International Tutorial DigitalWorld / GEOProcessing 2014

Digital Speedup Meets Long-term Requirements: Sciences, Knowledge, and Computing

The International Conference on Advanced Geographic Information Systems, Applications, and Services (DigitalWorld / GEOProcessing)
March 23–27, 2014, Barcelona, Catalunya, Spain

Dr. rer. nat. Claus-Peter Rückemann[1,2,3]

[1] Leibniz Universität Hannover, Hannover, Germany
[2] Westfälische Wilhelms-Universität Münster (WWU), Münster, Germany
[3] North-German Supercomputing Alliance (HLRN), Germany

ruckema(at)uni-muenster.de

Contents

Tutorial targets

Focus:

- Aspects of the speedup of digital developments with respect to the requirements for long-term sustainable knowledge and resources in many disciplines.
- Scientific progress demands for long-term persistent knowledge in research, documentation, and consecutive use.
- Processing and computing are becoming increasingly important in many disciplines, not only in disciplines spanning large time intervals like in geosciences or archaeology.
- Requirements for methodologies and systematics.
- Introduction to long-term knowledge handling and classification.
- High End Computing and storage resources used for processing and computing.
- Requirements and operation for advanced scientific computing environments.
- Advanced computing and collaboration.
- Basics of decision making and resources planning.
- Case studies for geoscientific and archaeological information systems: Knowledge and computing resources usage, operation and lifecycle aspects.
- Application scenarios: How the interests and needs of users and disciplines, services, and resource providers can be respected in order to generate long-term benefits from creating knowledge resources and using collaboration frameworks.

Focus questions

Some focus questions are:

- What is long-term knowledge from the sciences / geosciences / society view?
- How can we preserve, document, and handle various long-term information and knowledge?
- What does processing and computing mean in geosciences, life sciences, social sciences and other disciplines?
- Which "support for science itself" can we expect from "Big Data", "Exa-scale" and friends?

It is intended to have a dialogue with the audience on how the "speedup" of digital developments can be led to a "progress" for long-term sustainable knowledge.

Aspects and Challenges

Focus:

Increasing the overall long-term efficiency of

- gathering and using information, knowledge and computing,
- scientific research,
- related application scenarios,
- respecting the interests of users and disciplines, services, and providers of resources.

Red Line: Digital Speedup – Sciences, Knowledge, Computing

Aspects:

- Digital speedup and long-term requirements.
- Sciences - disciplines.
- Services and sevelopment.
- Providers.
- Knowledge and decision making.
- High End Computing.
- Application scenarios - consequences and participated parties.

Digital Speedup

Speedup aspects

- Application lifecycles,
- Data lifecycle,
- Performance,
- Data size,
- Availability,
- Ubiquity,
- Proprietary,
- Increasing complexity,
- Legalisation,
- Prices,
- ...

Sciences

Sciences requirements aspects

Items	Examples
Algorithms	Seismic stacking
Workflows	Knowledge discovery process
Methodologies	Supercomputing
Systematics	Processing
Measurements	Quantity and quality, Precision, Resolutions in space and time ...
Results, other	Evaluation
...	

Knowledge

Knowledge requirements aspects

Item	**Example**
Structure	Knowledge resources (e.g., LX)
Classification	Universal classification (e.g., Universal Decimal Classification, UDC)
Networks	Flexible networks (e.g., Software Defined Networks, SDN)
Intelligent components	Intelligent and autonomous (e.g., Multi-Agent System, MAS)
Complex systems	Collaboration frameworks (e.g., "Collaboration house")
Computational requirements	Integrated systems (e.g., Integrated Information and Computing System, IICS)
...	

Computing

Computing requirements aspects

Item	Example
Network	Technology, speed
HPC	Technology, performance
Big Data	Technology, algorithms
Infrastructure	Operation lifecycles
Heterogeneity	Hard- and software architectures
Services	Quality, staff
Interfaces	Ease of use
Application	Lifecycles
Application data sets	Data sizes etc.
Functions	Modifications, compatibility
Long-term archiving	Data size, storage IO
Data format	Deprecated features, compatibility
...	

Sciences, knowledge, computing

Snippets on questions and decisions

- Persistent knowledge?
- Essential data/knowledge extraction?
- Learning-knowledge extraction?
- Learning system components?
- Human/learning support?
- Interoperability?
- Sustainable funding?
- ...

Essentials and dependencies

Long-term essential tasks:

- Collect data (e.g., results, extended environment),
- Documentation,
- Vitality issues,
- ...

Long-term dependencies:

- Extend classifications,
- Port software,
- Update content,
- Optimize software,
- Rewrite software,
- Redesign algorithms,
- Reimplement system components,
- ...

Sciences, History, Archaeology, . . .

Classical, medieval, modern, . . .

Heron of Alexandria: (greek antique, "Steam Ball")
$\Longrightarrow$ "entertainment" but **not used as technology**.

Isidore of Seville: (encyclopedic, broad documentation)
$\Longrightarrow$ end of medieval phase, **not further used**.

Polyhistor: (Martin Fogel, broad knowledge)
$\Longrightarrow$ broad base, **not further used**.

In percentage we nearly know nothing about the past.

- Ancient and historical objects are mostly lost.
- Ancient and historical documentation is mostly lost.
- Ancient and historical technology is not fully understood.
- Context of past applications is not available.

Perception and hybris - how is the situation today

Perception of technology and exaggerated opinion of oneself (real sayings)

1996: **"It is not necessary to backup [internal] data because one can re-generate all data of mankind worldwide from the Internet."**

2000: **"All problems of software development can be solved with web services."**

2003: **"High Performance Computing can solve the remaining questions of mankind."**

2006: **"Grid Computing will drastically save investments by sharing resources."**

2008: **"Cloud Computing is just done by provisioning of services."**

2011: **"Certifying and optimising services will help reduce staff."**

2013: **"We have to provide all hardware architectures we can get hold of because user research requires them."**

2014: **"This supercomputer can store 8 billions of books."**

State of digital perception

Present trends (besides exceptions)

- **Big Data hype: People without any compact result data or sustainable case scenarios talking about the importance of huge amounts of whatever data.**
- **Exa-hype: People with access only to small TFlop/s and PFlop/s report about more than 1000× larger environments.**
- **Flop/s hype: People without any available application and experience present future application requirements for millions of cores per application.**
- **Optimisation hype: People propagating optimisation of codes will significantly help solve problems.**
- ...

Way (NOT) to go: Keep the speed on long-term

What others do: "Experts say: Fast is more important than long-term."

Let us take a look on what a virtual, "effective" institution will do.

"N"ewtoneless
"U"niversity
"T"echnology
"S"ervice

NUTS believe:

- **Technology speedup means increased knowledge**.
- **Digital media is driving knowledge**.
- **Technology development guarantees sustainability**.

NUTS live with:

- Primary requirements for sciences, knowledge, and computing are **technological requirements**,
- **Methodologies and systematics are "nice-to-have"**,
- **Knowledge will only be used within a project lifetime**.

NUTS set up, to "accelerate" long-term digital media use:

- **Digital-decision** in-house-group,
- **Decision-barricade** group ("malefiz management"),
- **Digital-cleanup competence** in-house working group.

Way to go: Cultural and Technological Development (Motivation)

Knowledge base:

Knowledge transfer is essential.

Over generations of objects and subjects, this requires:

- Knowledge recognition (expertise).
- Knowledge documentation, for any aspect of nature and society (sciences, literature, technical descriptions, tools, cultural heritage, mythology, songs, media, ...).
- Long-term means.

Long-term: Sciences - Disciplines

– View: Sciences - Disciplines –

Requirements

Needs and requirements from disciplines classically are in contrast with how resources and services are managed and operated.

Building services on this base typically polarises interests of participated groups.

From this point of view, most building processes regarding computing environments reveal a very small grade of efficiency.

Disciplines Involvement with High End Computing

Disciplines involvement goals, examples:

- Long-term knowledge creation (results, data, algorithms, computing instructions, etc.).
- Structure of knowledge.
- Reasoning (society and needs).
- Perception (grow with needs).
- Redundancy and availability.
- Formats, portability.
- System architectures.
- Batch-queue configuration.
- Workarounds and science / technology balance.

What does this mean for knowledge resources and transfer?

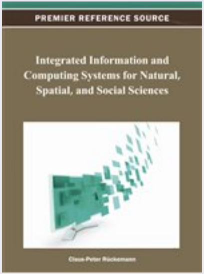
Integrated Information and Computing Systems

References

C.-P. Rückemann (ed.): *Integrated Information and Computing Systems for Natural, Spatial, and Social Sciences*, 21 chapters, IGI Global, Hershey, Pennsylvania, USA, 2012, Premier Reference Source, DOI: 10.4018/978-1-4666-2190-9, ISBN-13: 978-1-4666-2190-9 (hardcover), EISBN: 978-1-4666-2191-6 (e-book).

Topics:

- Integrated Systems, Information, Communication, and Computation
- Collaboration, Frameworks, and Legal Aspects
- Advanced Cognition, Intelligent Systems, and Security Management
- High End Computing, Storage, and Services
- Supercomputing, High Performance Computing, Computing Systems, Energy Efficiency, and Cloud
- Communication, Computation, Advanced Scientific Computing
- Advanced Applications, Modelling and Simulation in Natural Sciences, Geosciences, Medicine
- Big Data Exploration, Visualisation, Education, and Social Media
- Spatial Sciences, Social Sciences, Teaching, Learning, and Digital Media



http://www.igi-global.com/book/integrated-information-computing-systems-natural/67413
http://dx.doi.org/10.4018/978-1-4666-2190-9

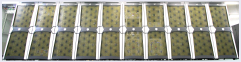
Disciplines

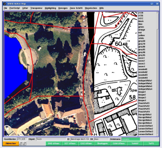
Increasing high end demands

Geosciences, planetology, climatology, physics, astrophysics, chemistry, engineering, oceanography, meteorology, geoinformatics, medicine, life sciences, archaeology, library sciences, . . ., processing, computing systems, information systems, search engines, criticality management, . . .

Privacy and Anonymity Target System Examples

Information and Computing Systems :: Data and Information





Privacy: **private**, **societal**, **economic**, **intellectual**

Discipline	Privacy	Anonymity
Geoscientific Information Systems	p, s, e, i	Individual, Society
Archaeology Information Systems	s	Society
Medical Data Information Systems	p, s, e	Individual, (Society)
Flight and Transport Systems	p	Individual, (Society)
Banking, Accounting, Billing Systems	p	Individual, (Society)
Exploration IS (energy, oil&gas)	e	Society
Environmental IS (pollution)	p, s, e	Individual, Society
Computing shared/distributed	e, i	Individual, Society
Navigation Systems	p	Individual
Recherche Systems, Search Engines	p, s, e, i	Individual
Georeferencing	p, s, e, i	Individual
Automation	p, s, e, i	Individual
Integration	p, s, e, i	Individual

Trust Systems: Present and Future Goals

What is the essence of protecting information and knowledge?

Present

© CPR

Future

Support trust! Protect data, information, and knowledge!

Minimise threats and misuse!

Separate security from management & administration!

Communicate: Any process needs communication!

Create modular technical-legal frameworks!

GMES/GEOSS/SEIS, GSDI/INSPIRE/GDI-DE, FDA/HIPAA, PSI/EPSI.

Collaboration frameworks reducing complexity!

"Collaboration house" framework.

Economic integration, accounting, billing!

Modular distributed systems like SGAS.

Examples: Legal Frameworks / Geo Information Systems

Examples:

- Global Spatial Data Infrastructure (GSDI)
  http://www.gsdi.org
- INfrastructure for SPatial Information in Europe (INSPIRE)
  http://www.ec-gis.org/inspire
- Geodateninfrastruktur Deutschland (GDI-DE)
  http://www.gdi-de.org
- European Public Sector Information (EPSI)
  http://www.epsiplus.net
- Global Monitoring for the Environment and Security (GMES)
  http://www.gmes.info
- Global Earth Observation System of Systems (GEOSS)
  http://www.earthobservations.org/geoss.shtml
- Group on Earth Observations (GEO)
  http://www.earthobservations.org
- Shared Environmental Information System (SEIS)
  http://ec.europa.eu/environment/seis/
- Geo Exploration and Information (GEXI)
  http://www.user.uni-hannover.de/cpr/x/rprojs/en/index.html#GEXI

Constraints for Geoinformatics Contributors and Participants

Legal Geo Data and Information Frameworks

Name	Framework and Reference
GMES	Global Monitoring for the Environment and Security http://www.gmes.info
GEOSS	Global Earth Observation System of Systems / GEO (Group on Earth Obs.) http://www.earthobservations.org/geoss.shtml
SEIS	Shared Environmental Information System http://ec.europa.eu/environment/seis/
GSDI	Global Spatial Data Infrastructure http://www.gsdi.org
INSPIRE	Infrastructure for Spatial Information in Europe directive (2007/2/EC) http://www.ec-gis.org/inspire
GDI-DE	Geodateninfrastruktur Deutschland http://www.gdi-de.org
PSI/EPSI	Public Sector Information directive / European Public Sector Information http://www.epsiplus.net

Laws and Legal Regulations Regarding Geo Data (national, DE)

- Copyright law (UrhG),
- Data security and privacy law (BDSG),
- Freedom of information law (IFG),
- Law on the reuse of information from public institutions (IWG),
- Environmental information law (UIG),
- Law on accessing digital geo data (GeoZG).

Examples: Applications and Tools

Examples: Applications, Tools, . . . corresponding to interfaces and architectures

- **Applications and libraries:** Mostly own code developments, commercial developments, community developments, e.g., BLAS, LAPACK, NAG, ATLAS, CPMD, MOLPRO, FEOM, Gaussian, NAMD, FFT, TAU, NWChem, VMD, EnSight, ABAQUS, ANSYS, FLUENT, STAR-CD . . .
- **Parallelisation:** MPI (SGI MPI / MPT, Intel MPI / . . .), OpenMP, MPICH, MVAPICH, SHMEM . . .
- **Profiling / Debugging:** Intel Threading & Tracing Tools, PerfSuite, PCP, TotalView, ddt, gdb . . .
- **Software Components:** SLES, CLE, SGI Tempo, Scali Manage, GPFS, Moab, Torque, Lustre, PP, C3, Ganglia, Grid tools . . .

View: Services and Developers

– View: Services and Developers –

Provided

In almost all cases the percentage of re-used knowledge over system generations is very small, leading to perpetuous "re-invention" and "re-discussion" for every cycle.

The suggested rate of re-use is below 10 percent.

Compute Portfolio

Services differ by physics and intention, especially:

- **Latencies and bandwidth:** Low segment: Latency 100 μs to several milliseconds (distributed), latency 1–2 μs (local), bandwidth 1.5–4 GB/s (local),
- **Distributed data transfer:** Data transfer for supercomputing is essential with any big (volume) data, physics provide limitation to economical distributed solution.
- **Distributed memory usage:** Shared memory usage for supercomputing is essential with shared memory algorithms, physics provide limitation to economical distributed solution.

HPC IO

High Performance Computing I/O Compute Resource (HPC Center)

- minimal entry level per job (logical AND):
- publicly funded research, no production jobs or industry users
- will be validated that the minimal access requirements are fulfilled and reasonable
- 512 nodes
- 4096 Cores (e.g., Intel)
- maximum: about double nodes and cores
- 3 GB memory usage per core
- already parallelized MPI read/write into 1-1000 files
- extensive and already optimised MPI I/O communication from mostly all cores used (max 1-2 TB per second overall)
- runtime per job 12 hours
- job must be batch system based
- environment fixed, depending on installation
- ssh access only
- programming, compilation, installation by user

HPC

High Performance Computing Compute Resource (HPC Center)

- minimal entry level per job (logical AND):
- publicly funded research, no production jobs or industry users
- will be validated that the minimal access requirements are fulfilled and reasonable
- 512 nodes
- 4096 Cores (e.g., Intel)
- maximum: about double nodes and cores
- 3 GB memory usage per core
- already parallelized MPI read/write into 1-100 files
- extensive MPI compute communication from mostly all cores used
- runtime per job 36 hours
- job must be batch system based
- environment fixed, depending on installation
- ssh access only
- programming, compilation, installation by user

HPC SMP and Shared Memory

High Performance Computing SMP Resource (HPC Center)

- minimal entry level per job (logical AND):
- publicly funded research, no production jobs or industry users
- will be validated that the minimal access requirements are fulfilled and reasonable
- 32 nodes
- 32x8 Cores (e.g., Intel)
- maximum: 2-3 jobs/tasks in parallel
- 512 GB memory usage per job
- OpenMP communication
- runtime per job 12 hours
- job must be batch system based
- environment fixed, depending on installation
- ssh access only
- programming, compilation, installation by user

Cloud and Grid

Cloud / Grid (Provider, Computing Center)

- minimal entry level per job:
- 1-64 nodes
- *n* (Intel or other) cores depending on architecture and provider
- small to medium sized memory usage per core
- small I/O (Giga-Bytes not TeraBytes overall)
- hundreds of cores
- loosely coupled, parallel jobs, task-parallel, moderate MPI parallel
- further services/anything else per-pay that scientific HPC Centers might not provide
- efficiency requirements depend on provider and customer agreements
- middleware and access depending on provider

View: Computing, Storage - Providers

– View: Providers –

Requirements

- Economical environment.
- Efficient operation.
- Sustainable investment.
- Defined policies.
- ...

High End Systems and facets besides knowledge

More than a tool? Always think of knowledge:

- Content.
- Context.

A knowledge base has to be multi-disciplinary and facetted:

- Disciplines,
- Services,
- Providers.

Architecture and Provisioning

Which architecture?

- Standalone / workstation,
- Cluster,
- Grid,
- Cloud,
- High Performance Computing (HPC),
- Other.

How do you provision services or resources?

- Institute,
- Alliance,
- Hosting,
- Housing,
- Other.

Type and state

Which type?

- Research
- Industry
- Mix
- Other

Which kind of usage?

- Interactive
- Batch
- Hybrid
- Other

Usage and Programming

How can the architecture be used efficiently?

- MPP (Massively Parallel Processing),
- SMP (Shared-Memory Parallel),
- Other.

Which model?

- Low Level: MPI (Message Passing Interface),
- Low Level: OpenMP,
- High Level: PGAS (Partitioned Global Address Space),
- Virtualisation: PVM (Parallel Virtual Machine),
- Other.

Productivity

How do you gather information about productivity?

- Profiling,
- Benchmarking,
- Polling,
- Quality of . . . "measurements",
- Other.

How do YOU gather knowledge?

. . . and are there differences?

- In general.
- Within disciplines.
- With High End Computing.

High End Computing Requirements Study and Disciplines

How to build long-term knowledge transfer?

- Requirements studies with user groups,
- Documentation of tender processes,
- Documentation of operation and service,
- ...

Disciplines: Natural sciences, spatial sciences, archaeology, geosciences, etc.

- Disciplinary,
- Inter-disciplinary,
- Multi-disciplinary,
- ...

Tender Process – How Requirements are Currently "Considered"

Multi-step cycle of 4-7 years:

Requirements:

- **Users / disciplines**
  $\Longrightarrow$ request users / disciplines for comments.
- **Infrastructure**
  $\Longrightarrow$ participate infrastructure planners, architects, administration, etc.
- **Legal regulations (non-discrimination / environment / procedures)**
  $\Longrightarrow$ participate lawyers.
- **Technical developments**
  $\Longrightarrow$ information from developers and industry.
- **Future planning**
  $\Longrightarrow$ participate hierarchy.
- ...

This should be drastically improved by PARTICIPATING experience and knowledge, practically experienced auditing, on-topic users, developers, and industry ...

Comparison of High End Systems

Can High End Systems be compared seriously? Remember:

- Every HEC / Supercomputing system is unique in it's overall hardware, software stack, and configuration.
- Development cyle is about 5 years.
- Most tests for the bleading edge components have to be done on final, entire systems.

Extraordinary With Singular Aspects: The Greatest, Biggest, Greenest

Top500 Top500 list with the "fastest" supercomputers in the world.
`http://www.top500.org`.
Only standard-benchmark: High Performance Linpack (HPL).
(2012-11 Blue Waters/NCSA system opts out of Top500 list due to Linpack.)

Green500 "Ecological" list going for performance in relation to energy consumption.
`http://www.green500.org`.
Only energy and only in operation.

Graph500 `http://www.graph500.org`.

...

Complex Systems

Supercomputing Resources – Examples

For the further dialog within the tutorial, the tutorial discusses some selected historical and up-to-date High Performance Computing systems and hardware and components used with Advanced Scientific Computing.

- Cray2
- JUMP
- BSC
- HLRB
- Shenzhen
- Jaguar
- Tianhe
- Sequoia
- Titan
- German supercomputing (SuperMUC, JUQUEEN, HLRN, and others)
- ⇒ Supercomputing and big data
- ⇒ Operation and infrastructure transition phases
- ⇒ Infrastructures, networks, and architectures
- ⇒ Major long-term and sustainability issues with infrastructures
- ...

`--- ABOVE EXAMPLES FOR DISCUSSION LEFT OUT HERE ---`

Long-term: Knowledge

Where knowledge is ...

Knowledge is created from a subjective combination of different attainments as there are intuition, experience, information, education, decision, power of persuasion and so on, which are selected, compared and balanced against each other, which are transformed and interpreted.

And the consequences ...

Authentic knowledge therefore does not exist, it always has to be enlived again. Knowledge must not be confused with information or data which can be stored. Knowledge cannot be stored nor can it simply exist, neither in the Internet, nor in computers, databases, programs or books.

Knowledge and Application

Processes

- Knowledge base creation,
- Knowledge base transfer over generations,
- Documentation of requirements respective algorithms,
- Documentation of context respective architectures,
- Usage development within tender processes.

Selection on Structure, Content, Context, and Computing

Theory and practice

- Structural deficits.
- Content can be described and even signed to a certain extend.
- Context cannot be handled to a comparable extent. (Users can sign a PDF document, but what about signing it's context?)
- Long-term issues are mostly out of sight. (What will signature validity mean to archiving and reuse?)
- What does this in general mean to long-term knowledge-based processes?

Classification implementation

LX Foundation Scientific Resources

As one of the elementary qualities, the LX Foundation Knowledge Resources allow to refer to any kind of references, therefore they also allow to refer to any kind of classification. If nothing special is mentioned then all the basic classification codes are used in an unaltered way. If any classification refers to a modified code then the authors of contributions have to notice and document the modifications explicitely.

License remark:

The examples and small unsorted excerpts of the knowledge resources objects features only refer to main UDC-based classes, which for this part of the publication are taken from the Multilingual Universal Decimal Classification Summary (UDCC Publication No. 088) released by the UDC Consortium under the Creative Commons Attribution Share Alike 3.0 license (first release 2009, subsequent update 2012):

- Multilingual Universal Decimal Classification Summary (2012). UDC Consortium, 2012, Web resource, v. 1.1. The Hague: UDC Consortium (UDCC Publication No. 088), Retrieved January 12, 2014, from http://www.udcc.org/udcsummary/php/index.php.
- Creative Commons Attribution Share Alike (2012). Creative Commons Attribution Share Alike 3.0 license, Retrieved January 12, 2014, from http://creativecommons.org/licenses/by-sa/3.0/.

Knowledge, Documentation, and Classification

Universal Decimal Classification (UDC)

The Universal Decimal Classification (UDC) is a general plan for the knowledge classification. UDC is a hierarchical decimal classification system that divides the main knowledge fields into 10 main categories (numbered from 0 to 9). Each field is in turn divided into 10 subfields, each subfield is inturn divided into 10 subsubfields, and so on. A more extensive classification code in general describes a more specific subject.

Faceted and multi-disciplinary context

"Facetted" and "multi-disciplinary" is synonym to the Universal Decimal Classification (UDC), `http://www.udcc.org`. UDC uses a "(...)" notation in order to indicate aspect. These descriptions are called facets. In multi-disciplinary object context a faceted classification does provide advantages over enumerative concepts.

The classification deployed for a universal documentation must be able to describe any object with any relation, structure, and level of detail. Objects include any media, textual documents, illustrations, photos, maps, videos, sound recordings, as well as realia, physical objects such as museum objects.

Documentation and Form

Form (UDC, excerpt, English)

```
(0.02)     Documents according to physical, external form
(0.03)     Documents according to method of production
(0.034)    Machine-readable documents
(0.04)     Documents according to stage of production
(0.05)     Documents for particular kinds of user
(0.06)     Documents according to level of presentation and availability
(0.07)     Supplementary matter issued with a document
(0.08)     Separately issued supplements or parts of documents
(01)       Bibliographies
(02)       Books in general
(03)       Reference works
(04)       Non-serial separates. Separata
(041)      Pamphlets. Brochures
(042)      Addresses. Lectures. Speeches
(043)      Theses. Dissertations
(044)      Personal documents. Correspondence. Letters. Circulars
(045)      Articles in serials, collections etc. Contributions
(046)      Newspaper articles
(047)      Reports. Notices. Bulletins
(048)      Bibliographic descriptions. Abstracts. Summaries. Surveys
(049)      Other non-serial separates
(05)       Serial publications. Periodicals
(06)       Documents relating to societies, associations, organizations
(07)       Documents for instruction, teaching, study, training
(08)       Collected and polygraphic works. Forms. Lists. Illustrations. Business publ.
(09)       Presentation in historical form. Legal and historical sources
(091)      Presentation in chronological, historical form. Historical presentation.
(092)      Biographical presentation
(093)      Historical sources
(094)      Legal sources. Legal documents
```

Documentation and Language

Languages, natural and artificial (UDC, excerpt, English)

```
=1      Indo-European languages of Europe
=11     Germanic languages
=12     Italic languages
=13     Romance languages
=14     Greek (Hellenic)
=15     Celtic languages
=16     Slavic languages
=17     Baltic languages
=2      Indo-Iranian, Nuristani (Kafiri) and dead Indo-European languages
=21     Indic languages
=29     Dead Indo-European languages (not listed elsewhere)
=3      Dead languages of unknown affiliation. Caucasian languages
=35     Caucasian languages
=4      Afro-Asiatic, Nilo-Saharan, Congo-Kordofanian, Khoisan languages
=5      Ural-Altaic, Palaeo-Siberian, Eskimo-Aleut, Dravidian and Sino-Tibetan
=521    Japanese
=531    Korean
=541    Ainu
=6      Austro-Asiatic languages. Austronesian languages
=7      Indo-Pacific (non-Austronesian) languages. Australian languages
=8      American indigenous languages
=81     Indigenous languages of Canada, USA and Northern-Central Mexico
=82     Indigenous languages of western North American Coast, Mexico and Yucatán
=84     Ge-Pano-Carib languages. Macro-Chibchan languages
=85     Andean languages. Equatorial languages
=86     Chaco languages. Patagonian and Fuegian languages
=88     Isolated, unclassified Central and South American indigenous languages
=9      Artificial languages
=92     Artificial languages for use among human beings. Int. aux. languages (interlanguages)
=93     Artificial languages used to instruct machines. Programming/computer languages
```

Creating Groups and References

UDC Operations

Standardised operations with UDC are, e.g.,

Operation	Symbol
Addition	"+"
Consecutive extension	"/"
Relation	":"
Subgrouping	"[]"
Non-UDC notation	"*"
Alphabetic extension	"A-Z"

besides place, time, nationality, language, form, and characteristics.

Examples

```
(0.02/.08)                Special auxiliary subdivision for document form
=1/=8                     Natural languages
=1/=2                     Indo-European languages
=9/=93                    Artificial languages
59+636                    Zoology and animal breeding
(7):(4)                   Europe referring to America
311:[622+669](485)        statistics of mining and metallurgy in Sweden
004.382.2:[902+550.8] CPR  Supercomputers ref. to archaeology and geosciences, CPR author
```

Obstacles

Obstacles reducing success and efficiency with the processes

- Time consumption (e.g., staff, project timelines),
- Documentation (e.g., low percentage of reusability),
- Classification (e.g., limited views),
- Tools (e.g., changing repeatedly),
- "Standards" (e.g., changing repeatedly),
- ...
- Different perception of goals, strategies, and completeness.

Complementary

Structure

- Must be able to contain and refer to any content.

Full text and keywords

- Groups, regular expressions, search functions, . . .

Soundex

- Algorithm for calculating codes from text strings, representing phonetic properties.
- Originally only used for names, in Englisch.
- The original algorithm mainly encodes consonants.
- Goal is to encode homophones with the same representation, minor spelling differences do result in the same representation.
- Various modifications for any language, topics, any kind of words, support for many programming environments.

Helpers – What you always need

Staff and resources are most important

- Quantity of Staff and Resources depends (sometimes due to economical aspects).
- Quality of Data (QoD) can optimise requirements for staff and resources.

Examples for Multi-Disciplinary Use

Multi-disciplinary status

- Medical Informatics,
- Geoinformatics,
- Legal Informatics,
- Geoforensics,
- Archaeology and Digital Archaeology,
- Medical Geology,
- Digital Forensics,
-
- . . .

Archaeology

Content

- Overall information is widely distributed.
- Sometimes very difficult and a long lasting challenge not only to create information but even to get access to a few suitable information sources.
- Digital and realia objects.
- All participating disciplines, services, and resources have to be prepared for challenges as big data, critical data, accessibility, longevity, and usability.

. . . digital and long-term issues

- Even best practice cannot preserve realia and data context.
- Context is often destroyed.
- Long-term issues.
- Currently neither a standard being used for one discipline nor an international standard . . .

Archaeology

Goal

- Need integrated knowledge base for archaeological and natural sciences.
- Necessary to collect data from central data centers or registers. Examples archaeological and geophysical data:
  - North American Database of Archaeological Geophysics (NADAG).
  - Center for Advanced Spatial Technologies (CAST).
  - Archaeology Data Service (ADS).
  - Records as with Center of Digital Antiquity.
  - Records as with the Digital Archaeological Record (tDAR).
- An integrated "Collaboration house" framework is designed to consider all aspects and to handle any kind of object.

Medicine

. . . digital and long-term issues

- Documentation.
- Natural sciences data integration?
- Catalogs (International Classification / Catalog of Diseases, ICD).
- Classification (Universal Decimal Classification, UDC).
- Data security.
- Privacy.
- Anonymity.
- . . .

Libraries

. . . digital and long-term issues

- Documentation.
- Catalogues.
- Classification (Universal Decimal Classification, UDC). Today about 150000 libraries are using UDC classification and implementing information systems herewith.
- Referencing.
- Search.
- Licensing.

Decision Making

Basics of Decision Making ("DM")

Decision making is the fundamental base for any process as well as decision making is a process and result itself.

Nevertheless it is very common

- ... to have deficits in decision making processes.
- ... to underestimate the value of knowledge creation.
- ... to have opposition due to historical and social development.

Aware of!

- No decision is an influence to the "selection", too!
- To shorten planned decision making processes means significant interaction.

Introduction to Decision Making

What we can learn from others (references):

http://www.cartoonstock.com/directory/d/decision-making.asp
http://www.decision-making-solutions.com/management_cartoons.html
http://search.dilbert.com/comic/Decision%20Making

--- ABOVE EXAMPLES FOR DISCUSSION LEFT OUT HERE ---

About Decisions

Lemma 1:

- **It is easy to do any decision without expertise.**

Lemma 2:

- A decision (making process) should be **fast and perfectly correct**.

  In case a decision cannot be fast **and** perfect,
  it should be fast **or** perfect.

  In **no** case should a decision be slow and wrong.

Base for Decision

Essential relation:

Decision making! $\Longleftrightarrow$ **Selection making!**

Classics: Ask for Decision

. . . prominent YES or NO decision example:

(Y/N)?

Problem Analysis

Description:

- Performance analysis (current status / resulting status),
- Problem / target identification (e.g., deviations from performance standard, causes, change of distinctive feature),
- Problem / target description,
- Distinguishing marks between what has been effected by a cause and what has not,
- Deduction of causes from relevant changes found with the problem analysis (identification),
- Cause to a problem is most likely the one that exactly explains the sum of facts.

Example Decision Making Process

Description:

- Establishing the objectives,
- Classification of objectives,
- Place classified objectives in order of importance,
- Development of alternative actions,
- Evaluation of alternatives against all the objectives,
- The tentative decision is that alternative being is able to achieve all the objectives,
- Evaluation of the tentative decision for possible consequences,
- Take decisive actions, take additional actions (prevent adverse consequences from becoming problems)
- Start problem analysis and decision making process iteratively,
- Steps for decision model in order to determine an optimal production plan and reduce conflict potential.

Decision Planning Process

Description:

For best practice, introduce a decision planning process to important decisions in order to result in the following benefits:

1. **Establish independent goals.**
   That means a conscious and directed series of choices.
2. **Aim to a standard of measurement.**
   The measurement should provide information on the distance to the goal.
3. **Convert values to action.**
   The resulting information should be used to support the planning.
4. **Commit limited resources in an orderly way.**
   Planning and commitments for any kind of resources, e.g., staff, money, time.

Example Decision Making Phases

Phases:

Orientation stage: Starting with kick-off or warm-up, exchange with all parties.

Conflict stage: Dispute, arguments, working on common denominators and positions.

Emergence stage: Vague positions and opinions being discussed.

Reinforcement stage: Decision making and justification.

Selected Decision Making Techniques

Techniques:

Rational decision making: List the pro and contra (advantages and disadvantages) of each option. Contrast the costs and benefits of alternatives.

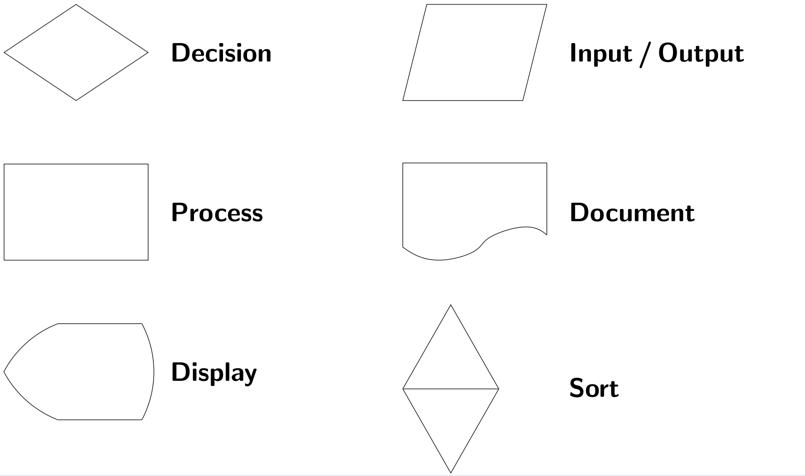
Elimination by aspects: Choosing alternatives by "mathematical psychology". Covert elimination process, comparing the available alternatives by aspects. Choose an aspect and eliminate the alternatives without the aspect. Repeat until one alternative remains.

Simple prioritisation: Choosing an alternative showing the highest probability-weighted utility from all alternatives, resulting from the decision analysis process.

Satisficing: The examination of alternatives is stopped as soon as an acceptable alternative is found.

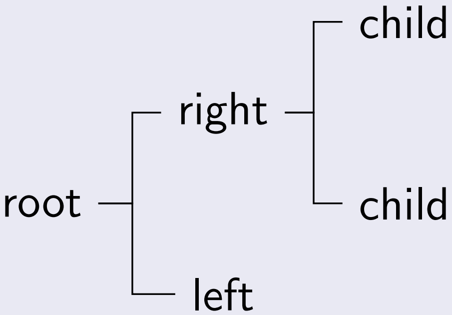
Visualising Flow Basics

Symbols



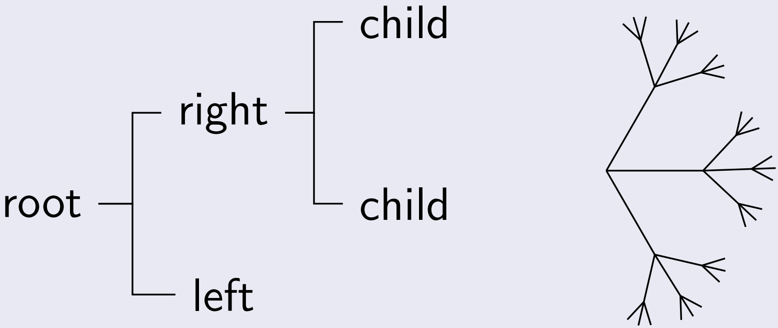
Trees and Forks

Trees and Forks



Trees and Forks

Trees and Forks



Areas of Automated Application

Commonly no tools with:

- Private,
- Evolution.

Prominent (support only) tools with:

- Environment,
- Catastrophy,
- Geostatistics,
- Military,
- Games,
- Exploration,
- Medicine,
- Traffic,
- Court,
- Contracts,
- Computer,
- Budget,
- Security,
- ...

Areas of Application

Examples for decision systems and support:

- **Environment:**
  - FLODIS: Sustainable Floodmanagement of the Oder river IS.
  - STEWARD: Support Technology for Environmental, Water and Agricultural Resource Decisions.
  - SDSS: Spatial Decision Support System.
  - LANDS: Land Analysis and Decision Support (system).
  - WEDSS: Whole Earth Decision Support System (international).
- **Catastrophy:**
  - WDSS: Warning Decision Support System (oceanography).
- **Geostatistics:**
  - MCDM: Multicriterion Decision Making.
- **Military:**
  - ADA: Applied Decision Analysis.
  - EOTDA: Electro-Optical Tactical Decision Aid.
- **Games:**
  - . . . For example, Chess, mathematical basics, defined alternatives.

Areas of Application and Why is Decision Support Imperfect

Why are there no systems for?:

- Natural sciences fundamentals,
- Informatics development,
- Basic algorithms,
- Geophysical data analysis,
- Computing architectures,
- Hardware systems development,
- ...

Add Forensics to the Decision (© CPR / LX / GEXI)

Invisible things can make a difference:

```




<EOF>
```

Add Forensics to the Decision (© CPR / LX / GEXI)

Invisible things can make a difference:

```
      
         
                  
     
<EOF>
```

The invisible seen here:

```
␣␣␣␣␣␣$
␣␣␣␣␣␣␣␣␣$
␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣$
␣␣␣␣␣$
<EOF>
```

Add Forensics to the Decision (© CPR / LX / GEXI)

Invisible things can make a difference:

```
      
         
                  
     
<EOF>
```

The invisible seen here:

```
␣␣␣␣␣␣$
␣␣␣␣␣␣␣␣␣$
␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣$
␣␣␣␣␣$
<EOF>
```

The translation is:

```
F
I
R
E
<EOF>
```

Long-term: Application Scenarios in Research and Education

– Application Scenarios –

Application scenarios, knowledge and computing support

The following case studies show simplified, practical application scenarios for

- separating essential knowledge
  (e.g., knowledge resources, structure)
- creating knowledge based components
  (e.g., Active Source)
- supporting increased decision potential
  (e.g., UDC classification)
- integrating high end resources
  (e.g., compute and storage)

Hardware Trace

View: Content and context

One view: *(classification)*

- Type: Poster,
- Format: Image,
- Content: Supercomputing system,
- Context: Type and size of resources,
- System: North-German Supercomputing Alliance (HLRN),
- Secondary information: PDF/image information (author, subject),
- Originary sources: long-term, LX Hardware Trace, created: 2008.

Another view: *(classification)*

- Content: Number of cores, compute nodes, disks, hardware architecture, massively parallel system, communication properties.
- Context: Supercomputing system,
- Usage and application: Geosciences, earth sciences, physics, ...
- System: HLRN-II, North-German Supercomputing Alliance (HLRN).

Integrated Systems

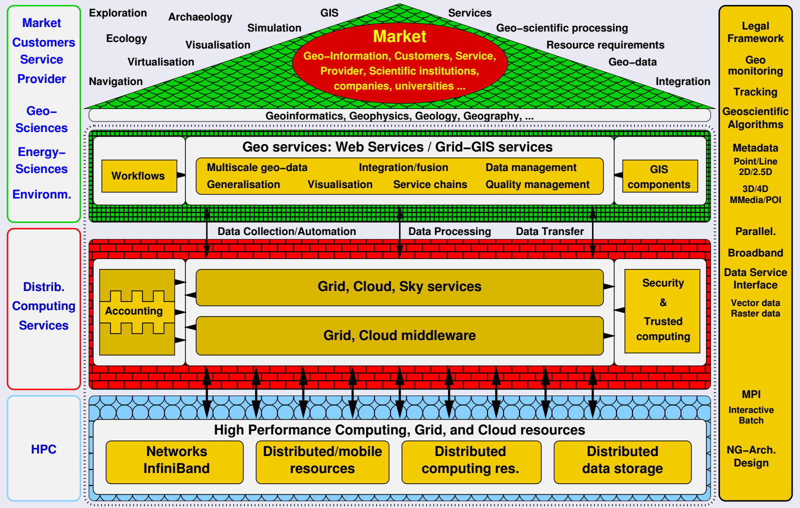
Frameworks supporting integration of:

- Information *(e.g., knowledge resources)*,
- Computation *(e.g., advanced scientific computing)*,
- Collaboration *(e.g., collaboration frameworks)*.

- Disciplines *(e.g., knowledge, collaboration, interfaces)*,
- Services *(e.g., policies, interfaces)*,
- Resources *(e.g., management, architecture, policies)*.

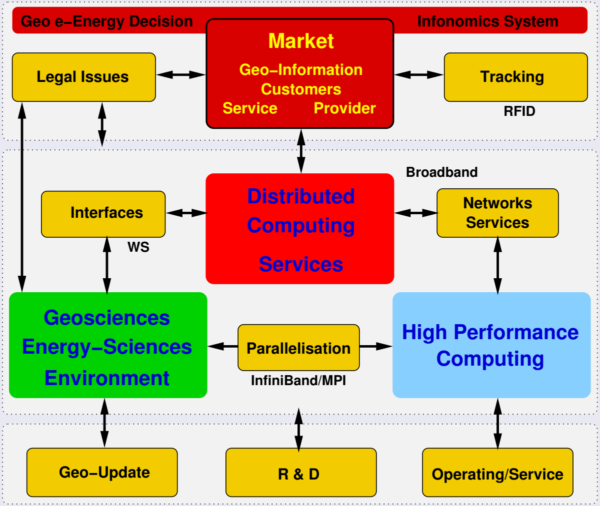
Information, computing, and collaboration

Collaboration house framework, integrating information, scientific computing © CPR



International Collaboration: Columns of the Infonomics System

Columns and Interactions © CPR / GEXI



Collaboration Case Studies

Case Studies (Geo Exploration and Information, GEXI)

- **Objects:** National and international cooperations and projects with participants from academia and industry.
- **Resources:** ZIVGrid, ZIVHPC, ZIVSMP, HLRN, D-Grid, ...
- **Frameworks:** GISIG (actmap-project), Grid-GIS framework, Actmap Computing Resources Interface (CRI),
- **Systems:** Information systems, processing and visualisation, dynamical distributed resource usage (introduced with Active Source) ...
- **Columns:** Geosciences and Disciplines, Distributed Computing and Services, High End Computing and HPC configuration,
- **Purpose:** Research, education, learning.

Dynamical Applications / Components and Configuration

Action Principle

- Components linked via Active Source,
- use of event bind ("Geoevents") calls, triggering batch calls,
- using computing resources,
- transferring results back local application,
- loading results into Active Source application.

Components Configured for HPC and Dynamical Visualisation (excerpt)

Component	**Software / Configuration**
Frameworks	GISIG, actmap, Grid-GIS Framework, Actmap Computing Resources Interface (Actmap CRI). Configuration for integration of data and applications, flexible transfer of data, secured execution of foreign Active Sources on demand, accounting as well as batch and interactive use of resources.
System	Linux / SLES, Storage, Filesystems, Lustre, Management Suites
Batch system	Moab, Torque
Networks	InfiniBand (MPI, I/O), Ethernet, Service networks
Message Passing	MPI, OpenMP, MPT
Transfer / interchange	Secure Shell / keys, `pdsh`
Security	Trusted Computing, Sandboxing, Tcl, Tcl Plugin
Policies	`home`, `javascript`, `trusted`

Batch System and Scheduling / Distributing Data

Automisation of Batch and Interactive Access

- Batch system, scheduling and resource management implemented on HLRN-II is based on Moab and Torque, PBS (Portable Batch System) resource specification language.
- Interactive use and calculation is depending on batch system features.
- Currently the end user application will have to do the job synchronisation. With a conventional system configuration the management of multi user operation is difficult.
- Synchronising and multi user operation work against interactive use.

Data Transfer and Communication

- Within event triggered jobs, MPI and batch means can be used for distributing and collecting data and job output. For distributing files automatically within the system e.g. `dsh`, `pdsh`, C3 tools, Secure Shell (`ssh` and `scp`) are used.
- Interactive communication is supported by the appropriate Secure Shell key configuration. It must be part of the system configuration to correctly employ authorisation keys and `crontab` or `at` features.

Accessing Computing Resources

Actmap Computing Resources Interface (CRI)

The Actmap CRI is an actmap library (`actlcri`) containing procedures for handling computing resources. It can hold functions and procedures and even platform specific parts in a portable way. It can be used by calling the source code library as well as the byte code library generated with a compiler like TclPro. With CRI being part of Active Source, parallel processing interfaces can be used, for example MPI (Message Passing Interface) and OpenMP using InfiniBand.

Active Source MPI (SGI MPT) Script / OpenMP Script

```
#!/bin/bash
#PBS -N myjob
#PBS -j oe
#PBS -l walltime=00:10:00
#PBS -l nodes=8:ppn=4
#PBS -l feature=ice
#PBS -l partition=hannover
#PBS -l naccesspolicy=singlejob
module load mpt
cd $PBS_O_WORKDIR
np=$(cat $PBS_NODEFILE | wc -l)
mpiexec_mpt -np $np ./dyna.out 2>&1
```

```
#!/bin/bash
#PBS -N myjob
#PBS -j oe
#PBS -A myproject
#PBS -l walltime=00:10:00
#PBS -l nodes=1:ppn=4
#PBS -l feature=xe
#PBS -l naccesspolicy=singlejob

cd $PBS_O_WORKDIR
export OMP_NUM_THREADS=4
./dyna.out 2>&1
```

Case Study / Interactive Components and Interfaces

Precalculation and Processing

Parallel data processing can be triggered from within an Active Map, e.g. processing of satellite data and images, as well as calculation and rendering of virtual reality scenes and raytracing. Precalculation of views can be automated from the application, processing several hundred views at a time using dedicated compute nodes for each calculation in order to create high level GIS views.

```
convert -scale 2400x1200 inview01.jpg outview01.jpg
convert -scale 2400x1200 inview02.jpg outview02.jpg
convert -scale 2400x1200 inview03.jpg outview03.jpg
...
```

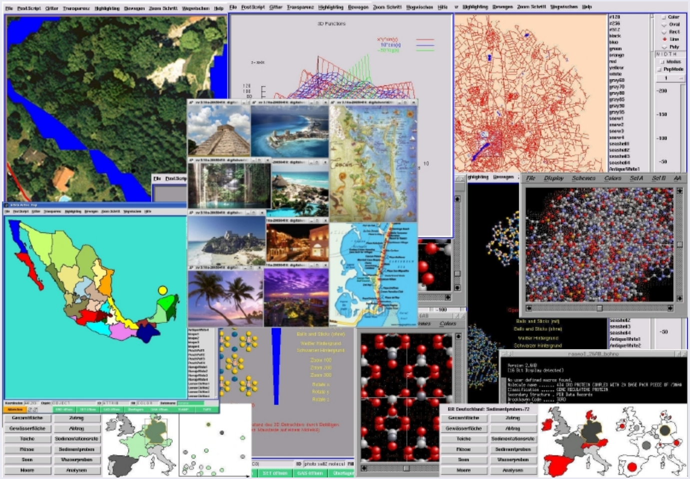
Binding of Precalculation Script

Event bindings can bind events to selective objects of a category. With Active Source it is possible to deliver any part of the application with support of distributed computing and storage resources, e.g. for simple cases via HTTP.

```
$w bind precalc_bio <Button-1> {exec precalc_bio.sh}
```

Introduction, application scenarios, trust

Scenarios / Data object subject to protection (GEXI case study) © CPR / GEXI



Data objects

- vector
- raster
- aerial
- photo
- spatial
- calculation
- measurement
- processing
- meta objects
- interactive
- commercial
- license
- ...

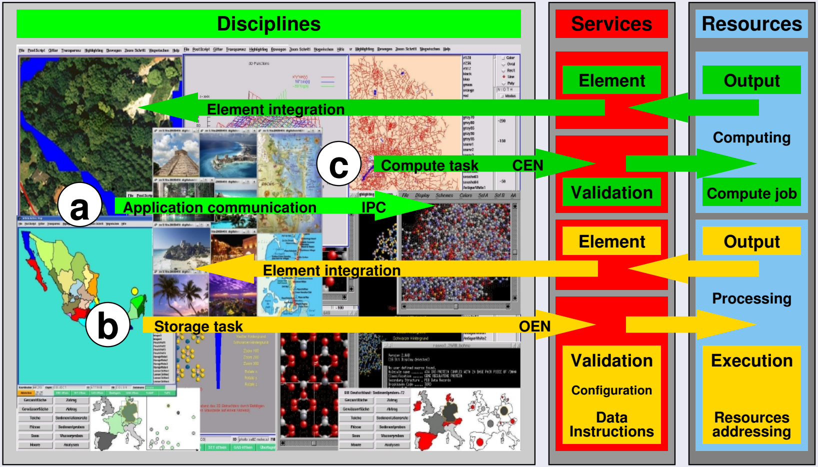
Drawbacks of existing common algorithms regarding integrated systems

Object handling based on existing concepts

- is not portable in between different file formats,
- does not respect meta-data of the information handled,
- does modify the original documents,
- is not intuitively extendable for information systems,
- and there is no flexible and open implementation available, and further on there are
- security issues associated with available products,
- the proprietary solution is not completely transparent,
- the XML has large overhead for huge object collections,
- huge transfer rates for large number of objects, and
- security issues with transfer actions to outer networks.

Implementation of Different Tasks for Advanced Scientific Computing

Integrated systems and resources for advanced scientific computing © CPR / GEXI



Envelope Components

Using the following concepts, we can, mostly for any system, implement:

- Application communication via IPC.
- Application triggering on events.
- Storage object requests based on envelopes.
- Compute requests based on envelopes.

Used for demonstration and studies with Integrated Systems:

- Active Source Information System components for
- Flexible implementation,
- Maximum transparency,
- Separate knowledge (Structure, UDC, CEN, OEN),
- Allowing OO-support (object, element) on application level,
- Multi-system support.

Application communication

Application communication with framework-internal and external applications (IPC)

```
catch {
 send {rasmol #1} "$what"
}
```

- Self-descriptive Tcl syntax.
- Inter-Process Communication send starting molecular graphics visualisation.
- Catching messages for further analysis by the components.

Application triggering and components

Application triggering, linking to application components

```
text 450.0 535.0 -tags {itemtext relictrotatex} -fill yellow -text "Rotate x" -justify center
...
$w bind relictrotatex <Button-1> {sendAllRasMol {rotate x 10}}
$w bind relictballsandsticks <Button-1> {sendAllRasMol {spacefill 100}}
$w bind relictwhitebg <Button-1> {sendAllRasMol {set background white}}
$w bind relictzoom100 <Button-1> {sendAllRasMol {zoom 100}}
```

Storage object requests

Generic Object Envelopes (OEN) © CPR / GEXI

```
<ObjectEnvelope><!-- ObjectEnvelope (OEN)-->
<Object>
<Filename>GIS_Case_Study_20090804.jpg</Filename>
<Md5sum>...</Md5sum>
<Sha1sum>...</Sha1sum>
<DateCreated>2010-08-01:221114</DateCreated>
<DateModified>2010-08-01:222029</DateModified>
<ID>...</ID><CertificateID>...</CertificateID>
<Signature>...</Signature>
<Content><ContentData>...</ContentData></Content>
</Object>
</ObjectEnvelope>
```

- OEN containing element structures, handling and embedding data / information.
- End-user public client application, implementation via browser plugin / services.
- Instructions embedded in envelopes, content-stream and content-reference.
- Respect any meta-data for objects, handle different object formats, staying transparent, portable, keep original documents unmodified, supports signed object elements and PKI, usable with sources and binaries like Active Source.

Compute requests

Generic Compute Envelope (CEN) © CPR / GEXI

```
<ComputeEnvelope><!-- ComputeEnvelope (CEN)-->
<Instruction><Filename>Processing_Batch_GIS612.pbs</Filename>
<Sha512sum>...</Sha512sum>
<DateCreated>2010-08-01:201057</DateCreated>
<DateModified>2010-08-01:211804</DateModified>
<CertificateID>...</CertificateID><Signature>...</Signature>
<Content><DataReference>https://doi...</DataReference></Content>
<Script><Pbs><Shell>#!/bin/bash</Shell>
<JobName>#PBS -N myjob</JobName>
<Oe>#PBS -j oe</Oe>
<Walltime>#PBS -l walltime=00:10:00</Walltime>
<NodesPpn>#PBS -l nodes=8:ppn=4</NodesPpn>
<Feature>#PBS -l feature=ice</Feature>
<Partition>#PBS -l partition=hannover</Partition>
<Accesspolicy>#PBS -l naccesspolicy=singlejob</Accesspolicy>
<Module>module load mpt</Module>
<Cd>cd $PBS_O_WORKDIR</Cd>
<Np>np=$(cat $PBS_NODEFILE | wc -l)</Np>
<Exec>mpiexec_mpt -np $np ./dyna.out 2>&1</Exec>
</Pbs></Script></Instruction></ComputeEnvelope>
```

- Compute requests for resources handled via CEN interfaces, self-descriptive, environment preconfigured, references parallel processed on various architectures.

Trust Case: Requirements for trust in information

Subject to handling and protection with digital signatures

- Allow object authors to set up a secure signing environment.
- Allow the consumer of the data object to validate the object concerning integrity and authentication of the signer.

Trust Case: Object Envelope

Object Envelope (OEN)

```
<ObjectEnvelope><!-- ObjectEnvelope (OEN)-->
<Object>
<Filename>GIS_Case_Study_20090804.jpg</Filename>
<Md5sum>...</Md5sum>
<Sha1sum>...</Sha1sum>
<DateCreated>2010-08-01:221114</DateCreated>
<DateModified>2010-08-01:222029</DateModified>
<ID>...</ID><CertificateID>...</CertificateID>
<Signature>...</Signature>
<Content><ContentData>...</ContentData></Content>
</Object>
</ObjectEnvelope>
```

OEN referencing signed data

```
... <Content><ContentDataReference>https://doi...</
 ContentReference></Content> ...
```

Trust Case: Envelope Benefits

Object Envelopes

- Benefit of *content-reference* with high performant distributed or multicore resources: references can be processed in parallel on these architectures.
- More flexible than sole XML signature standard (RFC 2807).
- Matching to the situation, scalable, transparent, open, portable, using general modular components.
- For qualified requests signatures/signature groups can be verified. For non-qualified requests signatures can be ignored.
- All OEN can be embedded into existing information and computing system components.
- Tools and algorithms for content or meta data can be handled very flexible, supporting encryption, check sums, integrity, authentication, reliability, confidentiality, and authorisation.

Trust Case: OEN embedded

OEN embedded with GISIG Active Source

```
proc create_country_mexico {} {
global w
$w create polygon 0.938583i 0.354331i 2.055118i ...
#BCMT-----------------------------------------------------------
###EN \gisigsnip{Object Data: Object Envelope (OEN)}
#ECMT-----------------------------------------------------------
#BOEN <ObjectEnvelope>
##OEN <Object>
##OEN <Filename>mexico_site_name_tulum_temple.jpg</Filename>
##OEN <Md5sum>251b443901d87a28f83f8026a1ac9191 *mexico_site_name_tulum_temple.jpg</Md5sum>
##OEN <Sha1sum>f0eb9d21cfe2c9855c033be5c8ad77710356c1eb *mexico_site_name_tulum_temple.jpg</Sha1sum>
##OEN <DateCreated>2010-08-01:221114</DateCreated>
##OEN <DateModified>2010-08-01:222029</DateModified>
##OEN <ID>...</ID><CertificateID>...</CertificateID>
##OEN <Signature>...</Signature>
##OEN <Content><ContentDataReference>http://.../ mexico_site_name_tulum_temple.jpg</ContentReference></Content>
##OEN </Object>
#EOEN </ObjectEnvelope>
... proc create_country_mexico_autoevents {} { ...
```

Introduction, elements and amazements, trust

Elements for data objects being subject to handling and protection

- Vector data and multi-dimensional data.
- Raster data (aerial, remote sensing, and photographic).
- Primary and secondary spatial information.
- Calculation, measurement, and processing results.
- Meta data, instruction and interactive information.
- Commercially provided or licensed data, . . .

Amazements

Most problems arise from

- complexity necessary to reflect the use cases and
- being built on prepackaged components each having own practical 'amazements' for integrated development
- and from content and context handling.

Trust in computing and trust in information

Trust in computing

- Currently "trust in computing" can cover the content aspects.
- Context aspects are out of scope with todays systems. For the three development layers this mainly states tasks for services and resources layers.

Trust in information

- Secure signing environment for object authors.
- Validation of objects for the consumer of the data object, concerning integrity and authentication of the signer.

Content and context

What can be controlled and/or signed

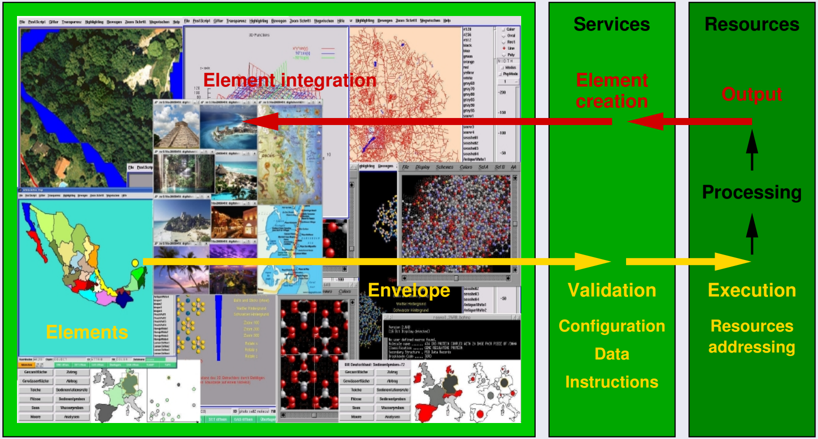
- Information and computing resources instructions.
- Links between the information and computing system.
- Prerequisites of the computing system.
- Processing directives and script elements.
- Input / output data necessary.
- ...

What cannot be fully validated

- Environment and network specifications.
- Nodes characteristics.
- State of the components of the system.
- ...

Workflow

Workflow with application scenarios from the GEXI case studies © CPR / GEXI

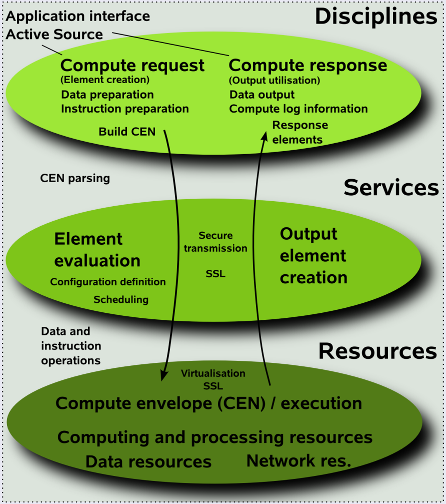


Compute access

Compute access for integrated systems © CPR / GEXI

Workflow Process:

Request



Workflow Process:

Response

Workflow: Request

The request process

- **Disciplines layer:** The user's compute request is started based on computation data and the instruction information necessary for processing. Elements for the CEN are created by the user. The CEN is built from the elements, supported by application functions.
- **Services layer:** The elements are evaluated and adapted by the system configuration definition. The instruction sets are prepared for scheduling.
- **Resources layer:** Data and instruction operations are handled by batch system or interactive use. Compute, data, network, and storage resources are used with elements and configuration by services layer definition.

Workflow: Response

The response process

- **Resources layer:** The resulting output will be handled as described in the CEN instructions. Very large data can be stored on appropriate storage resources for later use, smaller or interactive data can be directly delivered to the services layer.
- **Services layer:** Services functions handle the output and do create output elements, delivered to the user or interface defined in the original CEN envelope.
- **Disciplines layer:** The data from the output elements is delivered for utilisation to the user or interface, e.g. to be interactively integrated into the application.

Compute Envelope

Example CEN: Generic Compute Envelope data

```
<ComputeEnvelope><!-- ComputeEnvelope (CEN)-->
<Instruction>
<Filename>Processing_Bat_GIS515.torque</Filename>
<Md5sum>...</Md5sum><Sha512sum>...</Sha512sum>
<DateCreated>2010-08-01:231523</DateCreated>
<DateModified>2010-08-01:232734</DateModified>
<ID>...</ID><CertificateID>...</CertificateID>
<Signature>...</Signature><Content>...</Content>
</Instruction>
</ComputeEnvelope>
```

Embedded DataReference

```
...<Content><DataReference>https://doi...</DataReference><
 /Content>...
```

Integrated components in practice

Active Source and embedded CEN

```
proc create_country_mexico {} {
global w
#  Sonora
$w create polygon 0.938583i 0.354331i 2.055118i ...
#BCMT---------------------------------------------------
###EN \gisigsnip{Compute Data: Compute Envelope (CEN)}
#ECMT---------------------------------------------------
#BCEN <ComputeEnvelope>
##CEN <Instruction>
##CEN <Filename>Processing_Bat_GIS515.torque</Filename>
##CEN <Md5sum>...</Md5sum>
##CEN <Sha1sum>...</Sha1sum>
##CEN <Sha512sum>...</Sha512sum>
##CEN <DateCreated>2010-09-12:230012</DateCreated>
##CEN <DateModified>2010-09-12:235052</DateModified>
##CEN <ID>...</ID><CertificateID>...</CertificateID>
##CEN <Signature>...</Signature>
##CEN <Content>...</Content>
##CEN </Instruction>
#ECEN </ComputeEnvelope>
... proc create_country_mexico_autoevents {} { ...
```

Trust Case: Solution for use with integrated systems

Benefits and future objectives

- Needed: not only a signature standard and an envelope technology
- More: a generic extensible concept for information and computing system components.

Benefits for complex information and computing systems

- No overhead, minimising communication.
- Transparent handling.
- No proprietary algorithms.

Future objectives, combined with client components

- Channels for limiting communication traffic.
- Qualified signature services and accounting.
- Using signed objects without verification.
- Verify signed objects on demand.

Trust case: Evaluation

Primary benefits of OEN with signed objects. The algorithm is

- portable in between different object and file formats.
- It respects meta-data for the objects.
- Original documents can stay unmodified.
- The solution is most transparent, extendable, flexible, and scalable, for security aspects and modularisation.
- Guaranteed data integrity and authentication derived from the cryptographic strength of current asymmetric algorithms and digital signature processes.
- Flexible meta data association for any object and data type, including check sums and time stamps.

Main drawbacks

- Requirements for use outside the case studies: Interoperability between multiple PKIs, a global cryptosystem (Global PKI), special PKI-enabled clients to generate, store and manage certificates and associated data is not already implemented.
- Risks: Lost, destroyed, or compromised private keys and loss of primary verification for keyed object data.
- Inconveniences: Authors have to register at a CA and request digital certificates.

Trust Case: Envelope summary

Summary

- Security and verification of information content is an essential part of the challenge to build future integrated information and computing systems.
- Object Envelope techniques can help to establish a flexible and portable way for using content data.
- With implementation and legal issues, the security aspect are on the rise for any complex system.
- Even though PKI technology offers means to attest, identify, manage the exchange of encryption keys and secure transmission between parties, there has not been broad-based adoption of PKI technology by public and private organisation.
- A significant number of countries recognise digital signatures as legally binding. In case of security enhanced integrated information and computing system components object signing provides a robust solution to facilitate "trust in information" and to overall support "trust in computing". In order to put this implementation into international public practice there is a need for future PKI development and deployment offering a global public key cryptosystem for the Future Internet. This work showed that it is possible to bring complex information and computing systems to life, being able to create interfaces that can also be interfaces between the logical columns and interest groups.

Privacy Case: Archaeology Information Systems and Tourism

Subject/object privacy: Protection of archaeological sites © CPR / GEXI

Privacy Case: Archaeology Information Systems and Tourism

Subject/object privacy: Protection of archaeological sites © CPR / GEXI

Privacy Case: Archaeology Information Systems and Tourism

Subject/object privacy: Protection of archaeological sites © CPR / GEXI

Protect non-public location and existence-information.
Problem: Subject-related. Prevent lootings and illegal digging.
Economy: Promote education & individual tourism.

Privacy Case: Medical Data Information Systems

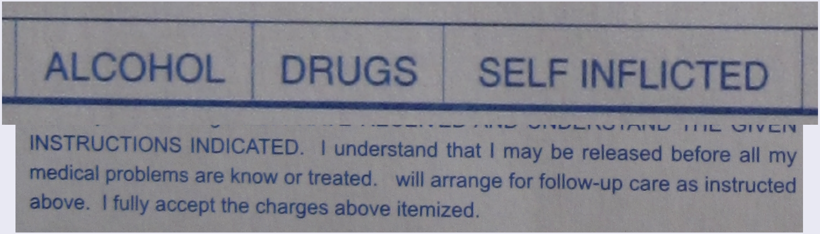
Individual privacy and anonymity: Protection of individual information © CPR / GEXI

Related to:	ALCOHOL	DRUGS	SELF INFLICTED	NO

I hereby acknowledge that I HAVE RECEIVED AND UNDERSTAND THE GIVEN INSTRUCTIONS INDICATED. I understand that I may be released before all my medical problems are know or treated. will arrange for follow-up care as instructed above. I fully accept the charges above itemized.

Privacy Case: Medical Data Information Systems

Individual privacy and anonymity: Protection of individual information © CPR / GEXI

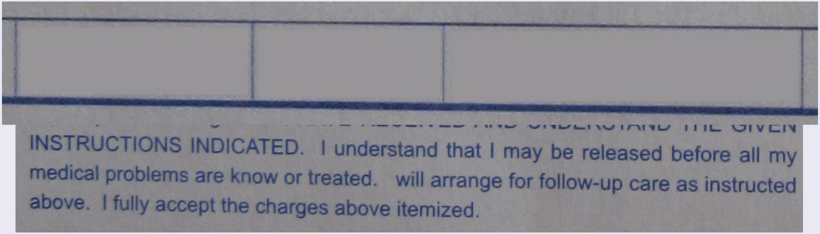


ALCOHOL	DRUGS	SELF INFLICTED

INSTRUCTIONS INDICATED. I understand that I may be released before all my medical problems are know or treated. will arrange for follow-up care as instructed above. I fully accept the charges above itemized.

Privacy Case: Medical Data Information Systems

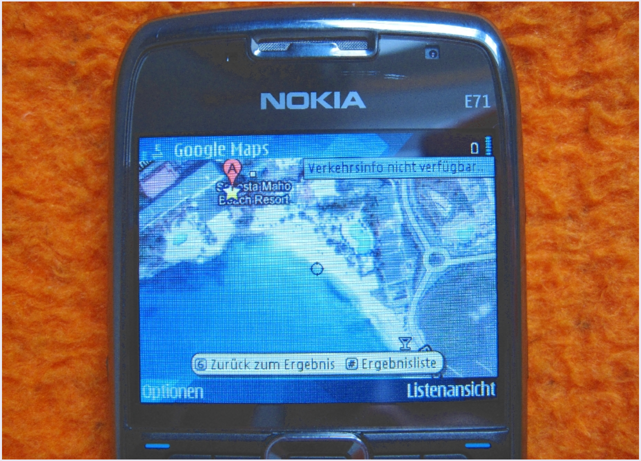
Individual privacy and anonymity: Protection of individual information © CPR / GEXI


INSTRUCTIONS INDICATED. I understand that I may be released before all my medical problems are know or treated. will arrange for follow-up care as instructed above. I fully accept the charges above itemized.

Protect individual information/categorisation.
Problem: Prevent misuse, data collection, data trade, . . .
Problem: Prevent digitalisation side effects.
Economy: Enable medical support, epidemology IS.

Privacy Case: Navigation Systems

Individual privacy and anonymity: Protect individual activities, habits, . . . © CPR / GEXI



Protect individual movement profiles.

Privacy Case: Distributed Computing Systems / High End Computing

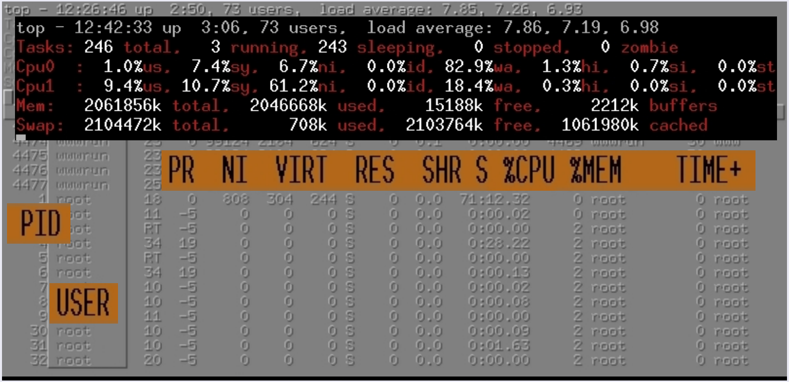
Privacy and anonymity: Real system base security and protection © CPR / GEXI

```
top - 12:26:46 up  2:50, 73 users,  load average: 7.85, 7.26, 6.93
Tasks: 247 total,   2 running, 245 sleeping,   0 stopped,   0 zombie
Cpu0  :  0.7%us,  9.3%sy,  0.0%ni,  0.3%id, 88.7%wa,  0.0%hi,  1.0%si,  0.0%st
Cpu1  :  2.3%us,  1.7%sy,  0.0%ni, 21.2%id, 74.8%wa,  0.0%hi,  0.0%si,  0.0%st
Mem:   2061856k total,  2045280k used,    16576k free,     3016k buffers
Swap:  2104472k total,      668k used,  2103804k free,  1068024k cached

  PID USER      PR  NI  VIRT  RES  SHR S %CPU %MEM    TIME+  PPID RUSER      UID GROUP
 4473 wwwrun    23   0 99124 2184  624 S    0  0.1   0:00.00  4469 wwwrun      30 www
 4474 wwwrun    23   0 99124 2184  624 S    0  0.1   0:00.00  4469 wwwrun      30 www
 4475 wwwrun    23   0 99124 2184  624 S    0  0.1   0:00.00  4469 wwwrun      30 www
 4476 wwwrun    23   0 99124 2184  624 S    0  0.1   0:00.00  4469 wwwrun      30 www
 4477 wwwrun    25   0 99124 2184  624 S    0  0.1   0:00.00  4469 wwwrun      30 www
    1 root      18   0   808  304  244 S    0  0.0  71:12.32     0 root         0 root
    2 root      11  -5     0    0    0 S    0  0.0   0:00.02     0 root         0 root
    3 root      RT  -5     0    0    0 S    0  0.0   0:00.00     2 root         0 root
    4 root      34  19     0    0    0 S    0  0.0   0:28.22     2 root         0 root
    5 root      RT  -5     0    0    0 S    0  0.0   0:00.00     2 root         0 root
    6 root      34  19     0    0    0 S    0  0.0   0:00.13     2 root         0 root
    7 root      10  -5     0    0    0 S    0  0.0   0:00.02     2 root         0 root
    8 root      10  -5     0    0    0 S    0  0.0   0:00.08     2 root         0 root
    9 root      11  -5     0    0    0 S    0  0.0   0:00.00     2 root         0 root
   30 root      10  -5     0    0    0 S    0  0.0   0:00.09     2 root         0 root
   31 root      10  -5     0    0    0 S    0  0.0   0:01.63     2 root         0 root
   32 root      20  -5     0    0    0 S    0  0.0   0:00.00     2 root         0 root
```

Privacy Case: Distributed Computing Systems / High End Computing

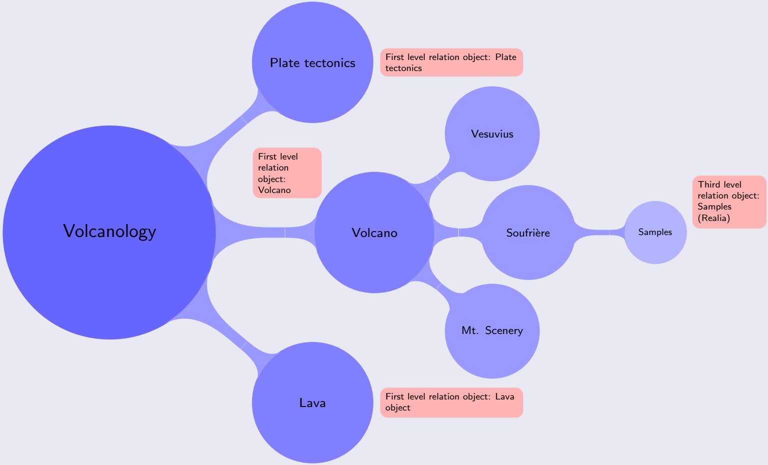
Privacy and anonymity: Real system base security and protection © CPR / GEXI



Protect individual properties.
Problem: Ensure privacy for investments and data.
Economy: HW and SW support, separating data, process load . . .

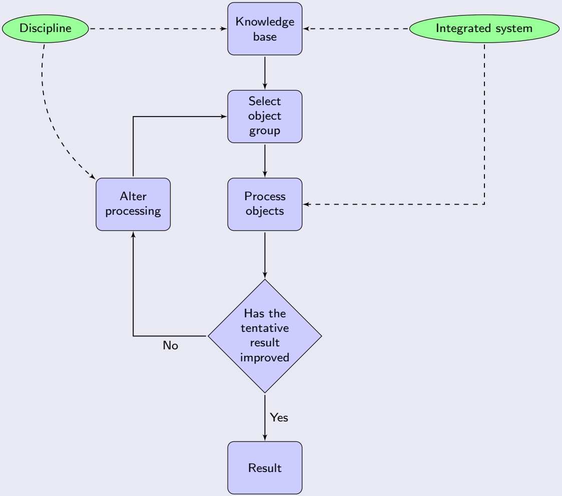
Knowledge Resources: Objects and Relations, Classification

Example: Search objects and realia references in volcanology dimension © CPR / GEXI



Knowledge Resources: Workflows

Example: Simple workflow with knowledge base © CPR / GEXI



Information and structure

State of the art, goals

- Integrating Information Systems and supercomputing resources is not trivial,
- There is currently no general solution available,
- Base for any reasonable results in this area is strong multi-disciplinary work,
- Collaboration framework,
- Dynamical access of suitable supercomputing resources,
- Advanced scientific computing resources and facilities,
- Enabling supercomputing support for scientific information systems
- Long-term classification, internationalisation,
- Goal to overcome many complex scientific impediments in prominent disciplines, requirements for mighty information systems,
- Studies show need for implementations of Integrated Information and Computing Systems (IICS),
- Multi-disciplinary context, advanced cognition,
- Interactive use shows up needing capabilities for dynamical computing.

Phonetic algorithms (multi-disciplinary, geosciences) (© CPR / LX / GEXI)

LX Soundex code (SNDX-standard) for La Soufrière volcano © CPR / LX / GEXI

```
L216:La_Soufriere
L216:La_Soufri{`e}re
L216:La_Soufrière
```

LX Soundex code (SNDX-standard) for Vesuvius volcano and comparables

```
V210:Vesuv
V210:Vesuvio
V212:Vesuvius
```

Phonetic algorithms (multi-disciplinary, archaeology

LX Soundex code (SNDX-standard) for Yucatán and comparables

```
Y235:Yucatan
Y235:Yucat'an
Y235:Yucatán
```

LX Soundex code (SNDX-standard) for Chichén Itzá and comparables

```
C250:Chichén
C253:Chich'en_Itz'a
C253:Chichen_Itza
C253:Chichén_Itzá
```

LX Soundex code (SNDX-standard) for Cobá and comparables

```
C100:Coba
C100:Cob'a
C100:Cobá
```

Individualised algorithms and objects

LX Soundex code (SNDX-latin) for 'Leibniz'-homophones (excerpt)

```
SNDX-latin:L152:Laipunitsu
SNDX-latin:L152:Lajbnic
SNDX-latin:L152:Leibnics
SNDX-latin:L152:Leibnitio
SNDX-latin:L152:Leibnitius
SNDX-latin:L152:Leibnits
SNDX-latin:L152:Leibnitz
SNDX-latin:L152:Leibnitzius
SNDX-latin:L152:Leibniz
SNDX-latin:L152:Leibnizius
SNDX-latin:L152:Leibnütz
SNDX-latin:L152:Leibnüz
SNDX-latin:L152:Leibnuzius
SNDX-latin:L152:Leibnüzius
SNDX-latin:L152:Lejbnic
```

The individualised algorithm has harmonised the L152, L153, L215 codes in homophonic pseudonym parts for L152. Objects can carry any references to these algorithms.

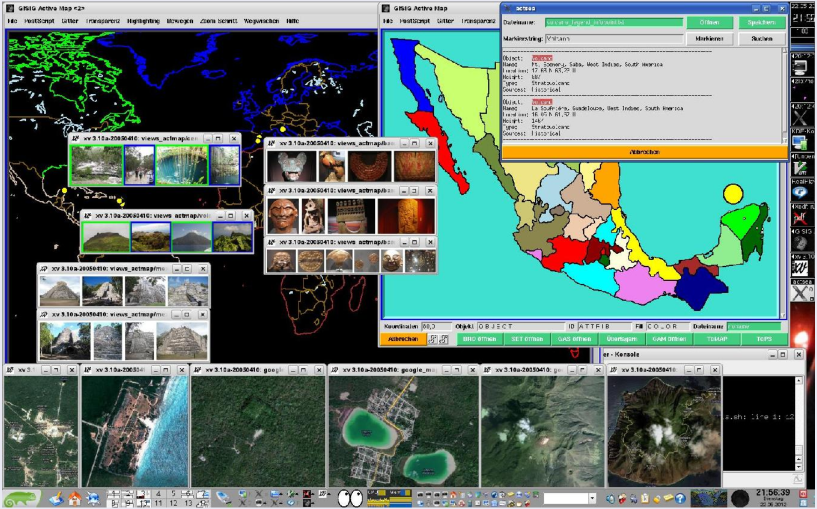
Keyword context

Keyword context data from a 'Leibniz'-object (excerpt): 'terra motus'-key

```
keyword-Context: KYW :: Leibniz, Korrespondent, Tschirnhaus
keyword-Context: TXT :: Venedig, Neapolis, Puzzolo, Grotta del Cane
keyword-Context: TXT :: Neapolis, welches nach Rom und Venedig eine der schönsten städten Italiae ist
keyword-Context: TXT :: schwöfel bäder, schweffel
keyword-Context: KYW :: Schwefel, Solfatara, Fumarole
keyword-Context: TXT :: Neapolis, den brennenden Berg Vesuvium
keyword-Context: TXT :: Grotta del Cane
keyword-Context: TXT :: Neapolis, den brennenden Berg Vesuvium
keyword-Context: KYW DE :: Vulkanismus, Vulkanologie, Vesuv, Vesuvius, Vesuvium, Erdbeben, Beben
keyword-Context: KYW EN :: volcanism, volcanology, Vesuvius, Vesuvium, earthquake, quake
...
link-Context: LNK :: http://www.gwlb.de/Leibniz/Leibnizarchiv/Veroeffentlichungen/III7B.pdf
keyword-Context: TXT :: terrae motu, Sicilien
keyword-Context: KYW :: Erdbewegungen, Erdbeben, Vulkane, terrae motu, terra motus, Sicilien, Sizilien
...
link-Context: LNK :: http://echo.mpiwg-berlin.mpg.de
keyword-Context: KYW DE :: Nicolaus Seelaender, Nicolaus Seeländer, Kupferplatten, Leibniz, Leibniz Einhorn, Einhornhöhle b. Scharzfeld im Harz
...
link-Context: LNK :: http://194.95.154.13/CiXbase/gwlbhss/
keyword-Context: TXT :: 1631/1632 16xx, terra motus, fogelius
keyword-Context: KYW DE :: Erdbeben, Seismologie, Seismik, Fogel, Fogelius, Vulkan, Vesuvius, CiXbase, cixbase
keyword-Context: KYW EN :: earthquake, seismology, seismics, Fogel, Fogelius, volcano, Vesuvius, CiXbase, cixbase
```

Integrated Information and Computing System (IICS)

Dynamical use of information systems and scientific computing (© CPR / LX / GEXI)



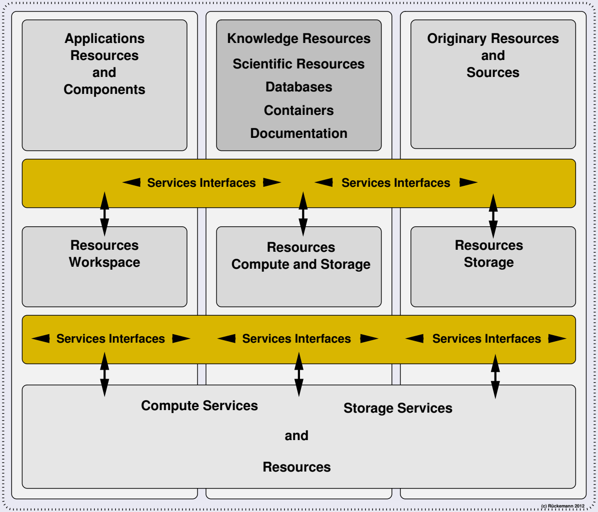
Sustainability and vitality

Situation and Motivation

- Data mining is not only an analysis step of knowledge discovery in databases based on informatics but much more general in data pools.
- It is an inter-disciplinary as well as multi-disciplinary field.
- It means discovering patterns in data pools implementing methods for statistics, classification, artificial intelligence, learning ... based on knowledge resources.
- The process targets to extract information from knowledge resources and gaining content and context, e.g., based on structure and references, in order to prepare for further use.
- Sustainable long-term strategies have to combine operation, services, and especially the knowledge resources.
- The available systems components integrate Resources Oriented Architectures (ROA), Services Oriented Architectures (SOA), and Knowledge Oriented Architectures (KOA).
- For long-term operation all three must be obtained from the creation and operation. Considering the entirety necessary for a successful long-term change management with future information technology structures. KOA is the most important complement as it contains highest percentage of the overall investments for the results and data that may even not be reproducable later on.

Architecture for documentation and development

Architecture: Columns of practical dimensions – economical long-term strategy



Three main columns:
Applications resources, Knowledge resources, Originary resources.

Workflows and algorithms

Components

- The **knowledge resources block** is the central resource in the long-term strategy.
- The knowledge resources can contain **any kind of content**. Application components can be migrated into the knowledge resources for documentation purposes and re-use.
- The **services** can access archived and historical data as well as live data and feed it into the workflows.
- Services interfaces allow to build **complex workflows** using arbitrary algorithms. The knowledge resources can be accessed from applications, which will extract suitable information and trigger the use of compute and storage resources. Objects can be selected by any algorithm, e.g., combinatory, search, and filter algorithms. The results can be delivered to a defined location or service.

Universal classification practically used

Spatial features and place

UDC Code	Description
UDC:(1)	Place and space in general. Localization. Orientation
UDC:(1-0/-9)	Special auxiliary subdivision for boundaries and spatial ...
UDC:(1-0)	Zones
UDC:(1-1)	Orientation. Points of the compass. Relative position
UDC:(1-19)	Relative location, direction and orientation
UDC:(1-2)	Lowest administrative units. Localities
UDC:(1-5)	Dependent or semi-dependent territories
UDC:(1-6)	States or groupings of states from various points of view
UDC:(1-7)	Places and areas according to privacy, publicness ...
UDC:(1-8)	Location. Source. Transit. Destination
UDC:(1-9)	Regionalization according to specialized points of view
UDC:(2)	Physiographic designation
UDC:(20)	Ecosphere
UDC:(21)	Surface of the Earth in general. Land areas in particular. ...
UDC:(23)	Above sea level. Surface relief. Above ground generally. ...
UDC:(24)	Below sea level. Underground. Subterranean
UDC:(25)	Natural flat ground (at, above or below sea level). ...
UDC:(26)	Oceans, seas and interconnections
UDC:(28)	Inland waters
UDC:(29)	The world according to physiographic features
UDC:(3)	Places of the ancient and mediaeval world
UDC:(32)	Ancient Egypt
UDC:(36)	Regions of the so-called barbarians
UDC:(37)	Italia. Ancient Rome and Italy
UDC:(38)	Ancient Greece
UDC:(4/9)	Countries and places of the modern world

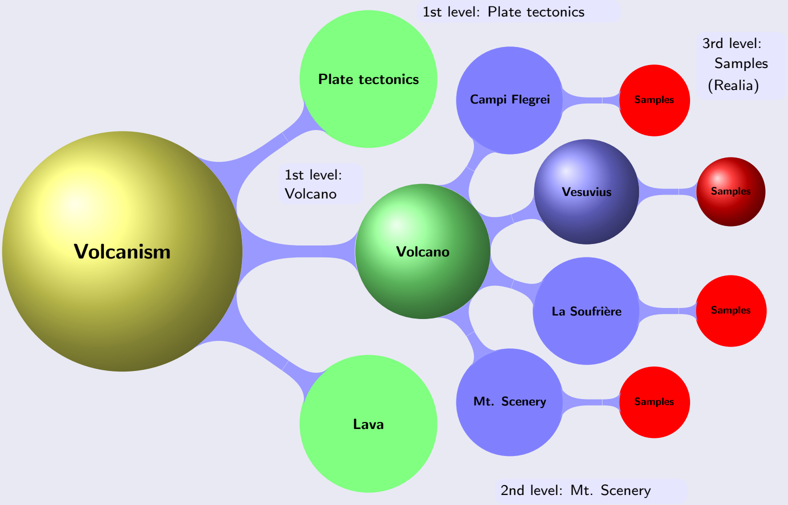
Universal classification practically used

Time

UDC Code	Description
UDC:"0/2"	Dates and ranges of time (CE or AD) ...
UDC:"3/7"	Time divisions other than dates in Christian ...
UDC:"3"	Conventional time divisions and subdivisions ...
UDC:"4"	Duration. Time-span. Period. Term. Ages ...
UDC:"5"	Periodicity. Frequency. Recurrence at ...
UDC:"6"	Geological, archaeological and cultural time divisions
UDC:"61/62"	Geological time division
UDC:"63"	Archaeological, prehistoric, protohistoric periods ...
UDC:"67/69"	Time reckonings: universal, secular, non-Christian ...
UDC:"67"	Universal time reckoning. Before Present
UDC:"68"	Secular time reckonings other than universal and ...
UDC:"69"	Dates and time units in non-Christian ...
UDC:"7"	Phenomena in time. Phenomenology of time
UDC:551.7+"61"	Cryptozoic aeon. Precambrian. 600+ MYBP ...
UDC:551.7+"616"	Archaean. Ur-gneiss formation. Ur-schiefer formation
UDC:551.7+"628"	Tertiary. 70-1 MYBP
UDC:551.7+"628.22"	Palaeocene
UDC:551.7+"628.24"	Eocene
UDC:551.7+"628.26"	Oligocene
UDC:551.7+"628.4"	Neogene
UDC:551.7+"628.42"	Miocene
UDC:551.7+"628.44"	Pliocene
UDC:551.7+"628.6"	Quaternary. 1 MYBP - Present
UDC:551.7+"628.62"	Pleistocene in general. Diluvium
UDC:551.7+"628.64"	Holocene. Postglacial in general

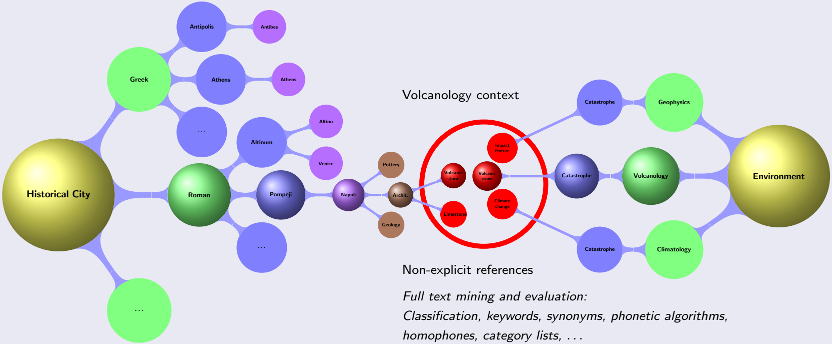
Knowledge Resources Object Carousels

Terrestrial "volcanism" context, computed volcano references / object relations



Computing object carousel connections

Historical city and environment object carousels, trees with computed references

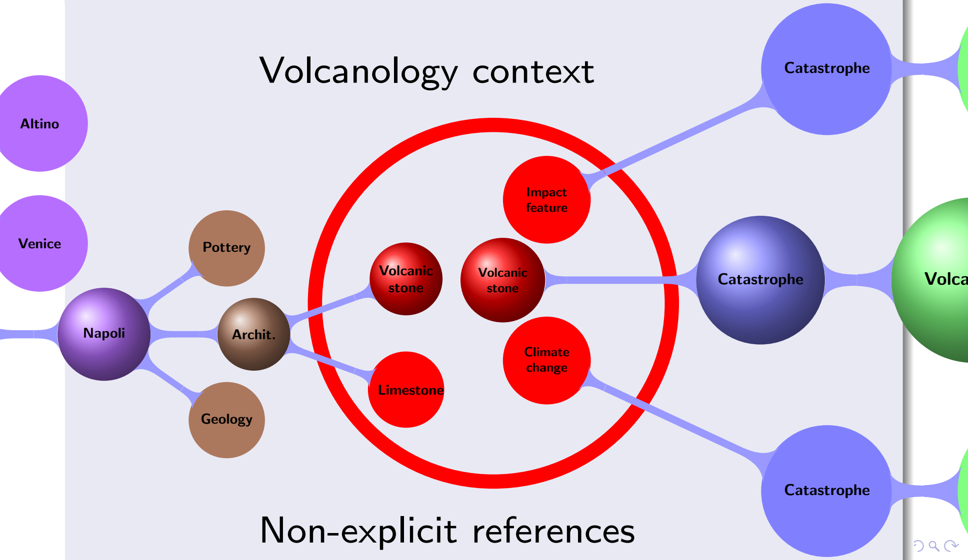


Full text mining and evaluation:
Classification, keywords, synonyms, phonetic algorithms, homophones, category lists, . . .

Carousel links, calculated via non-explicit references of comparable objects (red) from knowledge resources within trees. Starting topics are identified by large golden bullets. The two fitting lines within the object carousels are `Historical City : Roman : Pompeji : Napoli : Architecture : Volcanic stone` and `Environment : Volcanology : Catastrophe : Volcanic stone`. Fitting object term for historical city and environment is `Volcanic stone`. Excerpt of associated multi-disciplinary branch level objects: `Limestone`, `Impact feature`, `Climate change`.

Computing object carousel connections

Historical city and environment object carousels, trees with computed references

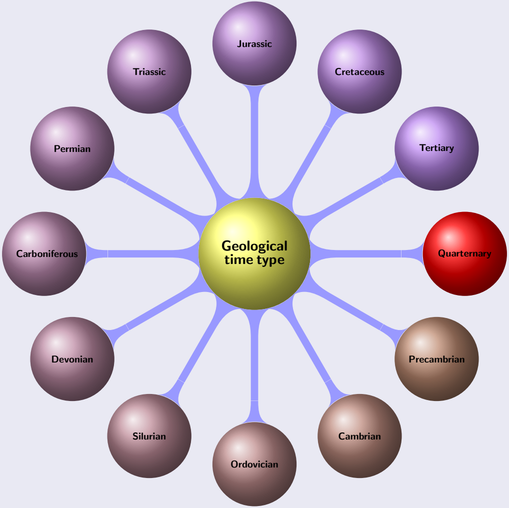


Knowledge Resources Object Carousels on Objects

Volcano and type references computed for terrestrial volcanism and references

Knowledge Resources Object Carousels on Objects

Geological spaces of time computed for terrestrial volcanoes, impacts, ...



Views, alternatives, and replacements

Replacement definition for relevant terms (LX resources)

```
Cretacious :: Cretaceous
Kreide :: Cretaceous
Trias :: Triassic
Carbon :: Carboniferous
Karbon :: Carboniferous
Silurium :: Silurian
Silur :: Silurian
Ordovicium :: Ordovician
Ordovizium :: Ordovician
Cambrium :: Cambrian
Kambrium :: Cambrian
Precambrium :: Precambrian
Präkambrium :: Precambrian
```

The example lists an excerpt of relevant terms and types of notation that can be considered equal for the target context.

Processing media citation references with UDC classified knowledge objects

Media citation set excerpt used with object "Vesuvius" (LX resources)

```
cite: YES 20070000 {LXK:Pompeii; Vesuvius; reconstruction; 3D; animation; Holocene} {UDC:...} {PAGE:----..----} LXCITE://Bonaventura:2007:My_DVD
cite: YES 20130000 {LXK:Pompeii; Vesuvius; Vesuvio; Holocene; postcard} {UDC:...} {PAGE:----..----} LXCITE://Guardasole:2013:Vesuvio_1270m
cite: YES 20070000 {LXK:Pompeii; Vesuvius; reconstruction; diorama} {UDC:...} {PAGE:----..----} LXCITE://Bonaventura:2007:Pompeii
cite: YES 20070000 {LXK:Pompeii; Vesuvius; bakery; mill stones; material; stone; volcanic lava; basalt; Holocene; diorama} {UDC:...} {PAGE:--56..--59} LXCITE://Bonaventura:2007:Pompeii
```

The examples are part of the "Vesuvius" and "volcanic mill stone" object references. The media citations refer to 3D video animations and dioramic reconstructions as well as even to postcards. These references resolve to full publications and data.

Classifications, editions, and objects

Development of “Tertiary” classification with UDC editions.

UDC Code (a)	UDC Code (b)	Description
UDC:"623"	UDC:"628"	Tertiary (70-1 MYBP)
UDC:"623.1"	UDC:"628.2"	Palaeogene (70-25 MYBP)
UDC:"623.5"	UDC:"628.4"	Neocene (25-1 MYBP)
UDC:551.77	UDC:551.7+"628"	Cenozoic (Cainozoic). Neozoic
UDC:551.78	UDC:551.7+"628"	Tertiary. 70-1 MYBP
UDC:551.781	UDC:551.7+"628.2"	Palaeogenic. Nummulitic
UDC:551.781.3	UDC:551.7+"628.22"	Palaeocene
UDC:551.781.4	UDC:551.7+"628.24"	Eocene
UDC:551.781.5	UDC:551.7+"628.26"	Oligocene
UDC:551.782	UDC:551.7+"628.4"	Neogene
UDC:551.782.1	UDC:551.7+"628.42"	Miocene
UDC:551.782.2	UDC:551.7+"628.44"	Pliocene

Target moved from (a) to (b) and
was also adapted to a new subgrouping (lower block).

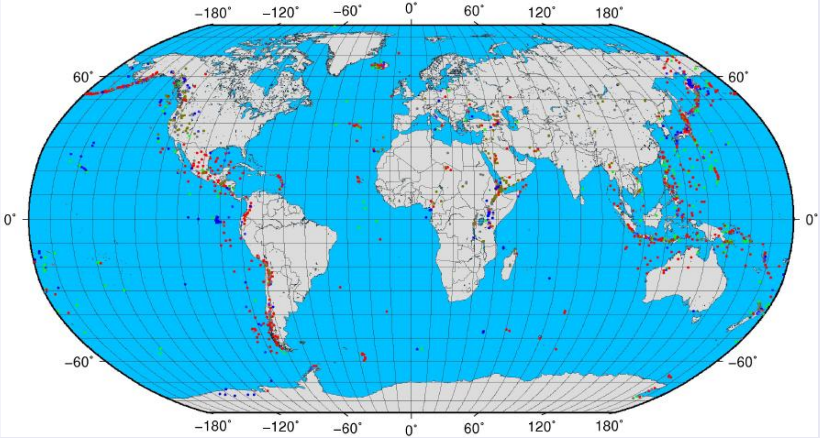
Result matrix: Computed systematical classification

Computed systematical classification of volcanological features (knowledge res.)

Volcano Type	Group	References Data Examples
Complex volcano	A	Vesuvius VNUM:0101-02= UDC:[551.21+911.2+55]:[902]"63"(4+23+24)... GPS:40.821N14.426E Quarternary VEI:VEI5
Compound volcano	A	Cayambe VNUM:1502-004 UDC:[551.21+911.2+55]:(8+23+24)... GPS:... Holocene ...
Somma volcano	A	Ebeko VNUM:0900-38= UDC:... GPS:... Quarternary ...
Submarine volcano	A	Campi Flegrei Mar Sicilia VNUM:0101-07= UDC:... GPS:... Quarternary ...
Subglacial volcano	A	Katla VNUM:1702-03= UDC:... GPS:... Quarternary ...
Unspecified type	A	– VNUM:– GPS:... – ...
Strato volcano	B	Vulcano VNUM:0101-05= UDC:... GPS:... Quarternary ...
Shield volcano	C	Etna VNUM:0101-06= UDC:... GPS:... Quarternary ...
Explosion crater	D	Larderello VNUM:0101-001 UDC:... GPS:... Quarternary ...
Caldera	D	Campi Flegrei VNUM:0101-01= UDC:... GPS:... Quarternary ...
Tuff cone	E	Tutuila VNUM:0404-02- UDC:... GPS:... Holocene ...
Scoria cone	E	Antofagasta de la Sierra VNUM:1505-124 UDC:... GPS:... Holocene ...

Computed spatial distribution

Volcanomap – . . . classified volcanological features from resulting object entries



Colour groups computed via the result matrix:
A: green, B: red, C: blue, D: lighter blue, E: grey, F: dark green.

Processing and computational numbers and issues

Tasks and times regarding the integrated information and computing systems

Item	Value / Description
UDC, number of classification items	70,000
Number of classification languages	50
Number of classification variations (50×70000)	3,500,000
Knowledge object subset, number of items	100,000
Number of terms	10,000,000
Number of object languages	2
Operations, number per subset result entry	50,000,000
Number per subset result entry, incl. keywords	500,000,000
Parallelisation (subset), wall time / num. of nodes	7,500 s / 1
Wall time / number of nodes	1,300 s / 10
Wall time / number of nodes	220 s / 100
Wall time / number of nodes (extrapolated)	4 s / 10,000

Requirements and significant effects

Knowledge context and parallelisation

Besides the large requirements per instance with most workflows there are significant effects by parallelising even within single instances. The following issues have shown to lead to advanced challenges and increased processing and computational times.

- Nomenclature, terms, and attributes tend to be at least partially different in different cultures and languages. For many discovery workflows as well as efforts to increase the quality of the result matrices it is necessary to consider more than one culture and language.
- Processing a classification numbering in decreasing numbering with increasing age or following in different directions is less consequent. For example, in geosciences it is natural to start spaces of time with Quarternary, followed by older stratigraphy.
- In addition to the existing singular spaces of time mapping most objects require appropriate different mappings to absolute dates, e.g., with Bronze Age having different absolute dates for different regions or cultures.
- The calculation with extensive composite classifications, facets, and respective ranges instead of native classifications can increase the computational requirements drastically as has been shown with the knowledge content from the Gottfried Wilhelm Leibniz resources.

Knowledge Oriented Architecture

Knowledge processing and universal classification

- **Knowledge Oriented Architecture** (KOA) of the resources based on flexible integration of documentation and development architecture. Utilising the **Collaboration house** framework for disciplines, services, and resources.
- Knowledge objects, here the geological and volcanic feature objects, can be used with **any of their attributes**. Therefore, any references to objects belonging or referring to any other objects can be computed from this.
- For an object referred to a **timescale of periods** other objects can be associated with the respective object, even beyond direct references. For example, "geological time type" can refer from "volcano type" to any other comparable spaces of time classification (e.g., geophysics, palaeontology, archaeology).
- Further, volcanic objects from the Quarternary can be **associated** to meteorite impact events from the Quarternary.
- With secondary steps **further information** can be integrated. This can include geophysical data, media data or associated objects.
- The resulting quality depends on an intelligent use of **context and classification**.
- A strong **classification support** is essential, the more as object and even many citations, media, and publications are not explicitly aware of the nomenclature of spaces of time used with specific content.

Architecture, documentation, and processing

Systematics and classification

Employing a universal classification with multi-disciplinary content this way, e.g., with volcanological content, expedites knowledge discovery as well as it targets on scientific discovery.

Regarding methodology it further allows to

- Support a **systematic documentation**,
- Define a **normative classification**,
- Define **cognitive interfaces**.

Regarding architecture and implementation it allows to

- Support **decision making** in complex systems,
- Implement **learning system components** and
- **Support components by intelligent systems**.

Structure

Simplified LX knowledge resources object entry used with IICS (© CPR / LX / GEXI)

```
Cenote Sagrado [Geology, Spelaeology, Archaeology]:
               Cenote, Yucatán, México.
               Holy cenote in the area of Chichén Itzá.
               ...
               %%UDC:[55+56+911.2]:[902+903+904]:
                [25+930.85]"63"(7+23+24)=84/=88
```

Classification

Classification set of UDC samples used with IICS (© CPR / LX / GEXI)

```
UDC:[902+903+904]:[25+930.85]"63"(7)(093)=84/=88
UDC:[902+903+904]:[930.85]"63"(23)(7):(4)=84/=88
UDC:[55+56+911.2]:[902+903+904]:[25+930.85]"63"(7+23+24)
 =84/=88
UDC:[25+930.85]:[902]"63"(7)(093)=84/=88
UDC:[911.2+55+56]:[57+930.85]:[902+903+904]"63"(7+23+24)
 =84/=88
UDC:[911.2+55]:[57+930.85]:[902]"63"(7+23+24)=84/=88
```

Communication and computing

Example of dynamical dataset, Active Source component (© CPR / LX / GEXI)

```
#BCMT--------------------------------------------------
###EN \gisigsnip{Object Data: Country Mexico}
#ECMT--------------------------------------------------
proc create_country_mexico {} {
global w
#  Sonora
$w create polygon 0.938583i 0.354331i 2.055118i ...
#BCMT--------------------------------------------------
###EN \gisigsnip{Compute Data: Compute Envelope (CEN)}
#ECMT--------------------------------------------------
#BCEN <ComputeEnvelope>
##CEN <Instruction>
##CEN <Filename>Processing_Bat_GIS515.torque</Filename>
##CEN <Md5sum>...</Md5sum>
##CEN <Sha1sum>...</Sha1sum>
##CEN <Sha512sum>...</Sha512sum>
##CEN <DateCreated>2010-09-12:230012</DateCreated>
##CEN <DateModified>2010-09-12:235052</DateModified>
##CEN <ID>...</ID><CertificateID>...</CertificateID>
##CEN <Signature>...</Signature>
##CEN <Content>...</Content>
##CEN </Instruction>
#ECEN </ComputeEnvelope>
...
proc create_country_mexico_autoevents {} {
global w
$w bind legend_infopoint <Any-Enter> {set killatleave [exec ./mexico_legend_infopoint_viewall.sh
  $op_parallel ] }
$w bind legend_infopoint <Any-Leave> {exec ./mexico_legend_infopoint_kaxv.sh }
$w bind tulum <Any-Enter> {set killatleave [exec $appl_image_viewer -geometry +800+400 ./
 mexico_site_name_tulum_temple.jpg $op_parallel ] }
$w bind tulum <Any-Leave> {exec kill -9 $killatleave }
} ...
```

Universal Decimal Classification (UDC)

Example excerpt of UDC codes used in the following case studies

UDC Code	Description (English)
UDC:55	Earth Sciences. Geological sciences
UDC:56	Palaeontology
UDC:911.2	Physical geography
UDC:902	Archaeology
UDC:903	Prehistory. Prehistoric remains, artefacts, antiquities
UDC:904	Cultural remains of historical times
UDC:25	Religions of antiquity. Minor cults and religions
UDC:930.85	History of civilization. Cultural history
UDC:"63"	Archaeological, prehistoric, protohistoric periods and ages
UDC:(7)	North and Central America
UDC:(23)	Above sea level. Surface relief. Above ground generally. Mountains
UDC:(24)	Below sea level. Underground. Subterranean
UDC:=84/=88	Central and South American indigenous languages

Dimension space and background classification

Dimension space

The classification deployed for documentation must be able to describe any object with any relation, structure, and level of detail. Objects include any

- Textual documents,
- Illustrations,
- Maps,
- Media, photos, videos, sound recordings, as well as
- Realia, physical objects such as museum objects ...

Background classification UDC operations

A suitable background classification is the UDC. The objects use preliminary classifications for multi-disciplinary content. Standardised UDC operations are, e.g.,

- Addition ("+"),
- Consecutive extension ("/"),
- Relation (":"),
- Subgrouping ("[]"),
- Non-UDC notation ("*"),
- Alphabetic extension ("A-Z"),
- Besides place, time, nationality, language, form, and characteristics.

Implementation case studies

State

- No documentation / usability from the object itself
- No support from secondary information (not by form, not by pattern recognition as related objects may be completely different, no geo-location dependency as the important relation may not be depending on close location and so on)
- Huge heterogeneity of objects
- Target is real complex multi-disciplinary context

Required basic prerequisites

- Scientific resources information (LX Foundation Scientific Resources),
- Structuring necessary (LX databases),
- Classification necessary (Universal Decimal Classification, UDC)
- Computation necessary (High End Computing, supercomputing resources), needs to compare millions of objects and classification
- Storage, computation, and processing needed for generating results (text, graphics, maps), thousands of resulting objects in parallel
- Algorithms for communication, computation,
- Functional requirements (Geo Exploration and Information, GEXI collaborations),
- Dynamical Information Systems and data objects (Active Source).
- Batch, scheduling env. (Condor, LoadLeveler, Grid Engine, Moab / Torque, ...)

Implementation case studies

Case study motivation

- Archaeological Information Systems needed for multi-disciplinary investigation,
- Huge potential of integrative benefits and even more pressing that archives are needed for multi-disciplinary records of prehistorical and historical sites while context is often being changed or destroyed by time and development.

Relevant categories of content

Commonly only three categories are relevant to archaeological projects,

1. project level metadata (e.g., keywords, site, dates, project information, geodata),
2. descriptive and resource level metadata (e.g., comprehensive description, documents, databases, geo-data), and
3. file level metadata (software, hardware, accompanying files).

From information science point of view by far not sufficient:

- Licensing and archiving restrictions, access, big data, long-term aspects,
- Precision restrictions,
- Network limitations,
- Context of environment, hardware, storage, and software,
- Hardware restrictions and long-term availability,
- Tools and library limitations and implementation specifics.

Information Matrix

Dimensions of the information matrix (excerpt)

Dimension	Meaning, Examples
Time	Chronology
Topic	Disciplines
	Purpose (tools, pottery, weapons, technology, architecture, inscriptions, sculpture, jewellery)
	Culture (civilisation, ethnology, groups, etymology)
	Infrastructure (streets, pathways, routes)
	Environment (land, sea, geology, volcanology, speleology, hydrogeology, astronomy, physics, climatology)
	Genealogy (historical, mythological documentation)
	Genetics (relationship, migration, human, plants)
	Biology (plants, agriculture, microorganisms)
	Trade (mobility, cultural contacts, travel)
Depth	Underground, subterranean
Site	Areal distribution, region
...	...
Data	Resources level, virtualisation

Dimension view (a)

Archaeological IICS (excerpt) (© CPR / LX / GEXI)

Topic	Purpose / Environment / Infrastructure	Ref.
Precolombian	Architecture	
Caribbean	Environment (volcanology, geology, hydrogeology)	
	La Soufrière Volcano, Guadeloupe, F.W.I.	OC
	Mt. Scenery Volcano, Saba, D.W.I.	OC
	Cenote Sagrado, Chichén Itzá, Yucatán, México	OC
	Ik Kil Cenote, Yucatán, México	OC
Arawak	Architecture	
Prehistory	Architecture	

Topic: architecture mythology environment infrastructure

Entity: Object Location: O On site, D Distributed; Object Media: C Compute, S Storage.

Compute: CONNECT REFERTO-TOPIC REFERTO-SPATIAL VIEW-TO VIEW-FROM

Dimension view (b)

Archaeological IICS (excerpt)

Topic	Purpose / Environment / Infrastructure	Ref.
Egypt	Architecture	
Rome	Architecture	
Catalonia	Architecture	
	Monument de Colom, Port, Barcelona, Spain	OC
Maya	Architecture	
	Kukulkán Pyramid, Chichén Itzá, Yucatán, México	OC
	Nohoch Mul Pyramid, Cobá, Yucatán, México	OC
	El Meco Pyramid, Yucatán, México	OC
	El Rey Pyramid, Cancún, Yucatán, México	OC
	Pelote area, Cobá, Yucatán, México	OS
	Pok ta Pok, Cancún, Yucatán, México	OS
	Templo del Alacran, Cancún, Yucatán, México	OS
	Port, Tulúm, Yucatán, México	OC
	Infrastructure	
	Sacbé, Chichén Itzá, Yucatán, México	OS
	Sculpture	
	Diving God & T. Pinturas, Tulúm, Yucatán, México	OC
	Diving God, Cobá, Yucatán, México	OC

SAMPLE (objects)

Example: Regional Pyramid of Maya, Yucatán, México (© CPR / LX / GEXI)

Kukulkán

Nohoch Mul

El Meco

El Rey

- **Function:** SAMPLE objects from a group and / or location.
- **Content / context**: compute and storage: objects pyramids, Maya, Yucatán region.
- **Computation:** Selection of media photo objects.

REFERTO-TOPIC and REFERTO-SPACE (chain classification)

Example: Diving god, Tulúm, Colom (© CPR / LX / GEXI)

Diving God, Cobá

Pinturas, Tulúm

Port, Tulúm

Colom

- **Function:** Objects and meaning cross-purpose REFERTO other objects.
- Building relation chains.
- **Content / context**: Two chains (cyan and magenta), interlinked, UDC:(7) : (4) relation:
  - 1) diving god refers to pinturas, pinturas refers to Tulum harbour, . . .
  - 2) Colom in Barcelona refers to Tulum harbour but points to India, Tulum harbour refers to pinturas, . . .).
- **Computation:** Selection of media photo objects and grouping.

VIEW-TO VIEW-FROM (in-purpose)

Example: Volcanoes and Cenotes (© CPR / LX / GEXI)

La Soufrière Mt. Scenery

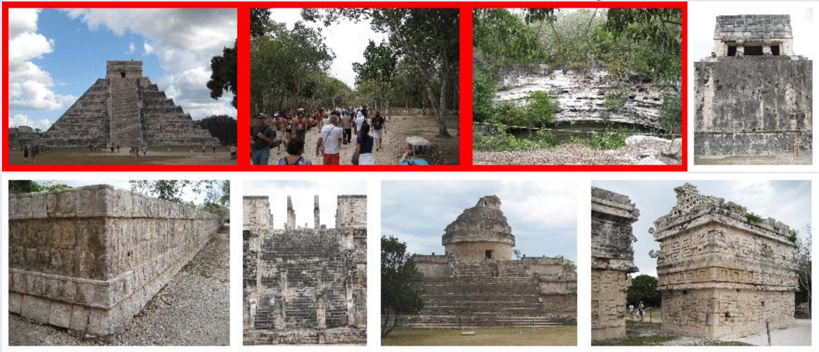
Cenote Sagrado Ik Kil

- **Function:** VIEW from (blue) an and towards (green) an object.
- **Content / context**: Objects volcanology / geology / hydrology, Caribbean, above and below sea level, UDC: "(23)", "(24)",
- **Computation:** Selection of media photo objects, grouping.

CONNECT (in-topic)

Ex.: Kukulkán, Cenote, connected by Sacbé (Chichén Itzá group) (© CPR / LX / GEXI)

Kukulkán Sacbé Cenote Sagrado



- **Function:** Objects CONNECT (marked red) from a group.
- **Content / context**: Objects can be computed by using the relation from classification, e.g., from groups, locations.
- **Computation:** Selection of media photo objects, grouping.

COLLECTION

Example: Precolombian Museum (© CPR / LX / GEXI)

Barbier-Mueller d'Art Precolombí

- **Function:** Any objects being part of a special COLLECTION.
- **Content / context**: Pre-Colombian archaeological objects from museum collection.
- **Computation:** Selection of media photo objects.

CONTEXT

Example: Pottery (amphores) (© CPR / LX / GEXI)

Pottery/Amphores

Comparison

Ship wreck (Valencia)

Anchor (Valencia)

- **Function:** Objects in a special CONTEXT from various locations.
- **Content / context**: Amphores context, comparison, wreck situation, wreck/anchor.
- **Computation:** Selection of media photo objects.

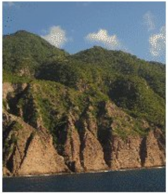
DISCIPLINE

Example: Geology (Caribbean limestone and tuff/volcanic ash) (© CPR / LX / GEXI)

Limestone, Caribbean

Limestone, Caribbean

Volc./Tuff, Carib.



Limest./Tuff, Carib.

- **Function:** Objects from a DISCIPLINE in a special context or location (various collections).
- **Content / context**: Caribbean region, geology and volcanism, limestone and tuff.
- **Computation:** Selection of media photo objects.

Natural sciences and Humanities (Geosciences and archaeology case)

Knowledge and resources

- Implementation shows that goal of integrating IICS components and advanced scientific computing based on structured information and faceted classification of objects has been successful,
- It provides a very flexible and extensible solution for multi-disciplinary applications from natural sciences and humanities, e.g., implementation case study of Archaeological Information Systems,
- Structuring and classification with LX and UDC have provided efficient and economic means using IICS components and supercomputing resources,
- Solution scales, (regarding references, resolution, view arrangements, . . .),
- The concept can be transferred very flexible to numerous applications,
- It has been demonstrated with the case studies that Archaeological IICS provide advanced multi-disciplinary information as from archaeology and geosciences by means of High End Computing resources,
- Atoms are: basic architecture created using the collaboration house framework, long-term documentation and classification of objects, flexible algorithms, workflows and Active Source components,
- Informatics approaches: Collaboration frameworks, Partitioned Global Address Space (PGAS) models, Parallel Virtual Machine (PVM),
- Future development consideration: “tooth system” for long-term documentation and algorithms used with IICS and exploitation of supercomputing resources.

Summary

– Summary –

Significance

- Digital / technological speedup is not coherent with the long-term requirements for knowledge creation.
- Knowledge resources and knowledge creation are of strategical long-term importance.
- Digital speedup supports disciplines and sciences mostly with data processing and computing for generating short-term 'ad hoc' data only.
- Sustainable solutions have to be driven by the disciplines / sciences and fostered by society. Technology on itself provides no sufficient solution.

Cognition attained

Cognition and recommendation

There is no single or isolated general definition for long-term knowledge resources.

Focus

There are long-term relevant results and knowledge.

Recommendation from case studies within last decaded

- **Funding triggered long-term consideration**
- **Researcher instead of administration/hierarchy funding**
- **Researcher-audited researcher-defined best practice**
- **Researcher-defined long-term implementation**

Lessons Learned

Sciences, Knowledge, Computing:

- Speedup: Frequent change in technology is one trigger for non-sustainable long-term knowledge creation and computing infrastructures.

Lessons Learned

Sciences, Knowledge, Computing:

- Speedup: Frequent change in technology is one trigger for non-sustainable long-term knowledge creation and computing infrastructures.
- Sciences: Funding on long-term knowledge creation is (still) not sustainable.

Lessons Learned

Sciences, Knowledge, Computing:

- Speedup: Frequent change in technology is one trigger for non-sustainable long-term knowledge creation and computing infrastructures.
- Sciences: Funding on long-term knowledge creation is (still) not sustainable.
- Knowledge: Long-term multi-disciplinary knowledge resources are required.

Lessons Learned

Sciences, Knowledge, Computing:

- Speedup: Frequent change in technology is one trigger for non-sustainable long-term knowledge creation and computing infrastructures.
- Sciences: Funding on long-term knowledge creation is (still) not sustainable.
- Knowledge: Long-term multi-disciplinary knowledge resources are required.
- Computing: Most computing environments are available for a very limited period of time. In many cases sustainable application should start with the scientific long-term goals.

Future Challenges

Following events:

Future long-term multi-disciplinary topics?

Future Challenges

Following events:

Future long-term multi-disciplinary topics?

Overall goals:

- Improve the long-term creation of knowledge.
- Improve multi-disciplinary documentation and work.
- Improve decision making processes and workflows.
- Improve the Quality of Data.
- Support integrated information and computing systems.
- Dissemination with processes, learning, and education.
- A "State of the art for long-term issues".
- Where we are: Content, Classification, . . . on its own.
- Mid- and long-term: Context.
- Where we go: Investments in sustainable usability.

Follow-up topics at this years' conference

Presentation: Knowledge and Processing

- **Tuesday, 2014-03-25, 17:00 – 18:45**

  GEOProcessing 4–Session, Discussion on:

  **Knowledge Processing for
  Geosciences, Volcanology, and Spatial Sciences
  Employing Universal Classification.**

  Program: `http://www.iaria.org/conferences2014/ProgramGEOProcessing14.html`

References

References and acknowledgements, see:

⇒ C.-P. Rückemann, "Knowledge Processing for Geosciences, Volcanology, and Spatial Sciences Employing Universal Classification," in *Proceedings of The Sixth International Conference on Advanced Communications and Computation (INFOCOMP 2013), Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2014), DigitalWorld 2014, March 23–27, 2014, Barcelona, Spain*. XPS Press, 2014, ISSN: 2308-393X, ISBN-13: 978-1-61208-326-1, pp. 76–82, URL: http://www.thinkmind.org/index.php?view=instance&instance=GEOProcessing+2014 [accessed: 2014-03-23], http://www.iaria.org/conferences2014/ProgramGEOProcessing14.html [accessed: 2014-02-28].

Networking

Thank you for your attention!

Wish you an inspiring conference and a pleasant stay in Barcelona!