Integrating Future High End Computing and Information Systems Using a Collaboration Framework Respecting Implementation, Legal Issues, and Security

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

***Abstract*—This paper gives an extended overview of challenges creating complex integrated information and computing systems. It covers implementation, legal, and security issues with these processes and how the overall complexity can be reduced using collaboration frameworks with Distributed and High Performance Computing resources in natural sciences disciplines for building integrated public / commercial information system components within the e-Society. Focus is on using a collaboration framework for implementing computing resources, interfaces for data and application interchange, based on current developments regarding infonomics systems, last years case studies within the long-term GEXI project, and Active Source. A suitable framework base has been created over the last years, being used for a number of scenarios in research environments, using High End Computing resources. Application of these methods for commercial service structures affords the consideration of various legal, security, and trust aspects. In collaboration with international partners from natural sciences, industry, economy, and education the framework has been found the solution to overcome the legal cooperation barrier. Established on this work, international cooperations are currently built. In addition, this paper presents two major implementation case studies in order to show the application of the collaboration framework, one for environmental and energy exploration information, computing, and resources management and another for epidemology information systems.**

Keywords–Legal Frameworks; Collaboration Management; Implementation; Legal Issues; Security; Distributed Systems; High Performance Computing; Grid-GIS house; e-Science; Geoscientific Information Systems.

I. Introduction

Todays information system design, development, implementation, and usage are in many cases characterised by dynamism and fast varying means, heterogeneous content, access and information security in complex environments, short-term financing, and individual architectures. About over a decade now, the amount of information available as well as the computing power has been continuously increasing, but there is no integrated information-computing system actually really bringing these vast resources together.

Research on overcoming these shortcomings for international collaboration management is going on for the last years [1]. Over the last years a long-term project, Geo Exploration and Information (GEXI) [2] for analysing case studies, has examined chances to overcome the deficits.

This paper presents the current results with a collaboration framework that has been developed and successfully used as a solution for various cases. It delivers the results collected from a study taken on participating national and international collaboration projects regarding the sections High Performance Computing (HPC), Distributed Computing (DC) and services, and natural sciences. There is a number of factors limiting the vigors that are devoted on the development of integrated systems. These are the problems with distribution of valuable resources like High End Computing (HEC) needed to be integrated on one hand and the legal diversities as with automation, personalisation, security, and differences in professional, national, and international context on the other. As resulting from the developments of a suitable framework for DC and HPC, this work provides the legal complement for the technical and scientific base currently used for various international collaborations, building information, processing and decision making systems. Two case studies are presented in order to discuss the different aspects of "trust in computing" and "trust in information" emphasis.

II. Motivation

Geoscientific information systems belong to the most advanced information systems available today. A driving force behind the development besides public interests is from applied natural sciences, exploration and energy, resources, and environmental management. Oil and gas, climatology, aerospace and automotive industry for example depend on privacy for their computations. Neither services nor security in general will change the advised behaviour within the next decade. It is uninviting to expect the driving force from the resources engineering approach only. It is a common misbelief that industry will use any distant foreign HEC resources for economically interesting or critical calculations, for computing or storage in context with their strategic data or with production. The research projects of the last years have shown that we need a legally integrated comprehensive collaboration framework for complex modular design and development [3] as well as new methods and algorithms on the geoinformation side [4] regarding distributed resources and secure communication.

This paper is organised as follows. Section three presents preliminary work and cases studies. Sections four and five describe legal frameworks, regulations and name the problems addressed. Section six describes the new collaboration framework. Section seven explains the status of the implementation regarding the participating key player topics. Section eight reports evaluation and consequences for primary topics on technology and legal issues and the collaboration portal. Section nine gives an extended presentation of selected case studies, implemented following the collaboration framework. The focus is on implementation, legal issues, and security aspects faced by the resources, services, and disciplines columns. Section ten shows the lessons learned and Section eleven summarises the conclusion and outlook on future work.

III. Preliminary work and case studies

This discussion of legal issues is the result of the successful work of the last years, in the disciplines of geosciences application architecture [5] (Active Source), preparing an interdisciplinary Grid-GIS house framework [6], project case studies, and the adjunctive configuration of various HEC (HPC, Distributed, Grid, Cloud, and GPU Computing) resources over the last three years [4]. Work is ongoing for developments in cooperation with international industry and economy partners [7], [3].

The analysis of these case studies showed that the coordinated cooperations have a strong need to address the legal and security base for handling critical data (e.g. business relevant development and exploration data), computing and geo-processing as well as components used.

At todays level of information integration, with the overall complexity of information and decision making systems, there is a neccessity for building flexible and extendable information systems for cognostic purposes that consider aspects of security and economy in complex environments. For future cooperations and implementations it is most important to focus on legal issues regarding the frameworks.

IV. Legal frameworks and regulations

As in this context, there exists no collaboration development framework, there is a number of partially interesting international and national legal topic-frameworks (Table I) for information systems, regarding content and structure.

Table I
Legal frameworks with geosciences.

Name	*Framework and Reference*	
GMES	Global Monitoring for the Environment and Security	[8]
GEOSS	Global Earth Observation System of Systems	[9]
	of the GEO (Group on Earth Observations)	[10]
SEIS	Shared Environmental Information System	[11]
GSDI	Global Spatial Data Infrastructure	[12]
INSPIRE	Infrastructure for Spatial Information in Europe directive (2007/2/EC)	[13]
GDI-DE	Geodateninfrastruktur Deutschland	[14]
PSI/EPSI	Public Sector Information directive / European Public Sector Information	[15]

Besides these frameworks there is a number of laws and legal regulations regarding geo data in Germany: copyright law (Urheberrechtsgesetz, UrhG), data security and privacy law (Bundesdatenschutzgesetz, BDSG), freedom of information law (Informationsfreiheitsgesetz, IFG), law on the reuse of information from public institutions (Gesetz über die Weiterverwendung von Informationen öffentlicher Stellen, Informationsweiterverwendungsgesetz, IWG), environmental information law (Umweltinformationsgesetz, UIG), law on accessing digital geo data (Gesetz über den Zugang zu digitalen Geodaten, Geodatenzugangsgesetz, GeoZG). These regulations do concern many aspects, contributors and participants of a modern information system, content as well as implementation, access, and usage.

V. Problems addressed

Currently there exist traditional information structures, old non-intelligent applications and data formats, but the data-barrier at various media is still omnipresent. Although the earliest of the fundaments of the named academic frameworks are dating back into the year 1998 there is still no information system available integrating the technical and legal diversities. But all of these frameworks can be seen complementary to comprehensive information system frameworks for geo-computing and geo-processing. What we target to, is legally conform modular applications and data formats for integrability of all resources.

A complementary effort is necessary for integrating industry, economy, and legal expertise into this process of creating a next generation information system framework at an international level. Further industrial and economic impulses originate from exploration, geosciences, energy-sciences, climatology, and education for handling data on resources management, observation, environment, biodiversity, weather, medicine.

In these disciplines many applications do need an integration of information systems with simulation and virtual reality. Not only that oil and gas industry, insurances, town planning, tourism industry and many others do have strong needs to be integrated into this processes. The integration will provide chances for new insights into complexity of the environmental systems. Therefore, primary goals are implementations based on the collaboration framework:

- integration of academia, industry, economy, law,
- modularisation of system components,
- structuring of data and information,
- integration of HEC and storage resources,
- legally conform georeferencing of data and objects,
- licensing (e.g. topography, remote sensing),
- personalisation of information and services.

VI. Collaboration Framework

Illustrating the directions of integrating and co-developing large collaboration target frameworks and applications for service-oriented DC and HPC, Figure 1 shows the columns of the infonomics system and Figure 2 shows the dependencies of market and services (green colour, shingle and cross pattern), computing services (red colour, brick pattern), HPC and distributed resources (blue colour, gravelly pattern), and resources to be provisioned or developed (gold colour). The proposed Computing Industry Alliance ([4], Leadership in Research consortium) will be a suitable umbrella organisation for distributed and HPC and geo-exploration sciences. The framework described is an example currently building the base for creating efficient interdisciplinary industry research cooperations for implementing the next generation of dynamical applications on distributed and HPC resources based on the "Grid-GIS house" [6]. Interests to force this development exist, not only in the Gulf of México region but as well in Russia and Saudi Arabia.

Resulting from the GEXI project [2] started in the year 1996 as a public and private support network, the components and mechanisms have been topic of several information science, HPC, Grid Computing cooperations and European activities of the last years [16].

Three key player collaboration sections from HPC and Distributed and Grid Computing, from services and technical development, and from geosciences and exploration are currently building the next generation of information and computation system.

With support of international partners from geosciences, high end computing industry, economy, and education an extended Grid-GIS house [6] for the geosciences and exploration disciplines has been created this year [3] and legal aspects are currently discussed for further cooperation.

VII. Status of the Implementation

A. Infonomics system and interactions

The entirety of the essential columns of the framework, geosciences and energy-sciences, Distributed Computing and services, and HPC forms a well balanced infonomics system. The necessary interactions for the information and computing systems build the interfaces for the columns of the infonomics system (Figure 1). For integration into the

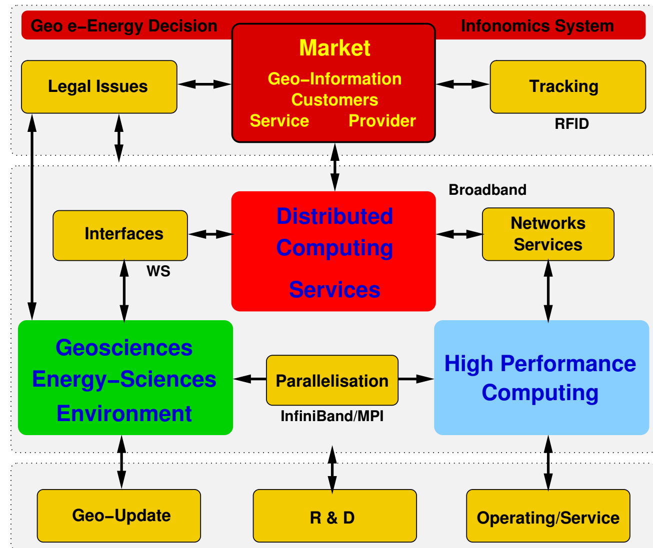

Figure 1. Columns of the infonomics system.

market, suitable services (e.g. Software as a Service, SaaS) will be provisioned as geosciences and e-Energy seek a complex information and decision making environment. Efficiency and effectivity is important from various perspectives: Fast networks (broadband, InfiniBand) are needed for distributed resources and HPC, interfaces for services and scientific applications, complementary to the simple OGC Web Processing Service (WPS), as well as parallelisation for geoscience algorithms is currently expedited with the collaboration framework in order to be employed on HPC resources.

Supplies for the natural sciences disciplines will be done with updates of data and algorithms and the computing resources do need a continuous operating as research and development are essential for interaction between the columns of the system. As the market does not only want to "trade" electronic goods, a suitable coupling with the information systems is necessary for physical identity tracking and monitoring, e.g., RFID for container cargo and Intelligent Transport Systems (ITS). Legal issues regarding all of these topics are omnipresent, for the columns and for the market.

B. Legal focus points

The entirety of these aspects describes a next generation "Information and Decision Making Environment" for the future internet, containing electronic and physical operations like information system components and logistics and tracking support for objects and goods.

Figure 3 shows the important legal focus points and dependency relations.

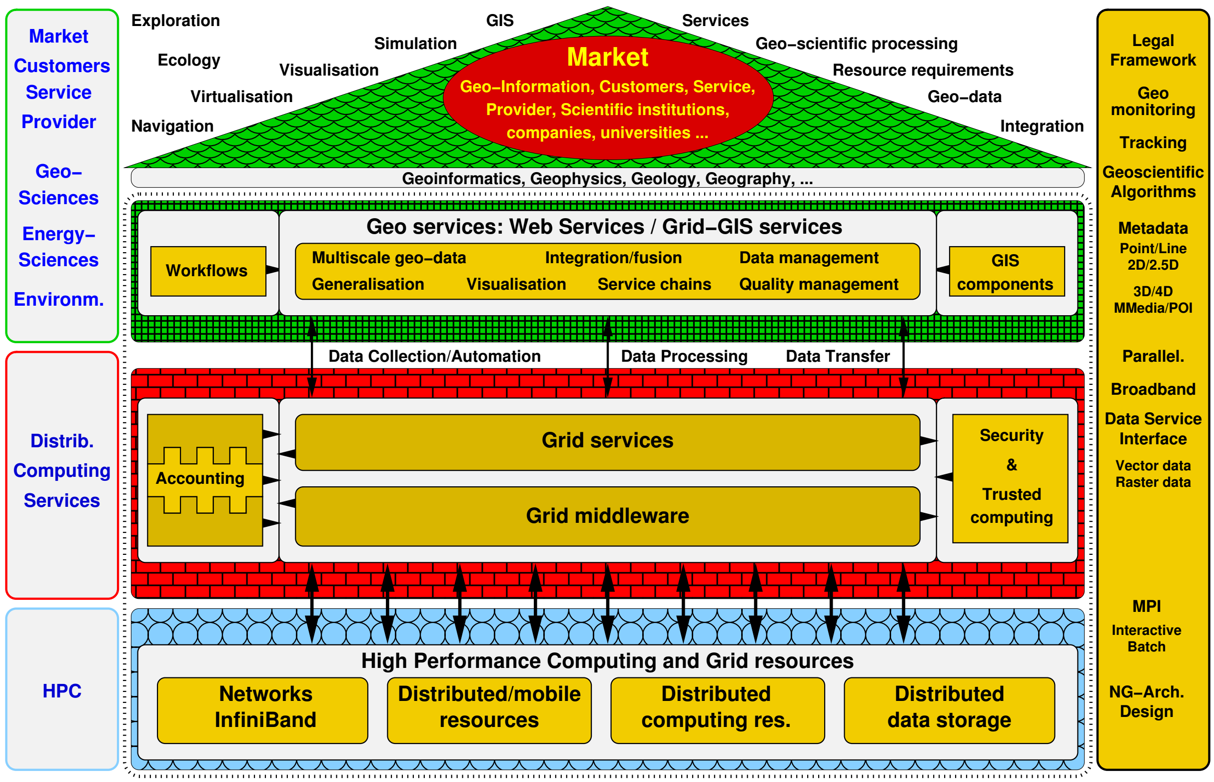

Figure 2. Collaboration framework for geosciences / exploration and HEC ("Grid-GIS house").

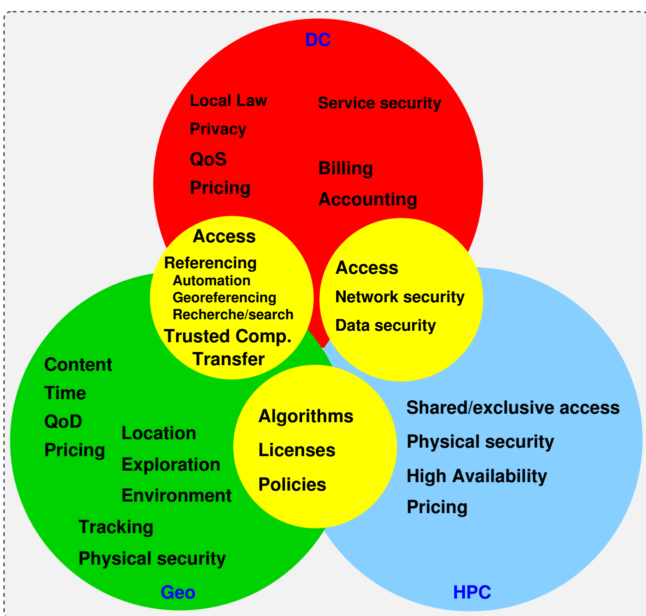

Figure 3. Legal focus points with GEXI.

As some related issues (yellow colour) strongly overlap with different columns they should be worked on in a collaboration development between disciplines, providers, and computing industry.

There have been implementations regarding a number of issues within case studies and future topics have been identified. Table II shows a compact overview of the implementation focus points and the current directions (✓ mostly done, ➠ future tasks within project). As the table points out, a number of issues is already done others will have to be worked on, for integrating new features needed for the next stage for an upheaval in system complexity. Some other issues will have to remain undone work with regard to standardisation as they afford an individual configuration, parametrisation, and optimisation, like access to HPC resources or because there are suitable solutions like MPI for parallelisation.

The number of legally safe application scenarios in focus, working with these resources and technologies should be increased by the participating academia and industry. These scenarios include commonly shared interactive distributed resources usage for computation and information storage and processing, distributed among international partners with different legal conception and regulations regarding privacy and anonymity.

There is a number of selected topics, being in the focus of current implementation and usability studies. The following passages show some topics important for the interdisciplinary context of the case studies.

C. Security

Physical and logical aspects and features, hazards and threats have been considered within the last years.

Table II
IMPLEMENTATION FOCUS POINTS AND DIRECTIONS.

Issue	*Implementation & Status*		*Future*
Access	individual	✓	➠ standardisation
	Distributed Comp.	✓	(—)
	High Perf. Comp.	✓	(—)
Trusted Comp.	sandbox, security policies	✓	➠ porting, PKI
Transfer	encryption / keys pairs	✓	➠ WebServices
	broadband	—	➠ industry/federal
QoS	categorisation	—	interdisciplinary
Content	composite data	—	interdisciplinary
	vector, raster	✓	➠ standardisation
	attribute data	✓	➠ standardisation
	event data	✓	➠ standardisation
	references	✓	➠ standardisation
Licensing	individual	—	➠ license mgmt.
QoD	level cognostics	—	interdisciplinary
Referencing	automation	—	➠ scripting
	georeferencing	—	➠ categories
	recherche / search	—	➠ spec. engines
Algorithms	parallelisation	✓	➠ standardisation
	loosely coupled	✓	(—)
	MPI, OpenMP, Java	✓	(—)
Tracking	phys. identification	—	➠ RFID
Accounting	non-commercial acc.	✓	➠ undisputable
	integrated solution, SGAS	✓	➠ modules
Pricing	individual, flatrate	✓	➠ compound units
Billing	individual, flatrate	✓	➠ cumulative

- Restrictions on hardware usage.
- Restrictions on access.
- Trust in information.
- Trust in computing.
- Trusted scripting.
- Inter Process Communication (IPC).
- Active Source methods.
- Accounting security, bidirectional and undisputable.
- Sandboxing and policies.

In the implementation case studies these aspects had to be handled in order to implement real-life systems. This will be discussed in some of the next sections.

D. Sandboxing policies

Various sandboxing mechanisms have been experimented with. The Tcl architecture supports very flexible features for sandboxing and policy implementation. With dynamical client-server applications the Tcl plugin supports Tcl/Tk applets, so called Tclets. The Tcl plugin implements the standard Safe-Tcl subset and defining new policies. For the Safe-Tcl interpreter various commands can be removed from the Tcl interpreter by configuration, used to run Tcl applets. A limited version of Tk has been added. This sandboxing can be used in the most flexible way with high level languages and other scripting languages. With the case studies, control for the following commands has been found most important, in order to gain trustable modules.

- exec (execute programs),
- load (dynamically load shared libraries implementing C or Tcl language commands),
- open (open a file, restricted open-read-only version available),
- send (send Tcl commands to other applications),
- cd (change directory),
- socket (open a network socket),
- source (load script files),
- exit (terminate a process).

Tk images cannot be created or read from files. Image create photo commands take strings of base64 encoded images instead. Further commands are handled with the Safe-Tcl.

- wm (window manager control),
- toplevel (create toplevel windows),
- menu (display a menu),
- tk (set and query Tk application names),
- tkwait (block on events),
- bell (ring terminal bell),
- clipboard (access the clipboard selection),
- glob (match file names in a directory),
- grab (grab the cursor),
- pwd (query present working directory).

These and additional functions and commands can be configured for the executing sandbox environments.

E. Accounting implementation

An integrated accounting and billing approach has been developed in the last years. The SweGrid Accounting System (SGAS) [17] has been considered most useful for Distributed and High Performance Computing. It supports scalable resources and capacity allocation [18], [19], decentralised concepts for fairshare scheduling [20], OGSA-based bank service [21], distributed usage logging [22], and support for federated cloud infrastructures [23]. The integrated accounting and billing approach supports market-ready secure, transparent, and flexible resources management.

The following section evaluates the current status and describes the lessons learned for implementation, technical and legal consequences.

VIII. EVALUATION AND LEGAL CONSEQUENCES

Some aspects like loosely coupled (with Grid, Cluster, and HPC) as well as MPI parallelisation for Massively Parallel Processing (MPP) and Symmetric Multi-Processing (SMP) have been successfully implemented and used for various purposes (e.g. in the projects Condor-network, ZIVGrid, ZIVHPC, ZIVSMP, HLRN) [6], [4]. It has been proven viable to use a collaborative implementation strategy to integrate individual solutions for making long-term investments sustainable. International work is currently done for parallelising application-triggered algorithms for interactive use on huge HPC resources. Focus is on management, exploration, and environmental applications for geosciences, energy management, and information sciences. Invited industry partners are currently implementing parallelised application suites. From the legal point of view, protocols and exclusive commercial use of resources have to be implemented for reliable use of HEC resources.

For accounting, pricing, and billing individual solutions mostly based on flatrates have been used (DC and HPC). Future focus will concentrate on compound units and modular solutions. An integrated distributed accounting and billing solution considering national and international legal regulations must be implemented for infonomics purposes.

Up to now various methods and technologies have been implemented for use with the new features. There will have to be strong standardisation efforts for secure and ergonomic access, quality management, and new types of content specification, allowing flexible separation of data, information, and functional parts. Considering legal aspects is crucial for the design of these specifications in order to create a suitable structure based on the legal frameworks.

Physical aspects for infonomics systems have often been neglected in the past. Data sizes have been sized small and only transferred for small distances. Prominent topics are broadband for public use with data transfer and physical identification for real world tracking. These will be needed for transfer of large amounts of computation data in national and international context, for monitoring and logistics in exploration and environmental management. Projects are experimenting with large data sizes (>100 TeraBytes) on international data transfer for use with applications. These are expected to be industry topics in the near future. Broadband networks will allow to transfer larger amounts of data using secure channels on external networks.

The implemented data handling is suitable for data types and applications currently used. The case studies of the last years have shown that there is need for standardised cognostic categories. Large size applications with composite data types do need data-categorisation for generalising, integration, automation, georeferencing, and search engine facilities in order to minimising conflicts with legal regulations.

Operation and updates will be interesting for providers and industry. Legal regulations demand a transparent least-invasive access and update concept for complex information and computing systems.

A. Primary topics on technology and legal issues

A number of primary topics resulting from the interactions within the columns of the infonomics system exist from user point of view: security, safety/privacy, consistency, international standards, legal issues, and primary associated laws and regulations, identified with the GEXI case studies (Table III). This table shows some of the most important components, that are not worked out ultimatively

Table III
RESULTING PRIMARY TOPICS AND FUTURE EMPHASIS.

Topic	*Column*	*Sec.*	*Saf.*	*Con.*	*Int.*	*Leg.*	*Law (DE)*
				Services			
Services	DG	(✓)	(✓)	—	—	—	BDSG GeoZG
Georeferencing	DG	—	—	—	—	✗	BDSG
Automation	DG	—	—	—	(—)	✗	BDSG
				Communication/Transfer			
Networking	DH	(✓)	(✓)	(✓)	(—)	(—)	BDSG
				Distributed Computing			
Accounting	DHG	(✓)	(✓)	(✓)	✗	✗	BDSG
Billing	DHG	(✓)	(✓)	(✓)	✗	✗	BDSG
				High Performance Computing			
Computing	HD	✗	✗	✗	✗	✗	BDSG UrhG
Networking/IB	HD	✗	✗	✗	✗	✗	BDSG UrhG
Storage	DH	(✓)	✗	✓	✗	✗	BDSG UrhG
				Disciplines			
Inf. Systems	GD	(✓)	(—)	(✓)	✗	✗	IFG GeoZG IWG
Geosciences	G	(—)	(—)	(—)	✗	✗	IFG GeoZG IWG
Exploration	OG	(—)	(—)	(—)	✗	✗	IFG UrhG IWG
Environment	OG	(—)	(—)	(—)	✗	✗	IFG UIG IWG
Medicine	OG	(—)	(—)	(—)	✗	✗	IFG BDSG IWG
e-Science	OGD	(—)	(—)	(—)	✗	✗	IFG BDSG IWG

(✓ partially done, ✗ worked on within interdisciplinary cooperations, G: Geo, H: HPC, D: DC, O: other). Only some topics like storage consistency can be considered done for mid-term, as there are means of creating suitable solutions. Other topics like safety within the productive employment of geosciences, exploration, medicine, and e-Science can be considered specific to their discipline. There are some topics most important as they concern several of the topics and disciplines: facilitate market use of HEC resources and creating collaboration frameworks for national to international use.

B. Collaboration portal

The GEXI case studies have shown the additional need for a single access point, a frontend portal. Resulting from the integration of the consolidated academic and industrial interests, Figure 4 shows a sketch of the prototype GEXI portal addressing the geo-exploration-energy market.

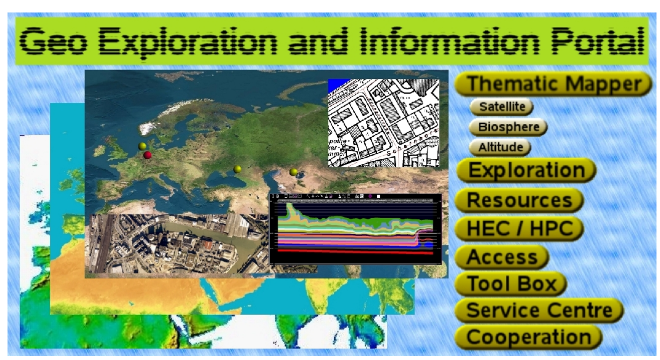

Figure 4. Sketch of the GEXI portal.

The portal is aimed to provide a collective collaboration platform for the columns of the infonomics system in order to prove the conjoint economic and legal feasibility. Clients for disseminating the provisioned services, information and distributed resources will be integrated with this portal. Such a portal is capable of providing a solution to technical-legal diversities as with access, services accounting, licensing, QoD, automation, personalisation, security, and internationalisation, handling focus points within and between the collaborating columns at the backend and reducing complexity.

IX. Implementation case studies

A. Application scenarios

The information and computing system components make use of various technologies, IPC, sandboxing, embedded applications, browser plugins, remote execution, network protocols, computing interfaces as well as public and sensible data. Figure 5 shows some of the basic application scenarios.

There exists a number of scenarios showing how "trust in computation" and "trust in information" can more easily be achieved by reducing complexity for the partners in otherwise very complex systems. The following sections give an extended presentation of selected case studies, implemented following the collaboration framework. The focus is on implementation, legal issues, and security aspects faced by the resources, services, and disciplines columns. The following collaboration matrices show what topics the columns Resources (R), Services (S), and Disciplines (D) had to take care while realising the components regarding

- implementation (i),
- legal issues (l),
- security (s).

The column partners have been responsible for the topics respecting the work packages designed respecting the collaboration framework.

B. Environmental information and computing

Various information resources are available for environmental and energy exploration. Mostly all of the implementations making use of these resources are standalone systems. Computing resources are not considered part of the implementations at all.

For the future, integrating information, monitoring, management, and computing systems is necessary for effectively and efficiently using these resources. As an example, the information on private, governmental, and industrial land use, national parks information, and energy and mineral resources exploration is very complex. The parameters of wind energy and solar energy are highly dynamical. Calculation of weather impacts, construction and simulation of new facilities offshore and onshore leads to new demands on information and computing.

The case study showed that "trust in computation", reliability and suitability of information, QoD, and security of critical investments are most important for the academic and industrial partners. As a result, for this scenario it has been regarded necessary for the services to implement and configure a complex combination of the following features:

- dedicated networks,
- firewalls, access lists, routing,
- sandboxes,
- trusted scripting,
- shared and non-shared use,
- queue limits,
- demilitarised zone,
- access control and keys,
- resources monitoring and accounting,
- local auditing, communication packet filtering, security management,
- job logging,
- encrypted data transfer and communication,
- services monitoring,
- on-site support and management.

Table IV gives a summarising excerpt of the collaboration matrix, showing columns (Resources, Services, Disciplines), topics (implementation, legal issues, security), the location where the activity is concentrating on, and the focus topics with their priority (highest priority is marked '!!!').

As the table shows, emphasised priority with this case is on computational aspects. No individual client applications have been regarded necessary for the users with this case study. The standard client is a Secure Shell client accessing resources and implementing individual automation facilities using access key pairs and batch system services. User groups will develop many of the algorithms and tools needed for processing. Public keys are not the primary means for authentication with this scenario. Data encryption is triggered by users. Decision flow and overall priority is on the users side. Security and reliability concentrates on computation and shared resources usage. Legal aspects are handled on disciplines/users side. The collaboration matrix shows that both the disciplines and services have had to concentrate on the legal topic-framework implementation as the services layer had to be staffed with privileged users as well as operated and managed by services groups.

With assignment of distributed resources, legal and security aspects of information security and access require transparent handling on services side. The Active Source concept has been used for the implementation, based on a modular server-client and services interfaces architecture. Figure 6 shows a screenshot from the environmental application case study, showing an Active Map of the National Parks information system and some distributed online information [24], [25]. The implementation targets on environment-geosciences information and energy exploration based on High Performance Computing (HPC) and Distributed Computing (DC) services.

For the current collaboration, [26] resources management

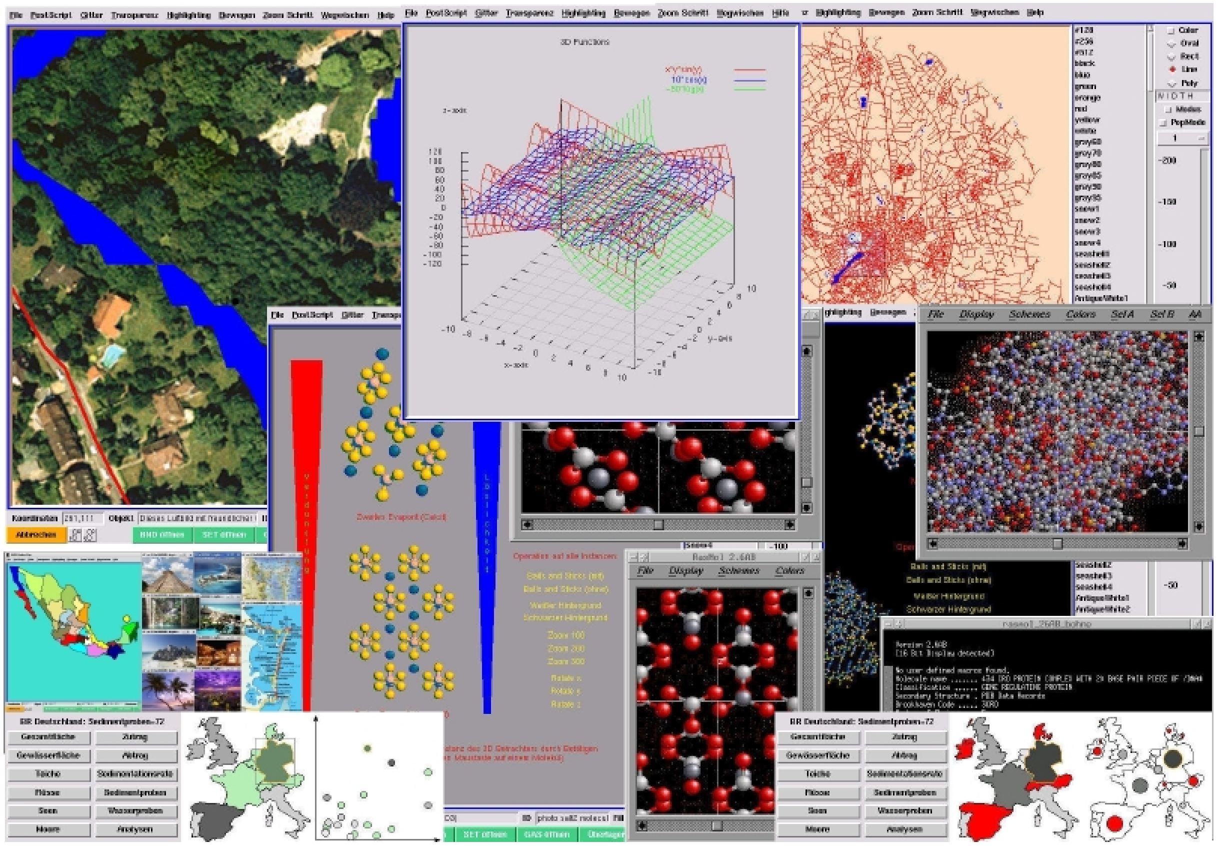

Figure 5. Basic application scenario implementations from the GEXI case studies examples.

for geosciences, energy sciences, and mobility is based on the collaboration framework. A couple of aspects of future Enterprise Resource Planning (ERP) systems are most relevant for supporting integrated HEC and Information Systems in industrial applications in this scenario:

- combining distributed and parallel techniques,
- transaction processing,
- highly parallel communications,
- loosely parallel communications of highly parallel applications,
- industrial systems,
- expanding scalability,
- improving the number of compute nodes for industry fields of application, currently less than 100 compute nodes,
- open up access to scientific scenarios with currently up to over 500 compute nodes,
- preparing new solutions for complex industrial information-computing systems.

These are objects for further implementation on the resources and services columns. The implementations for global monitoring software and expanding the physical distribution for PSI strongly depend on their availability.

The case study further showed that "trust in computing" and reliability are most important for the academic and industrial partners. It has been possible to transparently separate nearly all of the implementation aspects for the three columns. The most prominent conjoint implementation issues having to be worked on for the future is national laws and regulations as far as there is no general solution for securing critical data, and computation with distributed usage cannot be supported by signing some kind of black-box "computation objects".

C. Epidemology information and computing

Interdisciplinary research in the ecology and epidemology of vector-borne diseases produces huge amounts of data regarding to biological and epidemological processes [27], [28].

In epidemology a vector is an insect or any living carrier that transmits an infectious agent. Examples are hematophagous arthropod vectors such as mosquitos, ticks and flies which are responsible for transmitting protozoa, bacteria, and viruses between vertebrate hosts, causing diseases as Malaria, Lyme disease, and Sandfly Fever. Examples for processes are disease prevalence, abundance, and distribution of living organisms.

These parameters are highly dynamical and are influenced in time and space by a number of biotic (e.g. vegetation) and abiotic (e.g. temperature and humidity) factors.

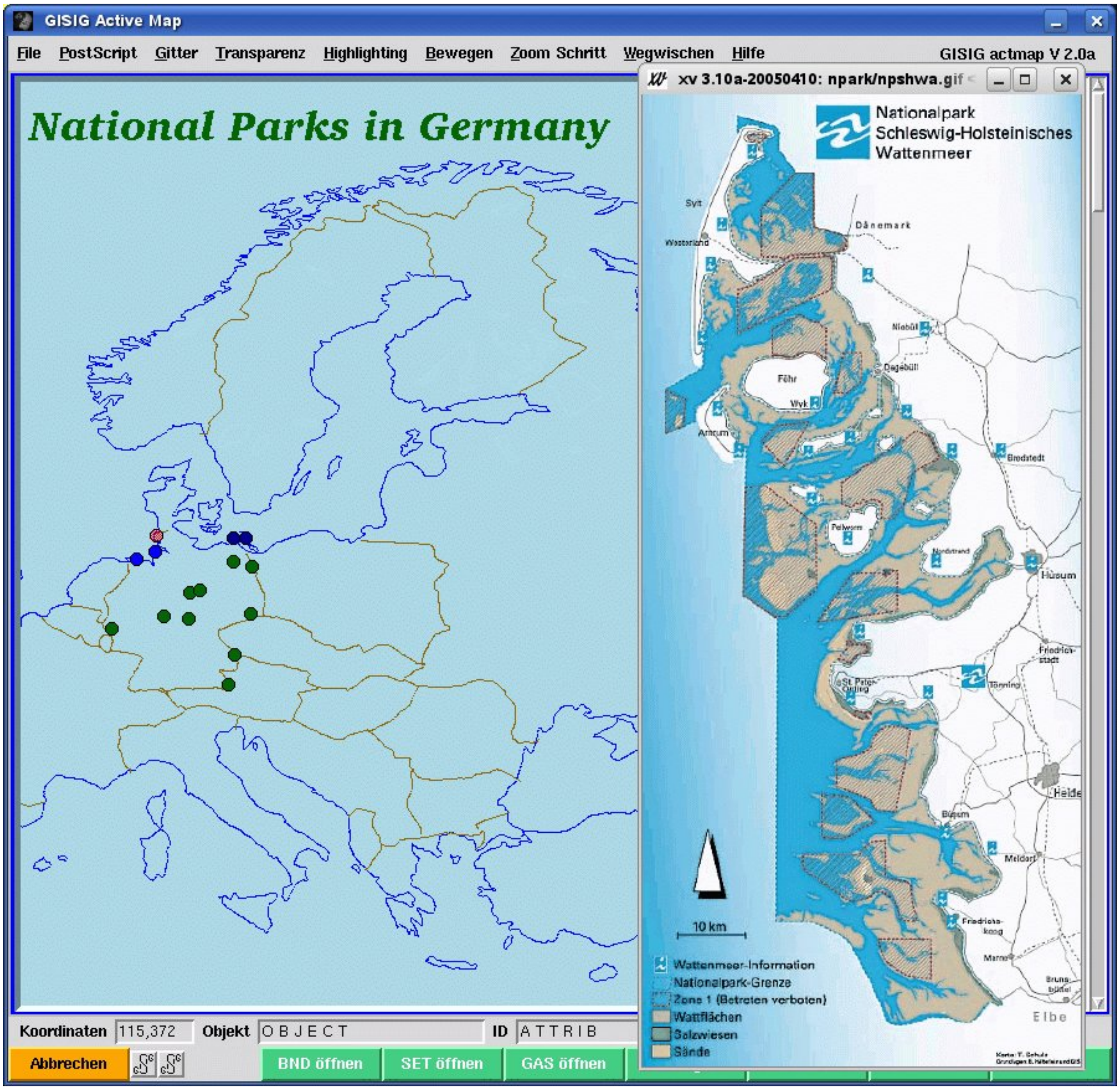

Figure 6. Environmental application case study showing a distributed multi-discipline Active Source implementation.

There are three main types of objects: epidemological objects, environmental objects, and sociocultural objects. The following listing showing the simplified example matrix.

- epidemological: disease case numbers, vector distribution, number of vectors with pathogen,
- environmental: landscape structure, meteorological conditions, vegetation coverage, micro climate change,
- sociocultural: interchange of organisms by human travel and commerce.

The case study showed that using security enhanced, PKI/CA and PKC [29] based integrated Information and Computing System components for scientific application including industry participation has been regarded most suitable. By these means evaluation of data, generating various views, querying distributed information and visually summarising information by calculating multi-dimensional views varying in time and spatial representation can be supported in a secure and flexible way. For locally distributed information spatial information processing algorithms can be employed. Partners from several disciplines will be able to evaluate and analyse this data, e.g., human health epidemologists, public health authorities, and physicians.

A well balanced infonomics system will allow for the input, storage, data manipulation, analysis, and visual presentation of georeferenced data and is particulary suitable for identifying local clusters of diseases, and for analysing spatial relationships between diseases and risk factors.

Table V gives a summary for the collaboration matrix. As the table shows, emphasised priority with this case is on content and utilisation aspects. Special client applications

Table IV
COLLABORATION MATRIX FOR THE ENVIRONMENTAL INFORMATION AND COMPUTING CASE STUDY.

Column	*Topic*	*Location*	*Focus : Priority*
		Physical	
R – –	i – s	provider	compute resources security : !!
R – –	i – s	provider	storage resources security : !!
		Application	
– – D	i – s	prov./user	content access client : !
– – D	i – s	user	services automation : !
– S –	i – s	provider	process communication security : !!!
– S –	i – s	user	execution security : !!!
		Computation	
R – –	i – s	prov./user	comp. res. availability : !!!
R – –	i – s	prov./user	resources access security : !!!
R – –	i – s	provider	interfaces security : !!!
R – –	i – s	provider	computation reliability : !!!
– S –	i – s	provider	power-on encryption : !!!
		Content Information	
– – D	– l –	user	pollution data legal regulations : !!
– – D	i – s	user	functional content security : !!
– – D	i – –	user	object data regulations : !!
– – D	i – s	user	event data security : !!
– – D	i – s	user	autoevent data security : !!
– – D	i – s	user	power-off encryption : !!
– – D	- l –	user	national legal context : !!
		Utilisation Information	
– S –	i – s	provider	content data transfer : !
– S –	i – –	provider	user and client access : !
– S –	i – s	provider	content storage security : !
– S –	i – –	provider	modification and transfer : !
– S –	i – –	provider	reliability of communication : !
– – D	l – –	user	information signing : !
– – D	i – –	user	distribution of information : !
– S D	– l –	provider	national laws and regulations : !

Table V
COLLABORATION MATRIX FOR THE EPIDEMOLOGY INFORMATION AND COMPUTING CASE STUDY.

Column	*Topic*	*Location*	*Focus : Priority*
		Physical	
R – –	i – s	provider	compute resources security : !!
R – –	i – s	provider	storage resources security : !!
		Application	
– S –	i – s	prov./user	content access client : !!!
– S –	i – s	prov./user	services automation : !!!
– S –	i – s	provider	process communication security : !!!
– S –	i – s	user	execution security : !!!
		Computation	
R – –	i – s	prov./user	comp. res. availability : !
R – –	i – s	prov./user	resources access security : !!
R – –	i – s	provider	interfaces security : !!
R – –	i – s	provider	computation reliability : !
– S –	i – s	provider	power-on encryption : !
		Content Information	
– – D	– l –	user	epidemiol. data legal regulations : !!!
– – D	i – s	user	functional content security : !!
– – D	i – –	user	object data regulations : !!
– – D	– l s	user	content data security : !!!
– – D	i – s	user	autoevent data security : !!
– – D	i – s	user	power-off encryption : !!
– – D	- l –	user	national legal context : !!
		Utilisation Information	
– S –	i – s	provider	content data transfer : !!!
– S –	i – –	provider	user and client access : !!!
– S –	i – s	provider	content storage security : !!!
– S –	i – –	provider	modification and transfer : !
– S –	i – –	provider	reliability of communication : !
– – D	l – –	user	information signing : !!!
– – D	i – –	user	distribution of information : !
– – D	– l –	provider	national laws and regulations : !

are necessary for authors and users with this case study. The user groups will not develop services on their own, neither algorithms nor tools needed for processing. The centre of the information system is the PKI/CA and PMI/AA infrastructure. Processing uses the signed objects within the services layer, only accessible via dedicated service interfaces.

Security therefore concentrates on information and more or less computation. The physical shared resources usage can be critical due to the storage/scheduling location being currently not predictable while being effective. The services column is responsible for the system and client security.

Legal aspects are handled on disciplines side. The collaboration matrix shows that the disciplines column has to concentrate on the legal topic-framework implementation.

With this scenario "trust in information" is twofold, regarding the content information domain and the utilisation information domain. It has been possible to transparently separate nearly all of the implementation aspects for the three columns.

The case study showed that for the application within the integrated information and computing system the three main types of objects need role-based data access for users and clients. Objects have to be signed with digital signature and timestamps of the originating authors and manipulation. For real life scenarios network transfer encryption has to be used. Due to a distributed storage environment host-side encryption has been regarded necessary.

Figure 7 shows the workpackage layers view. As well as in other disciplines, in epidemiology sciences on the one hand there is a strong need to assure accurate data objects all over the life-cycle of objects, thus for content guaranteeing "trust in information". On the other hand a suitable access control infrastructure has to be established.

Strong authentication and authorisation by means of cryptographic techniques specified as Public Key Certificates (PKC) and Attribute Certificates (AC) in Public Key Infrastructure (PKI) and Privilege Management Infrastructure (PMI) environments [30] provides a framework for adressing important security considerations of authentication, confidentiality, authorisation, and integrity (PKI) and allows for particularly controlled data access (PMI). In this regard the authority (CA) signs the public user keys in order to maintain the integrity of the public key, expiration information

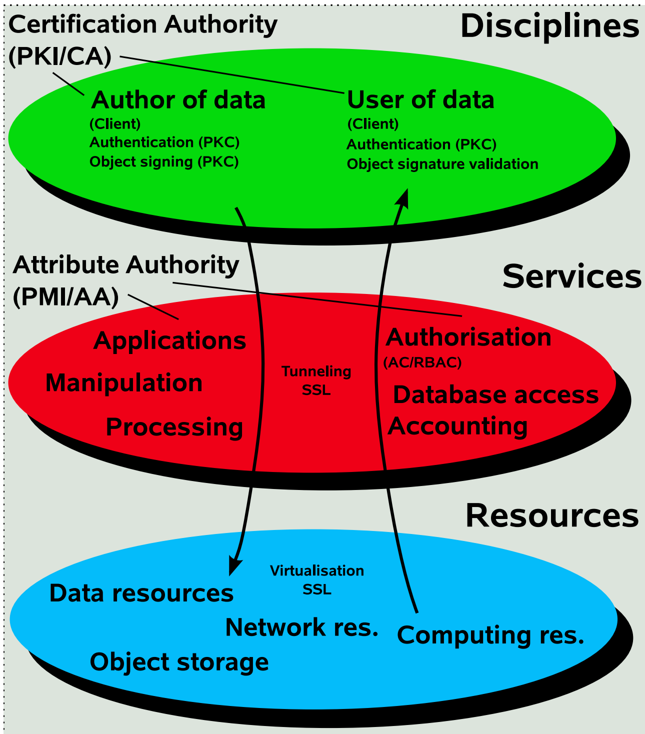

Figure 7. Workpackage layers for the epidemology information and computing case study development.

and other important information contained within the user certificates. Attribute Authority (AA) is the authority which assigns attributes (permissions and privileges) by signing the attribute certificate. Digital signatures are used in both PKI and PMI as the mechanism which binds the issueing certification authority to the certificate. CA and AA are separate authorities and should be established independently.

A sophisticated trust management system using X.509 Attribute Certificates [31] can be used to store the user roles, based on Role Based Access Control (RBAC) [32].

In case of multidisciplinary working and development groups (Figure 7), intending epidemiological analysis of infectious diseases, the procedure has been used as follows. For an operation example with the case study this means:

- **Disciplines layer:** The author of a data object, the originator, e.g. co-worker of a human health organisation, signs the created object, e.g. disease case numbers, with the private key of the author and according timestamps.
- **Services layer:** The object is processed by the application services.
- **Resources layer:** The processed object is stored to the distributed storage.
- **Disciplines layer:** An user, e.g. a member of a research team at an university, requests an object.
- **Services layer:** The user is authenticating at the authentication service via AC/RBAC, the signature of the corresponding author is validated. The authentication service requests the object or service operation.
- **Resources layer:** The object is collected from the distributed resources.
- **Services layer:** The object is calculated, accounted, and provisioned via services for the user client.

X. Lessons learned

The environmental case has been found more heterogeneous than the epidemological example. This is the result of the fact that in these disciplines many interest groups are dealing with algorithms, simulations and spatial planning for over the last decades. These disciplines will be able to develop some own services components and make broader use of sophisticated High End Computing resources. It has been more transparent to define interfaces for discussing and integrating legal frameworks and regulations into the multi-disciplinary implementation process. Nevertheless for the majority of use cases, smaller scenarios can be seen where strict separation of disciplinary work, services and development, and operation of resources will be lived.

This leads to the conclusion that in the future of integrated information and computing systems we will need to create means of securely submitting modular application components into the services pipeline.

A collaboration framework for development and operation in combination with and integrated information and computing systems handling these features should be able to address many of the heterogeneous conditions regarding implementation, legal issues, and security existing in national as well as international context. It will support stearing of data, information, and application workflows being conform with legal regulations and data security standards as well as obeying policies.

XI. Conclusion and Future work

Developing international cooperations within the fields of geosciences, exploration, and computing, based on an interdisciplinary collaboration framework is regarded a perfect solution for all partners in the GEXI study for the different case study scenarios, in order to modularise information system development and reduce complexity. Technology suitable for solving open problems with implementation, legal environment or trust is still in the genesis. For instance with trust and encryption regarding Distributed Computing and shared resources, full homomorphic encryption techniques are desirably. Currently basic algorithms are available.

The modular integration of services and disciplines within a collaboration framework has proven best results to be flexible and efficient for large international projects with various legal characteristics as separating technical and legal work packages. With this, the future allocation of responsibilities and integration of specialist frameworks has become a more transparent process.

The ongoing work is oriented towards the consequences resulting from the evaluation. The legal aspects for these topics will attend the next steps in order to elaborate the framework for use with a collaboration portal with support from legal working groups.

At this point it is essential for complex implementations to integrate the national legal regulations (in Germany e.g. IFG, GeoZG, IWG) with securely managing content and workflows into the international context as well as to refine the interactions within the collaborating sections:

1) For legal issues with geosciences and exploration, the aspects of data contents, cognostics, data combination and automation (georeferencing) and parallelisation for HPC and shared resource usage are in the focus. Integration with the national and international legal frameworks (e.g. GDI-DE, INSPIRE, GSDI) will have to be forced in order to accomplish a base for future information and decision making systems for commercial and educational purposes.
2) Regarding the Distributed Computing and services column, desktop user interfaces, services and security, networking, and undisputable distributed accounting have been set top on the working list.
3) With High Performance Computing and resources standardisation and a secure networking model with PKI, privacy, and encryption support for future HEC architectures development is priority.

Further on with implementation and legal issues, the security aspect are on the rise for any complex system. Even though PKI technology offers means to attest, identify, manage the exchange of encryption keys an secure transmission between parties, there has not been broad-based adoption of PKI technology by public and private organisation. After all, a significant number of countries recognise digital signatures as legally binding. In case of security enhanced integrated information and computing system components object signing provides a robust solution to facilitate “trust in information” and to overall support “trust in computing”. In order to put this implementation into international public practice there is a need for future PKI development and deployment offering a global public key cryptosystem for the Future Internet.

Preliminary work has created a common base for an ethical understanding of cross-disciplinary use of data. Various trust situations, important for services providers on the one hand as well as for disciplines on the other hand could be described, handled, and implemented with the separation of work packages. This work showed that it is possible to bring complex information and computing systems to life, being able to create interfaces that can also be interfaces between the logical columns and interest groups.

ACKNOWLEDGEMENTS

We are grateful to all national and international academic and industry partners in the GEXI cooperations for the innovative constructive work and especially for the environmental and energy exploration case study input, security advisor and representative Mrs. Birgit Gersbeck-Schierholz for implementing and coordinating the epidemology disciplines case study, and to the colleagues at the Leibniz Universität Hannover, at the IRI, the North-German Supercomputing Alliance (HLRN), WWU, ZIV, D-Grid and the participants of the postgraduate European Legal Informatics Study Programme (EULISP) for prolific discussion of scientific, legal, and technical aspects as well as to the staff at ZIV, ZIB, L3S and associated HPC companies for supporting this work by managing and providing HEC resources over the years.

REFERENCES

[1] C.-P. Rückemann, “Legal Issues Regarding Distributed and High Performance Computing in Geosciences and Exploration,” in *Proceedings of the International Conference on Digital Society (ICDS 2010), The International Conference on Technical and Legal Aspects of the e-Society (CYBERLAWS 2010), February 10–16, 2010, St. Maarten, Netherlands Antilles / DigitalWorld 2010, International Academy, Research, and Industry Association (IARIA)*. IEEE Computer Society Press, IEEE Xplore Digital Library, 2010, pp. 339–344, Berntzen, L., Bodendorf, F., Lawrence, E., Perry, M., Smedberg, Å. (eds.), ISBN: 978-0-7695-3953-9, URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5432414 (PDF) [accessed: 2010-12-26], (Best Paper Award).

[2] “Geo Exploration and Information (GEXI),” 1996, 1999, 2010, URL: http://www.user.uni-hannover.de/cpr/x/rprojs/en/index.html#GEXI (Information) [accessed: 2011-01-01].

[3] C.-P. Rückemann, *Accounting and Billing in Computing Environments*. Business Science Reference, IGI Global, Hershey, Pennsylvania, USA, Oct. 2009, 25 pages, Chapter X, in Pankowska, M. (ed.), Infonomics for Distributed Business and Decision-Making Environments: Creating Information System Ecology, 421 pages, ISBN: 978-1-60566-890-1, URL: http://www.igi-global.com/reference/details.asp?ID=34799 (Information) [accessed: 2010-12-26].

[4] C.-P. Rückemann, “Using Parallel MultiCore and HPC Systems for Dynamical Visualisation,” in *Proceedings of the International Conference on Advanced Geographic Information Systems & Web Services (GEOWS 2009), February 1–7, 2009, Cancun, Mexico / DigitalWorld 2009, International Academy, Research, and Industry Association (IARIA)*. IEEE Computer Society Press, IEEE Xplore Digital Library, 2009, pp. 13–18, Dragicevic, S., Roman, D., Tanasescu, V. (eds.), ISBN: 978-0-7695-3527-2, URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4782685&isnumber=4782675 (PDF) [accessed: 2010-12-26], (Best Paper Award).

[5] C.-P. Rückemann, “Beitrag zur Realisierung portabler Komponenten für Geoinformationssysteme. Ein Konzept zur ereignisgesteuerten und dynamischen Visualisierung und Aufbereitung geowissenschaftlicher Daten,” Dissertation, Westfälische Wilhelms-Universität, Münster, Deutschland, 2001, 161 (xxii + 139) Seiten, Ill., graph. Darst.,

Kt., URL: http://wwwmath.uni-muenster.de/cs/u/ruckema/x/dis/download/dis3acro.pdf [accessed: 2010-12-26].

[6] C.-P. Rückemann, "Geographic Grid-Computing and HPC empowering Dynamical Visualisation for Geoscientific Information Systems," in *Proceedings of the 4th International Conference on Grid Service Engineering and Management (GSEM), September 25–26, 2007, Leipzig, Deutschland, co-located with Software, Agents and services for Business, Research, and E-sciences (SABRE 2007)*, R. Kowalczyk, Ed., vol. 117. GI-Edition, Lecture Notes in Informatics (LNI), Gesellschaft für Informatik e.V. (GI), 2007, pp. 66–80, ISBN: 78-3-8579-211-6, ISSN: 1617-5468.

[7] 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: 2010-12-26].

[8] "Global Monitoring for the Environment and Security (GMES)," 2010, URL: http://www.gmes.info [accessed: 2011-01-01].

[9] "Global Earth Observation System of Systems (GEOSS)," 2010, URL: http://www.earthobservations.org/geoss.shtml [accessed: 2011-01-01].

[10] "Group on Earth Observations (GEO)," 2009, URL: http://www.earthobservations.org [accessed: 2011-01-01].

[11] "Shared Environmental Information System (SEIS)," 2010, URL: http://ec.europa.eu/environment/seis/ [accessed: 2011-01-01].

[12] "Global Spatial Data Infrastructure (GSDI)," 2010, URL: http://www.gsdi.org [accessed: 2011-01-01].

[13] "INfrastructure for SPatial Information in Europe (INSPIRE)," 2010, URL: http://www.ec-gis.org/inspire [accessed: 2011-01-01].

[14] "Geodateninfrastruktur Deutschland (GDI-DE)," 2010, URL: http://www.gdi-de.org [accessed: 2010-12-26].

[15] "European Public Sector Information (EPSI)," 2010, URL: http://www.epsiplus.net [accessed: 2011-01-01].

[16] C.-P. Rückemann, *Fundamental Aspects of Information Science, Security, and Computing (Lecture)*. EULISP Lecture Notes, European Legal Informatics Study Programme, Institute for Legal Informatics, Leibniz Universität Hannover (IRI / LUH), 2009, URL: http://www.eulisp.de [accessed: 2011-01-01].

[17] "SGAS – SweGrid Accounting System," 2010, URL: http://www.sgas.se/ [accessed: 2011-01-01].

[18] P. Gardfjäll, E. Elmroth, L. Johnsson, O. Mulmo, and T. Sandholm, "Scalable Grid-wide Capacity Allocation with the SweGrid Accounting System (SGAS)," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 18, pp. 2089–2122, 2008, URL: http://www.cs.umu.se/~elmroth/papers/sgas_revised_aug_2007.pdf [accessed: 2010-12-26] (Preprint).

[19] T. Sandholm, P. Gardfjäll, E. Elmroth, L. Johnsson, and O. Mulmo, "A Service-oriented Approach to Enforce Grid Resource Allocations," *International Journal of Cooperative Information Systems*, vol. 15, no. 3, pp. 439–459, 2006, URL: http://www.cs.umu.se/~elmroth/papers/SGASIJCIS_2006.pdf [accessed: 2010-12-26].

[20] E. Elmroth and P. Gardfjäll, "Design and Evaluation of a Decentralized System for Grid-wide Fairshare Scheduling," *e-Science 2005, First IEEE Conference on e-Science and Grid Computing*, pp. 221–229, 2005, URL: http://www.cs.umu.se/~elmroth/papers/fsgrid.pdf [accessed: 2010-12-26].

[21] E. Elmroth, P. Gardfjäll, O. Mulmo, and T. Sandholm, "An OGSA-based Bank Service for Grid Accounting Systems," *State-of-the-art in Scientific Computing*, vol. 3732, pp. 1051–1060, 2006, URL: http://www.cs.umu.se/~elmroth/papers/egms_para04.pdf [accessed: 2010-12-26].

[22] E. Elmroth and D. Henriksson, "Distributed Usage Logging for Federated Grids / Future Generation Computer Systems," *The International Journal of Grid Computing: Theory, Methods and Applications*, submitted 2009.

[23] E. Elmroth, F. Galán, D. Henriksson, and D. Perales, "Accounting and Billing for Federated Cloud Infrastructures," *In: Proceedings of the Eighth International Conference on Grid and Cooperative Computing (GCC 2009), J. E. Guerrero (ed.)*, pp. 268–275, 2009, URL: http://www.cs.umu.se/~elmroth/papers/eghp_gcc2009.pdf [accessed: 2010-12-26].

[24] "Nationalpark Schleswig-Holsteinisches Wattenmeer," 2010, URL: http://www.sh-nordsee.de/nationalpark/ [accessed: 2010-12-26].

[25] "National Parks, Germany," 2010, URL: http://www.nationalparke.de/ [accessed: 2010-12-26].

[26] C.-P. Rückemann, "Future Geo-Exploration Information and Computing Systems Created by Academia-Industry Collaboration," in *Proceedings of the 9th International Scientific-Practical Conference 2010 (HTFR 2010), April 22–23, 2010, Saint Petersburg, Russia*. Saint Petersburg: Saint Petersburg University Press, 2010, pp. 254–258, ISBN: 978-5-7422-2558-4, URL: http://htfr.org [accessed: 2010-12-26].

[27] C. G. Moore, "Interdisciplinary research in the ecology of vector-borne diseases: Opportunities and needs," *Journal of Vector Ecology*, vol. 33, no. 2, pp. 218–224, 2008, DOI: 10.3376/1081-1710-33.2.218, URL: http://www.sove.org/Journal%20PDF/December%202008/1-Moore%2008-56.pdf [accessed: 2010-12-26].

[28] S. I. Hay, A. Graham, and D. J. Rogers, Eds., *Global Mapping of Infectious Diseases: Methods, Examples and Emerging Applications*. Academic Press, Edition: Pap/Dvdr R, 2007, iSBN-10: 0-12-031764-8, ISBN-13: 978-0-12-031764-6, ISBN-13: 978-0-12-031766-0 (DVD), ISBN-10: 0-12-031766-4 (DVD).

[29] B. F. S. Gersbeck-Schierholz, "Trustworthy Communication by Means of Public Key Cryptography," 2010, URL: http://www.rrzn.uni-hannover.de/fileadmin/it_sicherheit/pdf/pki2010_gersbeck.pdf [accessed: 2010-12-26].

[30] "ITU-T Recommendation X.509 ISO/IEC 9594-8, The Directory: Authentication Framework," 2000, URL: http://wwwitu.int/itu-t/recommendations [accessed: 2010-12-26].

[31] D. W. Chadwick and A. Otenko, "The PERMIS X.509 role based privilege management infrastructure," *Future Generation Computer Systems (FGCS)*, vol. 19, pp. 277–289, 2003, DOI: 10.1016/S0167-739X(02)00153-X, URL: http://portal.acm.org/citation.cfm?id=770786 [accessed: 2010-12-26].

[32] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman, "Role-Based Access Control Models," *IEEE Computer*, vol. 29, no. 2, pp. 38–47, 1996, DOI: 10.1109/2.485845, ISSN: 0018-9162, URL: http://csrc.nist.gov/rbac/sandhu96.pdf [accessed: 2010-12-26].