Integrating Information Systems and Scientific Computing

Claus-Peter Rückemann
Leibniz Universität Hannover,
Westfälische Wilhelms-Universität Münster (WWU),
North-German Supercomputing Alliance (HLRN), Germany
Email: ruckema@uni-muenster.de

***Abstract*—This paper presents the practical results and challenges from the real-life implementation of interactive complex systems in High End Computing environments. The successful implementation has been made possible using a new concept of higher-level data structures for dynamical applications and configuration of the resources. The discussion shows how an implementation of integrated information systems, compute and storage resources has been achieved. The implementation uses techniques ensuring to create a flexible way for communication with complex information and advanced scientific computing systems. Besides Inter-Process Communication these are mainly Object Envelopes for object handling and Compute Envelopes for computation objects. These algorithms provide means for generic data processing and flexible information exchange. Targeting mission critical environments, the interfaces can embed instruction information, validation and verification methods. The application covers challenges of collaborative implementation, legal, and security issues as well as scientific documentation and classification with these processes. The focus is on integrating information and computing systems, Distributed and High Performance Computing (HPC) resources, for use in natural sciences disciplines and scientific information systems. Implementing higher-level data structure frameworks for dynamical applications and resources configuration has led to scalable and modular integrated public / commercial information system components.**

Keywords–Integrated Systems; Information Systems; Advanced Scientific Computing; Geosciences; High Performance Computing; Computing Systems; Documentation; Classification.

I. Introduction

Based on the implementation of components for Integrated Information and Computing Systems (IICS) and resources, several aspects have been studied [1]. Besides the challenges of complex integrated systems, geoscientific and technical issues, the aspect of multi-disciplinary international projects managing and efficiently operating complex systems has been in the focus.

Regarding the feasibility of dynamical interactive applications and high end resources the goal is definitively a matter of "capability" computing. Whereas the interactive tasks should be considered to represent massively dynamical interactive requirements. Classical capability computing requirements can only be considered of secondary level importance compared to the dynamical interactive challenges of these application scenarios.

The prominent "Information System" components still completely ignore these advanced aspects and abilities for integration and computing. In High Performance Computing (HPC) supercomputers, that means computer systems at the upper performance limit of current technical capabilities for computing, are employed to solve challenging scientific problems. In consequence there is no general or common architecture and configuration for HPC resources as in the lower parts of the performance pyramid. Within the last decades a large number of implementations of information systems, computing and storage systems and other resources have been created. Nearly all of these implementations lack features for extending information systems with the various resources available. Thus, the integration can open advances using larger resources, interactive processing, and reducing time consumption for assigned tasks. Most of these applications and resources are very complex standalone systems and used that way, neglecting, that for many sophisticated use cases conjoint applications are desirable.

This paper is organised as follows. Section II presents motivation, challenges and complexity with the implementation. Section III shows the previous work continued with this research. Section IV summarises the targeted properties and challenges with IICS, distributed information, structuring information as well as unstructured information, large data volume, event multiplicity and context implementation issues. Section V presents two complex application scenarios. Sections VI and VII describe the prerequisites and basic resources configuration for the implementation. Sections VIII and IX show the components and dependencies for integrated systems and resources. Section X discusses the time and response dependence of the integrated solutions. Section XI describes the system implementation, and Section XII presents the evaluation. Sections XIII and XIV summarise the lessons learned, conclusions, and future work.

II. Motivation

With the implementation use cases for Information Systems, the suitability of Distributed and High Performance Computing resources have been studied. These use cases have focus on event triggered communication, dynamical cartography, compute and storage resources. The goal has been, to bring together the features and the experiences for creating and operating a flexible, multi-disciplinary IICS. An example that has been implemented is a spatial information system with hundreds of thousands possible ad-hoc simulations of interest, being used for geomonitoring as well as for geoprocessing. This can be implemented as a map for which any of a large number of Points of Interest (PoI) can be linked with on-demand visualisation and media generation, e.g., on-site videos on location, geology, weather. Within these interactive systems as many "next information of interest" as possible can be dynamically calculated in parallel, near real-time, in order to be of any practical use. These compute intensive tasks can, e.g., be triggered by interactive mouse movement and require to precompute the secondary information for all PoI currently seen on a screen in order to usable in time. In the following passages, we will show environmental components exactly using this implementation for many thousands of PoI. Due to the complexity of IICS, we have applied meta-instructions and signatures for algorithms and interfaces. For these cases, envelopes and Inter-Process Communication (IPC) has been used to provide a unique event and process triggered interface for event, computing, and storage access. The focus with massively interactive applications is naturally on the number of parallel requests. I/O could not been tested within the presented case study scenarios. The interactive handling of 10000 parallel application requests requires on one hand a large number of available system resources and on the other hand latencies as low as possible. For complex systems there are important properties to consider and therefore the components require some overall features, not limited to, but essentially for:

- Structuring,
- Referencing,
- Standardisation,
- Portability,
- Versioning,
- Extendability,
- Certification,
- Meta and object level,
- Ability for distributed application scenarios,
- Connectivity and interfaces.

The following sections discuss all of these aspects in detail and explain how solutions have been created.

III. Previous work

The next generation of systems necessary for providing profound means for communication and computation will have to gather methods evolved by active interdisciplinary interchange, grown with the requirements of the last decades: The information and computing disciplines need means for "in praxi" collaboration from disciplines, structural sciences, computer science, computing science, and information science. Examples are computing intensive interactive environmental monitoring and information systems or simulation supported dynamical geoinformation systems. In this manner, an efficient development and operation can be put into practice for making interactive use of systems with tenthousands of thousands of nodes, ten to hundred thousands of compute cores and hundred thousands to millions of system parts like memory bars and hard disk drives. Methodological sciences means sciences of developing methods for using resources and techniques for gathering new scientific insights. For years, "methodological sciences" or more precise "methodological techniques" have been commonly propagated to solve the problems of High End Computing. It has been commonly experienced that this is not true as there are no applicatory results regarding the essential requirements of complex and integrated high end systems. The available "methods and techniques" is not what we can use for supercomputing where every application and system architecture is different. Unfortunately up to now, this difference is implicit with common tender and operation strategy for so called "efficiency and economical" reasons.

The experiences with integrated systems have been compiled in various projects over the last years [2], [3] and legal issues and object security have been internationally discussed [4], [5]. The architecture of the framework and suitable components used, have been tested by implementing various integrated systems. The following sections show components of an integrated geoscientific information and computing systems developed in one of these case studies that can be used for environmental monitoring or feeding expert systems. None of the participating parties from industry and scientific disciplines can or will create one single application from the components discussed here. The goal is to enable the necessary operation and software stack, nevertheless, the components are modular entities. For the last years practical solutions to various requirements for communication requests have been implemented in a number of projects and case studies using various resources [6], [7], [8], [9], [10]. The most important communication facilities for IICS are:

- Communication requests with applications (example: Inter-Process Communication, IPC).
- Storage object requests (framework example: Object Envelopes, OEN).
- Compute requests (framework example: Compute Envelopes, CEN).

The idea of dynamical event management [11] and envelope-like descriptive containers has been inspired by the good experiences with the concept of Self Contained applications (SFC). Based on these components an integrated solution has been built, for use with local HPC resources supported by distributed information and compute resources. From the point of view of resources providers and integrators of HPC resources it would make very little sense to describe the application components here. Applications details have been published for several components before. For the core issues the conceptional results are by far the most important.

IV. Targeted properties and challenges

The presented approach for Integrated Systems delivers a set of features targeting essential properties regarding all aspects for the components, like with information, algorithms, and dynamical applications:

- Distributed information,
- Documentation and structuring information,
- Unstructured information,
- Large data volume,
- Event multiplicity,
- Resources integration,
- Context,
- Abstraction with context,
- Legal issues,
- Experience and qualification,
- Protection and privacy,
- Complex application scenarios.

The concept and implementation are aimed to provide a feasible long-term perspective and support a holistic level of sustainability.

A. Distributed information

The overall information is widely distributed and it is sometimes very difficult and a long lasting challenge not only to create information but even to get access to a few suitable information sources. The goal for these ambitions is an integrated knowledge base for archaeological geophysics. It will be necessary to collect data from central data centers or registers [12]. From the information point of view, an example for suitable archaeological and geophysical data that has been collected is from the North American Database of Archaeological Geophysics (NADAG) [13] and the Center for Advanced Spatial Technologies (CAST) [14] as well as with the work of the Archaeology Data Service (ADS) [15]. It must be emphasised that there is neither a standard being used for one discipline nor an international standard. All participating disciplines, services, and resources have to be prepared for challenges as big data, critical data, accessibility, longevity, and usability. The concept of this framework is designed to consider these aspects and in order to handle any objects as with the Center of Digital Antiquity [16] and with the Digital Archaeological Record (tDAR) [17], the United States' largest digital store of global archaeological data. In some cases, even concepts for active and smart cities have needed large efforts for collaboration and policies, e.g., with the Rio Operations Center, the public information management center for Rio de Janeiro, Brazil. It integrates and interconnects information from multiple government departments and public agencies in the municipality of Rio de Janeiro in order to improve city safety and responsiveness to various types of incidents, such as flash floods and landslides [18].

These efforts of using information can be supported by integration approaches based on semantics to be used for many purposes, e.g., as a guide [19], [20]. For all components presented, the main information, data, and algorithms are provided by the LX Foundation Scientific Resources [21], containing all the necessary structure and information to support any kind of implementation.

B. Structuring information

The Universal Decimal Classification (UDC) [22] is a hierarchical, multi-lingual and already widely established classification implementing faceted analysis with enumerative scheme features, allowing to build new classes by using relations and grouping. UDC is by definition multi-disciplinary. In multi-disciplinary object context [23] a faceted classification does provide advantages over enumerative concepts. Composition / decomposition and search strategies do benefit from faceted analysis. It is comprehensive and flexible extendable. A classification like UDC is necessarily complex but it has proved to be the only means being able to cope with classifying and referring to any kind of object. It is working with international functional applications, e.g., in applied sciences and medicine. Copies of referred objects can be conserved and it enables searchable relations, e.g., for comparable items regarding special object item tags. The UDC enables to use references like for object sources, may these be metadata, media, realia, dynamical information, citation as via BibTeX sources or Digital Object Identifier (DOI) [24] as well as for static sources. With interactive and dynamical use for interdisciplinary research the referenced objects must be made practically available in a generally accessible, reliable, and persistant way. A DOI-like service with appropriate infrastructure for real life object services, certification, policies and standards in Quality of Service, for reliable long-term availability object, persistency policies should be available. Therefore, for any complex application, these services must be free of costs for application users. It would not be sufficient to build knowledge machines based only on time-limited contracts with participating institutions. These requirements include the infrastructure and operation so data availability for this long-term purpose must not be depending on support from data centers providing the physical data as a "single point of storage".

C. Unstructured information

Unstructured information, the data variety, is one major complexity. For relational databases, a lot of players providing offerings in this space go through the cycle of what the needs are for structured data. As one can imagine, a lot of that work is also starting for unstructured or semi-structured data with Integrated Systems. Data access and transfer for structured data, unstructured data, and semi structured data may be different and may to a certain extend need different solutions for being effective and economic [25].

D. Large data volume

Effective handling of large data volume is increasingly important for geosciences, archaeology, and social networks in any complex context with IICS and High End Computing [5]. The large data challenge is immanent for scientific as well as for social network application scenarios.

Efficient parallel high end access on large volume data from thousands of cores is essentially depending on hardware appliances. These will be I/O appliances, internal networks (like InfiniBand architectures), and large highly parallel filesystems. This is a matter of scalability of hardware and respective software application. Splitting large volume data into many smaller files for highly parallel access can even exacerbate the problem. So with such physical constraints a general overall efficient and cost-effective solution cannot exist today.

There are scientific application scenarios where data volume cannot be reduced and where no data can be filtered and deleted and otherwise there are communication and networking implementations that may make use of filtering and data reduction. So, for complex situations we will certainly need both strategies, advancing resources and technologies development and data reduction. With the original development of dynamical components like Active Source “Kacheln” (tiles) have first been used [11] with data material, mostly based on precomputed static data, supporting the dynamic increase of the number of server resources. This precomputation is an important means in order to provide scalable services. The same aspect of Cloud scalability holds true for communication data required for application scenarios, e.g., with Meta Data Services (MDS).

Regarding High End Computing requirements, sciences can be distinguished in data intensive and compute intensive. Besides to computation with denser grids and the goals to compute for more details, the increased sensor data and mobility have increased data volume drastically, too [26]. Simple and homogeneous infrastructures for computing and I/O are widely considered best practice with most kind of research, even with ambitioned projects. Researchers and people operating technology must be aware that otherwise the infrastructure is likely to get the greatest challenge. Especially for various complex application scenarios large data volumes, compute power, and staff are needed.

Distributed resources, e.g., “Clouds”, “Nebulas”, “Superclouds”, “Skies”, are suitable for a wide number of applications [27], [28], [29] but they do have a major problem with data intensive sciences: data transfer. Not only that with large volumes this is expensive, in many case it is even technologically not feasible in any economic way.

E. Event multiplicity

Data and information velocity is to a certain extend depending from physical constraints. In many cases the end-to-end workflow from gathering data to interactively working with the information has to be considered. Table I shows the number of massively interactive communication events (n_e), number of cores (n_c), and overall response time (t_r) for integrated systems as has been implemented with this research and which is a strong aspect of motivation to use high end resources with interactive communication.

Table I
INTERACTIVE COMMUNICATION, RESOURCES, AND RESPONSE TIME AS BEING THE RESULT OF THE IMPLEMENTATION.

n_e	n_c	t_r (sec) (event multiplicity×event time)
100000	10000	10×5
10	100	1×5
1	1	1×5

This has been implemented and successfully verified for various dynamical applications using parallel compute and storage requests for geoprocessing triggered from a geospatial application. The parallel architecture of the interactive components, software and hardware, allows to precompute objects needed for interactive mode in very short time. Event multiplicity means the number of parallel events in relation to the number of available resources. Related to the event multiplicity, there are various possible cases, for example regarding cores with single core instances. The simplest one-one case, one event – one core, needs 5 seconds for the application to deliver the result. In cases with spare capacity needed, e.g., mission critical buffer capacities, these must be reserved for appropriate usage scenarios. In some cases the spare capacity will be needed per instance.

In High End Computing the multi-user performance is physically depending on hardware. If, with the above single event response time, for 10 events one hundred cores are available, there is spare capacity for several instances. In case of an event overload, which can easily be reached with integrated dynamical systems, e.g., interactive processing events bound to locations in spatial information system interfaces, the response time will get longer. The number of 100000 events is quite a realistic number per application instance. So even with 10000 cores available the interactive use will get limited response times.

F. Resources integration

For seamless operation the required resources have to be integrated into the application scenario. Availability of integrated resources can be provided by policies. The constraints, e.g., with hardware resources, energy efficiency

[30], or sustainable funding are bound to these, so being requirements that cannot be solved with case studies.

With High End Computing and Supercomputing it is the hardware defining the granularity. In general, it is bad for the overall system to splinter into hardware heterogeneity. Heterogeneity at the hardware level can result in multiple granularity as, e.g., with different architectures, separate clusters, various CPU and GPU architectures, concepts of separate physical memory, separate physical networks and so on. Whereas the theoretical overall compute power may be high, what is much more important is that the implementation and porting for these resources is most ineffective. The same holds true for the operation of heterogeneous resources. So in most cases it will be responsible to support homogeneous resources.

Batch and Scheduling has been implemented with many tools available although there is no general best solution because this will depend on the application scenario and resources available, with cluster, Grid, and Cloud [31], [32], [33], [34], [35], [36]. Parallelisation has been implemented both loosely coupled and using application triggering, OpenMP and MPI [37]. For example, components requiring wave analysis can utilise MPI or OpenMP support from additional frameworks and parallelisation support, e.g., IWAVE for the construction of regular grid finite difference and finite element methods for time-dependent partial differential equations [38]. Application performance can be analysed, e.g., with support of integrated tools [39]. A general support of High End Computing resources though is not possible with todays heterogeneous technology [40] so as there is no unifying framework, besides advanced scientific computing applications virtualisation as with the Parallel Virtual Machine (PVM), we have to live with and support several strategies. The appropriate information system resources and frameworks depend on the main focus of the application scenario. For example, in addition to geoscientific and geophysical processing, with environmental and spatial information in international context the GEOSS and GMES [41], [42], [43], [44] have been considered.

G. Context

Context is any information that can be used to characterise the situation of an entity. An entity is a person, place, or object that is considered relevant for the interaction between a user and an application, including the users and applications themselves. For long-term usage, software and hardware context cannot be certified in any suitable way today.

H. Abstraction with context

Future applications in advanced scientific computing need a much higher level of abstraction for communication and parallelisation. For many advanced scientific computing applications virtualisation as with the Parallel Virtual Machine (PVM) system is, instead of often long-winded system-specific rewriting of basic Message Passing use, widely preferred for portable and efficient use of shared and local-memory multiprocessor systems, in case of heterogeneous high end resources. Most applications can be integrated into present applications, using various interfaces, e.g., directly from Active Source or with the tkpvm interface. In addition, very complex applications supporting PVM are available, for example for seismic processing, scientific calculation and visualisation, raytracing and many more fields of advanced scientific computing.

I. Legal issues

In order to find solutions to those issues of Integrated Systems, large data volume, legal aspects of Cloud and Distributed Computing [45], as well as monitoring and forensics are actually discussed [46].

J. Experience and qualification

Experience and expertise are a most important value [47]. In High End Computing expert experience from scientific fields with application scenarios is getting increasingly significant. Non-scientific and low level skills, like parallel programming, operation or administrative experience is drastically loosing importance from the discipline point of view. Accordingly, optimisation is not feasible in most cases. The term that will describe the process best is configuration with complex and scientific applications:

- High level applications available.
- Easy to use frameworks, exploiting parallel computing resources.
- Scientists' working time is more important than optimisation. This is even more prominent if a long verification phase for algorithms and results is the consequence.
- Scientific staff is not funded for low level tasks.

K. Protection and privacy

Securing privacy protection of information for future Information and Computing Systems, e.g., with archaeological sites, is a crucial challenge [4] and requirements even increase with the increasing data volume [5], [48]. So data security has been recognised a key issue for future IICS development and components of collaboration. It needs multi-disciplinary efforts in order to implement a global framework considering privacy in such complex environments.

V. Complex application scenarios

With the implementation of use cases, especially regarding complex Information Systems, the suitability of Distributed and High Performance Computing resources supporting processing and computing of geoscientific data have been studied, e.g., with geosciences, archaeology, and environmental sciences. One of the most prominent examples for a future integration is archaeological geophysics [16], [17], [13], [14], [15].

A. Archaeology and geosciences

Currently basic analysis is being worked on for a principle solution with Archaeological IICS considering the software stack, documentation, structure, classification, and hardware. If so individually available, this should be possible without restructuring complex data all the time when migrating to different architectures and to be prepared for future resources. In case of archaeological, cultural heritage, and geoscientific information and computing systems, there is strong need for integration of different data, information, and sources with scientific computing capabilities, e.g., object information (multi-medial, photographical, textual, properties, sources, references) from natural sciences and humanities. The dynamical system components for this application scenario must enable advanced features like

- weave n-dimensional topics in time,
- use archaeological information in education,
- integrate sketch mapping,
- implement n-dimensional documentation,
- provide support by multi-disciplinary referencing and documentation,
- do discovery planning,
- and structural analysis,
- as well as multi-medial referencing.

In addition, using large data volume resources in an integrated environment like with archaeological geophysics can help to advance the state of the art in integrated systems and accelerates the pace of development for multi-disciplinary solutions and technologies as well as it increases the potential for discoveries and cognitive facilities.

B. Geoprocessing

Geoscientific data processing, "geoprocessing", means processing of geophysical and other data related to the earth or earth-like objects. Today without appropriate background in natural sciences the term is sometimes erroneously reduced to spatial data. Geophysical data processing does have completely different requirements than needed for handling of spatial data and may require a very compute intensive workflow of processes from preprocessing up to postprocessing. In many cases like Seismics or Ground Penetrating Radar although there is compute power available [49] this cannot be done automatically as geological and other information has to be analysed and evaluated. So professional experience is a most important part of the workflow that even cannot be fully automated itself.

VI. System prerequisites

System prerequisites can be considered to be part of the environment, which may be optimised for the application scenario. In almost any case with provisioning of high end resources this cannot be done by user groups. Therefore, the resources providers have been included into the process, as this is a core aspect of the "Collaboration house" framework, which has been developed for this kind of scenarios and which has been presented here (Figure 3).

For implementation and testing a suitable system architecture and hardware had been necessary. A single local system had to fulfill the following minimal criteria:

- Capacity for more than 5000 subjobs per job.
- At least one compute core available per subjob.
- Interactive batch system.
- No distributed compute and storage resources.
- Fast separate InfiniBand networks for compute and I/O.
- Highly performant parallel filesystem.
- Available for being fully configurable.

A system provided being fully configurable means especially configuration of hardware, network, operating system, middleware, scheduling, batch system. At this size this normally involves a time interval of at least three to six months.

It should be obvious that there are not many installations of some reasonable size and complexity that could be provided, configured and operated that way if in parallel to normal operation and production.

The available HPC and distributed resources at ZIV and HLRN [6], [7], [8], [9], [10]. as well as commercially provided High End Computing installations have been sufficient to fulfill all the necessary criteria.

VII. Basic resources configuration

With the systems used for the implementation some operating systems, middlewares, and communication have been available. In almost any case with high end resources in production this cannot be done by user groups. The resources providers have been included into the process.

Elementary operating system components on the resources involved are: AIX, Solaris, and various Linux distributions (SLES, Scientific Linux). Elementary middlewares, protocols, and accounting systems used for the integrated components are: Globus Toolkit, SGAS, DGAS. Unicore, SAGA, SOAP, and many others can be integrated, too. For communication and parallelisation MPI (Open-MPI [37], MPI from SGI, Intel, HP, IBM), OpenMP, MPICH, MVAPICH and other methods have been used along with IPC regarding to the type of operation and optimisation of algorithms needed. For scheduling and batch systems the resources used Moab, Torque, LoadLeveler, and SGE.

All these "tools" are only middleware, protocols, interfaces or isolated applications. They are certainly used on the system resources but they cannot integrate anything, not on the disciplines / application level, not on the services level, not on the resources level. So we want to concentrate on the important high-level issues for the further advanced view of components.

VIII. 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.

For demonstration and studies flexible and open Active Source Information System components have been used for maximum transparency. This allows OO-support (object, element) on application level as well as multi-system support. Listing 1 shows a simple example for application communication with framework-internal and external applications (Inter-Process Communication, IPC).

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

Listing 1. Application communication (IPC).

This is self-descriptive Tcl syntax. In this case, the IPC `send` is starting a molecular graphics visualisation tool and catching messages for further analysis by the components.

Listing 2 shows an example of how the communication triggering can be linked 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}}
```

Listing 2. Application component triggering.

Tcl language objects like `text` carry tag names (`relictrotatex`) and dynamical events like `Button` events are dynamically assigned and a user defined subroutine `sendAllRasMol` is executed, triggering parallel visualisation. Storage object requests for distributed resources can be done via OEN. Listing 3 shows a small example of a generic OEN file.

```
<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>
```

Listing 3. Storage object request (OEN).

OEN are containing element structures for handling and embedding data and information, like `Filename` and `Content`. An end-user public client application may be implemented via a browser plugin, based on appropriate services. With OEN instructions embedded in envelopes, for example as XML-based element structure representation, content can be handled as content-stream or as content-reference. Algorithms can respect any meta-data for objects and handle different object and file formats while staying transparent and portable. Using the content features the original documents can stay unmodified.

The way this will have to be implemented for different use cases depends on the situation, and in many cases on the size and number of data objects. However the hierarchical structured meta data is uniform and easily parsable. Further it supports signed object elements (`Signature`), validation and verification via Public Key Infrastructure (PKI) and is usable with sources and binaries like Active Source.

Compute requests for distributed resources are handled via CEN interfaces [50]. Listing 4 shows a generic CEN file with embedded compute instructions.

```
<ComputeEnvelope><!-- ComputeEnvelope (CEN)-->
<Instruction>
<Filename>Processing_Batch_GIS612.pbs</Filename>
<Md5sum>...</Md5sum>
<Sha1sum>...</Sha1sum>
<Sha512sum>...</Sha512sum>
<DateCreated>2010-08-01:201057</DateCreated>
<DateModified>2010-08-01:211804</DateModified>
<ID>...</ID>
<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>
```

Listing 4. Compute request (CEN).

Content can be handled as content-stream or as content-reference (`Content`, `ContentReference`). Compute instruction sets are self-descriptive and can be pre-configured to the local compute environment. In this case, standard PBS batch instructions like `walltime` and `nodes` are used. The way this will have to be implemented for different use cases depends on the situation, and in many cases on the size and number of data objects.

An important benefit of content-reference with high performant distributed or multicore resources is that references can be processed in parallel on these architectures. The number of physical parallel resources and the transfer capacities inside the network are limiting factors.

IX. INTEGRATED SYSTEMS WITH COUPLED RESOURCES

Figure 1 shows the applied integration of the information and communication systems with coupled computation resources, namely compute resources and storage resources. For integrating the features of information and communication systems with powerful compute resources and storage, it has been necessary to implement interfaces and software applications being able to efficiently use the benefits of High End Computing resources.

Following the results of the long-term case studies [51] three columns namely disciplines (as geosciences), services (as middleware and compute services), resources (computing and storage) had to be figured out for this scenario.

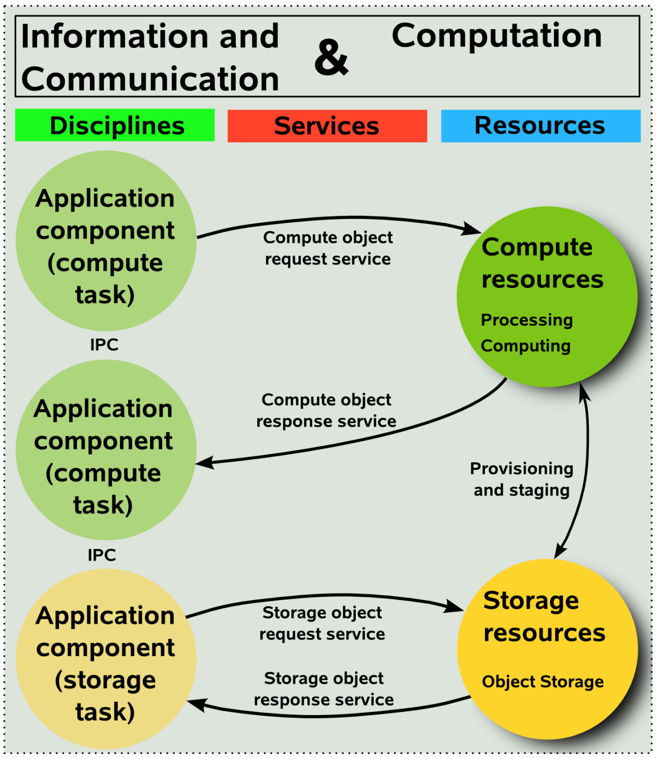

Figure 1. Integrated systems: disciplines integrating resources.

The discipline column shows application components with the state for a compute task and an application component with state for a storage task. Local tasks, ordinary communication between the applications without the need for external computing power, can as usual be done using a local service, for example using Inter-Process Communication (IPC).

Using services, requests can be sent to the configured compute object request service for compute intensive tasks. Results delivered from the computation are delivered for the compute object response service, giving the desired information back the one of the application components. Compute Envelopes (CEN) can be used for exchange of the compute requests. The resources column does provide compute resources for processing and computing as well as storage resources for object storage. Commonly, these resources are separated for backend use with high end applications customised on the compute resources. Application components may trigger storage tasks using a storage object request service. Data objects are handled by the service and delivered to the storage resources. Request for retrieval from the storage are handled by the storage object response service. Object Envelopes (OEN) can be used for exchange of the object requests. For enabling overall scalable integrated systems, mostly for large data volume, the computing and storage resources can communicate for using stored data from within compute tasks and for provisioning and staging of data. These services are so far using a loosely coupled parallel computing, parallelised on the application component level. Each single task can itself contain scalable and loosely to highly parallel computing jobs running on the available compute resources. MPI and OpenMP can be used here. The CEN Envelopes are used to transfer the tasks and their description. The user has to ensure that with using the resources the interactivity and latencies of the integrated system still result in appropriate and usable comprehensive system. Among the compute and storage resources a provisioning and staging mechanism for data and resources requests and responses can be used. Therefore, triggering of computing for storage operations and triggering of storage operations for computing are available.

X. TIME DEPENDENCE

The same reason why opening large resources for information system purposes is desirable, there is still a dependence on time consumption for interactive and batch processing. Table II shows the characteristic tasks and times that have been considered practical [51] with the current information system applications, for example with environmental monitoring and information system components.

Table II
TASKS AND TIMES REGARDING THE OVERALL INTEGRATED SYSTEMS.

Task	*Compute / Storage Times*
In-time events requests	1–3 seconds
Interactive requests	up to 3 minutes
Data processing	1–24 hours
Processing data transfer	n days
Object storage interval	n weeks
Object archive interval	n years

The different tasks afford appropriate policies for interactive and batch use inside the resources network. Besides that, the user and the developer of the application components can use the computing and storage interfaces in order to extend the application facilities using these non-local resources.

Nevertheless, for configuring the system and for implementing new operations the decisions have to be made, which type of implementation would be more suitable.

Interactive request are mostly not acceptable when response times are longer than a few minutes, even for specialised information systems. HPC systems have shown a good performance for parallelisation of interactive subjobs, being in the range of minutes. Whereas distributed resources are much less scalable and provide less performance due to smaller and mostly different resource architecture types and non-exclusive use. Compute times for 1 to 24 hours will force to decide about the field of operation of the system application, when assigning the tasks and events. For example, those compute resources doing computation on large jobs are the computational bottleneck for interactive use. On the other hand, for information system purposes, for example needing visual updates within longer intervals, like for special monitoring purposes for environmental, weather, volcano or catastrophes monitoring and using remote sensing, this scenario is very appropriate. Storage resources and object management can reduce the upload and staging times for objects that can be used more than once. Service providers are confronted with the fact that highly performant storage with reliable and long time interval archiving facilities will be needed at a reasonable price.

XI. Implemented system

The system implemented integrates the component features described from the projects and case studies. Figure 2 shows the implementation of the integrated systems and resources. As Resources Oriented Architectures (ROA) and Services Oriented Architectures (SOA) in themselves are not sufficient for a sustainable long-term development, an important aspect here are the disciplines and their application scenarios with a definable but loosely coupling of services and resources. The components were taken from the GEXI case studies and the well known actmap components [51], [52]. These components handle information like spatial and remote sensing data, can be used for dynamical cartography and trigger events, provide IPC and network communication, and integrate elements from remote compute and storage resources as available with existing compute resources [8], [9], [10]. Processing and computing tasks can for example consist of raytracing, seismic stacking, image transformation and calculation, pattern recognition, database requests, and post processing. The modularisation for development and operation of advanced HPC and application resources and services can improve the multi-disciplinary cooperation. The complexity of operation and usage policies is reduced.

A. Application components

The integrated system is built in three main columns, application components in use with scientific tasks for various disciplines, meaning the conventional scientific desktop and information system environment, services, and resources. These columns are well understood from the "Collaboration house" framework (Figure 3). As well as analysing and separating the essential layers for building complex integrated systems, it is essential that these allow a holistic view on the overall system, for operation, development, and strategies level. The collaboration framework developed [53] and studied for integrated information and computing is the "Collaboration house" [54], [55] (with Grid Computing and GIS initially known as Grid-GIS house and GIS house).

The present framework is considering the respective state of the art resource provisioning for establishing IICS based on multi-disciplinary international collaboration (Figure 3). Implementations in general are complex and can only be handled with a modular architecture being able to separate tasks and responsibilities for resources (HPC, HEC with Grid, Cloud, Sky Computing), services, and disciplines. Thus, the framework supports multiple tenants and groups as well as sole tenants. In opposite to the conventional isolated usage scheme, interaction and communication is not restricted to happen inside the disciplines and resources columns only. Non-isolated usage and provisioning can speed up the development of new components and the modification of existing components in complex environments. The workflows with the application scenarios (Figure 2) are:

a) Application communication.
b) Storage task.
c) Compute task.

These tasks can consists of a request, triggered by some event, and a response, when the resources operation is finished. The response can contain data with the status or not, in case that for example an object has been stored on the resources. Based on this algorithm, task definition can be reasonably portable, transparent, extendable, flexible, and scalable.

B. Application communication

a) Request: The internal and framework-external application is triggered from within the framework components (rasmol is used in the example). From within an actmap component a task to an application component is triggered. IPC calls are used with data and information defined for the event.

Response: A framework-external application is started (rasmol locally on the desktop). The external application can further be triggered from the applications available.

C. Storage task

b) Request: From within an actmap component a storage task is triggered. The stored OEN definition is used to transmit the task to the services. The services do the validation, configuration checks, create the data instructions and initiate the execution of the object request and processing for the resources.

Response: The processing output is transmitted to the services for element creation and the element (in this example a photo image) is integrated into the actmap component.

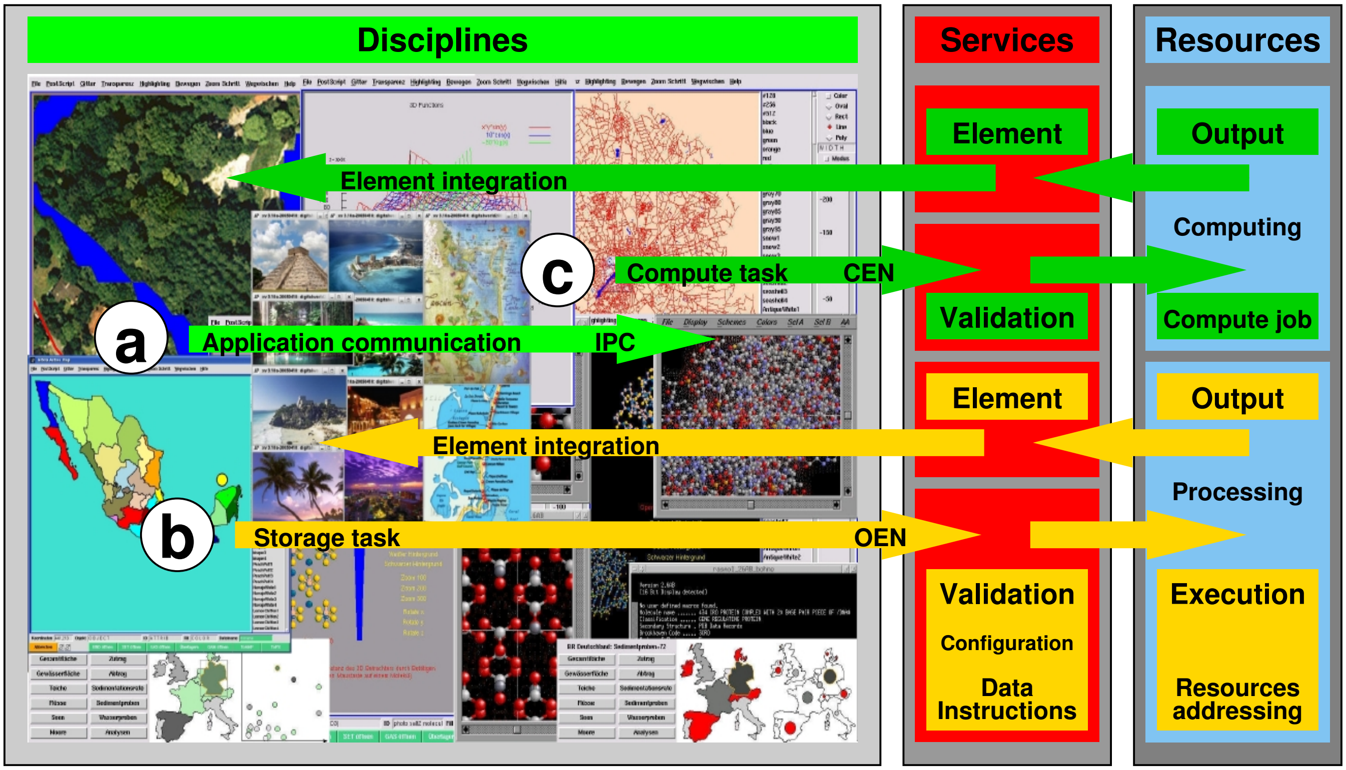

Figure 2. Implementation of the different tasks with integrated systems and resources for advanced scientific computing, utilising the disciplines, services, and resources columns. The disciplines column is triggering the different tasks from the application components (screenshots showing examples). Application communication (local), compute tasks (remote), and storage tasks (remote) are using different resources. If remote high end resources are available these can be used without additional effort. If these are not available then locally available resources can be configured.

D. Compute task

c) Request: From within an actmap component a compute task is triggered. The stored CEN definition is used to transmit the task to the services. The services do the validation (configuration checks, create the compute instructions and initiate the execution of the compute request and compute job for the resources.

Response: The processing output is transmitted to the services for element creation and the element (in this example a remote sensing image and vector object) is integrated into the actmap component.

XII. EVALUATION

The target has been to integrate application communication, computing, and storage resources for handling computing requests and content for distributed storage within one system architecture. The technical details of the components have been discussed in several publications and used in applications publically available. The case study has demonstrated that existing information systems and resources can be easily integrated using envelope interfaces in order to achieve a flexible computing and storage access. As the goal has been to demonstrate the principle and for the modular system components used and due to the previous experiences, the services necessary for integration afforded minimal scripting work. Modern information and computing systems object management is a major challenge for software and hardware infrastructure. Resulting from the case studies with information systems and compute resources, signed objects embedded in OEN can provide a flexible solution. The primary benefits shown from the case studies of this implementation are:

- Build a defined interface between dedicated information system components and computing system components.
- Uniform algorithm for using environment components.
- Integration of information and computing systems.
- Speed-up the development of new components and the modification of existing components in complex environments.
- Portable, transparent, extendable, flexible, and scalable.
- Hierarchical structured meta data, easily parsable.
- OO-support (object, element) on application level.
- Multi-system support.
- Support for signed object elements, validation and verification via PKI.
- Usable with sources and binaries like Active Source.
- Portable algorithms in between different object and file formats, respecting meta-data for objects.

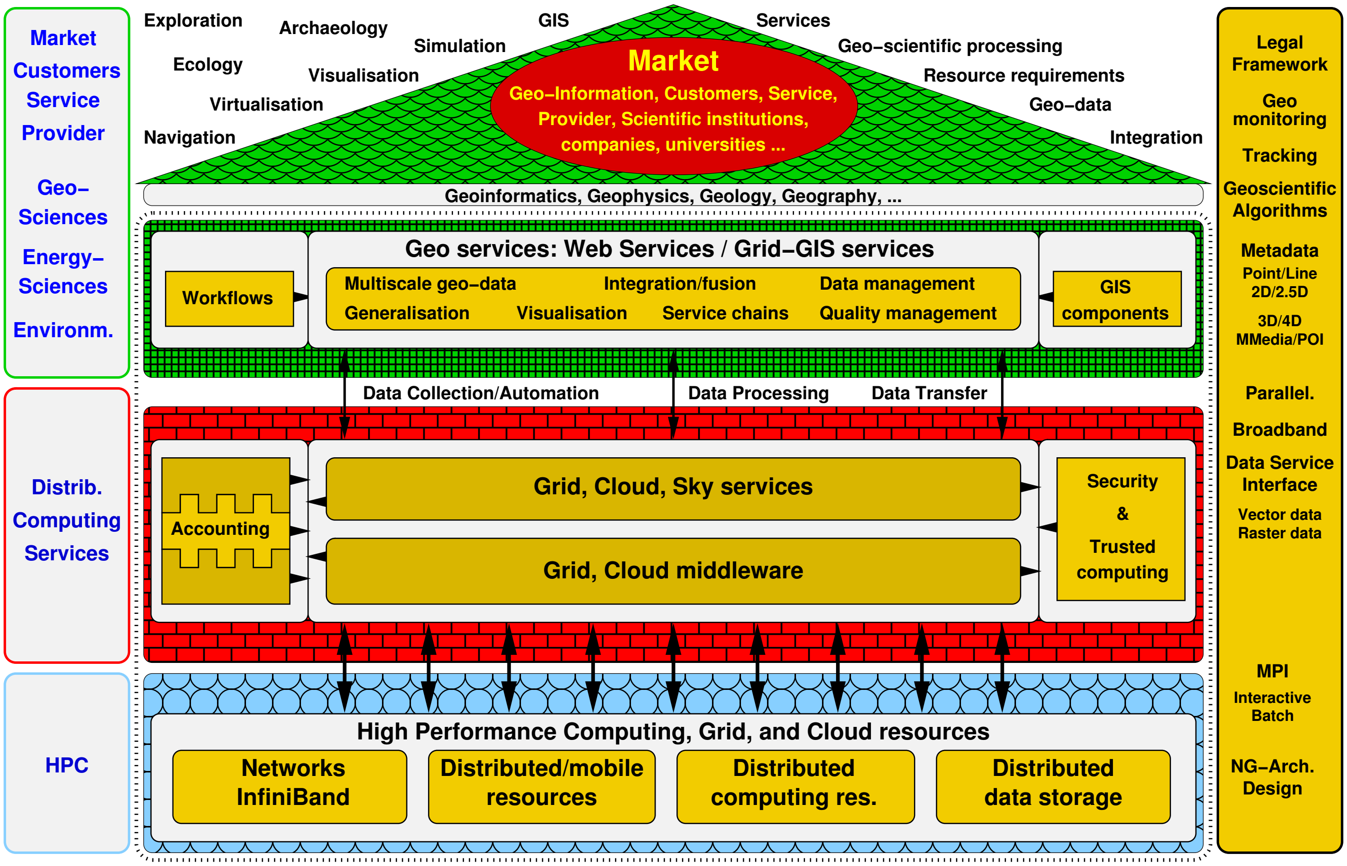

Figure 3. Framework for integrating information systems and scientific computing: This framework has been successfully used with the efficient implementation of all the components presented with this research. Application components have been developed within disciplines (top). System prerequisites and resources configuration have been supported by the resources providers (HPC and distributed computing services) operating the resources, compute and storage. The rightmost column shows examples for the features that have been implemented.

- Original documents and sources can stay unmodified.
- The solution is most transparent, extendable, flexible, and scalable, for security aspects and modularisation.
- Handling of cooperation and operation policies is less complex [56].
- 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 are:

- Additional complexity due to additional resources and system environment features like batch scripting (Condor [32], Moab / Torque [57]) and using verification/PKI.
- Complexity of parsing and configuration.
- Additional software clients might come handy to handle resources and generate, store and manage associated data and certificates.

The context is an important aspect, though it cannot be called "drawback" here. With closed products, e.g., when memory requirements are not transparent, it is difficult for users to specify their needs. Anyhow, testing is in many cases not the answer in productive environments. Separate measures have to be taken to otherwise minimise possible problems and ease the use of resources in productive operation.

Even in the face of the drawbacks, for information systems making standardised use of large numbers of accesses via the means of interfaces, the envelopes can provide efficient management and access, as programming interfaces can.

XIII. Lessons Learned

The integration of Information Systems and scientific computing has been successfully implemented for various case studies. By using information system components and external resources it has been possible to provide a very flexible and extensible solution for complex application scenarios. OEN and CEN, based on generic envelopes, are very flexible and extensible for creating portable, secure objects handling and processing components with IICS. This way context for complex application scenarios can be addressed regarding legal issues, operation, security, and privacy,

A comprehensive review of the most up-to-date results within the GEXI [51] project has been necessary. The most prominent aspects of dynamical integrated systems regarding information systems and disciplines, resources, and computing have been studied. All these issues of implementation and usage have been implemented considering the respective state of the art [55]. The efficiency of the solution is depending on the environment and system architecture provided, so resources should be accessible and configurable in a most user-friendly way.

Case studies including structuring information as with LX [58], with the use of classification like UDC, distributed information, and unstructured information, including large data volume have show that IICS are scalable to a high extend due to their flexible means of configuration. As an advanced target is to cope with the challenges of structuring and classification of information in complex systems, the LX resources have proved a valuable means providing structured information along with UDC having shown excellent applicability for multi-disciplinary classification of any object to be documented. For multi-disciplinary context as well as for special disciplines, knowledge resources integrated with UDC can provide an excellent means of guidance for using knowledge resources and building manual or automated decision making and processing workflows.

The case studies demonstrated that very different kinds of object data structures and instruction sets may be handled with the envelopes, in embedded or referenced use. Meta data, signatures, check sums, and instruction information can be used and customised in various ways for flexibly implementing information and computing system components. Support for transfer and staging of data in many aspects further depends on system configuration and resources as for example physical bottlenecks cannot be eliminated by any kind of software means.

For future IICS an interface layer between user configuration and system configuration would be very helpful. From system side in the future we need least operation-invasive functioning operating system resources limits, e.g., for memory and a flexible limits management. Homogeneous compute and storage resources and strong standardisation efforts for the implementation could support the use of high end resources regarding economic and efficient operation and use.

On the side of system resources integration, for the next generation of system resources we need I/O thresholds being defined and under control of the operating and management system. Memory management and limits are essential for providing the necessary high availability solutions for the resources providers. Regarding many challenges, service providers try to reduce planning and development requirements for their business, e.g., with unstructured data it is possible to avoid file system limitations by using scale-out Network Attached Storage (NAS) or object storage systems [59] while moving problems to other critical challenges like scalability issues, which tend to be moved to different facilities in the scenario. It is possible to proceed this way for business reasons but research in overall solutions does not carry advantages from this.

Benefits of the developments shown will be a convergence of technologies for integrated intelligent system components, as with Multi-Agent Systems (MAS) [60], advanced criticality management, and an ease of use for the overall ecosystem of use and development, services, and operation.

XIV. Conclusion and future work

It has been demonstrated that IICS can be successfully built from information system and advanced scientific computing components, employing a flexible and portable collaboration and envelopes framework.

All the enlisted challenges have been addressed with the implementation, using the collaboration framework, dynamical components, and structured and classified information. For those cases where it is not possible to provide general solutions for all these issues, the infrastructure concept is provided for building complex efficient systems. In some case like physical constraints, we even have to see advanced technology being developed besides any optimised algorithms.

For this implementation, Object Envelopes, Compute Envelopes, and IPC have been used. The application strongly benefits from an integration of supercomputing resources with the information system components. This is especially important for application scenarios with high event multiplicity. For the dynamical interactive application scenarios there are no implicit requirements for application checkpointing features. Other applications running in batch on a large number of cores for long runtimes should have checkpointing features in order to be candidates for High Performance Computing resources. Mission critical application scenarios need to define the multiplicity and the resources. This can be handled transparently by a dynamical event management as with Active Source.

For the case study Active Source components and Distributed and High Performance Computing resources provided the information system and computing environment. Objects necessary with the information system components have been documented and classified in order to support the automation of processes. With IICS the following main results have been achieved.

Local and inter-application communication can be done using IPC. Object Envelopes can be natively used for handling objects and implementing validation and verification methods for communication. Compute Envelopes can be used in order to define information system computation objects and embed instruction information. These algorithms provide means for generic data processing and flexible information exchange.

It has been shown that the collaboration concept is least invasive to the information system side as well as to the resources used whereby being very modular and scalable. The services in between can hold most of the complexity and standardisation issues and even handle products that are meant to be commercially used or licensed.

In the future, we will have to integrate the features for computing, storage, envelopes, structuring, and classification into a global framework for communication purposes and defining standardised interfaces. UDC will be integrated for classification and reference for objects from multi-disciplinary context for creating Archaeological and Universal IICS [61].

The state of the art implementation has demonstrated a flexible basic approach in order to begin to pave the way and show the next aspects to go on with for future IICS [61] for multi-disciplinary applications.

Acknowledgements

We are grateful to all national and international academic and industry partners in the GEXI cooperations for the innovative constructive work and to the scientific colleagues at the Leibniz Universität Hannover, the Institute for Legal Informatics (IRI), and the Westfälische Wilhelms-Universität (WWU), sharing their experiences working on resources on ZIV, HLRN, and D-Grid and for participating in fruitful case studies as well as the participants of the EULISP Programme for prolific discussion of scientific, legal, and technical aspects over the last years. My thanks go to the committee members and participants of the INFOCOMP conferences in Barcelona and Venice as well as the DigitalWorld conferences for inspiring and fruitful discussions on multi-disciplinary scientific documentation, classification, knowledge discovery, management, and application.

References

[1] C.-P. Rückemann, "Implementation of Integrated Systems and Resources for Information and Computing," in *Proceedings of the International Conference on Advanced Communications and Computation (INFOCOMP 2011), October 23–29, 2011, Barcelona, Spain*. XPS, Xpert Publishing Services, 2011, pp. 1–7, Rückemann, C.-P. and Christmann, W. and Saini, S. and Pankowska, M. (eds.), ISBN-13: 978-1-61208-009-3 (CDROM), ISBN-13: 978-1-61208-161-8, URL: http://www.thinkmind.org/download.php?articleid=infocomp_2011_1_10_10002 [accessed: 2012-05-27].

[2] C.-P. Rückemann, "Using Parallel MultiCore and HPC Systems for Dynamical Visualisation," in *Proceedings, GEOWS 2009, Cancun, Mexico*. IEEE CSP, IEEE Xplore, 2009, pp. 13–18, ISBN: 978-0-7695-3527-2, URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4782685&isnumber=4782675 [accessed: 2011-02-20].

[3] C.-P. Rückemann, "Envelope Interfaces for Geoscientific Processing with High Performance Computing and Information Systems." XPS, 2011, pp. 23–28, ISBN: 978-1-61208-003-1, URL: http://www.thinkmind.org/download.php?articleid=geoprocessing_2011_2_10_30030 [accessed: 2011-03-20].

[4] *ICDS-GEOProcessing-CYBERLAWS Joint International Panel, Panel on Privacy Invasion and Protection in Digital Society at DigitalWorld 2011, February 26, 2011, Gosier, Guadeloupe, France, The International Conference on Digital Society (ICDS 2011), The International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2011), The International Conference on Technical and Legal Aspects of the e-Society (CYBERLAWS 2011), February 23–28, 2011, Gosier, Guadeloupe, France / DigitalWorld 2011, [Lecture]*, 2011, URL: http://www.iaria.org/conferences2011/filesICDS11/ICDS_2011_PANEL.pdf [accessed: 2012-04-03].

[5] *GEOProcessing International Panel, Panel on Challenges in Handling Large Data Volume for GEO Processing, January 31, 2012, Valencia, Spain International Panel GEOProcessing, The International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2012), Polytechnic University of Valencia, January 30 – February 4, 2012, Valencia, Spain / DigitalWorld 2012, [Lecture]*, 2012, URL: http://www.iaria.org/conferences2012/filesGEOProcessing12/GEO_2012_PANEL.pdf [accessed: 2012-03-04], URL: http://www.iaria.org/conferences2012/ProgramGEOProcessing12.html (Programme) [accessed: 2012-03-04].

[6] "D-Grid, The German Grid Initiative," 2008, URL: http://www.d-grid.de [accessed: 2012-12-09].

[7] ZIVGrid, "ZIV der WWU Münster – ZIVGrid," 2008, URL: http://www.uni-muenster.de/ZIV/Server/ZIVGrid/ [accessed: 2008-12-23].

[8] "ZIVHPC, HPC Computing Resources," 2011, URL: https://www.uni-muenster.de/ZIV/Technik/ZIVHPC/index.html [accessed: 2011-07-10].

[9] "HLRN, North-German Supercomputing Alliance (Norddeutscher Verbund für Hoch- und Höchstleistungsrechnen)," 2011, URL: http://www.hlrn.de [accessed: 2011-07-10].

[10] "ZIVSMP, SMP Computing Resources," 2011, URL: https://www.uni-muenster.de/ZIV/Technik/ZIVHPC/ZIVSMP.html [accessed: 2011-07-10].

[11] C.-P. Rückemann, "Beitrag zur Realisierung portabler Komponenten für Geoinformationssysteme. Ein Konzept zur ereignisgesteuerten und dynamischen Visualisierung und Aufbereitung geowissenschaftlicher Daten," Diss., Westfälische Wilhelms-Universität, Münster, Deutschland, 2001, 161 (xxii+139) S., OPAC, OCLC: 50979238, URL: http://www.user.uni-hannover.de/cpr/x/publ/2001/dissertation/wwwmath.uni-muenster.de/cs/u/ruckema/x/dis/download/dis3acro.pdf [accessed: 2012-01-15], URL: http://wwwmath.uni-muenster.de/cs/u/ruckema/x/dis/download/dis3acro.pdf [accessed: 2012-01-15].

[12] N. P. Service, "National Register of Historic Places Official Website, Part of the National Park Service (NPS)," 2012, NPS, URL: http://www.nps.gov/nr [accessed: 2012-03-18].

[13] "North American Database of Archaeological Geophysics

(NADAG)," 2012, University of Arkansas, URL: http://www.cast.uark.edu/nadag/ [accessed: 2012-04-08].

[14] "Center for Advanced Spatial Technologies (CAST)," 2012, University of Arkansas, URL: http://www.cast.uark.edu/ [accessed: 2012-04-08].

[15] "Archaeology Data Service (ADS)," 2012, URL: http://archaeologydataservice.ac.uk/ [accessed: 2012-04-08].

[16] "Center for Digital Antiquity," 2012, Arizona State University, USA, URL: http://www.digitalantiquity.org/ [accessed: 2012-01-08].

[17] "The Digital Archaeological Record (tDAR)," 2012, URL: http://www.tdar.org [accessed: 2012-01-08].

[18] IBM, "City Government and IBM Close Partnership to Make Rio de Janeiro a Smarter City," *IBM News room - 2010-12-27, USA*, 2012, URL: http://www-03.ibm.com/press/us/en/pressrelease/33303.wss [accessed: 2012-03-18].

[19] D. Cacciagrano, E. Merelli, L. Vito, A. Sergiacomi, and S. Carota, "Semantics on the Cloud: Toward an Ubiquitous Business Intelligence 2.0 ERP Desktop," in *Proceedings of The Sixth International Conference on Advances in Semantic Processing (SEMAPRO 2012), September 23–28, 2012, Barcelona, Spain*. XPS, Xpert Publishing Services, 2012, Cacciagrano, D. R., Dini, P. (eds.), pages 42–47, ISBN-13: 978-1-61208-240-0, URL: http://www.thinkmind.org/download.php?articleid=semapro_2012_2_40_50119 [accessed: 2012-12-01].

[20] E. Bartocci, D. Cacciagrano, M. R. Di Berardini, E. Merelli, and L. Vito, "UBioLab: a web-LABoratory for Ubiquitous in-silico experiments," *Journal of Integrative Bioinformatics*, vol. 9, no. 1, 2012, pages 1–20, ISSN: 1613-4516, DOI: 10.2390/biecoll-jib-2012-192 [accessed: 2012-12-01].

[21] "LX-Project," 2012, URL: http://www.user.uni-hannover.de/cpr/x/rprojs/en/#LX (Information) [accessed: 2012-02-26].

[22] "Universal Decimal Classification Consortium (UDCC)," 2012, URL: http://www.udcc.org (Information) [accessed: 2012-02-19].

[23] C.-P. Rückemann, *Current Research on Multidisciplinary Information and Communication Systems*. Publisher of the University of Economics in Katowice, Poland, 2012, pp. 146–168, in: Dyduch, W. and Pankowska, M. (eds.), Internet Communication Management, International Week, URL: http://www.intweek.ue.katowice.pl/, Monograph, 192 pages, ISBN: 978-83-7246-776-8 (softcover), (Peer Review).

[24] "Digital Object Identifier (DOI) System, The International DOI Foundation (IDF)," 2012, URL: http://www.doi.org (Information) [accessed: 2012-02-19].

[25] N. Hemsoth, "IBM Big Data VP Surveys Landscape," *Datanami*, 2012, March 09, 2012, URL: http://www.datanami.com/datanami/2012-03-09/ibm_big_data_vp_surveys_landscape.html [accessed: 2012-03-18].

[26] N. Hemsoth, "A Floating Solution for Data-Intensive Science," *Datanami*, 2012, March 12, 2012, URL: http://www.datanami.com/datanami/2012-03-12/a_floating_solution_for_data-intensive_science.html [accessed: 2012-03-18].

[27] "Hpc cloud targets oil and gas space," HPC in the Cloud, November 21, 2012, 2012, URL: http://www.hpcinthecloud.com/hpccloud/2012-11-21/hpc_cloud_targets_oil_and_gas_space.html [accessed: 2012-12-01].

[28] "Cyrusone, dell, r systems deliver oil & gas cloud-based solution."

[29] T. Trader, "Gartner Lays Out Five-Part Cloud Strategy," *HPC in the Cloud*, 2012, April 03, 2012, URL: http://www.hpcinthecloud.com/ct/uz5657276Biz13083207 [accessed: 2012-04-08].

[30] EU, "European Code of Conduct for ICT, EU Code of Conduct for Data Centres," 2012, European Commission Joint Research Center, URL: http://re.jrc.ec.europa.eu/energyefficiency/html/standby_initiative_data_centers.htm [accessed: 2012-03-18].

[31] "IBM Tivoli Workload Scheduler LoadLeveler," 2005, URL: http://www-03.ibm.com/systems/software/loadleveler/ [accessed: 2010-10-10].

[32] "Condor, High Throughput Computing," 2011, URL: http://www.cs.wisc.edu/condor/ [accessed: 2010-12-26].

[33] "Sun Grid Engine," 2010, URL: http://gridengine.sunsource.net/ [accessed: 2010-10-10].

[34] "Moab Workload Manager Administrator's Guide," 2010, URL: http://www.clusterresources.com/products/mwm/ [accessed: 2009-11-16], URL: http://www.adaptivecomputing.com/resources/docs/mwm/ [accessed: 2012-12-09].

[35] "Torque Admin Manual," 2010, URL: http://www.clusterresources.com/torquedocs21/ [accessed: 2010-10-10].

[36] T. Trader, "Adaptive Computing Revs Up Moab," *HPC in the Cloud*, 2012, March 20, 2012, URL: http://www.hpcinthecloud.com/hpccloud/2012-03-20/adaptive_computing_revs_up_moab.html [accessed: 2012-03-25].

[37] "Open-MPI," 2011, URL: http://www.open-mpi.org [accessed: 2011-07-12].

[38] "IWAVE," 2012, URL: http://www.trip.caam.rice.edu/software/iwave/doc/html/index.html [accessed: 2012-03-18].

[39] "HPCToolkit," 2012, URL: http://hpctoolkit.org [accessed: 2012-03-18].

[40] M. Wolfe, "The Heterogeneous Programming Jungle," *HPCwire*, 2012, March 19, 2012, URL: http://www.hpcwire.com/hpcwire/2012-03-19/the_heterogeneous_programming_jungle.html [accessed: 2012-03-25].

[41] "Group on Earth Observations (GEO)," 2012, URL: http://www.earthobservations.org [accessed: 2012-03-25].

[42] "Global Earth Observation System of Systems (GEOSS)," 2012, URL: http://www.earthobservations.org/geoss.shtml [accessed: 2012-03-25].

[43] "Shared Environmental Information System (SEIS)," 2012, URL: http://ec.europa.eu/environment/seis/ [accessed: 2012-03-25].

[44] "Global Monitoring for the Environment and Security (GMES)," 2012, URL: http://www.gmes.info [accessed: 2012-03-25].

[45] D. Snead, "Cloud Computing: Legal Traps to Look Out For," *Tutorial, January 30, 2012, Valencia, Spain International Panel GEOProcessing, The International Conference on Advanced Geographic Information Systems,*

Applications, and Services (GEOProcessing 2012), Polytechnic University of Valencia, January 30 – February 4, 2012, Valencia, Spain / DigitalWorld 2012, [Lecture], 2012, URL: http://www.iaria.org/conferences2012/ProgramGEOProcessing12.html [accessed: 2012-02-26].

[46] S. Naqvi, "Digital Investigations and Forensic Analysis - Practices and Technologies," *Tutorial, January 30, 2012, Valencia, Spain International Panel GEOProcessing, The International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2012), Polytechnic University of Valencia, January 30 – February 4, 2012, Valencia, Spain / DigitalWorld 2012, [Lecture]*, 2012, URL: http://www.iaria.org/conferences2012/ProgramGEOProcessing12.html [accessed: 2012-02-26].

[47] D. Malzahn, "The Electronic Silverback, Absorb Knowledge Loss in Industry by Social Network Approach," in *Proceedings International Conference on Information, Process, and Knowledge Management (eKNOW 2012) Polytechnic University of Valencia (UPV), January 30 – February 4, 2012, Valencia, Spain / DigitalWorld 2012*. XPS, Xpert Publishing Services, 2012, pp. 47–52, ISBN 978-1-61208-181-6.

[48] C.-P. Rückemann, "Securing Privacy for Future Information and Computing Systems," *Privacy Invasion and Protection in Digital Society at DigitalWorld 2011, February 26, 2011, ICDS-GEOProcessing-CYBERLAWS Panel, The International Conference on Digital Society (ICDS), The International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing), The International Conference on Technical and Legal Aspects of the e-Society (CYBERLAWS), February 23–28, 2011, Gosier, Guadeloupe, France*, 2011, URL: http://www.iaria.org/conferences2011/filesICDS11/ICDS_2011_PANEL.pdf [accessed: 2012-03-25].

[49] HPCwire, "HPCwire: GRC and Super Micro Partner on Geophysical Processing Cluster," *HPCwire*, 2012, March 22, 2012, URL: http://www.hpcwire.com/hpcwire/2012-03-22/grc_and_super_micro_partner_on_geophysical_processing_cluster.html [accessed: 2012-03-25].

[50] C.-P. Rückemann and B. Gersbeck-Schierholz, "Object Security and Verification for Integrated Information and Computing Systems," in *Proceedings, CYBERLAWS 2011, Gosier, Guadeloupe, France*. XPS, 2011, pp. 1–6, ISBN: 978-1-61208-003-1, URL: http://www.thinkmind.org/download.php?articleid=cyberlaws_2011_1_10_70008 [accessed: 2012-11-25].

[51] "Geo Exploration and Information (GEXI)," 1996, 1999, 2010, 2012, URL: http://www.user.uni-hannover.de/cpr/x/rprojs/en/index.html#GEXI (Information) [accessed: 2012-12-06].

[52] "Applications with Active Map Software, Screenshots," 2005, URL: http://wwwmath.uni-muenster.de/cs/u/ruckema/x/sciframe/en/screenshots.html [accessed: 2011-02-20].

[53] C.-P. Rückemann, "Legal Issues Regarding Distributed and High Performance Computing in Geosciences and Exploration," in *Proceedings of the Int. Conf. on Digital Society (ICDS 2010), The Int. Conf. on Technical and Legal Aspects of the e-Society (CYBERLAWS 2010), February 10–16, 2010, St. Maarten, Netherlands Antilles*. IEEE Computer Society Press, IEEE Xplore Digital Library, 2010, pp. 339–344, ISBN: 978-0-7695-3953-9, URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5432414 [accessed: 2011-02-20].

[54] C.-P. Rückemann, "Dynamical Parallel Applications on Distributed and HPC Systems," *International Journal on Advances in Software*, vol. 2, no. 2, 2009, ISSN: 1942-2628, URL: http://www.iariajournals.org/software/ [accessed: 2012-11-25].

[55] C.-P. Rückemann, *Queueing Aspects of Integrated Information and Computing Systems in Geosciences and Natural Sciences*. InTech, 2011, pp. 1–26, Chapter 1, in: Advances in Data, Methods, Models and Their Applications in Geoscience, 336 pages, ISBN-13: 978-953-307-737-6 (hardcover), DOI: 10.5772/29337 [accessed: 2012-12-01], (Peer Review), URL: http://www.intechopen.com/download/pdf/pdfs_id/25377 [accessed: 2012-12-02], URL: http://www.intechopen.com/articles/show/title/queueing-aspects-of-integrated-information-and-computing-systems-in-geosciences-and-natural-sciences [accessed: 2012-12-02] URL: http://dx.doi.org/10.5772/29337 [accessed: 2012-12-02].

[56] "EULISP Lecture Notes, European Legal Informatics Study Programme, Institute for Legal Informatics, Leibniz Universität Hannover (IRI / LUH)," 2011, URL: http://www.eulisp.de [accessed: 2012-11-25].

[57] "Moab Admin Manual, Moab Users Guide," 2011, URL: http://www.clusterresources.com/products/mwm/moabdocs/index.shtml [accessed: 2011-02-20].

[58] C.-P. Rückemann, *Integrated Information and Computing Systems for Advanced Cognition with Natural Sciences*. Premier Reference Source, Information Science Reference, IGI Global, 701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA, 2012, pages 1–26, in: Rückemann, C.-P. (ed.), Integrated Information and Computing Systems for Natural, Spatial, and Social Sciences, 543 (xxiv + 519) pages, ISBN: 978-1-4666-2190-9, EISBN: 978-1-4666-2191-6, LCCN: 2012019278, DOI: 10.4018/978-1-4666-2190-9 [accessed: 2012-12-01].

[59] P. Speciale, "Big Unstructured Data and the Case for Optimized Object Storage," *Datanami*, 2012, March 20, 2012, URL: http://www.datanami.com/datanami/2012-03-20/big_unstructured_data_and_the_case_for_optimized_object_storage.html [accessed: 2012-03-25].

[60] U. Inden, S. Tieck, and C.-P. Rückemann, "Rotation-oriented Collaborative Air Traffic Management," in *Proceedings of the International Conference on Advanced Communications and Computation (INFOCOMP 2011), October 23–29, 2011, Barcelona, Spain*. XPS, 2011, pp. 25–30, iSBN-13: 978-1-61208-009-3, URL: http://www.thinkmind.org/download.php?articleid=infocomp_2011_2_10_10089 [accessed: 2012-05-27].

[61] C.-P. Rückemann, "Enabling Dynamical Use of Integrated Systems and Scientific Supercomputing Resources for Archaeological Information Systems," in *Proceedings of the International Conference on Advanced Communications and Computation (INFOCOMP 2012), October 21–26, 2012, Venice, Italy*. XPS, Xpert Publishing Services, 2012, Rückemann, C.-P. and Dini, P. and Hommel, W. and Pankowska, M. and Schubert, L. (eds.), pages 36–41, ISBN-13: 978-1-61208-226-4, URL: http://www.thinkmind.org/download.php?articleid=infocomp_2012_3_10_10012 [accessed: 2012-12-01], URL: http://www.thinkmind.org/index.php?view=article&articleid=infocomp_2012_3_10_10012 [accessed: 2012-12-01].