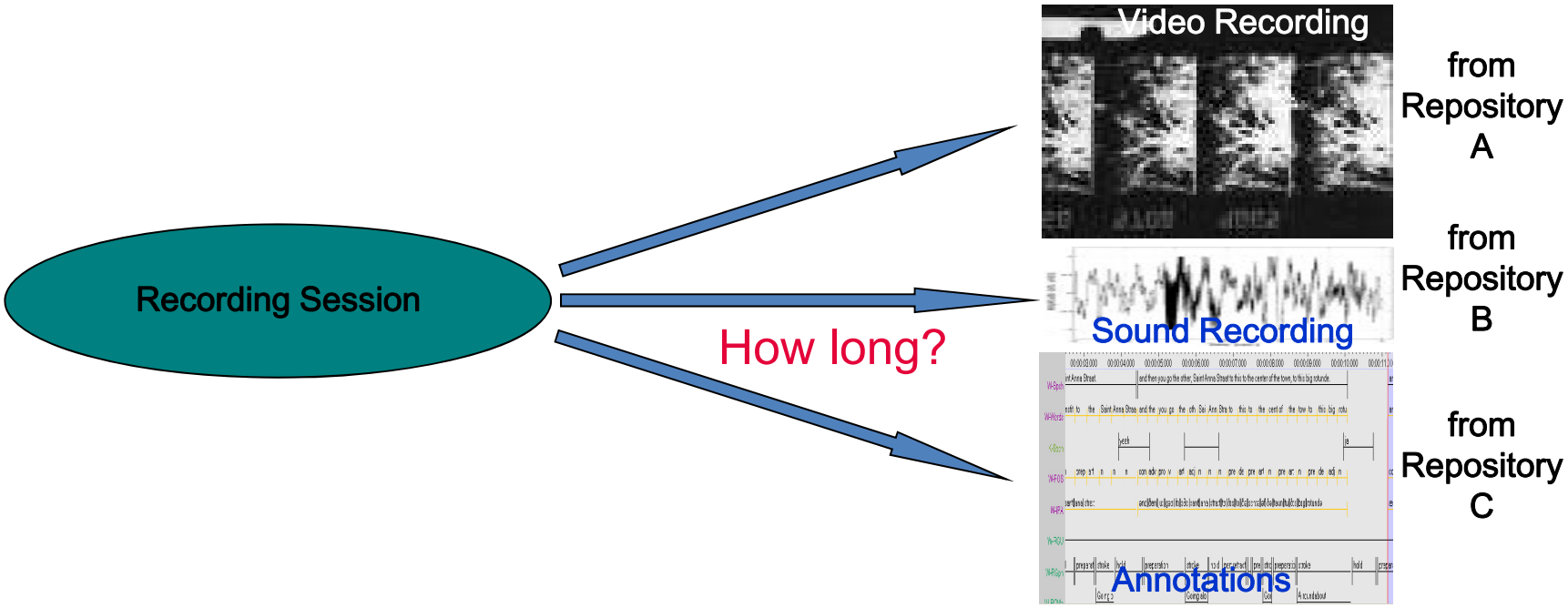
- Service Provider Federation
  - Agreement 1
  - n centers members
- Link up with national IdFs
  - Agreement 2
  - DFN De
  - HAKA Fi
  - SURFnet NI
- 1 Mio pot. Users-id
- currently more countries and centers

Content

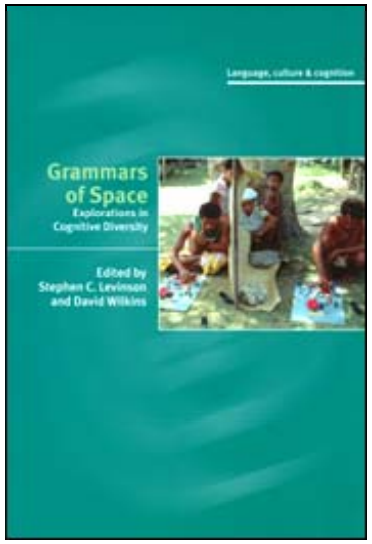
1. What is e-Science/e-Research?
   a) Data
   b) Operations
2. What do other communities do?
3. What does CLARIN do?
   a) Federation
   b) PIDs
   c) Metadata
   d) Web - Services
4. General Issues

What is it: a simple example

- assume that we have a recording of an extinct language and some annotations that tell us what someone said about medicine etc
- researchers create relations that need to be preserved

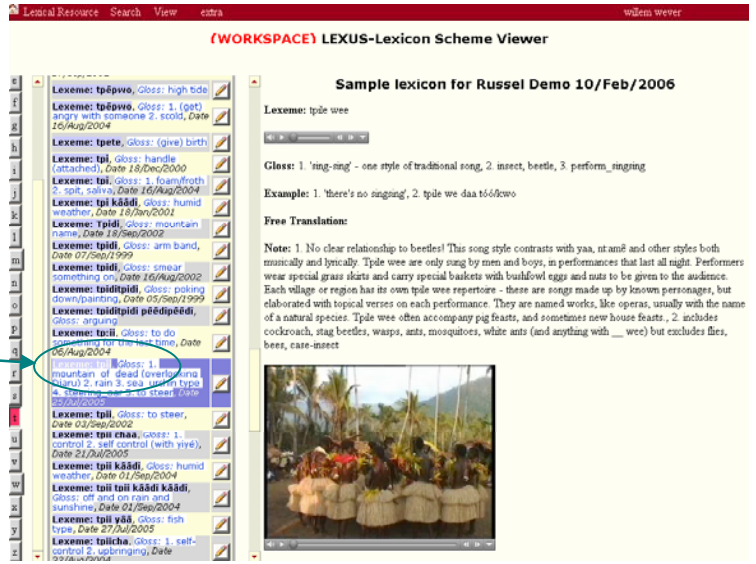


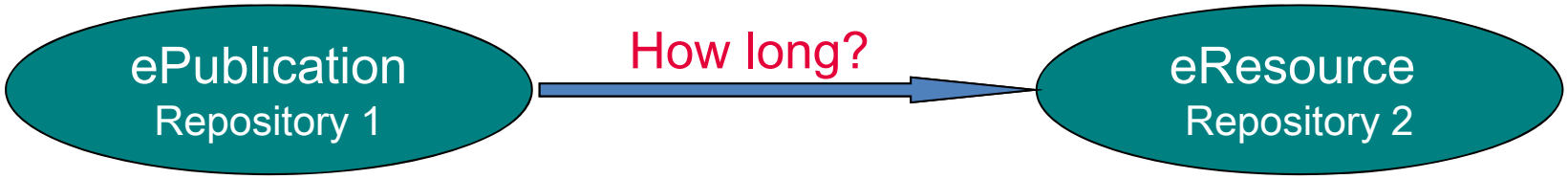
another simple example



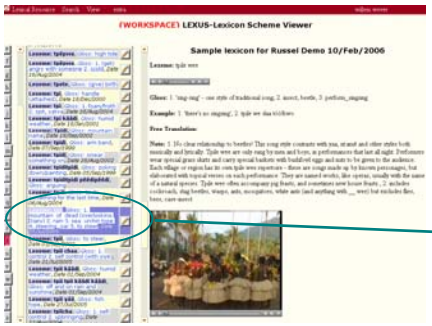
Biological and cultural processes have evolved together, in a symbiotic spiral; they are now indissolubly linked, with human survival unlikely without such culturally produced aids as clothing, cooked food, and tools. The twelve original essays collected in this volume take an evolutionary perspective on human culture, examining the emergence of culture in evolution and the underlying role of brain and cognition. The essay authors, all internationally prominent researchers in their fields, draw on the cognitive sciences -- including linguistics, developmental psychology, and cognition -- to develop conceptual and methodological tools for understanding the interaction of culture and genome. They go beyond the "how" -- the questions of behavioral mechanisms -- to address the "why" -- the evolutionary origin of our psychological functioning. What was the "X-factor," the magic ingredient of culture -- the element that took humans out of the general run of mammals and other highly social organisms?

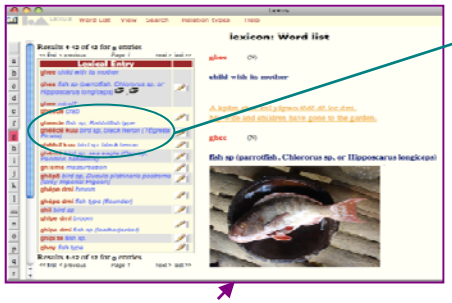
Several essays identify specific behavioral and functional factors that could account for human culture, including the capacity for "mind reading" that underlies social and cultural learning and the nature of morality and inhibitions, while others emphasize multiple partially independent factors -- planning, technology, learning, and language. The X-factor, these essays suggest, is a set of cognitive adaptations for culture.

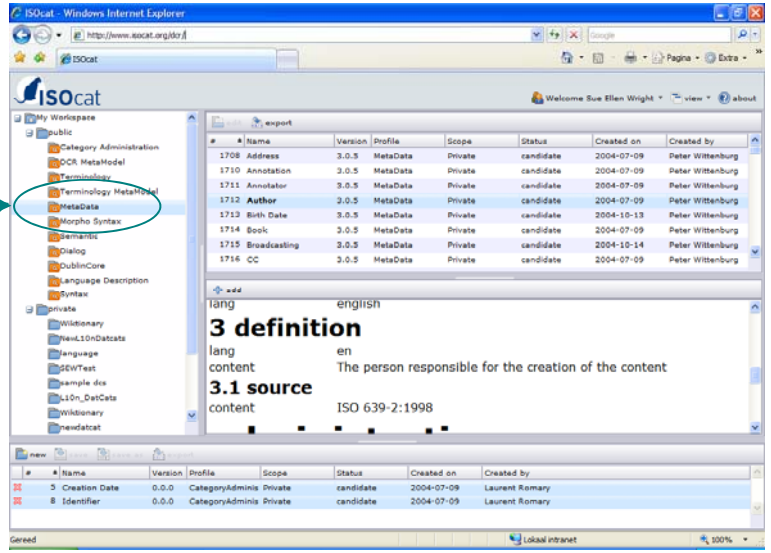




still another simple example







eResource1
Repository 1

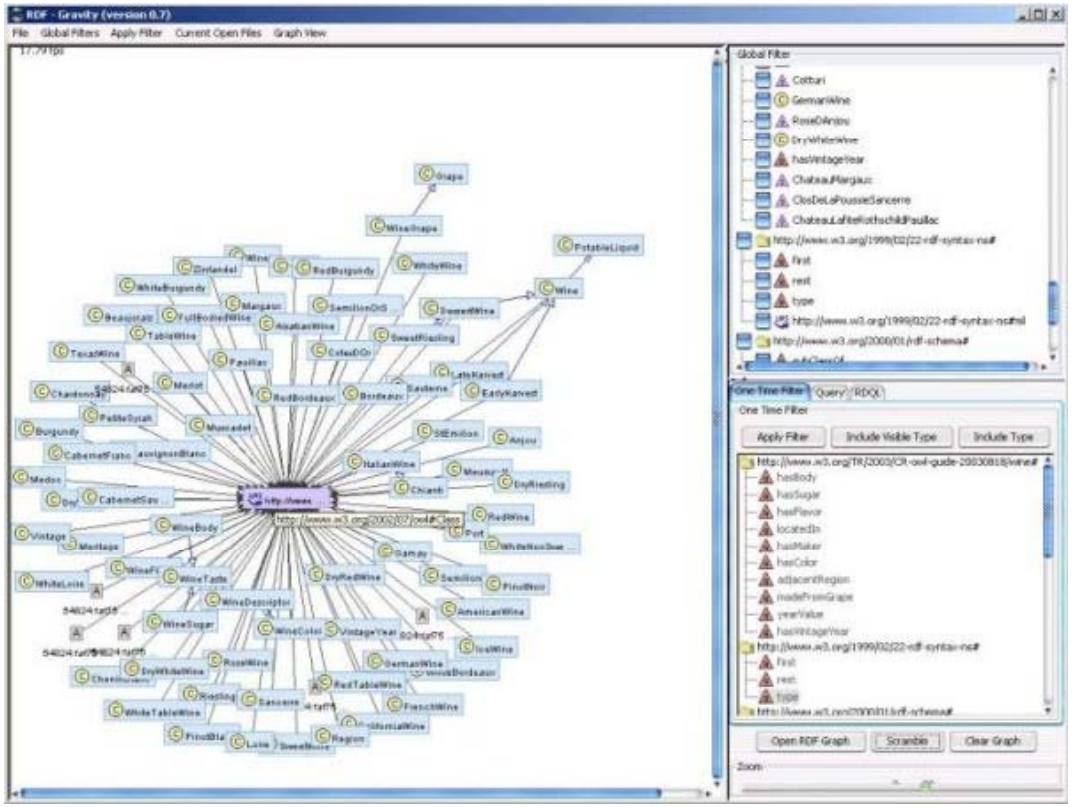
eResource2
Repository 2

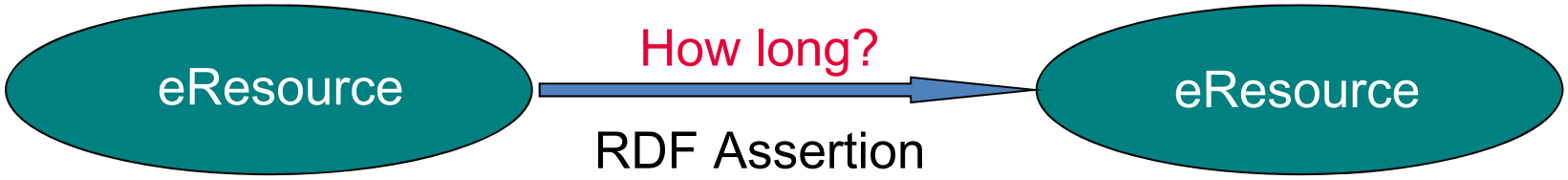
How long?

Ontology
open registry

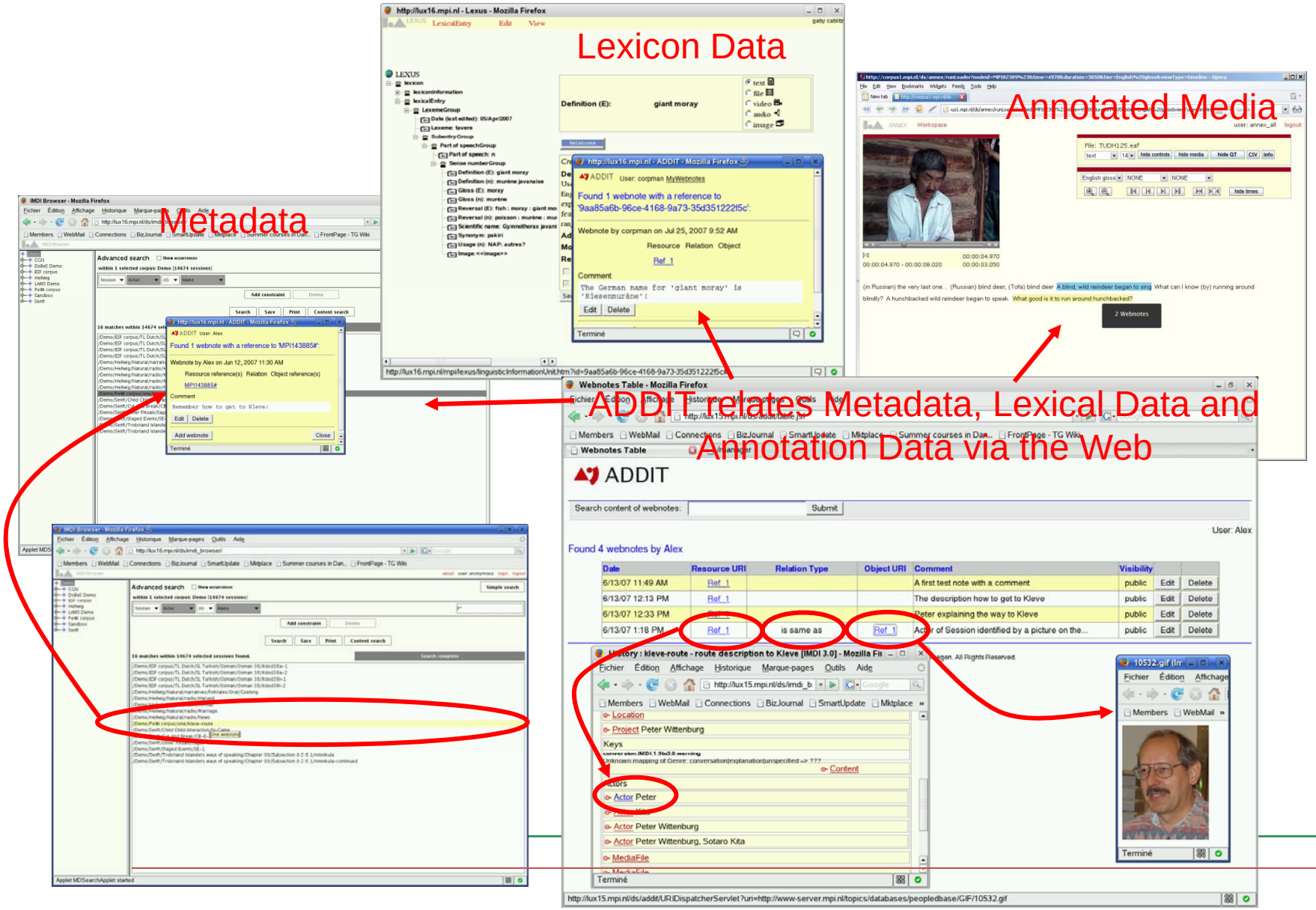
still another simple example

The Language Archive



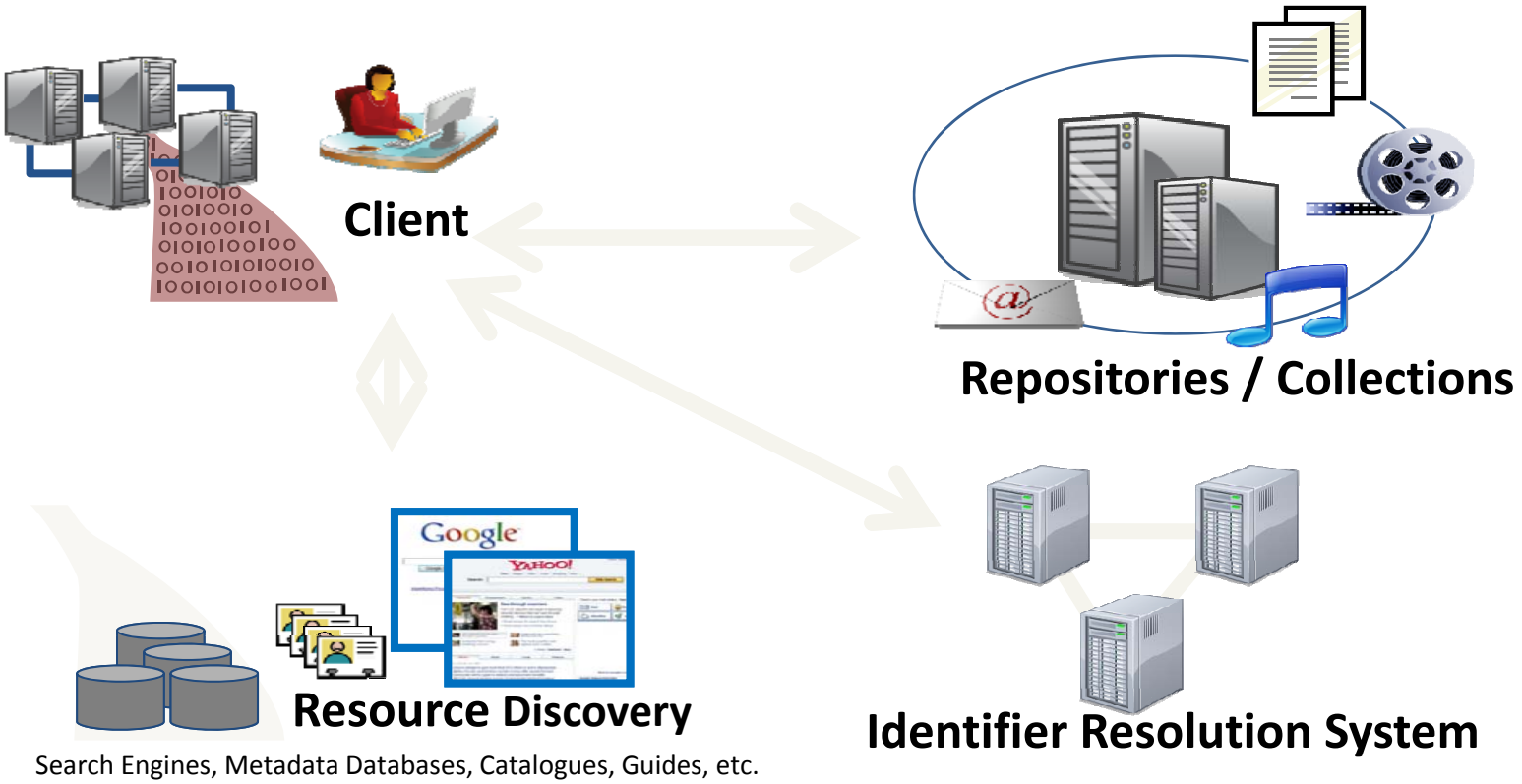


And a last example



is there a problem?

- could use Cool URIs as the W3C TAG suggests
- URLs change too often and we cannot influence that
- perhaps some exceptions such as
  http://www.isocat.org/datcat/DC-1708



PID Information

- tradition in SSH is to include samples in publications as proof and claim that you have the data
- eResearch is different: you need to make your data available and identifiable with the help of a PID
- thus:
  - make it explicit by depositing in a trusted repository
  - will register a unique and persistent identifier (PID)
  - PID will be associated with
    - checksum to proof authenticity
    - time stamp
    - pointers to metadata record
    - pointers to copies
- thus data is citable and identifiable

Currently almost 1 Mio PIDs

LAT

The Language Archive

PIDs are mentioned in the metadata descriptions at MPI

```
<?xml version="1.0" encoding="UTF-8"?>
<METATRANSCRIPT ArchiveHandle="hdl:1839/00-0000-0000-0005-82B0-2"
    Date="2006-07-18" FormatId="IMDI 3.0"
    Originator="Editor - Profile:SESSION.Profile.xml" Type="SESSION" Version="1"
     xmlns="http://www.mpi.nl/IMDI/Schema/IMDI"
     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
     xsi:schemaLocation="http://www.mpi.nl/IMDI/Schema/IMDI ./IMDI_3.0.xsd">
 <Session>
       <Name>DBD_RIF_14_12_01_064</Name>
       <Title>Dutch Bilingualism Database, Ethnic Dutch, Session 64</Title>
      …….

<MediaFile>
<ResourceLink ArchiveHandle="hdl:1839/00-0000-0000-0004-DC6B-0">
     http://corpus1.mpi.nl/qfs1/media-archive/dbd_data/boumans/T-
     Cult/Metadata/../Media/dbd_rif_14_12_01_064.wav</ResourceLink>
     …….
```

PID Services

- EPIC (GWDG/CSC/SARA) is ready and can be used
  - based on **Handles**
  - designed for millions of PIDs (MPI has about 1 million)
  - there is an API for online registration
  - business model defined by researchers
- at least one other well-known Service
  - DOI/IDF based on **Handles**
  - designed for publications
  - business model defined by big publishers

http://www.pidconsortium.eu

Content

1. What is e-Science/e-Research?
   a) Data
   b) Operations
2. What do other communities do?
3. What does CLARIN do?
   a) Federation
   b) PIDs
   c) Metadata
   d) Web - Services
4. General Issues

Metadaten Experience

- in linguistic domain about 15 years of experience
  - IMDI - first structured "set" defined by linguists to support research
  - OLAC - extended DublinCore set to allow general search
  - many other sets in almost all research disciplines

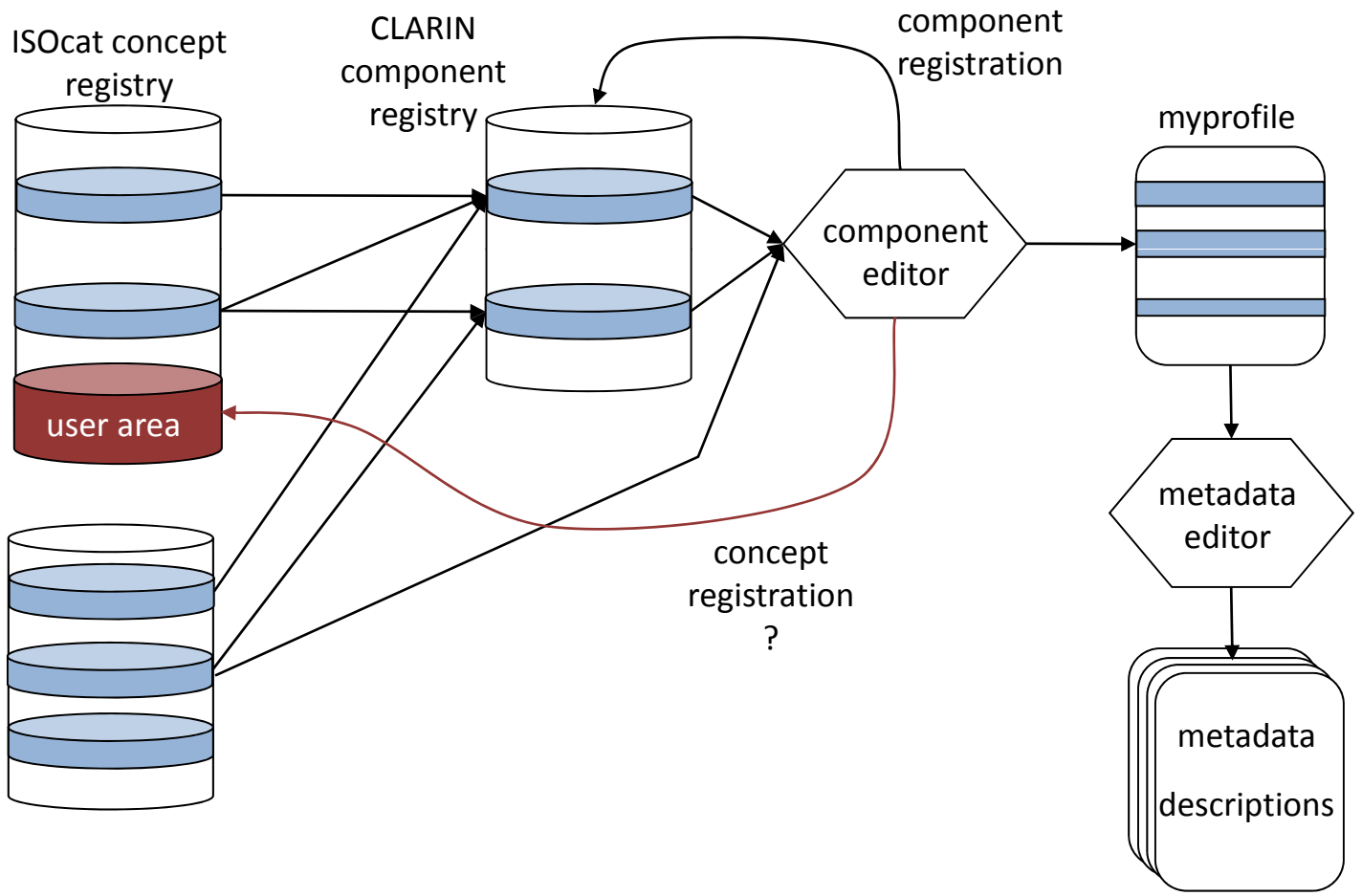
- metadata is indeed the glue that keeps all data organized
- IMDI not only used for browsing & searching, but also for management

- widely accepted now:
  - metadata needs to describe the data world of a discipline
  - there is no one structure since data world is very rich & complex
  - you need to be able to describe aggregations as well
  - metadata is increasingly important for machine processing
    machine processing requires HQ metadata

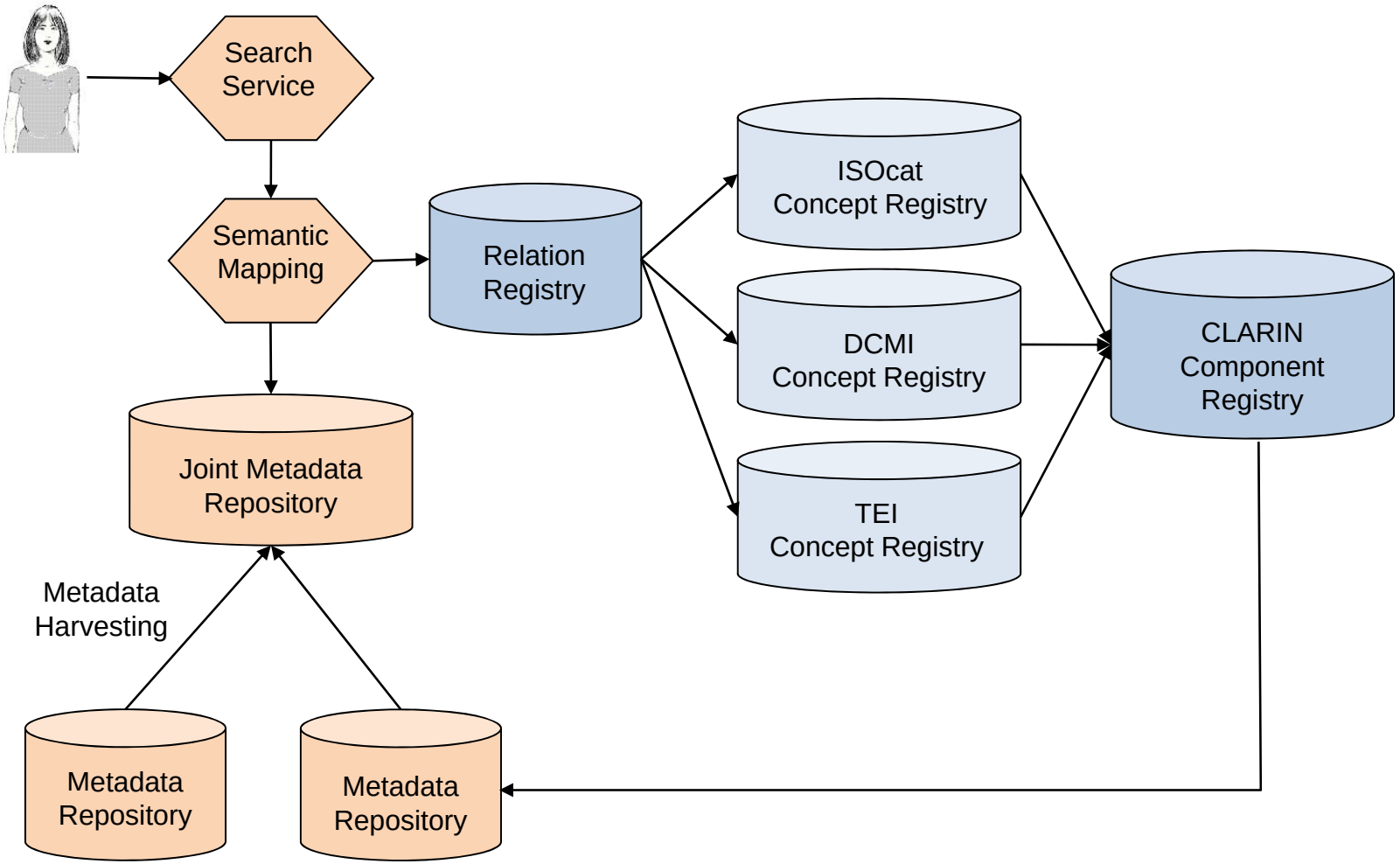
Metadaten Conclusions

- a proper metadata framework needs
  - to be modular -> components based on XML - syntactic interop
  - to build on registered element definitions - semantic interop
  - to make use of PIDs as references
- thus
  - for every particular resource type a profile
  - for every sub-discipline usage a profile
  - for tools/services tailored profiles using the same semantics
- but
  - take care to not get an indefinite proliferation
  - element semantics should not be context dependent

Component Metadata Infrastructure

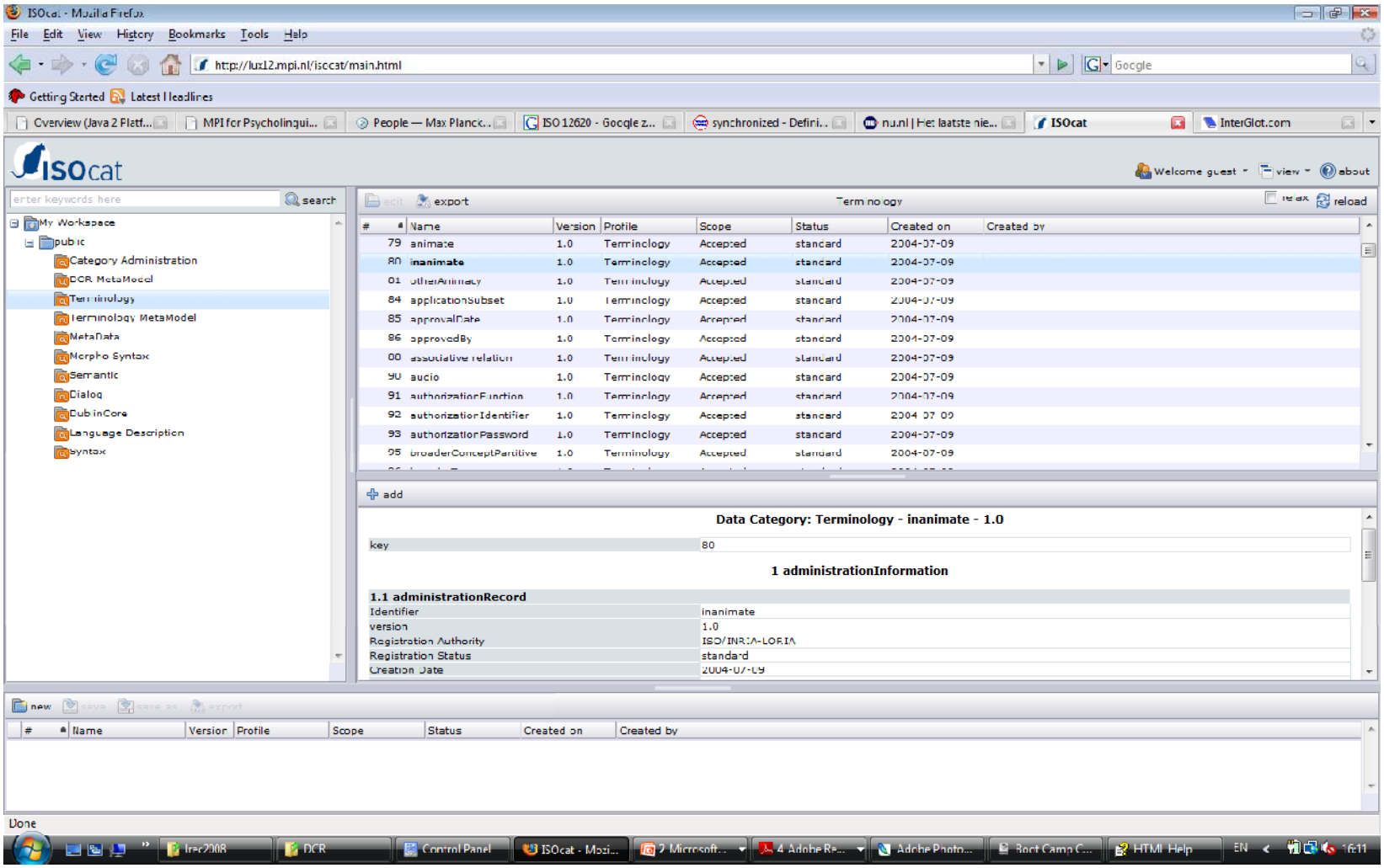


CMDI execution infrastructure



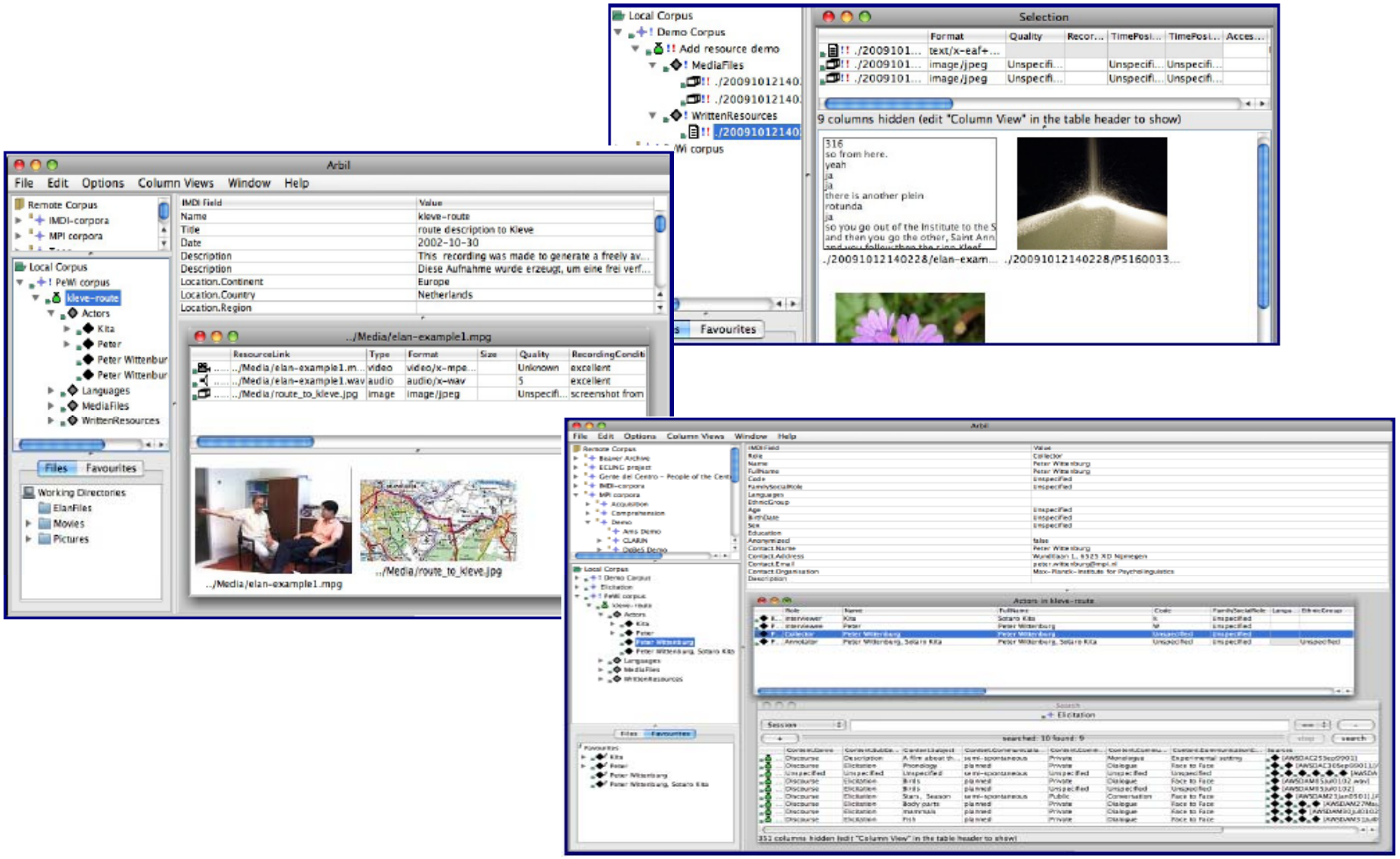
ISO 12620 & ISOcat

The Language Archive

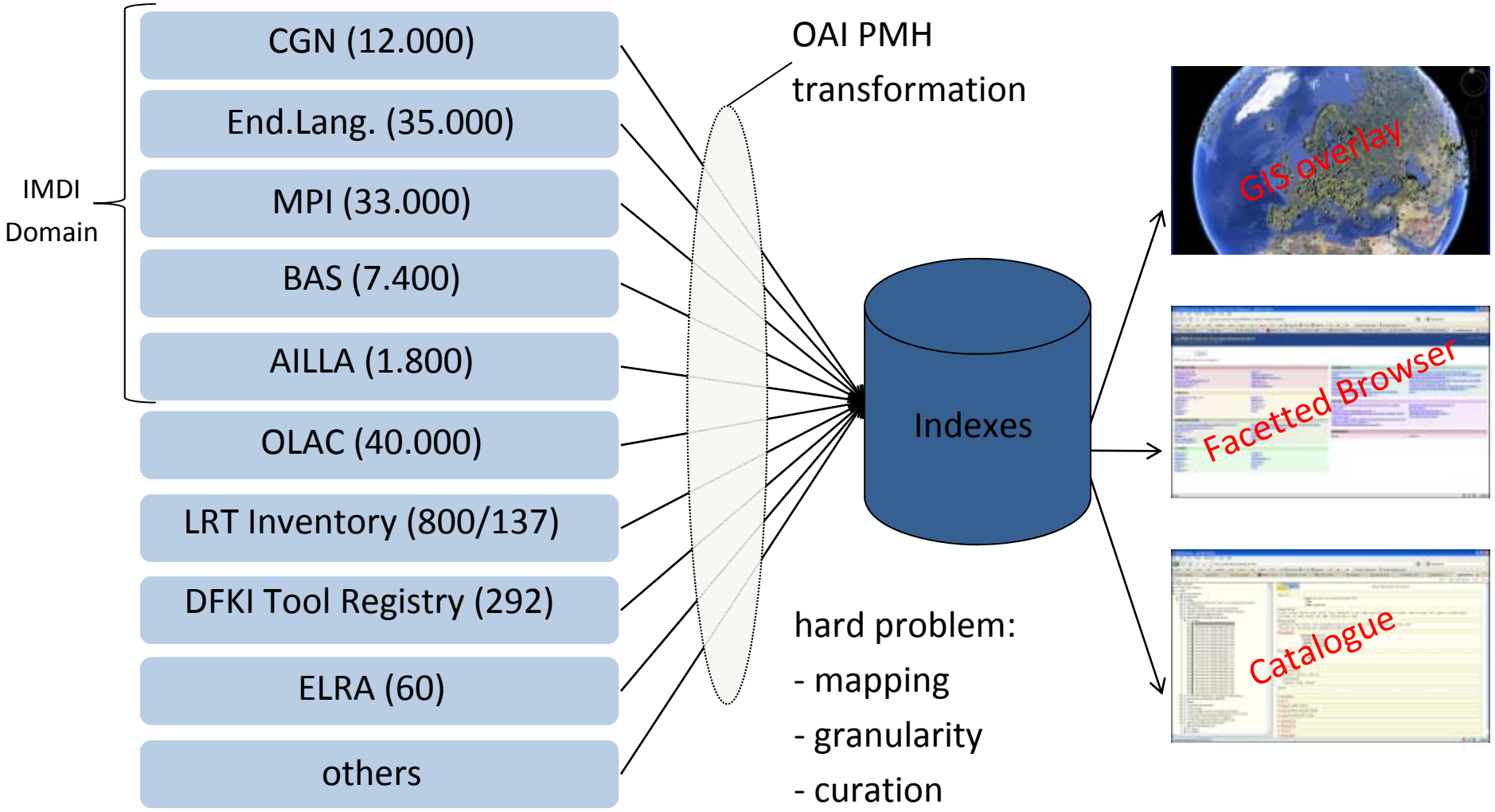


New Arbil Metadata tool

The Language Archive

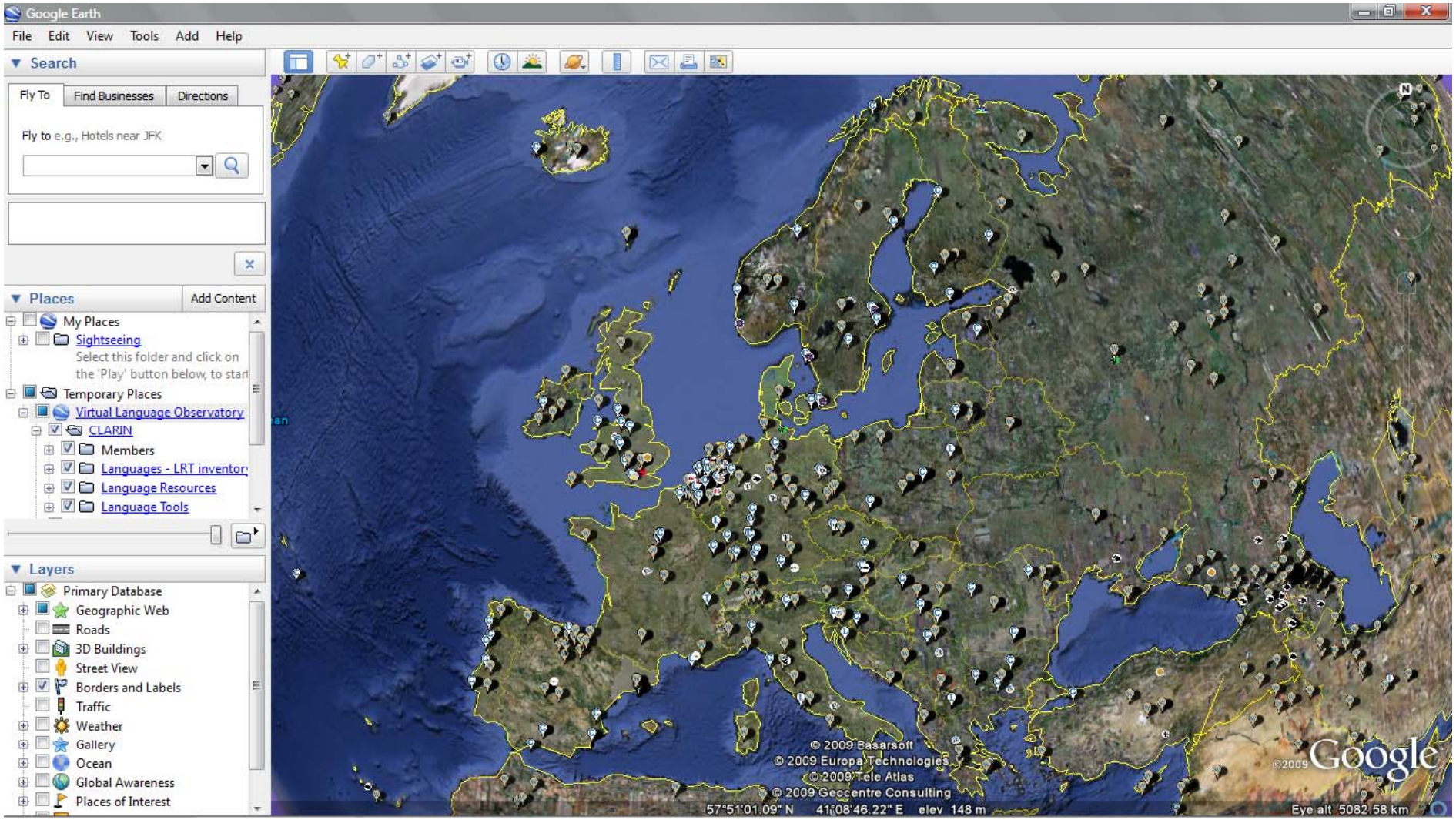


Virtual Language Observatory



Virtual Language Observatory

The Language Archive

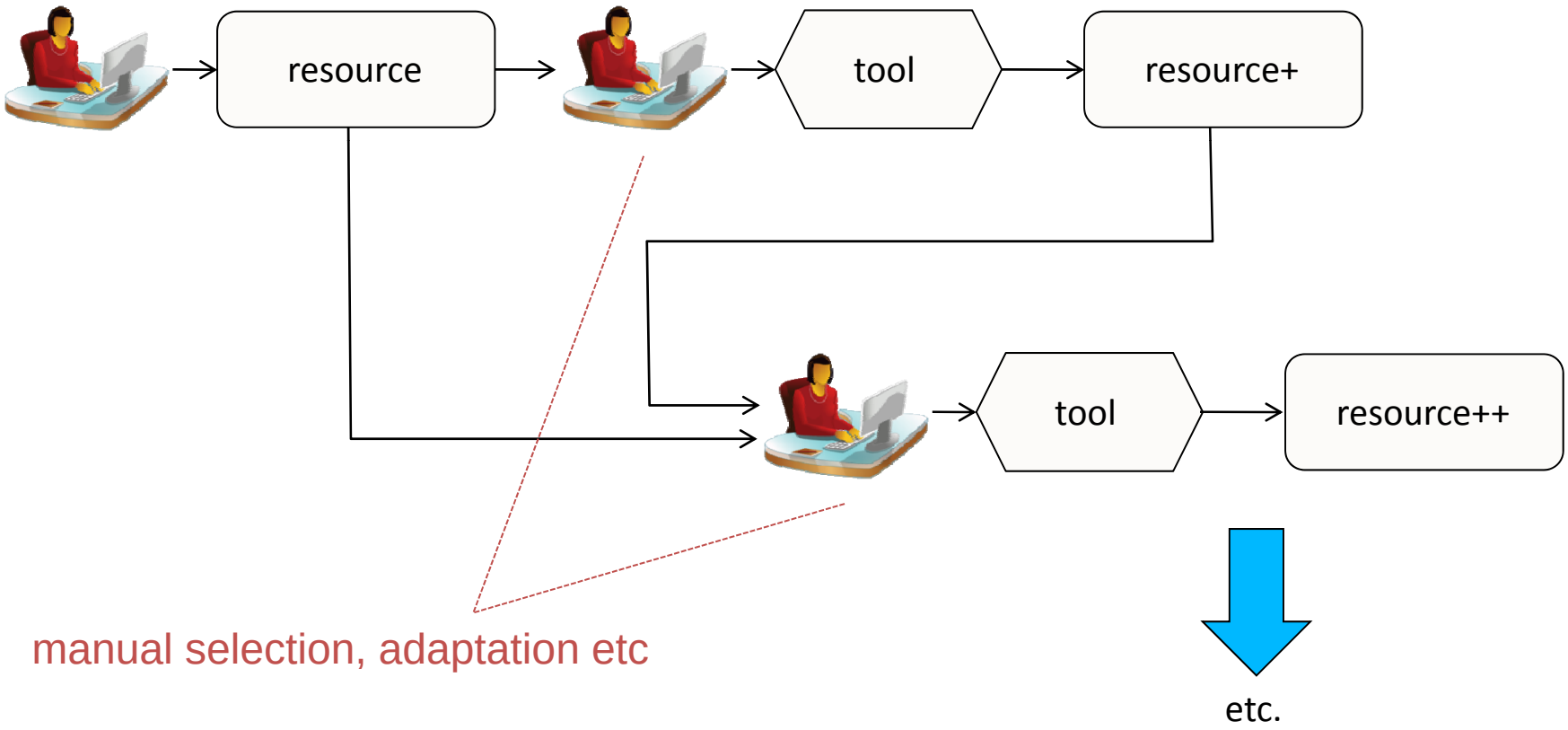


Content

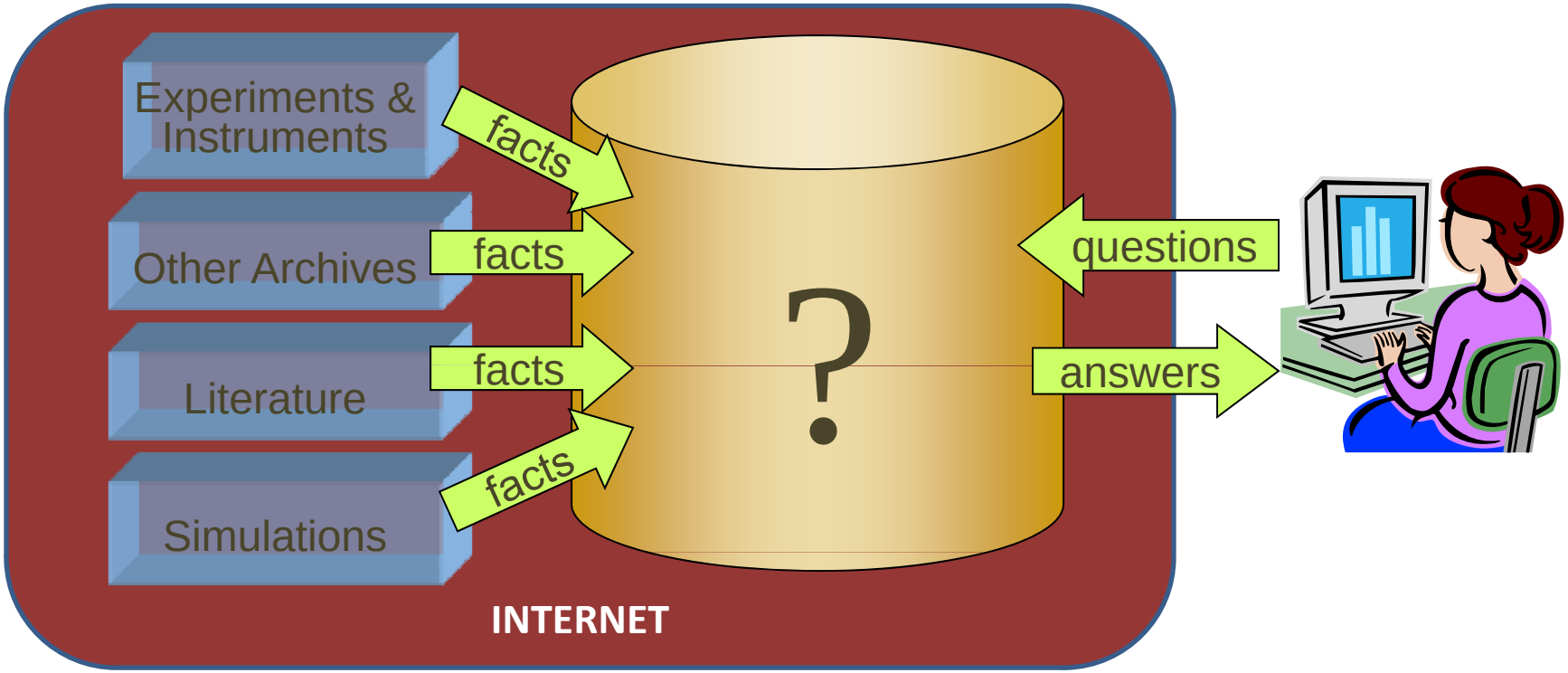
1. What is e-Science/e-Research?
   a) Data
   b) Operations
2. What do other communities do?
3. What does CLARIN do?
   a) Federation
   b) PIDs
   c) Metadata
   d) Web - Services
4. General Issues

Current Way of Acting

traditional workflow

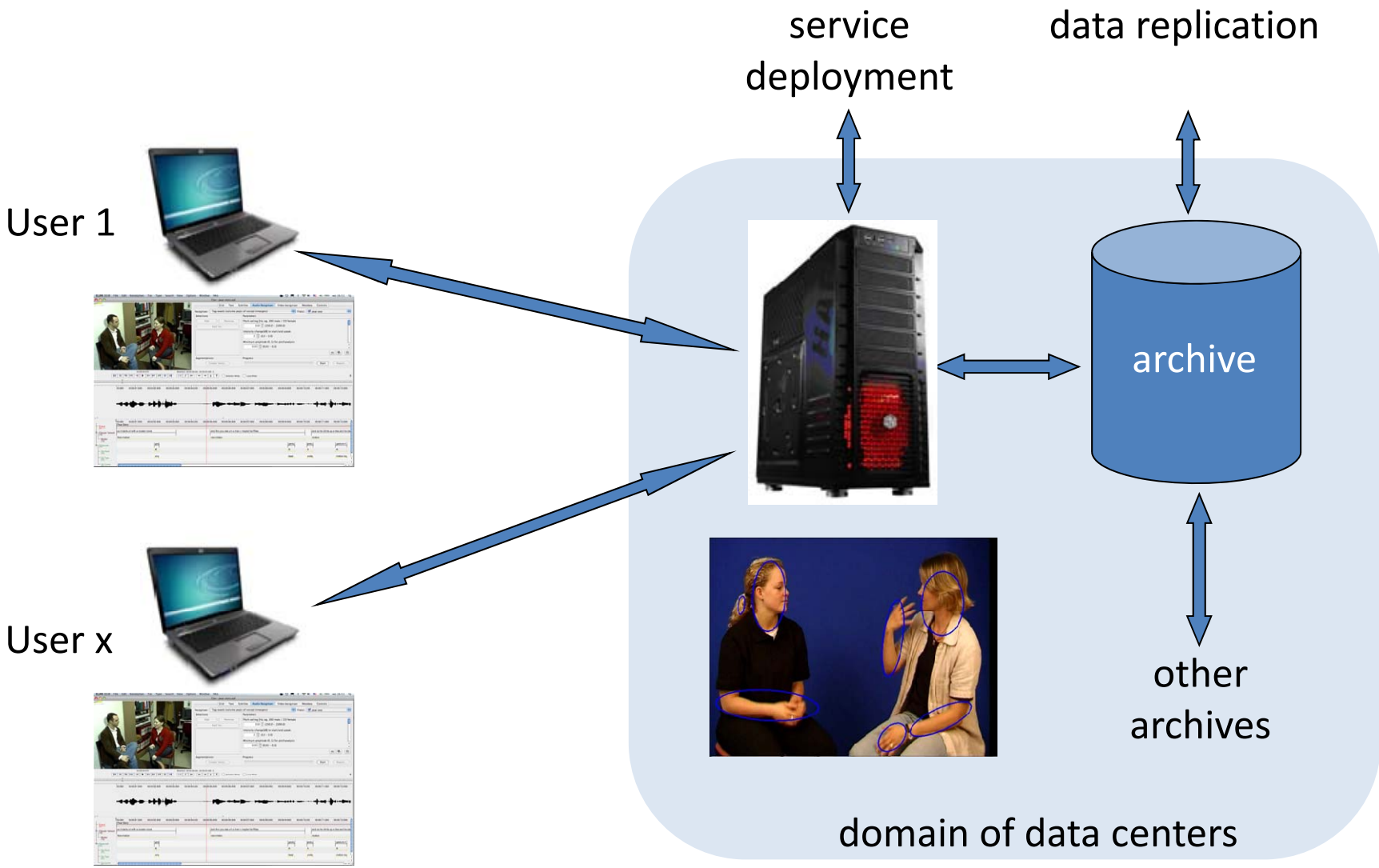


Dream of the "eResearcher"



- does not only need to be data that is available
- could also be services of various sorts
- machine translation - speech recognition etc
- language recognition (see RACAI)

Services Landscape



WebLicht Workflow Engine

The Language Archive

Stuttgart

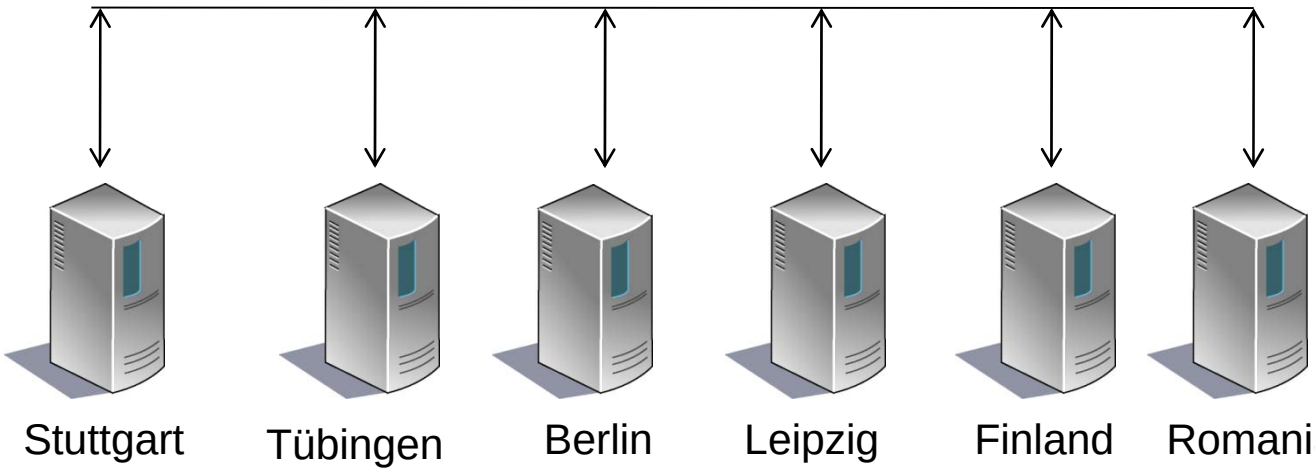
Standard-conformant
Text Corpus Encoding

Tübingen

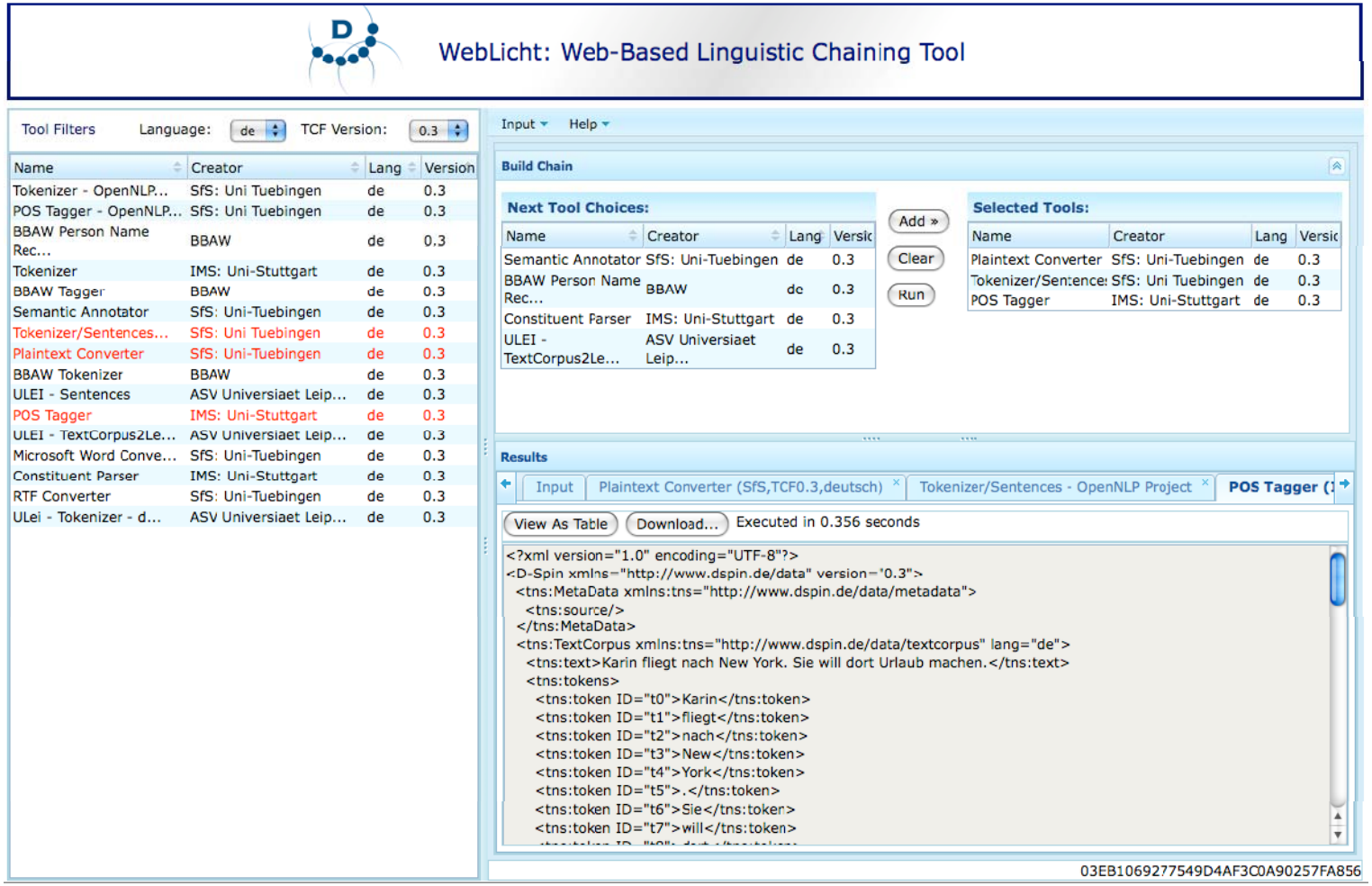
Web 2.0 Application for
Tool Chaining
and Execution

Leipzig

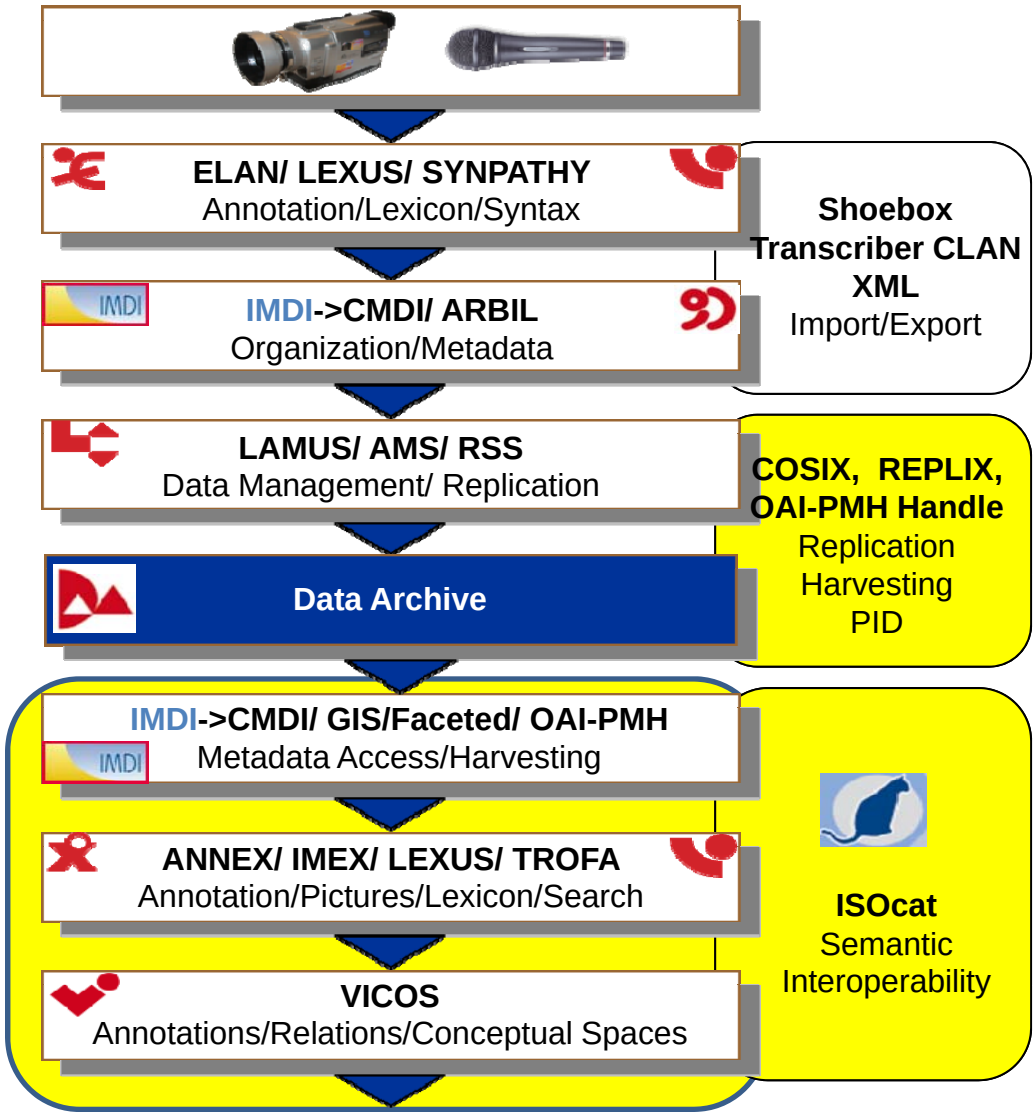
Repository



WebLicht - Screenshots



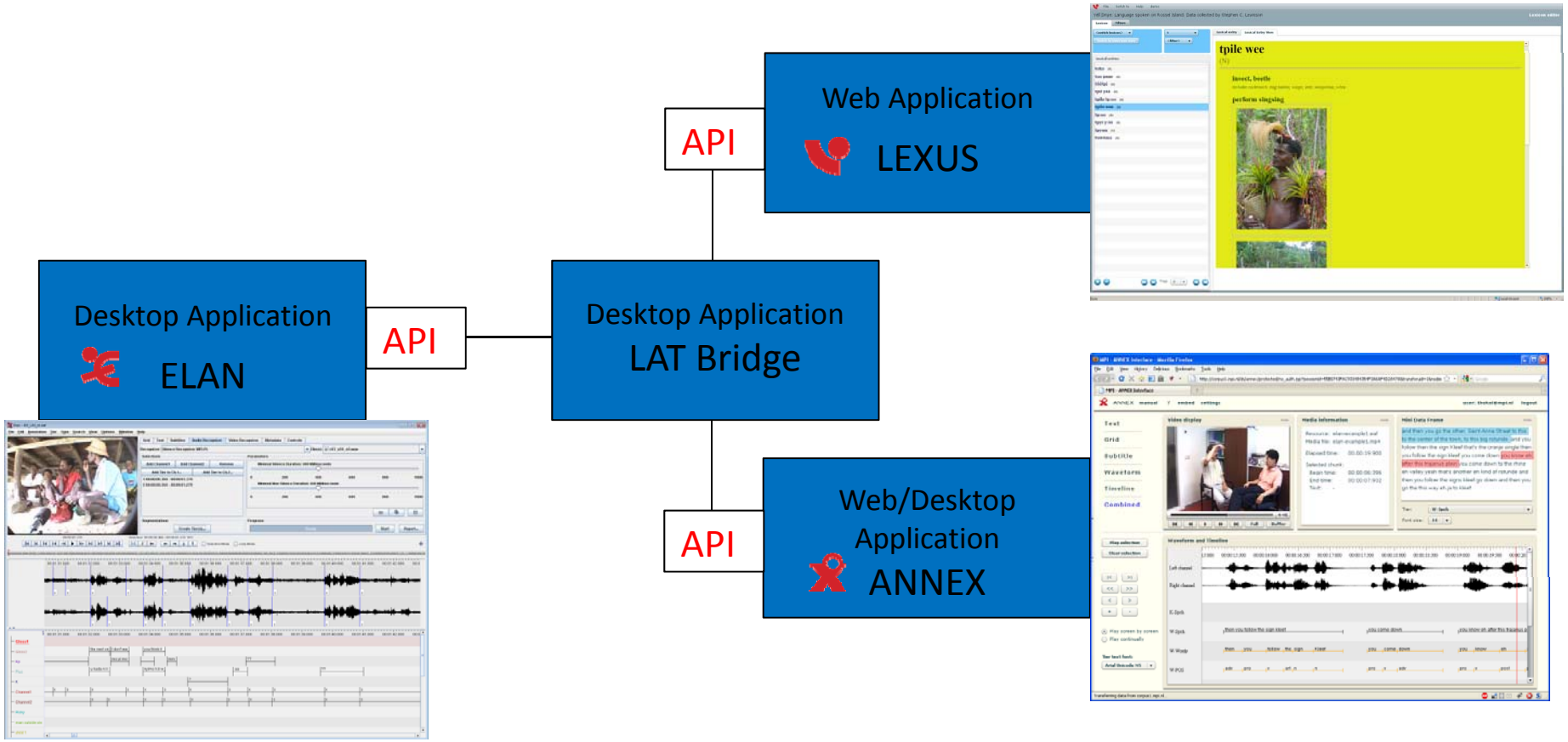
LAT Software to be available as WS



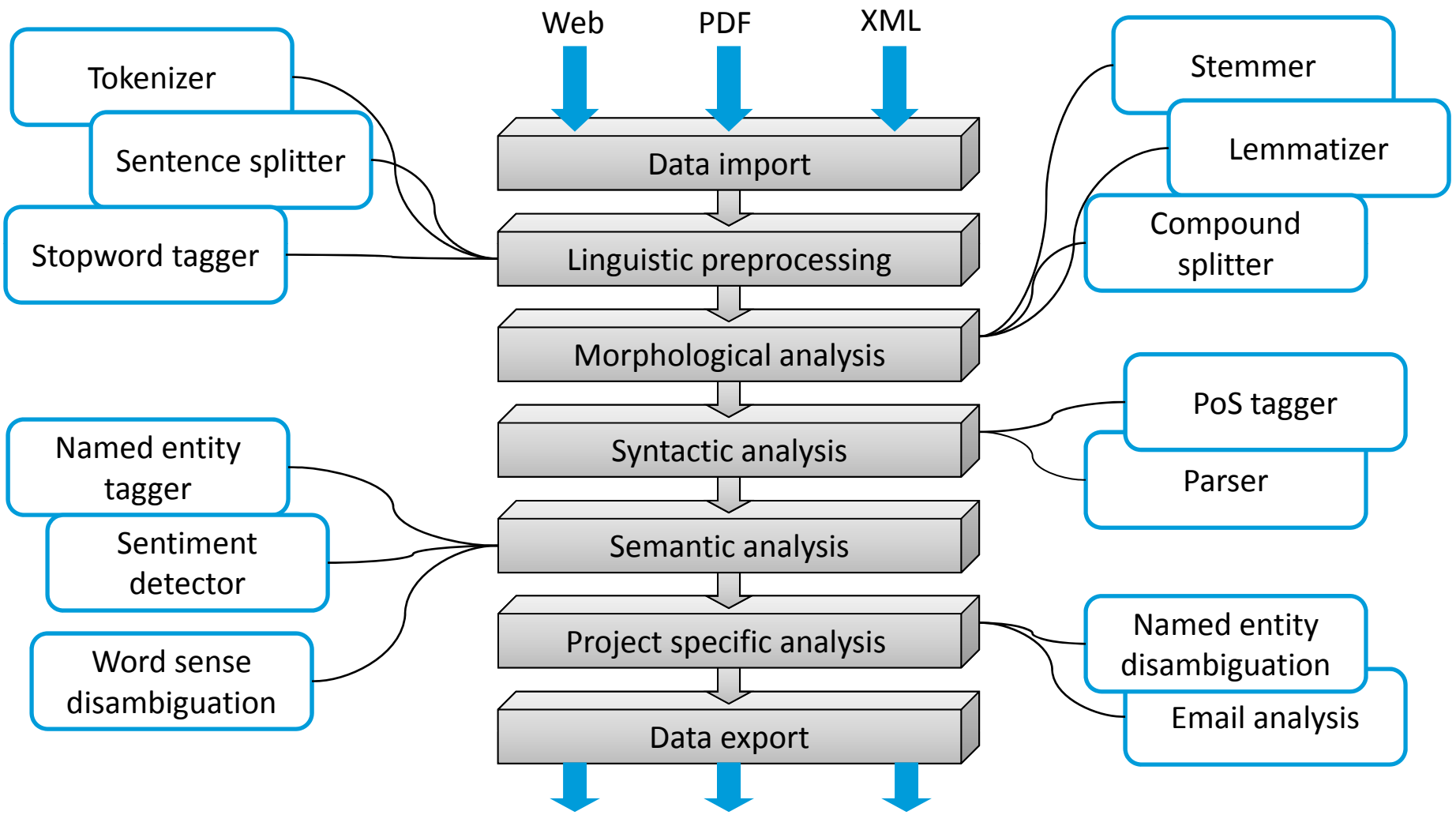
- full Lifecycle Support from data creation to semantic web like "exploitation"
- standards-based where possible
- modular design - all Java
- ELAN for example one of the most widely used annotation tools in the world
- data grid extensions
- ISO based interoperability extensions

Adapt the structure of LAT Tools

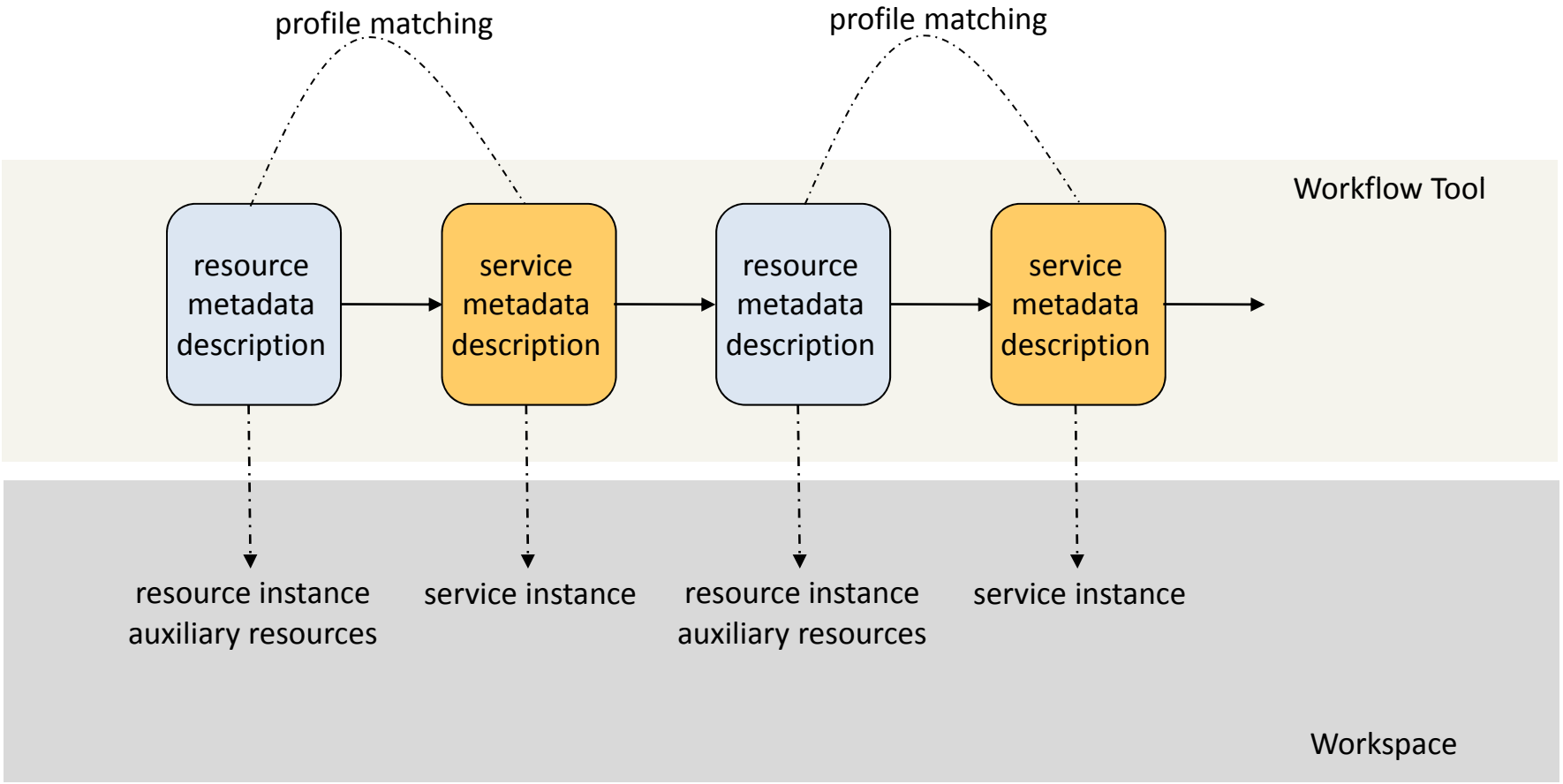
The Language Archive



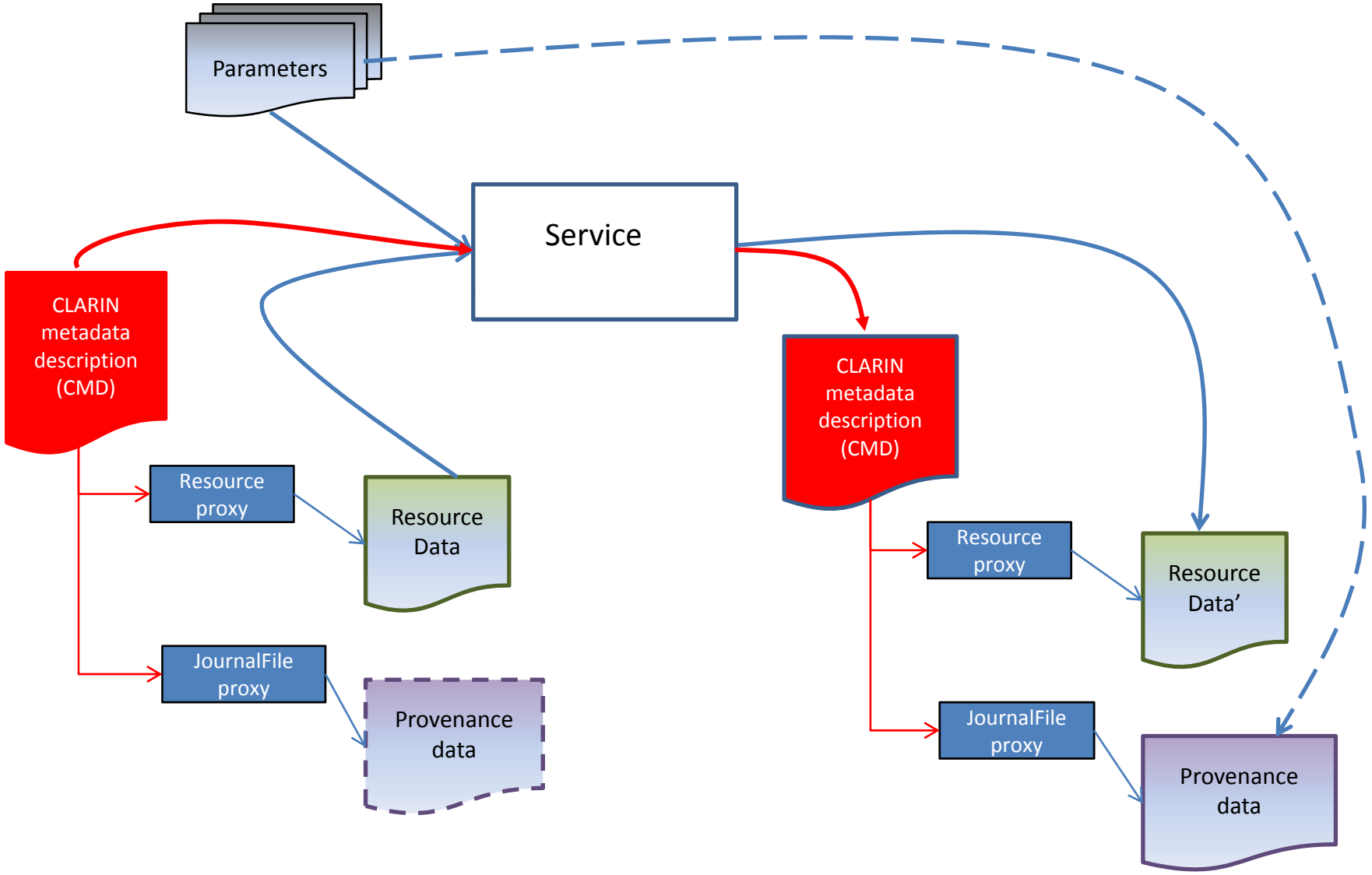
Darmstadt Knowledge Processing Software Repository (DKPro)



MD in workflow chain



MD creation


Parameters
Service
CLARIN metadata description (CMD)
Resource proxy
Resource Data
JournalFile proxy
Provenance data
CLARIN metadata description (CMD)
Resource proxy
Resource Data'
JournalFile proxy
Provenance data

Content

1. What is e-Science/e-Research?
   a) Data
   b) Operations
2. What do other communities do?
3. What does CLARIN do?
   a) Federation
   b) PIDs
   c) Metadata
   d) Web - Services
4. General Issues

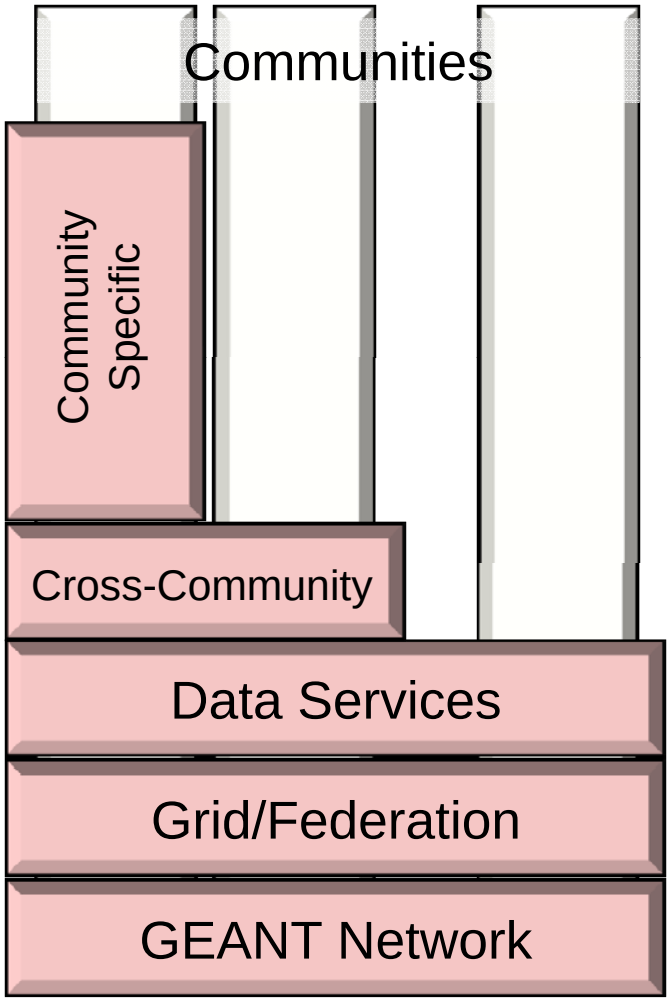
Basic IT Principles

The Language Archive

- all needs to be based on a few generic IT principles

- create and maintain atomic objects
  - don't mix different types of information
  - earlier granularity discussion
- use explicit syntax
  - obvious but still not common practice and not tool supported
- declare semantics
  - define the concepts you are using
  - ISOcat is a start - will it really take up?
- use PIDs
  - register all references explicitly
  - cool URIs may work for some purposes
- use stand-off principles

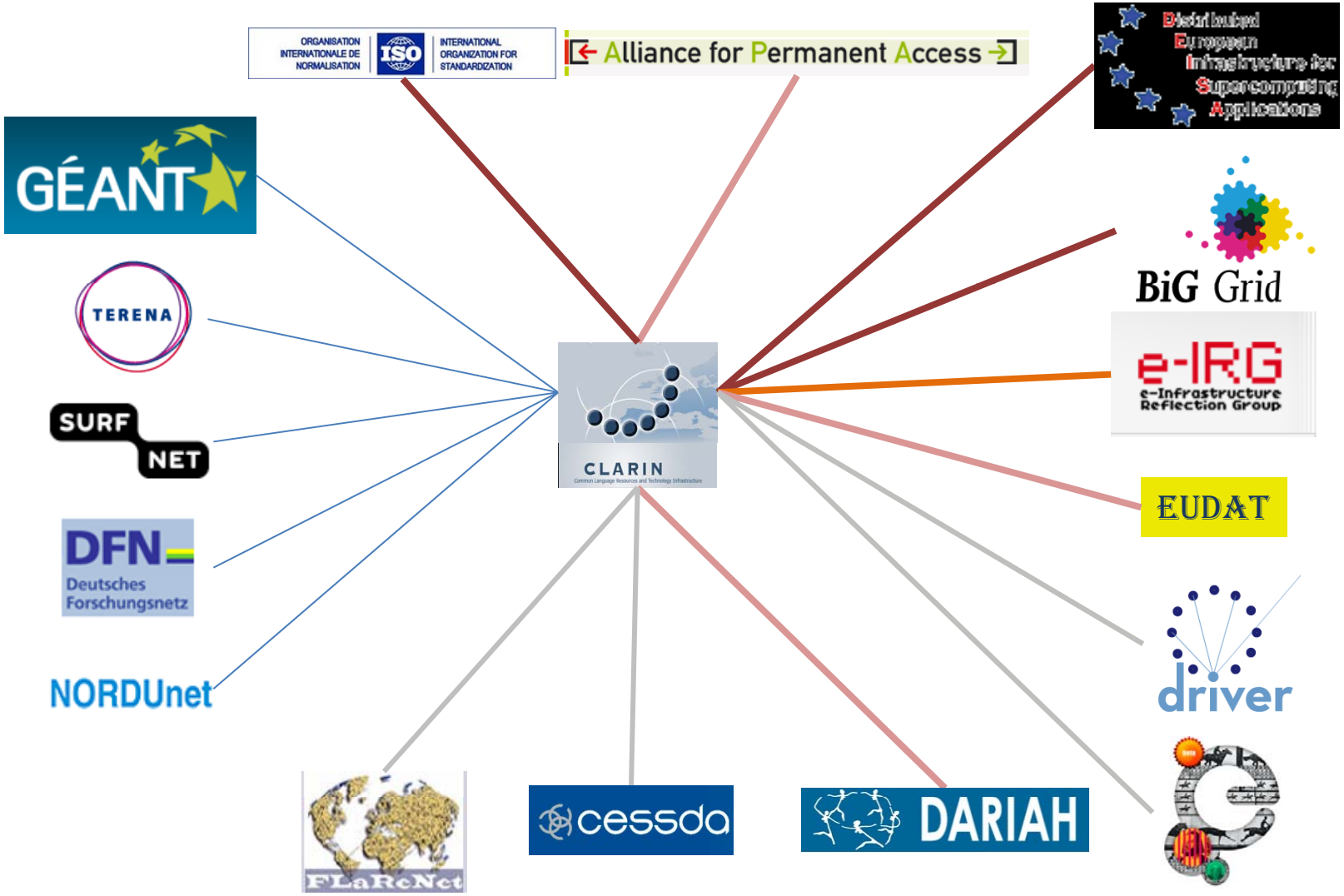
There are e-Infrastructures



- we all use networks (Email, Web, etc)
- physics people are using compute Grids
- we start using federation services
- MPI is using Data Services
- there will be a European Data Services Infrastructure (Data Preservation)

lots of interactions

The Language Archive



Falls nicht to end in Babylonish scenario nous avons still algo time om sistemas te improve.

Thanks for your attention.