International Tutorial DigitalWorld / GEOProcessing 2013

Advances in Geosciences, Natural Sciences, and Humanities: Significance of Knowledge, Processing, and Computing

The International Conference on Advanced Geographic Information Systems, Applications, and Services (DigitalWorld / GEOProcessing)

February 24 – March 1, 2013, Nice, France

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

Westfälische Wilhelms-Universität Münster

Contents

Tutorial targets

Focus:

- Aspects of knowledge, processing, and computing.
- Significance for geosciences, geoscientific application and information systems.
- Increasingly important since multi-disciplinary information from natural sciences and humanities is getting involved.
- Introduction to long-term knowledge handling and classification.
- High End Computing resources used for processing and computing.
- Requirements and operation for advanced scientific computing environments, basics of decision making and resources planning for advanced collaboration.
- Knowledge and computing resources usage, case studies for geoscientific and archaeological information systems.
- Operation and lifecycle aspects, existing application scenarios, interests and needs of users and disciplines, services, and resources providers.
- 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 geosciences view and how can we handle information and knowledge?
- What does processing and computing mean in geosciences and other disciplines?
- How can information content and context be preserved for long-term sustainability?
- Which diversity of knowledge, workflows, and resources does exist in geosciences?

It is intended to have a live discussion and feedback on the consequences for various topics and applications.

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

Aspects:

- Learn from history and failures.
- Motivation.
- Knowledge.
- Processes.
- Techniques.
- Overview of High End Computing (HEC).
- Consequences for participated parties.
- Consequences for selected application scenarios.

Sciences, Archaeology, and History

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.

Way (NOT) to go: Ignorance and Delegation Principle

What others do: “Experts say: Best practice is for theory.”

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

“N”ewtoneless
“U”niversity
“T”echnology
“S”ervice

NUTS think:

- **All Knowledge can be generated from the Internet.**
- **Heterogeneity means something for everyone.**
- **Defining tools is better than using standards.**

NUTS live with:

- Knowledge is just data but **pastel colours** in tables are more important,
- **Experiences can be considered an “add-on”**,
- **Good and best practice are formal issues.**

NUTS set up, to "accelerate and shorten" processes:

- **Executive chairing** in-house-group,
- **Administrative expert** in-house-group,
- **Technical delegation** somewhere-group to make it ‘cheap’.

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.

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.

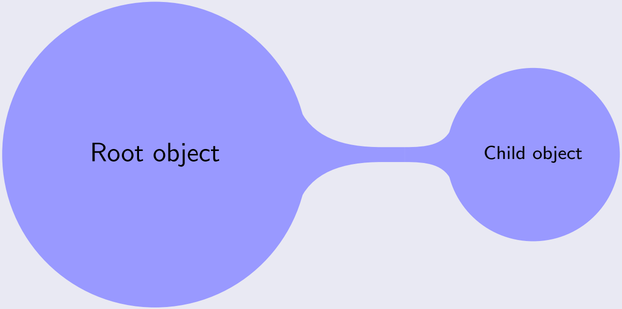
Knowledge and Application

Processes

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

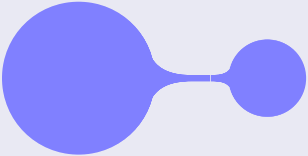
Objects and Relations

Objects and Relations



Object, Relations, and Quality (Mindmapping)

Quality of Relations

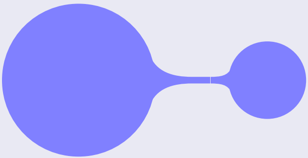


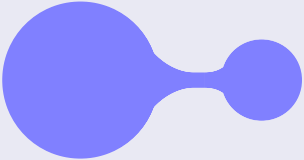
Object, Relations, and Quality (Mindmapping)

Quality of Relations

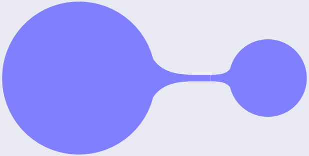
Object, Relations, and Quality (Mindmapping)

Quality of Relations





Quality of Objects



Object, Relations, and Quality (Mindmapping)

Quality of Relations

Quality of Objects

Knowledge: Prejudice and getting the right meaning

Wrong terms can be very persistent:

- Sunrise (earth is flat?),
- Sunset (from dusk till dawn?),
- Malaria (and prejudice is ahead of scientific results?).
- ...

Knowledge: Perception

Examples

- Depiction, traffic signs and their description different.
- Companies do try critical products in countries with reduced privacy perception.
- Overall personal security will mean insecurity for society.
- Color perception is different by society.

Description

- "*Standardisation*" and "*internationalisation*".
- Foreign word "*privacy*".
- Trend for *hidden security*.
- Words for new colors have been *added* to languages and perception.

Knowledge: Cultural Background

International and other differences

- Privacy perception,
- Different terminology,
- Legal regulations,
- Legal frameworks.

Knowledge: Language and Small Things

Even off-topic items can make a difference (not only in Harvard):

Small things can completely change the meaning without changing a word!

- Tweetie, Tom and Jerry do have different opinions.
- Tweetie, Tom, and Jerry do have different opinions.

Knowledge: Language and Small Things

Even off-topic items can make a difference (not only in Harvard):

Small things can completely change the meaning without changing a word!

- Tweetie, Tom and Jerry do have different opinions.
- Tweetie, Tom, and Jerry do have different opinions.

- Fishes, mammals and birds are three different kind of genus.
- Fishes, mammals, and birds are three different kind of genus.

Knowledge: Symbolisation and Language

Classical one: Who said this?

 +  + 

Knowledge: Symbolisation and Language

Classical one: Who said this?

+

+

Easier, English terms: Egg + Box + Girl.

Knowledge: Symbolisation and Language

Classical one: Who said this?

 +  + 

Easier, English terms: Egg + Box + Girl.

Hint: Try in different languages.

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?

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.

High End Computing

Basics and prerequisites

- Real goals. Define the goals, different views.
- Need for basic understanding and knowledge base for HEC.
- Prominent HEC and collaboration aspects decision making processes are necessary for.
- Separate the topics (disciplines, resources, ...).
- Gather the real requirements for the analysis.
- Up-to-date resource policies in theory and practice.
- Interesting fields of application are processes within disciplines.
- Future deployment of integration and classification with components of complex systems.

Decisions and Computing

Components (all areas, no sort order):

- Architecture,
- Operating System,
- Applications,
- Programming languages,
- Tools,
- System modeling,
- Vendors,
- Strategy,
- Targets,
- Staff,
- Operation,
- Services,
- System management,
- Complex licensing,
- Policies,
- Governance,
- ...

Decisions and High End Computing

Components (all areas, no sort order):

- Components (all areas) with strong focus on
- Applicability, efficiency,
- Architecture applicability,
- Operating System applicability,
- Efficient applications,
- Programming languages,
- Tools,
- System modeling,
- Vendors,
- Strategy,
- Targets,
- Staff,
- Operation,
- Services,
- System management,
- Complex licensing,
- Policies,
- Governance,
- ...

High Performance Computing / Advanced Scientific Computing

Overview

- Requirements
  - Fast CPU.
  - Parallel processing.
  - Large memory.
  - Fast I/O.
- Hardware / resources
- System / software / configuration
- Applications
- Examples?

- High Performance Computing. Base, parallel developments are integrated for HPC.
- Cluster computing, optimised utilisation of heterogeneous resources (Condor).
- Cloud and Grid Computing (e.g., Globus Toolkit, UNICORE).

- Cloud and Grid islands: Different companies – different technology and terms.

Architecture and Security

Architecture and Security

• Hardware / Computing.	
◦ MPP (Massively Parallel Processing).	MPP compute nodes
◦ SMP (Symmetric Multi-Processing).	SMP compute nodes
• System software.	
◦ Operating systems.	Login server, admin server
◦ Cluster management.	Management server
◦ Storage management.	Storage server
◦ File management.	File server
• Networks.	
◦ InifiniBand for I/O.	
◦ InifiniBand for Message Passing Interface (MPI).	
◦ NumaLink.	
◦ Service networks.	
• Parallel filesystems (GPFS, Lustre, ...).	MDS server, OSS server
• Batch system, scheduling, load balancing. (Moab, Torque, ...).	Batch server
• Accounting ...	
• Data handling, archive / backup.	Archive / backup server
• Optional Grid, Cloud services level.	

Best Practice Recommendations

Common ones

- Use abstraction,
- Use standards,
- Use high-level,
- Use modularisation,
- Use process and workflow modeling,
- Consider legal regulations,
- Define policies,
- Iterate processes,
- Implement an independent auditing,
- (consult other best practice, if available),
- ...

Isolated Approaches

Isolated Approaches in computing:

- Internationalisation,
- Transliteration,
- Syntax unification,
- Semantics,
- ...

Isolated Approaches in programming:

- High level languages,
- Object-oriented paradigms,
- Literate programming,
- Standardisation,
- ...

Intermediate Status

Anyhow, on the current base today:

Has someone seen an implementation that has been done perfectly?

Information Technology

Selection processes need to be made for:

- Purpose and usage,
- Budget,
- Components,
- Content and data security,
- Science, research, and staff,
- Policies,
- Access,
- Security,
- Operation and staff,
- ...

Selection Process

For which purpose do we need a selection process regarding:

- Resources,
- Information Mining and Management,
- Broadband networks,
- Fibre channel networks,
- Mobile Services,
- ...

Cloud and Grid Islands

Terms, brands, names (historical excerpt)

- Sun:
  - Cluster Grids
  - Enterprise Grids
  - Global Grids
- HP:
  - Utility Computing
- IBM:
  - Autonomic Computing, resources, dynamic VO
  - Grid + provisioning via Cloud Computing (SaaS, DaaS, AaaS ...)
- ...

Science, long-term, and service

For discussion:

Has anyone attending here implemented or continuously used some architecture on mid- or long-term (at least 10 to 15 years?) or something like Science as a Service?

Governance and IT-Governance

Information Technology Governance

- IT-Governance is a subset discipline of Corporate Governance focused on IT systems and their performance and risk management.
- http://en.wikipedia.org/wiki/Governance
- http://en.wikipedia.org/wiki/Information_technology_governance

Legal aspects and effects on software development / IT-governance:

- Rising interest in IT-Governance is partly due to compliance initiatives, as well as the acknowledgment that information technology systems, operation, and projects (e.g., networks, computing, cloud) can be hard to control and can heavily affect the overall performance of institutions and organisations.
- Compliance initiatives:
  - Basel II, Europe.
  - Sarbanes-Oxley Act (SOX), USA.

IT and Information Security, ISO/IEC 27000, 27001, 27002

ISO and IEC

- http://en.wikipedia.org/wiki/ISO/IEC_27000
- http://en.wikipedia.org/wiki/ISO/IEC_27001
- http://en.wikipedia.org/wiki/ISO/IEC_27002
- ...

Legal Issues with High End Computing

Challenges with:

- International collaboration,
- Frameworks and standards,
- Services provisioning,
- License models,
- Software Licenses (core numbers, floating, etc.),
- Not hardware, not software (firmware),
- Third parties,
- Operation,
- Maintenance,
- Non-deniability,
- Security,
- ...

Security

Aspects

- Infrastructure security (power on/off security),
- Information technology security (power on/off security),
- Data security,
- Privacy,
- ...

Categories:

- Scientific security research, encryption, in-silico security,
- Low level security (end user application, end user devices and algorithms),
- Day to day / trivial services and support issues,
- ...

Operating System Protection

Operating System Protection Profile (OSPP)

- Local auditing,
- Crypto-communication,
- Access control,
- Communication packet filtering,
- Security management.

((Practical/theoretical security.))

Data and Content Security

Issues:

- Research – Industry,
- Homomorphic application environments,
- Policies,
- Encryption,
- Signatures,
- Privacy,
- Anti-privacy,
- Plagiarism prevention and detection,
- ...

Content and Plagiarism in Information Science and Technology

Categories:

- Plagiarism prevention.
- Plagiarism detection.

Principal, legal, and technical issues:

- Authors/student work before copyright signed in review.
- Databases of third parties, reliability issues.
- Introduction into databases of third parties.
- Problem with double blind / blind review.
- Removal of references and / or citations.
- Self-plagiarism only for a small extend.
- Graphics cannot be checked.
- Sources checked with and without permission of authors?
- Sources might even not be published by authors.
- Too heterogeneous conditions in practice.
- Very time intensive process for authors, reviewers, editors, and publishers.
- Restrictions due to workflow, widely used procedures and structures.
- ...

Documentation

Aspects

- Trivial documentation,
- Technical and applications' documentation,
- Scientific documentation,
- Structure,
- Classification,
- Re-use.

License Models and Patents

Long-term problems?

- Public Domain,
- Freeware,
- GPL (and derivatives),
- Charityware (vim),
- Postcardware,
- Giftware,
- ...
- Open Access model, Open Access publishing,
- ...
- Open Source is trademark but this does not mean products labeled with Open Source are provided without limitations of any kind.
- Bilsky Case.
- Patent Absurdity.

Modeling and applications

Unified Modeling Language (UML)

The Unified Modeling Language (UML) can be used for various purposes with information sciences, software development, and even independent from information sciences, e.g. in economics and business context:

- “business model”
- classes
- messages, objects in their timing sequence
  - coarse overview
  - dynamic
  - parallel processes
  - distributed systems

UML Diagrams

UML Diagrams

- Use-case diagram
- Class diagram
- Package diagram
- Interaction diagram
- State diagram
- Activity diagram
- Implementation diagram

Basics on 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.

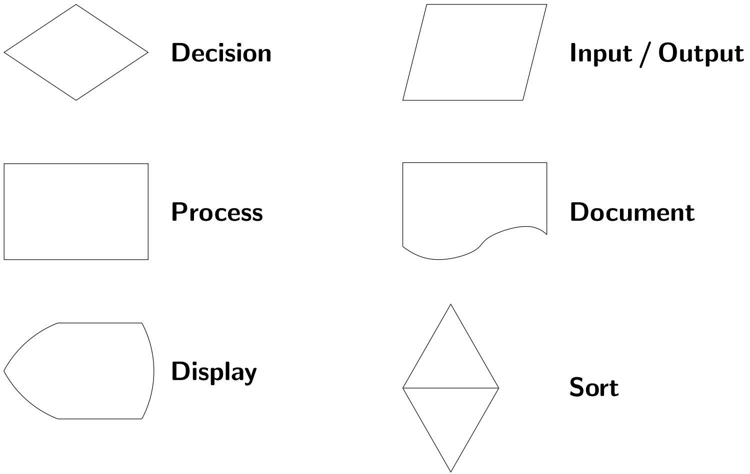
Elimination by aspects: Choosing alternatives by "mathematical psychology". Covert elimination process, comparing the available alternatives by aspects. Choose an aspect and eliminate the alternatives without the aspect. Repeat until one alternative remains.

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.

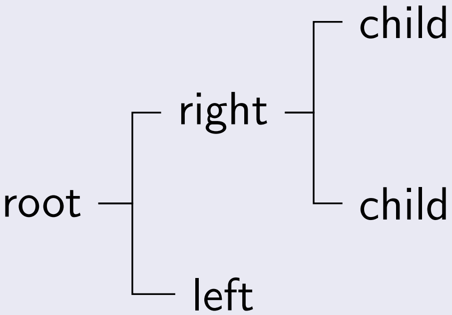
Visualising Flow Basics

Symbols



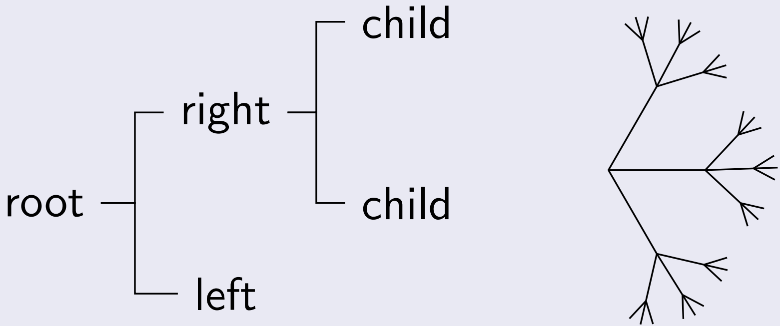
Trees and Forks

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>
```

View: Disciplines

– View: 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?

Infonomics / Information Management

References:

- Infonomics for Distributed Business and Decision-Making Environments
  http://www.igi-global.com/reference/details.asp?ID=34799
- Infonomics Society
  http://www.infonomics-society.org
- Infonomics Internet- and database recherche
  http://www.infoseeking.de
- AIIM Infonomics
  http://www.aiim.org/infonomics/
- Infonomics.nl
  http://www.infonomics.nl
- Infonomics.at
  http://www.infonomics.at
- Infonomics.com.au
  http://www.infonomics.com.au

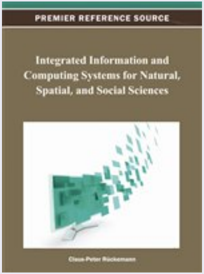
Integrated Information and Computing Systems

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

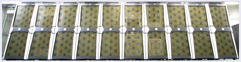
Disciplines

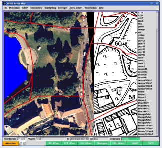
Increasing high end demands

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

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.

© CPR

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: 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
- SuperMUC
- JUQUEEN
- Sequoia
- Titan
- and others ...

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

Application Scenarios in Research and Education

– Application Scenarios –

Application scenarios and decision making 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

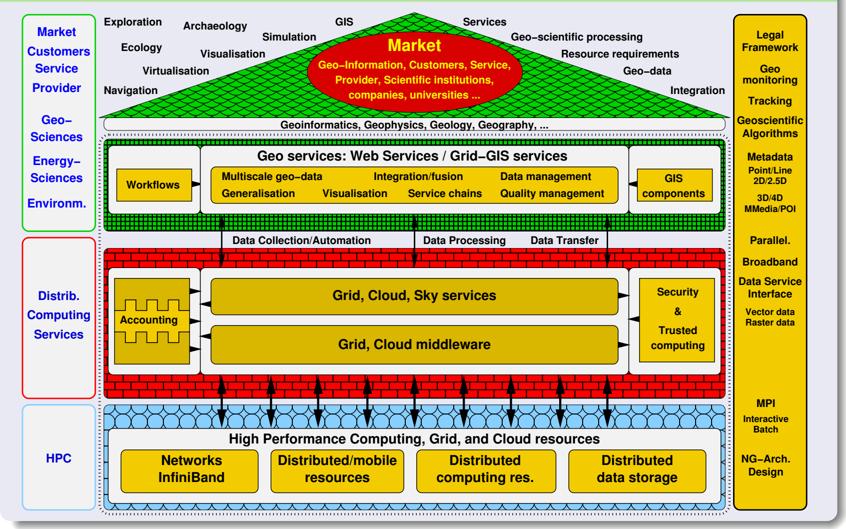
Frameworks supporting integration of:

- 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)*.

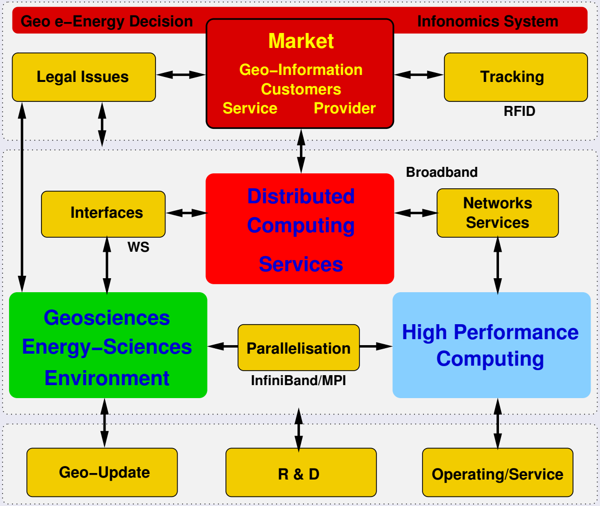
Information, computing, and collaboration

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
...
```

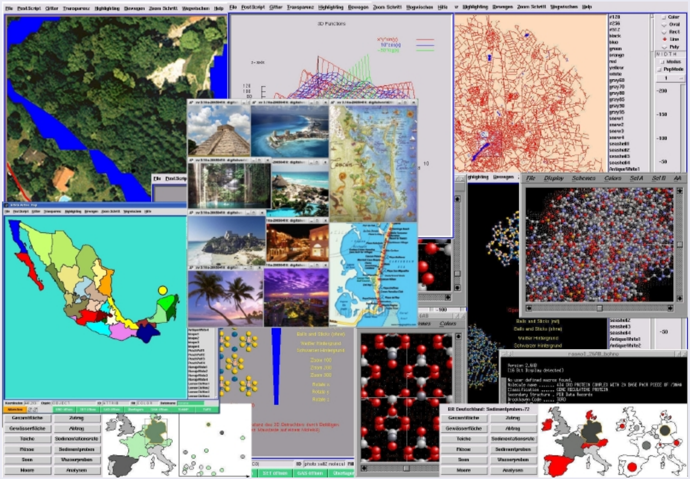
Binding of Precalculation Script

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



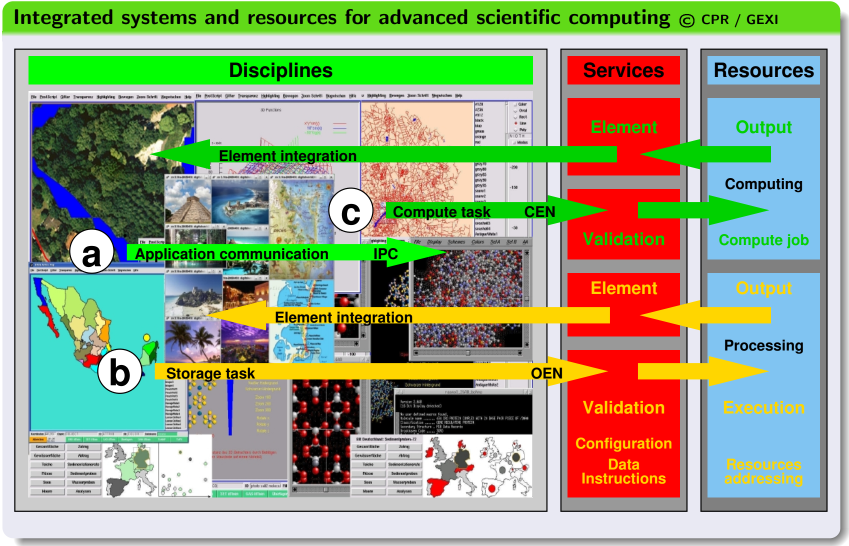
Data objects
vector
raster
aerial
photo
spatial
calculation
measurement
processing
meta objects
interactive
commercial
license
...

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 the Different Dynamical Tasks



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

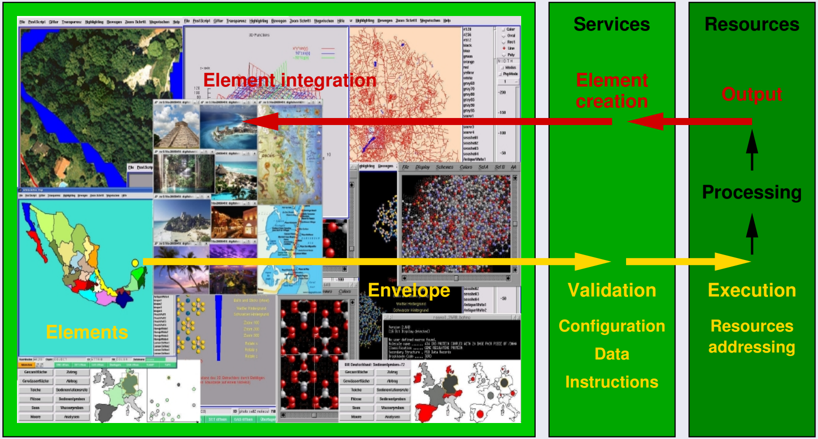
- Information and computing resources instructions.
- Links between the information and computing system.
- Prerequisites of the computing system.
- Processing directives and script elements.
- Input / output data necessary.
- ...

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

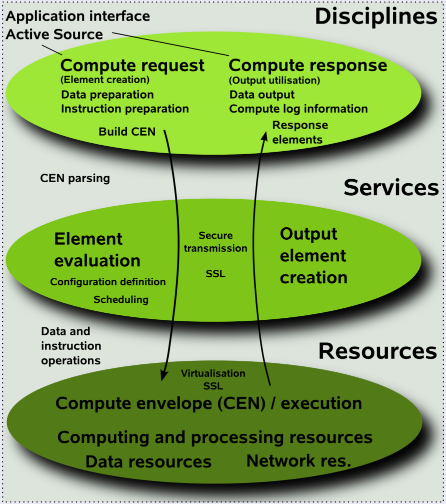


Compute access

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

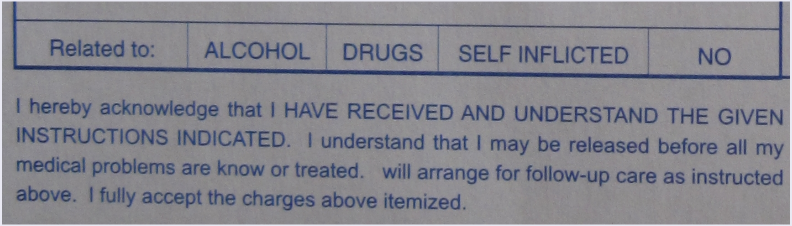
Privacy Case: Archaeology Information Systems and Tourism

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

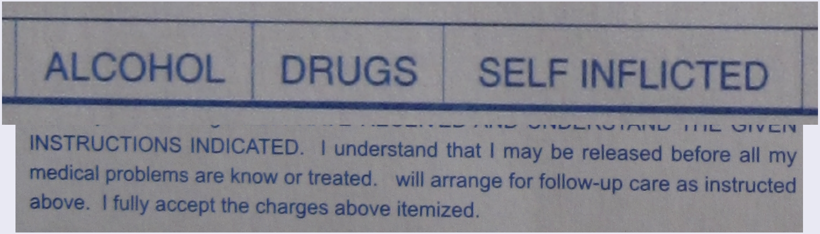


Related to:	ALCOHOL	DRUGS	SELF INFLICTED	NO

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

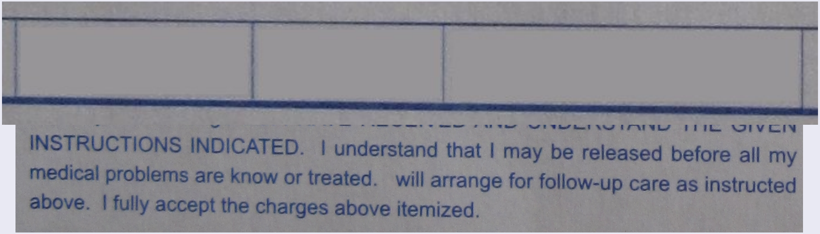


ALCOHOL | DRUGS | SELF INFLICTED

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

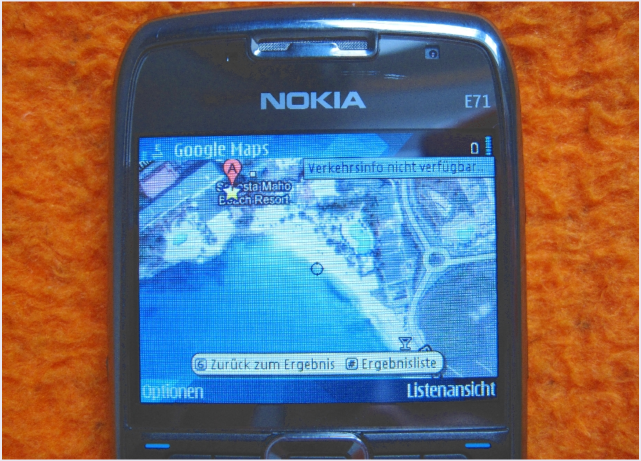
Individual privacy and anonymity: Protection of individual information © CPR / GEXI


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

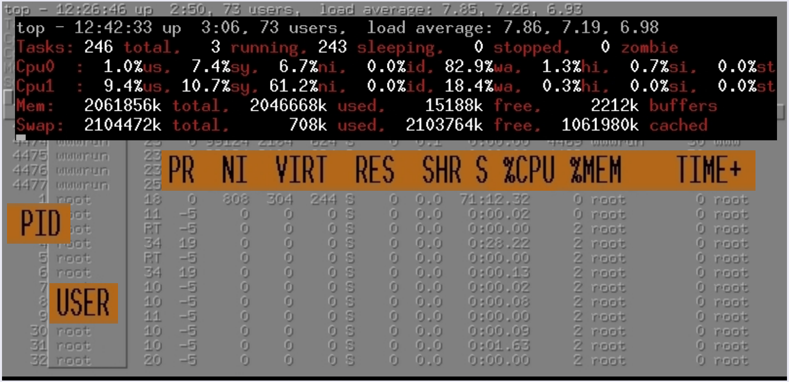
Privacy and anonymity: Real system base security and protection © CPR / GEXI

```
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

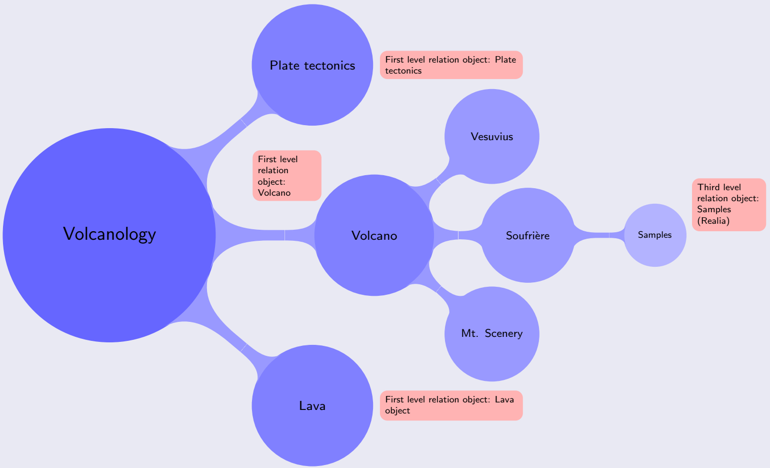
Privacy and anonymity: Real system base security and protection © CPR / GEXI



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

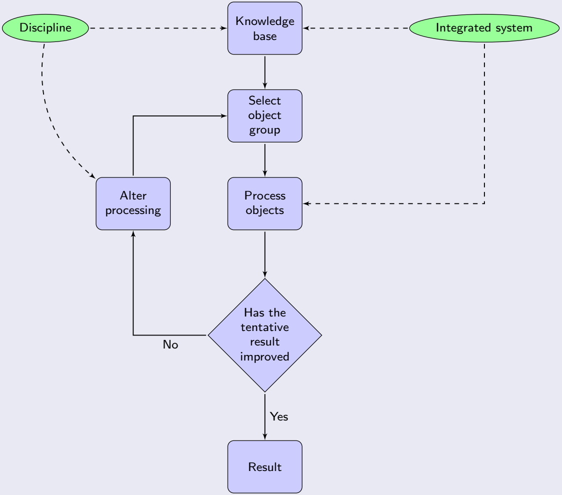
Knowledge Resources: Objects and Relations, Classification

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.

How to integrate special information

Manuscripts, Copperplates, . . .

European Cultural Heritage Online (ECHO)
and
Gottfried Wilhelm Leibniz Bibliothek (GWLB) Hannover

http://echo.mpiwg-berlin.mpg.de/content/copperplates

http://www.gwlb.de

http://www.leibnizcentral.de/CiXbase/gwlbks/

http://www.leibnizcentral.de/CiXbase/gwlbhss/

Processes with knowledge resources: Transliteration, Transcription, Translation, Classification, Content Description, Context Description, Sources Description, References Description, Bibliographical Description, Object Description, Material Description, Media Objects, Media Object Description, . . . (© CPR / GEXI)

How to integrate special information

Manuscripts – besides physical measures and digital media © CPR / GEXI

Processes:

Transliteration

Transcription

Translation

Classification

Content Descr.

Context Descr.

...

— COPYRIGHT EXAMPLES LEFT OUT HERE —

How to integrate special information

Copperplate: Historical documents and unicorns

```
1   Kupferstichplatten
2   Titel:  K 220 Einhorn und versteinerter Zahn
3   Beschriftung:   Tab. XII; Dens animalis marini
     Tidae prope Stederburgum e colle limoso
    effossi. Figura Sceleti prope Qvedlinburgum
    effossi.
4   Stecher:        Seeländer [signiert: N.
    Seelaender sc.]
5   Format:         318x196 mm
6   Bemerkung:      Abzug unter cua stark
    beschädigt. - Liste 1727, Nr. 23; Liste 1729
    a, Nr. 10. Abzug (ohne Tafelnummer) auch in
    Noviss. 56: IV,3, Bl. 12. Lt. Manuskript
    XXIII, 23b, Bl. 57' u. 57a, sollte dies
    ursprünglich Tafel X sein.
7   Abdruck:        Leibniz, Protogaea, Taf. XII,
    Text dazu S. 64 [über den Fund bei
    Quedlinburg]:Testis rei est Otto Gerikius,
    Magdeburgensis Consul, qui nostram aetatatem
    novis inventis illustravit [...] Gerikius
    igitur libro de vacuo edito, per occasionem
    narrat, repertum Sceleton unicornis in
    posteriore corporis parte, ...
8   Nachgestaltung:         Nachstich in Leibniz,
    Opera omnia, studio L. Dutens, 1768. -
    Wallmann, Abhandlung von den schätzbaren
    Alterthümern zu Quedlinburg, 1776, Tafel ...
9   Literatur:      ...
10  Signatur:       cup 4048
11  Signatur(Abzug):        cua 3203
```

— COPYRIGHT EXAMPLES LEFT OUT HERE —

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.

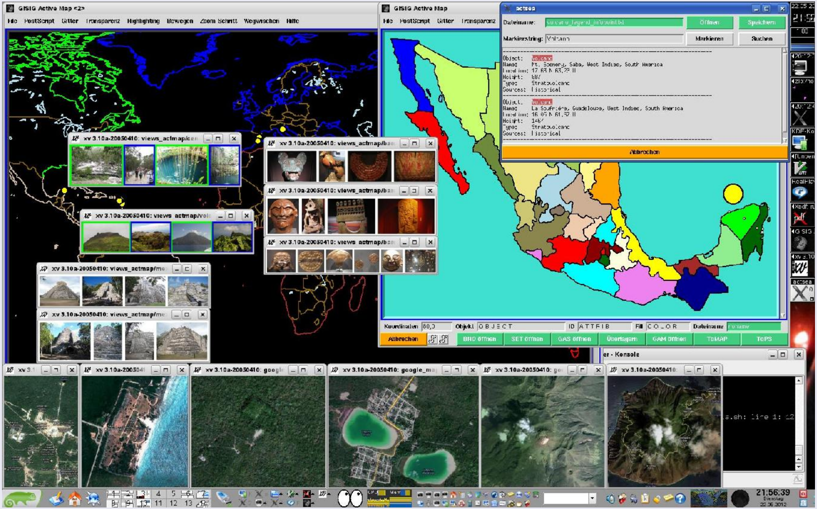
Keyword context

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)



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, 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

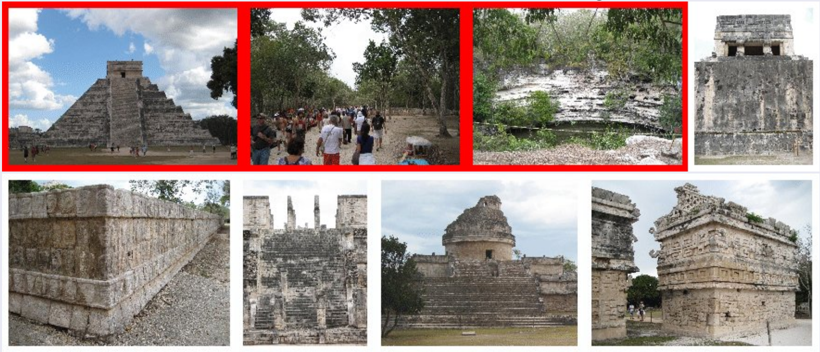
Cenote Sagrado   Ik Kil

- **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.

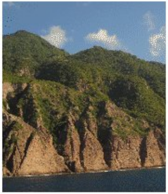
DISCIPLINE

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

- **Gathering and saving reusable knowledge is of strategical long-term importance in geosciences.**
- **Processing geo-data is essential for data analysis and knowledge creation.**
- **Computing facilities are mandatory for advanced knowledge discovery, modeling, and simulation.**

Lessons Learned

Geosciences, Knowledge, Processing, Computing:

- (Funding is not sustainable.)

Lessons Learned

Geosciences, Knowledge, Processing, Computing:

- (Funding is not sustainable.)
- Knowledge: Most results are still not long-term persistent and reusable.

Lessons Learned

Geosciences, Knowledge, Processing, Computing:

- (Funding is not sustainable.)
- Knowledge: Most results are still not long-term persistent and reusable.
- Processing: Most processing implementations are volatile and individually built.

Lessons Learned

Geosciences, Knowledge, Processing, Computing:

- (Funding is not sustainable.)
- Knowledge: Most results are still not long-term persistent and reusable.
- Processing: Most processing implementations are volatile and individually built.
- Computing: Most computing environments are available for a very limited period of time.

Lessons Learned

Geosciences, Knowledge, Processing, Computing:

- (Funding is not sustainable.)
- Knowledge: Most results are still not long-term persistent and reusable.
- Processing: Most processing implementations are volatile and individually built.
- Computing: Most computing environments are available for a very limited period of time.
- Changes in knowledge, processing, and computing infrastructures are very frequent.

Lessons Learned

Geosciences, Knowledge, Processing, Computing:

- (Funding is not sustainable.)
- Knowledge: Most results are still not long-term persistent and reusable.
- Processing: Most processing implementations are volatile and individually built.
- Computing: Most computing environments are available for a very limited period of time.
- Changes in knowledge, processing, and computing infrastructures are very frequent.
- Knowledge, processing, and computing are mostly not sufficiently documented.

Lessons Learned

Geosciences, Knowledge, Processing, Computing:

- (Funding is not sustainable.)
- Knowledge: Most results are still not long-term persistent and reusable.
- Processing: Most processing implementations are volatile and individually built.
- Computing: Most computing environments are available for a very limited period of time.
- Changes in knowledge, processing, and computing infrastructures are very frequent.
- Knowledge, processing, and computing are mostly not sufficiently documented.
- Differences in space and time!

Lessons Learned

Different ...

- resolutions,
- methods and algorithms,
- grid densities,
- context,
- equipment, ...

Lessons Learned

Countermeasures . . .

- Modularisation,
- Long-term documentation,
- Standardisation,
- Open development,
- Community models,
- Digital archives,
- Collaboration frameworks,
- Discipline-Services-Resources concept,
- Knowledge resources,
- . . .

Future Challenges

Following events:

Are there aspects for future multi-disciplinary topics?

Future Challenges

Following events:

Are there aspects for future multi-disciplinary topics?

Overall goals:

- **Improve long-term creation of knowledge.**

Future Challenges

Following events:

Are there aspects for future multi-disciplinary topics?

Overall goals:

- Improve long-term creation of knowledge.
- Improve decision making processes.

Future Challenges

Following events:

Are there aspects for future multi-disciplinary topics?

Overall goals:

- Improve long-term creation of knowledge.
- Improve decision making processes.
- Improve Quality of Data.

Future Challenges

Following events:

Are there aspects for future multi-disciplinary topics?

Overall goals:

- **Improve long-term creation of knowledge.**
- **Improve decision making processes.**
- **Improve Quality of Data.**
- **Improve multi-disciplinary work.**

Future Challenges

Following events:

Are there aspects for future multi-disciplinary topics?

Overall goals:

- Improve long-term creation of knowledge.
- Improve decision making processes.
- Improve Quality of Data.
- Improve multi-disciplinary work.
- Support integrated systems.

Future Challenges

Following events:

Are there aspects for future multi-disciplinary topics?

Overall goals:

- Improve long-term creation of knowledge.
- Improve decision making processes.
- Improve Quality of Data.
- Improve multi-disciplinary work.
- Support integrated systems.
- Dissemination with processes, learning, and education.

Future Challenges

Following events:

Are there aspects for future multi-disciplinary topics?

Overall goals:

- **Improve long-term creation of knowledge.**
- **Improve decision making processes.**
- **Improve Quality of Data.**
- **Improve multi-disciplinary work.**
- **Support integrated systems.**
- **Dissemination with processes, learning, and education.**
- **A "State of the art for long-term issues".**

Future Challenges

Following events:

Are there aspects for future multi-disciplinary topics?

Overall goals:

- Improve long-term creation of knowledge.
- Improve decision making processes.
- Improve Quality of Data.
- Improve multi-disciplinary work.
- Support integrated systems.
- Dissemination with processes, learning, and education.
- A "State of the art for long-term issues".
- Where we are: Content, Classification, Modelling.

Future Challenges

Following events:

Are there aspects for future multi-disciplinary topics?

Overall goals:

- Improve long-term creation of knowledge.
- Improve decision making processes.
- Improve Quality of Data.
- Improve multi-disciplinary work.
- Support integrated systems.
- Dissemination with processes, learning, and education.
- A “State of the art for long-term issues”.
- Where we are: Content, Classification, Modelling.
- Mid- and long-term: Context.

Follow-up topics at this years' conference

Presentation: Computing and Documentation

- **Tuesday, 2013-02-26, 15:15 – 17:00**

  GEOProcessing 4–Session, Discussion on:
  Advanced Scientific Computing and Multi-Disciplinary Documentation for Geosciences and Archaeology Information.

  Program: `http://www.iaria.org/conferences2013/ProgramGEOProcessing13.html`

International Panel GEOProcessing 2013

- **Tuesday, 2013-02-26, 19:00 – 20:30**

  International Expert Panel on Geosciences in the Age of Knowledge:
  Tackling the Complex and Challenging World of Future Geo-application Scenarios.

  Program: `http://www.iaria.org/conferences2013/ProgramGEOProcessing13.html`

Networking

Thank you for your attention!

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