International Tutorial Infoware

Advanced Computing and Data-Centricity: Saving Value and Taming Complexity

InfoWare 2016

The Eleventh International Multi-Conference on Computing in the Global Information Technology (ICCGI 2016)

November 13–17, 2016, Barcelona, Spain

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

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

ruckema(at)uni-muenster.de

Introduction

Advanced Computing and Data-Centricity: Saving Value and Taming Complexity

- **Data** is the core reason for computing. Investments in data are **multi-layered**. Extravagant data may require expensive, specialised solutions. The **value** of data hides in multi-layers, e.g., some data have to be created over long periods of time and cannot be created a second time. And, data does have **more than economical value**. Result oriented solutions can **benefit from data complexity**. Therefore, complexity can also carry values.
- It is beneficial to take a closer look at the details of the respective **relations and conditions**. **Centricity**, as in "data-centric", "knowledge-centric", and "computing-centric", is a significant aspect for understanding, choosing, and creating advanced solutions.
- This tutorial is addressed to all interested users and creators of data, disciplines, geosciences, environmental sciences, archaeology, social and life sciences, as well as to users of advanced applications and providers of resources and services for High End Computing.

Tutorial targets

Focus with aspects of centricity:

- Different types of data and organisation.
- Different types of computing and storage architectures.
- Different methods.
- Different goals.

Focus questions

Some focus questions are:

- What does **value** mean for '**holistic**' scenarios?
- Where to target **value** and where **complexity**?
- What is the meaning of **centricity** and are there reasons to think about centricity details?
- What is the **discipline/users' view**? Are there choices and how?
- What are examples of **knowledge creation, discovery, and workflows**?
- What are **benefits and tradeoffs** and how can issues like long-term relevant data, complexity, portability be handled?
- Which **architectures and scenarios** can be considered?
- Are there **different types of Big Data** and **can Big Data be data-centric**?
- Which cases require **high end solutions** and which **High Performance Computing** architectures?
- What are the **consequences of centricity**?

It is intended to have a concluding dialogue with the participants on practical scenarios and experiences.

Way (NOT) to go: Value Comes From What Management Does . . .

What others do: "Experts say: Complexity is a Staff Problem."

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

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

NUTS' initiative:

- **Have some non-researchers for deciding and organising research data management,. value, and centricity.**

NUTS' strategy:

- **Value is resulting from management activities (only).**
- **Staffs' task is to save the value.**
- **Complexity is resulting from staff activities.**
- **Management has to curtail complexity.**

NUTS' results and recommendations:

- **Extend management and administrative activities to increase value.**
- **Have staff do low level activities to reduce complexity.**
- **Restrict advanced computing to technical level only.**
- **Go for computing-centricity to minimise data-centricity.**
- **Data security and rights are decisions of the provider.**

Computer, Computer Science, and Information Science

Computer

Computer: (lat.) computare = calculate. A device applicable for universal automatic manipulation and processing of data.

Computer Science / Information Science

Computer Science / Information Science is the science of systematic processing of data / information, especially the automatic processing making use of computing installations.

Data and computing are interlinked in many ways

Computing is not possible without data

- Data :: electronic documentation
- Data :: storage
- Data analysis :: processing, computing
- Mobile/communication data :: digital communication
- Astrophysical research data :: collecting and analysing
- Physics data :: collecting and analysing
- Environmental data :: collecting and analysing
- Dynamical components :: dynamical processing
- Near real time data :: preview, streaming
- Simulation :: computing
- Modelling :: computing . . .

Data is becoming extravagant, specialised solutions are the consequence

Different types of Big Data may prefer different high end solutions

- Structured data resources.
- Unstructured data resources.
- Central data resources.
- Distributed data resources.
- ...

Different High Performance Computing applications prefer different data handling

- Documentation.
- Storage.
- Communication.
- Transfer.
- Computing architectures ...

First View: Centricity – Data

Data-centric

- The term "data-centric" refers to a focus, in which data is most relevant in context with a purpose.

First View: Centricity – Database

Database-centric

- The term "database-centric" refers to an architecture based on a database concept, which is used for data handling. In this scenario the database plays a crucial role. In some cases the terms "data" and "database" are mixed up.

Examples:

- File-based data structures and access methods as well as general-purpose database management. (A distinction is outdated.)
- Dynamic, table-driven logic, directed by the "contents" of a database, dynamic programming languages.
- Shared database, communication between parallel processes, distributed computing application components.
- Stored procedures that run on database servers. In complex systems this can include Inter Process Communication (IPC) and other methods.

There is not one single preferred case or solution. No single method will in general enhance security, fault-tolerance, scalability and so on.

First View: Centricity – Programming

Data-centric programming

- The term data-centric programming language refers to programming languages, with the primary purpose for management and manipulation of data. This includes accessing data, lists, structures, tables and so on, especially with data-intensive computing. Sometimes this goes along with dataflow orientation and declarative character.

Examples:

- Structured Query Language (SQL).
- Architecture of MapReduce. (Hadoop Pig . . .).
- High Performance Computing Cluster / Enterprise Control Language (HPCC /ECL).

Working on the content itself is even much more important and much more data-centric!

Centricity context

Relations and conditions: Causalities?

Understanding (data) centricity/locality/layout is significant

- for understanding,
- choosing, and
- creating

advanced solutions, "data-centric", "knowledge-centric", "computing-centric", ...

What means centricity?

Examples / scenarios

- **Data-centric:** Data is fetched from a data resource by processes and delivered to the computing. Data is continuously in creation and development process.
- **Knowledge-centric:** Knowledge is in the focus. Content is carrying knowledge data. Computing is a tool. Knowledge is continuously in creation and development process.
- **Computing-centric:** Processes communicate data to where the computing is taking place. Parametrisation and initial data are the start for computing results.
- **Integrated:** Any. In many overall cases data/knowledge-centric.

Which major scenarios exist?

Different conditions: Scales, data, and goals

- Capability / Turnaround Computing: Grand Challenge computing.
- Capacity Computing: Production runs.
- ...
- (Big) Volume Data
- (Big) Velocity Data
- (Big) Variability Data
- (Big) Vitality Data
- (Big) Veracity Data, ...
- ...
- Libraries (data-centric)
- Knowledge resources (data-centric/knowledge-centric)
- Computational modelling (computing-centric)
- Seismic processing (computing-centric)
- Combinations ...

From discipline/users' view, what are the choices and how?

Caring, . . .

- for the data.
- for data long-term aspects.
- for the Time to Solution (overall).
- for computing access.
- for computing architectures.
- for portability.
- . . .

Why is it important to think about centricity details?

For . . .

- Long-term aspects.
- (Real) projects.
- Project efficiency.
- Project sustainability.
- Job efficiency.
- . . .

Example Scenario

Research project: Data and parties (common scenario)

1) Seismic data	(e.g., SEGY)	computing-centric
2) Geological data	(stratigraphic data)	data-centric
3) Historical data	(data on bibliographic and	data-centric
	other realia objects)	data-centric
4) Archaeological data	(site data)	data-centric
	(simulation data)	computing-centric
5) Multi-disciplinary site data	(knowledge resources)	data-centric
6) Dynamical site data	(referenced data)	computing-centric

---

a) Geophysicist	(project-funded)
b) Geologist	(project-funded)
c) Archaeologist	(project-funded)
d) Information scientist	(project-funded)
e) Third party	(industry)
f) Someone coordinator	
g) Different data creators	different ownership / one contract

Example Data Characteristics

Data characteristics (common scenarios)

Discipline / Application Type	*Size / Range*	*Handling*
Seismic data	GB to TB	Groups of larger homogeneous data sets
Environmental data	MB to TB	Smaller and larger heterogeneous data sets
Knowledge object data	kB to TB	Small to huge arbitrary data sets
Health care data	kB to GB	Small to large and combinations

Why should users take a closer look at their data and workflows?

Example motivation

- Demands for longer data lifecycles, increasing.
- Lifecycles for computing architectures are decreasing.
- Lifecycles for computing services are decreasing.
- ‘Recycling’ data and workflows (availability, compatibility, . . .).
- . . .

Example Lifecycle Data and Computing

Cycles, small and large

- Research task long-term (many decades)
- ...
- Fundamental research
- ...
- Project funding (years)
- Researchers (3-5 years)
- **Data gathering, documentation, usage, discovery, analysis**
- Processing and computing / resources life-cycle (5 years)
- Dissemination, publication, (research data management)
- ...
- Project funding (years)
- Researchers (3-5 years) – different researchers
- **Data gathering, documentation, usage, discovery, analysis – same and comparable data**
- Processing and computing / resources life-cycle (5 years) – different resources
- Dissemination, publication, (research data management)
- ...
- **Long-term data gathering, documentation, usage, discovery, analysis**

High End Content

Knowledge

- **Knowledge** is created from a subjective combination of different attainments, which are selected, compared and balanced against each other, which are transformed, interpreted, and used in reasoning, also to infer further knowledge. Therefore, not all the knowledge can be explicitly formalised. Knowledge and content are multi- and inter-disciplinary long-term targets and values. In practice, powerful and secure information technology can support knowledge-based works and values.

Source: Result of the Delegates' Summit, Symposium on Advanced Computation and Information in Natural and Applied Sciences (SACINAS), ICNAAM, 2015.

Rückemann, C.-P., F. Hülsmann, B. Gersbeck-Schierholz, P. Skurowski, and M. Staniszewski: Knowledge and Computing. Post-Summit Results, Delegates' Summit: Best Practice and Definitions of Knowledge and Computing, September 23, 2015, The Fifth Symposium on Advanced Computation and Information in Natural and Applied Sciences, The 13th International Conference of Numerical Analysis and Applied Mathematics (ICNAAM), September 23-29, 2015, Rhodes, Greece, 2015. Knowledge in Motion / Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), Germany; Silesian University of Technology, Gliwice, Poland; International EULISP post-graduate participants, ISSC, European Legal Informatics Study Programme, Leibniz Universität Hannover, Germany.

High End Content Organisation

Knowledge organisation

- Organisation of knowledge Knowledge requires a universal organisation in order to establish a practical long-term implementation for knowledge objects, which can be flexibly used for varying computing requirements.

High End Computing

Computing

- **Computing** goes along with methodologies, technological means, and devices applicable for universal automatic manipulation and processing of data and information. Computing is a practical tool and has well defined purposes and goals.

Source: Result of the Delegates' Summit, Symposium on Advanced Computation and Information in Natural and Applied Sciences (SACINAS), ICNAAM, 2015.

Rückemann, C.-P., F. Hülsmann, B. Gersbeck-Schierholz, P. Skurowski, and M. Staniszewski: Knowledge and Computing. Post-Summit Results, Delegates' Summit: Best Practice and Definitions of Knowledge and Computing, September 23, 2015, The Fifth Symposium on Advanced Computation and Information in Natural and Applied Sciences, The 13th International Conference of Numerical Analysis and Applied Mathematics (ICNAAM), September 23-29, 2015, Rhodes, Greece, 2015. Knowledge in Motion / Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), Germany; Silesian University of Technology, Gliwice, Poland; International EULISP post-graduate participants, ISSC, European Legal Informatics Study Programme, Leibniz Universität Hannover, Germany.

High End Infrastructure

High Performance Computing (HPC) / Supercomputing

In High Performance Computing, supercomputers -i.e., computer systems at the *upper performance limit of currently feasible processing capacity*- are employed to solve challenging scientific problems.

HPC, Grid, and Cloud

User Level – for some cases

Grid Computing and Cloud Computing can be seen as an user level so to make resources (e.g., computing resources, storage resources) available to a defined extend.
For common use, specific HPC resources can be made available via Grid Computing.

Definition of what Grid Computing is (was)

Grid is a hardware and software infrastructure that allows service oriented, flexible, and seamless sharing of heterogeneous network resources for compute and data intensive tasks and provides faster throughput and scalability at lower costs.

High Performance Computing / Advanced Scientific Computing

Overview

- Requirements
  - Fast Central Processing Unit (CPU).
  - Parallel processing.
  - Large memory.
  - Fast Input/Output (I/O).
  - Powerful communication / networks.
- Hardware / resources
- System / software / configuration
- Applications
- Configuration, optimisation, scaling, ...

Alternatives?

- High Performance Computing.
- Cluster computing.
- Grid Computing.
- Cloud Computing.

Challenge: Which architectures can be considered?

Example products and marketing: Can the differences be named and defined?

- Various:
  - Supercomputing, High Performance Computing 'products'
  - Distributed Computing 'products'
- Sun:
  - Cluster Grids
  - Enterprise Grids
  - Global Grids
- HP:
  - Utility Computing
  - Hybrid Solutions
- IBM:
  - Autonomic Computing, resources, dynamic VO
  - Grid + provisioning via Cloud Computing (SaaS, DaaS, AaaS ...)
- MS:
  - High Productivity Computing
- ...

Challenge: Parallel computing, Software

Different levels can be distinguished on software level:

Job: Whole jobs run parallel on different processors. With this scenario there is no or little interaction between the jobs. Results are better computer utilisation and shorter real runtimes. (Example: workstation with several processors and multitasking).

Program: Parts of a program run on multiple processors. Results are shorter real runtimes. (Example: parallel computer).

Command: Parallel execution between the phases (instructions) of command execution. Result is accelerated execution of the whole command. (Example: serial computer / single processors).

Arithmetic, Bit-level: Hardware-parallel of integer arithmetics and Bit-wise parallel, but not necessarily word-wise serial access on memory or vice versa. Result is less clock cycles for working an instruction.

The levels of parallel computing given here can occur in combination, too.

Challenge: Parallel computing, Hardware

Different levels can be distinguished on hardware level:

Pipelining: Segmentation of operators which are worked consecutively (relevant for vector computers).

Functional units: Different functional independent units for working on (different) operations, e.g., super scalar computers can execute additions, multiplications, and logical operations in parallel.

Processor arrays: Arrays of identical processor elements for parallel execution of (similiar) operations. Example: MasPar computer with 16384 relatively simple processors, systolic arrays for image processing.

Multi processing: Several independent processors with own instruction sets each. Parallel execution is possible up to whole programs or jobs.

Challenge: Architectures, SMP, MPP, MPI . . .

Architecture

SMP: Symmetric Multi-Processing.

MPP: Massively Parallel Processing.

MPI: Message Passing Interface.

OMP: OpenMP, „open“ implementation, SMP/MPI, `http://www.openmp.org/`.

MPICH: MPICH Implementation.

Hybrid: MPI/OpenMP.

. . .

PGAS: Partitioned Global Address Space.

GASPI: Global Address Space Programming Interface.

Challenge: Filesystems

What about Hadoop and Lustre for Supercomputing and Cloud?

Filesystem type	*Examples*
Distributed	NFS, AFS, NCP, CIFS/SMB, XtreemFS, Ceph, Btrfs, HDFS/Hadoop, Tachyon
Shared	SAN, CXFS, GFS, Polyserve, StorNext FS, QFS
Parallel	GPFS, Lustre, PVFS, IBRIX, OneFS, PanFS, NFS/pNFS, BeeGFS

Challenge: Different types of computing and storage architectures

Sides: Computing / storage architectures and data

- High Performance Computing architectures.
- Distributed computing, Grid, Cloud.
- Highly parallel filesystems.
- Large I/O and meta-data systems.
- Highly parallel communication networks.
- Accelerator systems, Graphic Processing Units, ...
- Combinations.
- ...

  $\Longrightarrow$ Data-centricity.

  $\Longrightarrow$ Data-locality.

  $\Longrightarrow$ Data-layout.

Challenge: Different implementations and methods

Sides: Implementation architectures and methods

- Message Passing
- Shared Memory Processing
- ...

  $\Longrightarrow$ Algorithms

  $\Longrightarrow$ Workflows ...

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, Shenzhen, Jaguar, Tianhe, Sequoia, Titan, German supercomputing (HLRB, SuperMUC, JUQUEEN, HLRN, and others) ...
- ⇒ Supercomputing and big data
- ⇒ Operation and infrastructure transition phases
- ⇒ Infrastructures, networks, and architectures
- ⇒ Major long-term and sustainability issues with infrastructures
- ...
- **(All existing supercomputing resources are "individuals" – and different.)**

```
--------------------------------------------------------------------
-- ABOVE EXAMPLES AND OTHER MATERIAL FOR DISCUSSION --
----------- ORIGINALLY ON FOLLOWING PAGES -----------
------------------- LEFT OUT HERE --------------------
```

Challenge: Components, User / Data View

Example Components

- 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, Aries, . . .
  - Service networks.
- Parallel filesystems (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.

Challenge: User / Data, Tender Process and Resulting Lifecycles

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 ...

Challenge: Data employments and life style

Data

- Where data stays
- Where data travels (communication)
- Where data works (computing)
- Where data sleeps
- ...

Challenge: Data employments.

Essentials of Data Employments:

(The following four presentation slides contained copyrighted / proprietary photos, which have been removed for this publication. The textual information from the slides has been summarised on this slide.)

- Where data stays:
  Storage (e.g., Disk Storage Unit)
- Where data travels (communication):
  Networks (e.g., Cabling and Switches)
- Where data travel is channelled:
  Networks, Interconnects, . . . (e.g., Fibre network)
  Remark: Physics Nobelprize 2009 on fibre optics: Charles K. Kao (China). For the groundbreaking achievements concerning the transmission of light in fibers for optical communication. Willard S. Boyle (USA), George E. Smith (USA). For the invention of an imaging semiconductor circuit – the CCD sensor.
- Where data works (computing):
  System resources (e.g., compute nodes: cores and memory)
- Where data sleeps:
  Archive . . .
- . . .

Difference of locality and centricity

Locality

- **Locality:** Place to be at a time.
  - Different character of data: Some like to be at home, others like to travel. Some work alone, others work in groups.
  - Whatever is to be done, there is some central feature or attribute associated with a data character.

Centricity

- **Centricity:** The centre/task where a (more comprehensive) concept is focussing on.
  - If the centre/task is computing then a concept/implementation/architecture is called computing-centric.
  - If the centre/task is the data itself then a concept/implementation/architecture is called data-centric.

Central question and answer

Central question and answer . . . regarding the challenge(s):

(The following five presentation slides contained copyrighted / proprietary photos, which have been removed for this publication. The textual information from the slides has been summarised on this slide.)

- **Question: What does make the essential difference?**

  Examples: HLRN-II: Front Side ICE & UltraViolet Racks

- **Lemma-Question: Is the difference visible?**

  Examples: Different MPP and SMP racks, inside

- **Answer: Architecture and implementation make the difference!**

  Yes, the difference is even visible,
  when you know what to look for!

  Examples: Here, look for the cabling and components.

Disciplines and sample fields

Fields of demand:

- Geophysics, Geosciences, Particle Physics, Cosmology, . . .
- Environmental Sciences, Ocean Modelling, . . .
- Engineering, Computational Mechanics, Computational Fluid Dynamics, Material Sciences, . . .
- Life Sciences, Computational Chemistry, Biology . . .

Examples:

- Seismic Processing, Knowledge Discovery, Molecular Dynamic Structure Analysis, Quantumchemical Simulation, Laminar-Turbulent Transition, Flow Fields, Solar Convection Modelling, Chemical Reactions, Ab-Initio Simulations, 3-D Simulation, Calculation of the Decay, Calculation of Heavy Quark Masses, Climate Modelling, Sound Propagation of Machinery, Hydrodynamics, Global Climate System Effects, Quantum Chromo Dynamics, Molecular Dynamics Simulations, CFD Engineering, Heat Flow Calculation, Aerodynamics, Molecular Dynamics Simulations, Protein Decomposition, Ecosystem Modelling, Simulation of Atmospheres, Calculation of Metal Structures, Laser Material Processing, Sedimentary Modelling, . . .

User perspective on computing resources and tools

Can user/groups easily overview and handle “their” issues:

- Computing, heterogenous resources and configuration?
- Code porting and handling?
- Efficient programming (parallelisation, optimisation, scripting)?
- Data locality, porting, and optimisation?
- Input/output requirements and analysis?
- Memory requirements and analysis?
- Network requirements and analysis?
- Checkpointing on applications?
- Resources policies and exceptions?
- Functional archiving restrictions?
- Data long-term issues?
- Library issues?
- ...

User perspective on data and long-term significance

Sciences and disciplines: Statements from knowledge-and-IT experts:

- **“Persistent data are alpha and omega of scientific research and beyond.”** *Dr. Friedrich Hülsmann, Gottfried Wilhelm Leibniz Bibliothek (GWLB) Hannover, Germany, Knowledge in Motion (KiM) long-term project, DIMF.*
- **“Intelligently structured digital long-term resources can help protect against colateral damages to knowledge such as mankind experienced from the destruction of the library of Alexandria.”** *Dipl.-Biol. Birgit Gersbeck-Schierholz, Leibniz Universität Hannover, Germany, Knowledge in Motion (KiM) long-term project, DIMF.*
- **“Content is the primary long-term target and value and we need powerful and secure information technology to support this on the long run.”** *EULISP post-graduate participants, European Legal Informatics Study Programme, Leibniz Universität Hannover, Germany.*

Resulting Definitions: Data-centricity and Big Data

Data-centric and Big Data (Delegates and other contributors)

- **“The term data-centric refers to a focus, in which data is most relevant in context with a purpose. Data structuring, data shaping, and long-term aspects are important concerns. Data-centricity concentrates on data-based content and is beneficial for information and knowledge and for emphasizing their value. Technical implementations need to consider distributed data, non-distributed data, and data locality and enable advanced data handling and analysis. Implementations should support separating data from technical implementations as far as possible.”**
- **“The term Big Data refers to data of size and/or complexity at the upper limit of what is currently feasible to be handled with storage and computing installations. Big Data can be structured and unstructured. Data use with associated application scenarios can be categorised by volume, velocity, variability, vitality, veracity, value, etc. Driving forces in context with Big Data are advanced data analysis and insight. Disciplines have to define their ‘currency’ when advancing from Big Data to Value Data.”**

Citation: *Rückemann, C.-P., Kovacheva, Z., Schubert, L., Lishchuk, I., Gersbeck-Schierholz, B., and Hülsmann, F. (2016): Post-Summit Results, Delegates' Summit: Best Practice and Definitions of Data-centric and Big Data – Science, Society, Law, Industry, and Engineering; Sep. 19, 2016, The Sixth Symposium on Advanced Computation and Information in Natural and Applied Sciences (SACINAS), The 14th Internat. Conf. of Numerical Analysis and Applied Mathematics (ICNAAM), Sep. 19–25, 2016, Rhodes, Greece.*

Delegates and contributors: *Claus-Peter Rückemann, Knowledge in Motion / Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), Germany;Zlatinka Kovacheva, Middle East College, Department of Mathematics and Applied Sciences, Muscat, Oman; Lutz Schubert, University of Ulm, Germany; Iryna Lishchuk, Leibniz Universität Hannover, Institut für Rechtsinformatik, Germany; Birgit Gersbeck-Schierholz, Friedrich Hülsmann, Knowledge in Motion / Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), Germany*

Example: High End Content - Knowledge

Think of millions of references/objects/media associated with this object.

```
Vesuvius [Volcanology, Geology, Archaeology]:
         (lat.) Mons Vesuvius.
         (ital.) Vesuvio.
         Volcano, Gulf of Naples, Italy.
         Stratovolcano, large cone (Gran Cono) ...
         VNUM: 0101-02=,
         Summit Elevation: 1281\UD{m}. ...
         Syn.: Vesaevus, Vesevus, Vesbius, Vesvius
         s. volcano, super volcano, compound volcano
         s. also Pompeji, Herculaneum, seismology
         %%IML: UDC:[911.2+55]:[57+930.85]:[902]"63"(4+37+23+24)
          =12=14
         %%IML: GoogleMapsLocation: http://maps.google.de/maps?hl=
          de&gl=de&vpsrc=0&ie=UTF8&ll=40.821961,14.428868&spn
          =0.018804,0.028238&t=h&z=15
         ...
         Object:             Volcanic material.
         %%IML: media: ... {UDC:(0.034)(044)770} LXDATASTORAGE:
          //.../img_2401.jpg
```

Object carries names, synonyms, in different lang., dyn. usable geocoordinates, UDC classification ..., incl. geoclassification (UDC:(37), Italia. Ancient Rome and Italy).

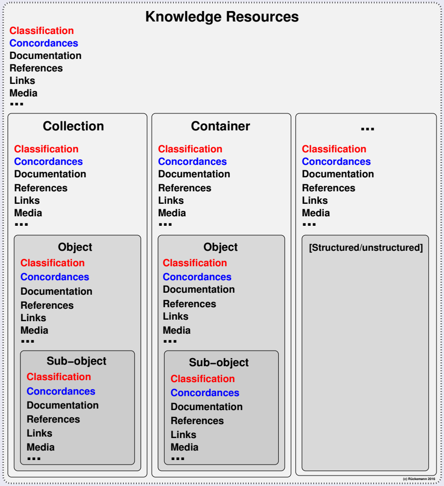
Example: High End Content – Geoscientific Knowledge Resources

Collection and Container References Types used for Processing (excerpt).

References Types		*Group and Implementation Example*
Classification	O & C	UDC
Concordance	O & C	UCC
In-object documentation	O & C	Text
Factual data	O & C	Text, data
Georeference	O & C	Geocoordinates
Keyword	O & C	Text
See	O & C	Text
Reference link	O & C	URL
Reference media	O & C	Link
Citation	O & C	Cite, bib
Content Factor	O & C	CONTFACT
Realia	O & C	Text
Language	O & C	EN, DE
Content-linked formatting	O & C	Markup, LaTeX

Example: High End Content Organisation

Knowledge organisation example: Two-dim. representation, attr./ref. structure



Example: High End Content Organisation

Knowledge organisation example: Two-dim. representation, attr./ref. structure

Knowledge Resources

Classification
Concordances
Documentation
References
Links
Media
...

Collection

Classification
Concordances
Documentation
References
Links
Media
...

Object

Classification
Concordances
Documentation
References
Links

Container

Classification
Concordances
Documentation
References
Links
Media
...

Object

Classification
Concordances
Documentation
References
Links

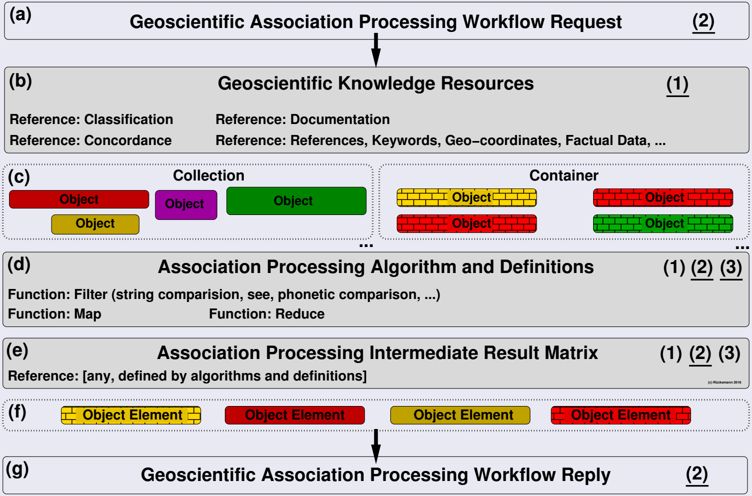
...

Classification
Concordances
Documentation
References
Links
Media
...

[Structured/unstructured]

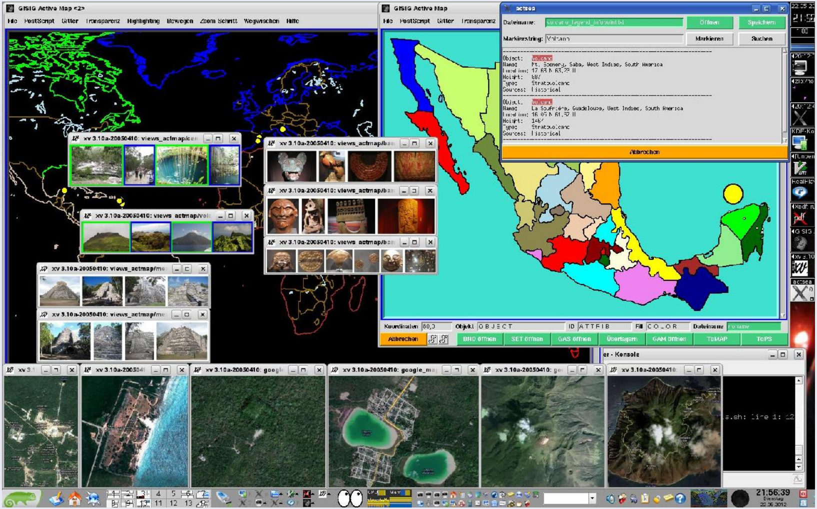
Example: High End Computing – Integration of workflows

Computing: Geoscientific association processing



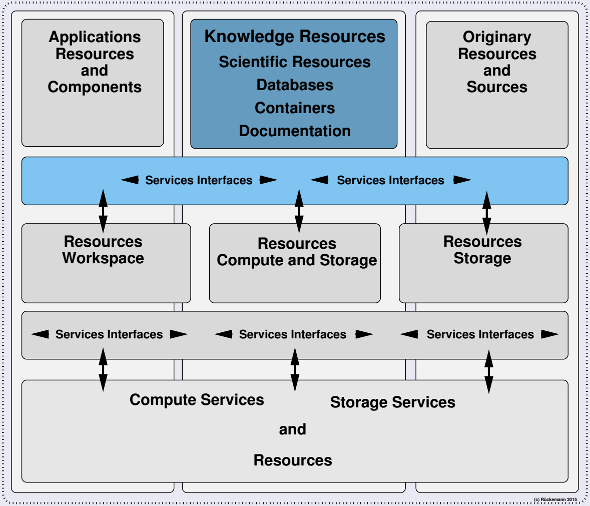
Example: HEC – Integrated Information and Computing System (IICS)

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



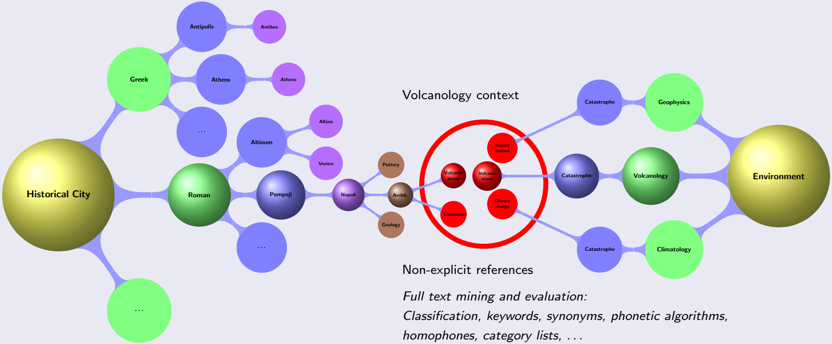
Example Long-term Architecture, Implementation, and Resources

Long-term architecture: Central component: Knowledge resources



Knowledge Discovery Example: Computing object carousel connections

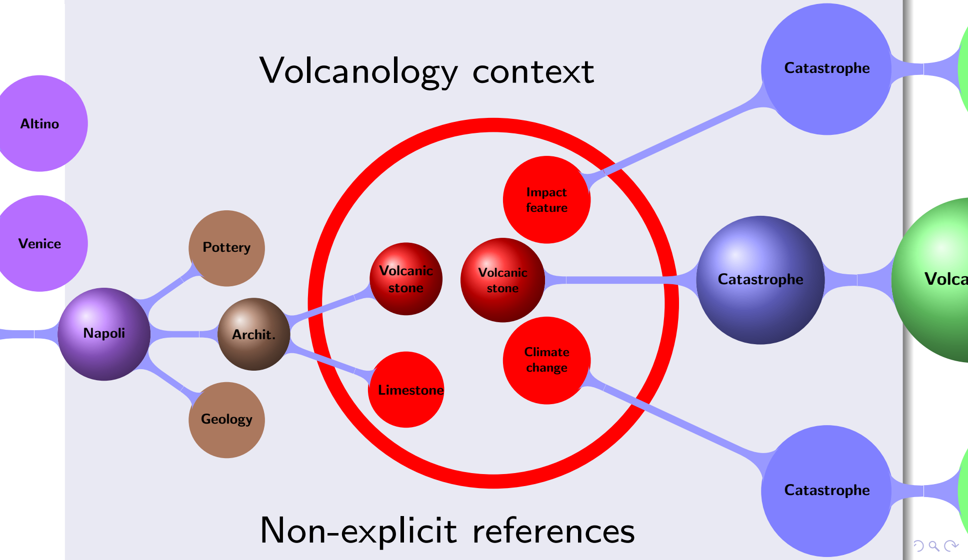
Historical city and environment object carousels, trees with computed references



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

Knowledge Discovery Example: Computing object carousel connections

Historical city and environment object carousels, trees with computed references



Can Big Data be data-centric?

What we can learn from this question

- No. Big Data can rarely be handled for long-term ...
- Yes. We need to consider "data-centric" in the same way we consider data (many "V") and solutions.

What are the consequences of centricity?

Improvements on

- Investments in chances
- Sustainability
- Long-term support
- Overall efficiency
- Data layout
- Documentation
- Re-use
- Data-structure quality
- Focus on individual requirements (solutions?)
- Funding long-term projects (best practice) / dissemination
- Data management

How to handle issues like long-term relevant data, complexity, portability

. . . and what are benefits and tradeoffs?

- Consider lifecycles of data and creation.
- Care for data, knowledge (conceptual, . . .), structure.
- Do not be frightened by complexity (multi-disciplinary, multi-lingual, . . .).
- Portability into the future is in many cases more important than to different present architectures.
- Beneficial: High quality content and structure.
- Tradeoffs: Pretentious learning curve.

Conclusions and Lessons Learned

Centricity, data, and computing:

- What does **value** mean for '**holistic**' scenarios?: *Value beyond economic investments. Value based on content not applicability.*

Conclusions and Lessons Learned

Centricity, data, and computing:

- What does **value** mean for '**holistic**' scenarios?: *Value beyond economic investments. Value based on content not applicability.*
- Where to target **value** and where **complexity**?: *Increase value at data side; decrease complexity at computing side.*

Conclusions and Lessons Learned

Centricity, data, and computing:

- What does **value** mean for '**holistic**' scenarios?: *Value beyond economic investments. Value based on content not applicability.*
- Where to target **value** and where **complexity**?: *Increase value at data side; decrease complexity at computing side.*
- What is the meaning of **centricity** ...?: *Centricity is the centre/task where a concept is focussing in (e.g., in context with data processing).*

Conclusions and Lessons Learned

Centricity, data, and computing:

- What does **value** mean for '**holistic**' scenarios?: *Value beyond economic investments. Value based on content not applicability.*
- Where to target **value** and where **complexity**?: *Increase value at data side; decrease complexity at computing side.*
- What is the meaning of **centricity** ...?: *Centricity is the centre/task where a concept is focussing in (e.g., in context with data processing).*
- What is the **discipline/users' view**? *Care for data, long-term, time to solution, decisions by researchers, . . .*

Conclusions and Lessons Learned

Centricity, data, and computing:

- What does **value** mean for '**holistic**' scenarios?: *Value beyond economic investments. Value based on content not applicability.*
- Where to target **value** and where **complexity**?: *Increase value at data side; decrease complexity at computing side.*
- What is the meaning of **centricity** ...?: *Centricity is the centre/task where a concept is focussing in (e.g., in context with data processing).*
- What is the **discipline/users' view**? *Care for data, long-term, time to solution, decisions by researchers, . . .*
- What are examples of **knowledge creation, discovery, and workflows**?: *Knowledge resources, discovery processes, processing.*

Conclusions and Lessons Learned

Centricity, data, and computing:

- What does **value** mean for '**holistic**' scenarios?: *Value beyond economic investments. Value based on content not applicability.*
- Where to target **value** and where **complexity**?: *Increase value at data side; decrease complexity at computing side.*
- What is the meaning of **centricity** ...?: *Centricity is the centre/task where a concept is focussing in (e.g., in context with data processing).*
- What is the **discipline/users' view**? *Care for data, long-term, time to solution, decisions by researchers, . . .*
- What are examples of **knowledge creation, discovery, and workflows**?: *Knowledge resources, discovery processes, processing.*
- What are **benefits and tradeoffs** . . .?: *Concentrate on long-term insight, utilise complexity, do not focus to optimise on technical aspects.*

Conclusions and Lessons Learned

Centricity, data, and computing:

- What does **value** mean for '**holistic**' scenarios?: *Value beyond economic investments. Value based on content not applicability.*
- Where to target **value** and where **complexity**?: *Increase value at data side; decrease complexity at computing side.*
- What is the meaning of **centricity** ...?: *Centricity is the centre/task where a concept is focussing in (e.g., in context with data processing).*
- What is the **discipline/users' view**? *Care for data, long-term, time to solution, decisions by researchers, . . .*
- What are examples of **knowledge creation, discovery, and workflows**?: *Knowledge resources, discovery processes, processing.*
- What are **benefits and tradeoffs** . . .?: *Concentrate on long-term insight, utilise complexity, do not focus to optimise on technical aspects.*
- Which **architectures and scenarios** can be considered?: *Any, e.g., High End architectures for long-term data and High Performance architectures for other.*

Conclusions and Lessons Learned

Centricity, data, and computing:

- What does **value** mean for '**holistic**' scenarios?: *Value beyond economic investments. Value based on content not applicability.*
- Where to target **value** and where **complexity**?: *Increase value at data side; decrease complexity at computing side.*
- What is the meaning of **centricity** ...?: *Centricity is the centre/task where a concept is focussing in (e.g., in context with data processing).*
- What is the **discipline/users' view**? *Care for data, long-term, time to solution, decisions by researchers, . . .*
- What are examples of **knowledge creation, discovery, and workflows**?: *Knowledge resources, discovery processes, processing.*
- What are **benefits and tradeoffs** . . .?: *Concentrate on long-term insight, utilise complexity, do not focus to optimise on technical aspects.*
- Which **architectures and scenarios** can be considered?: *Any, e.g., High End architectures for long-term data and High Performance architectures for other.*
- Are there **different types of Big Data** and **can Big Data be data-centric**?: *Yes, V. . . Yes, with the solid situational understanding of "data-centric".*

Conclusions and Lessons Learned

Centricity, data, and computing:

- What does **value** mean for '**holistic**' scenarios?: *Value beyond economic investments. Value based on content not applicability.*
- Where to target **value** and where **complexity**?: *Increase value at data side; decrease complexity at computing side.*
- What is the meaning of **centricity** ...?: *Centricity is the centre/task where a concept is focussing in (e.g., in context with data processing).*
- What is the **discipline/users' view**? *Care for data, long-term, time to solution, decisions by researchers, . . .*
- What are examples of **knowledge creation, discovery, and workflows**?: *Knowledge resources, discovery processes, processing.*
- What are **benefits and tradeoffs** . . .?: *Concentrate on long-term insight, utilise complexity, do not focus to optimise on technical aspects.*
- Which **architectures and scenarios** can be considered?: *Any, e.g., High End architectures for long-term data and High Performance architectures for other.*
- Are there **different types of Big Data** and **can Big Data be data-centric**?: *Yes, V. . . Yes, with the solid situational understanding of "data-centric".*
- What are the **consequences of centricity**?: *Care for and invest in data, use computing as a standard tool.*

Future Challenges

Following events:

How can the concentration on benefits of understanding centricity (data-centricity, ...) be fostered?

Future Challenges

Following events:

How can the concentration on benefits of understanding centricity (data-centricity, ...) be fostered?

Overall goals:

- **Invest in a solid situational understanding of centricity, data-centric ...**
- **Concentrate on respective focus/task, not on "products".**
- **Consider data, complexity, long-term aspects as value.**
- **Data fate should become a must in best practice for management processes and funding.**
- **Foster the long-term creation of knowledge and improve the Quality of Data.**
- **Foster multi-disciplinary documentation and work.**

References

References and acknowledgements, see:

⇒ C.-P. Rückemann, "Advanced Association Processing and Computation Facilities for Geoscientific and Archaeological Knowledge Resources Components," in *Proceedings of The Eighth International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2016), April 24 – 28, 2016, Venice, Italy*. XPS Press, 2016, ISSN: 2308-393X, ISBN-13: 978-1-61208-469-5, URL: http://www.thinkmind.org/index.php?view=instance&instance=GEOProcessing+2016 [accessed: 2016-04-24], http://www.iaria.org/conferences2016/ProgramGEOProcessing16.html [accessed: 2016-04-24].

⇒ C.-P. Rückemann, "Enhancement of Knowledge Resources and Discovery by Computation of Content Factors," in *Proceedings of The Sixth International Conference on Advanced Communications and Computation (INFOCOMP 2016), May 22–26, 2016, Valencia, Spain*. XPS Press, 2016, ISSN: 2308-393X, ISBN-13: 978-1-61208-478-7, URL: http://www.thinkmind.org/ [accessed: 2016-03-28], http://www.iaria.org/conferences2016/ProgramINFOCOMP16.html [accessed: 2016-03-28], (in press).

⇒ C.-P. Rückemann, "Fundamental Aspects of Information Science, Security, and Computing," 2007–2015, *(Univ. Lectures). ISSC, EULISP Lecture Notes, European Legal Informatics Study Programme*. Institut für Rechtsinformatik (IRI), Leibniz Universität Hannover, URL: http://www.eulisp.org [accessed: 2016-03-28].

Networking

Thank you for your attention!

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