Chapter 3

Project of "Pôle GLC" (Génie du Logiciel et de la Connaissance)

3.1 Permanent members (01/2012)

	position	pedr or pes	teaching dep.	A/D	administrative duties
Blay-Fornarino Mireille	PR2 27		info.IUT		In charge of AL speciality of IFI Master (2007-)
Boudaoud Karima	MCF 27		R&T.IUT		
Buffa Michel	MCF 27		info. UFR sc.		
Collet Philippe	MCF 27	PEDR 2008-	info. UFR sc.		
Crescenzo Pierre	MCF 27		info. UFR sc.		Head of MIAGE, UFR Sc. (2008/-)
Da Costa Pereira Célia Cristina	MCF 27		stid.IUT	2010/-	
Dartigues Christel	MCF 27		stid.IUT		
Faron-Zucker Catherine	MCF 27		Info.EPU		In charge of KIS speciality of IFI Master (2007/-)
Gaignard Alban	IE2 CNRS				
Lahire Philippe	PR2 27		info. UFR sc.		vice-dean of the faculty of sciences
Lavirotte Stephane	MCF 27	PES 2009-	IUFM		In charge of IAM speciality of IFI Master (2009/-)
LeThanh Nhan	PR1 27		info. UFR sc.		In charge of LPro SIL
Lingrand Diane	MCF 61		info.EPU		
Mirbel Isabelle	MCF 27 HDR		info. UFR sc.		
Montagnat Johan	DR2 CNRS				
Occello Audrey	MCF 27		info.EPU		
Pallez Denis	MCF 27		stid.IUT		
Pasquier Claude	CR2 CNRS			2012/-	
Pasquier Nicolas	MCF 27		info. UFR sc.		
Pinna-Dery Anne-Marie	MCF 27		info.EPU		In charge of HCI speciality of IFI Master (2007/-)
Renard Helene	MCF 27		info.EPU		
Renevier-Gonin Philippe	MCF 27		info. UFR sc.		In charge of L3 MIAGE, UFR sc. (2007/-)
Rey Gaetan	MCF 27		info.IUT		Head of first and second year CS dept, IUT Nice (2009/-)
Rigault Jean-Paul	PR1 27		info.EPU	2012/-	
Riveill Michel	PREX 27	PEDR 1994-	info.EPU		Vice-chair of UNS (campus Sophia Antipolis)
Sander Peter	PR2 27		info.EPU		In charge speciality of IFI Master localized in Vietnam (2008/-)
Tigli Jean-Yves	MCF 27	PEDR 2007-	info.EPU		

3.2 Self evaluation

In order to perform the self evaluation of the GLC pôle, we first make an analysis of progress with regards to the remarks made in the previous evaluation report. Then a risk analysis is performed and envisaged solutions are described, ranging from human resource management, funding and partnership to research lines.

In the evaluation report issued in November 2007, the following remarks were made to the GLC pôle (translated from french):

"This pôle is composed of two teams: Kewi and Rainbow. It scientific domain is software engineering of distributed systems and distributed knowledge over the Web. These two teams are new in their current boundaries.

Rainbow is a quite large team, which takes over the old team of the same name, itself of recent creation, as well as the active part of the old OCL project. This merger was not recommended in the previous evaluation: the old OCL project was of insufficient level and could jeopardize the still under restructuration Rainbow team. Nevertheless, for reasons of thematic proximity and of strong motivation, including from the active part of OCL, the Rainbow team has chosen this merger and precisely argues for it in the report. The risk still remains, but the potential is clear. It is now up to the team to digest its growth.

The Kewi team works on distributed knowledge engineering, typically on the Web. Its research field is quite large and we understand this team as a group under evolution, whose topics and boundaries are not well established today. This team seems an evolving basis to us. The team production is interesting, but it has some progression margin: the team lacks a real scientific dynamics, one or more leaders that could help current team members to express more efficiently their scientific potential. We recommend the lab not to let this team grow, unless a leader can be appointed. We recommend the pôle leaders to make a regular and specific follow-up on the activities of the team and of its members. We also recommend them to remain open to evolutions in structures and boundaries."

Two years and a half later, we can assert that the situation has noticeably evolved.

- The GLC pôle has a real identity and is well positioned in the lab. It makes a place for living and exchange and stands as a genuine team. The pôle has helped the weakest teams to have some support both at the scientific and financial levels.
- From the large Rainbow team, two teams, Modalis and Rainbow, have emerged. The first one is more focused on computational grid for medical imaging, while the second one targets ubiquitous computing. The links between the two teams are strong even if their research topics are slowly diverging. The scientific objectives are well established today and, in each of the two teams, future leaders are arising.
- On the other hand, Kewi gave birth to two teams: Keia and Kewi. Keia is a new team built around Nicolas Pasquier, who should obtain his HDR soon. This team has been strengthened by a MCF 27 position (in September 2010) and by a CNRS researcher (CR) in Biology, with whom the team has been collaborating for several years. On the other hand the Kewi team has been widely supported by INRIA through several delegations and temporary postings ("détachement") within the Edelweiss project, with which Kewi was maintaining informal relationships. Some plan for a joint team-project is undergoing construction, and a full professor position was opened this year, with unfortunately a few applicants within the profile (the only external candidate that came to the hearings was ranked first, but he has preferred staying in his university where a promotion was also possible for him).

We now perform a risk analysis following the SWOT [1] decomposition.

1. S for strengths (internal), W for weaknesses (internal), O for opportunities (external) and T for threats (external).

	Helpful	Harmful
Internal	**Strengths** - Very strong relationships with companies - Strong group solidarity among members - Very strong implication in teaching structures with close proximity with students - Very strong activity recovery since the previous contract	**Weaknesses** - Effort-consuming partnerships to establish and maintain - Effort-consuming empirical validation needed through large software prototyping or technology transfer - Lack of leading senior staff (full professors with (inter-)national scientific leadership) - Numerous MCF in position to obtain HDR, which creates competition risk in future promotions - Low number of *equivalent-full-time* researchers (13 against an average of 16.25 in other pôles, despite a permanent staff number similar to them) - High number of associate members
External	**Opportunities** - Excellent relations with the INRIA Sophia Mediterranean Center (delegation, temporary posting, common teams even without an EPC) - Creation of the STIC campus - New teaching program "smart building" could reinforce research activities - High volume of tenure (mainly senior) positions to be open, if re-openings are effective	**Threats** - Bidding and coordination of research projects, effort consuming - Specific industrial fabric (no industry, no local headquarters to get firm decisions, only start-ups that oblige to be reactive) - Administrative duties overload for some members - Overload due to involvement in teaching department, for some members - Low attractiveness of the area (high cost of living, long inter-campus distance)

Facing those risks and opportunities, we have already established some guidelines to better drive people and teams. Regarding people, the GLC pôle is striving to improve their research environment by :

- maintaining the current good working atmosphere;
- helping associate members to get back to research (not all of them will, but lowering the number of associates is a focused objective);
- smartly managing careers of junior members (MCF) to keep on with the current dynamics;
- building upon faced opportunities (new campus. . .) to attract junior and senior researchers with excellent records.

As for partnership and funding, the GLC pôle has demonstrated diverse and strong partnerships with companies of all sizes. A large amount of various funding sources (local, national, EU level) have been and are still going to be sought by all pôle members. While providing salient use cases, these efforts have also led to appropriate engineering means and increasing numbers of PhD studentships. Despite the effort spent on these tasks, the pôle needs to continue to push in that direction as use cases and validation from real industrial scenarios are the best way to pursue excellence in our research lines, while getting an adapted funding level. Hopefully we expect the next bidding and project coordinations to be less time-consuming as the pôle members are becoming more and more hardened to these activities.

As demonstrated at the beginning of this self-evaluation, the pôle is also striving to help team restructuring. But it also supports each team in identifying research lines that should be closed, either because we will never have the force to have a proper visibility or because we do not have enough expertise in the field. Besides the pôle also helps determining activities that should be started or deepened because of identified opportunities at mid or long terms. For example, since November 2007, the pôle has voluntarily:

- **stopped** activities related to *attack detection* and to relocate research activities related to security on more focused objectives according to the available expertise and man power.
- **stopped** activities bound to *data mining in economics*, as the MCF that was leading them has been promoted in another university. Data mining activities were then **refocused** on scientific domains of the STIC campus, i.e. computational biology and sustainable development.

– **started** a broader activity on Human-Computer Interaction (HCI), focusing on compositions of all aspects of an application, from its core to its presentation.
– **deepened** the remaining research lines on middleware, composition, semantic web, security and data mining.

3.3 Project and objectives

3.3.1 Middleware

Main contributors *P. Collet, A. Gaignard, P. Lahire, S. Lavirotte, D. Lingrand, J. Montagnat, H. Renard, G. Rey, M. Riveill, J.-Y. Tigli*

Over the coming years, the pressure to address distributed computing problems will keep growing. Seamless networking has become a reality and a growing number of devices will rely on external services and interoperability protocols for proper operations. Efficiently and seamlessly distributing workload over many processing units will be increasingly needed as the scale of scientific data analysis problems grows. With the higher number of processing units involved in the computations, system failures will become a norm and self-adapting or self-healing workload managers will be required.

The *middleware* research activity just emerged at the beginning of the last four-years period in I3S-GLC and it has now matured and acquired good international visibility. It is planed to take advantage of the momentum created over the next period, continuing to invest significantly in this area without major changes of the challenges addressed, although some significant evolutions of the techniques used have to be noted. The double expertise of the members of I3S-GLC in the areas of *Services/Component Architectures* and *High Throughput Computing* is a competitive advantage to tackle these issues in the coming years. Links that have been established between these communities, in particular with the creation of the MODALIS research team in 2008, will be strengthen. Techniques from the area of *Autonomic Computing* are also being adopted to tackle the natural unreliability of large scale distributed systems. *Semantic Web* technologies are also considered to abstract the middleware from the technological platforms supporting it, and delivering to end-users services that are better instrumented to deal with a specific domain of science. From a practical point of view, we also maintain an active contribution to the production and research grid infrastructures developed at the European and National levels (European Grid Initiative - InsPIRE project and Institut des Grilles du CNRS) that are fundamental to support our research activities.

The current focus on adaptable middleware for network of devices is put on the development of new weavers approaches for Aspects of Assembly and mechanisms to improve relevance of adaptation. These new weavers approaches aim at improving modularity (multi-cycle weaving techniques, PhD thesis started in 2008) and interaction between concerns of multiple domains (multiple weavers for Aspects of Assembly, PhD thesis started in 2009). Because the complexity of the resulting software architecture also need to be tackled, we need to improve the relevance of adaptation. We aim at developing contextual filters before weaving aspects for that purpose (one PhD thesis started in 2010) or we use semantic information to select aspects more efficiently (part of ANR project Continuum started in 2008).

The effort on software adaptation control will be completely redirected towards autonomic management and addressing challenges of large scale distributed systems. With the growth in scale of distributed system, their complexity challenges all modern software development techniques, leading to non-fully stabilized middleware stacks. Moreover, hardware components are subjects to failures and the Mean Time Between Failures becomes a critical metrics which impact is observed daily. Since failures *cannot* be completely avoided, their happening has to be taken into account in the design of such systems. Self-adaptation and principles to make software components autonomously resilient in case of failures will be developed. A challenge is to address all unknowns potentially leading to failure conditions. Generic methodologies to cope with broad class of failures are needed. In this perspective, relations with reasoning on software component-based adaptation will be investigated.

The problem of workload distribution will be tackled under two angles: (i) modeling and simulation of distributed systems and (ii) integration of knowledge on the computing processes described. Probabilistic models of large scale distributed systems will be extended through the exploitation of exhaustive usage traces. This will be made possible by the emergence of the *Grid Observatory*, a grid telescope developed for gathering statistical data on grid usage. The extended trace sets acquired will be used to refine model (adapt to specific execution context, evolve through time...) and refine evaluation of such models. Joint work with the *SimGrid* simulator development team (INRIA Nancy) is also planed to validate models using realistic simulation scenarios. The models will be exploited primarily to observe the behavior and determine optimal usage strategies of the workload manager. To adapt to different kind of distributed infrastructures (homogeneous cluster, grids, clouds...) both strategies based on batch-oriented resources acquisition and dedicated resources reservation will be considered. The semantics of the domain-specific computing process will also be considered. Domain data analysis services will be semantically annotated and instrumented so that a workflow designer can exploit reasoners to (i) check the coherency of data analysis procedures developed and (ii) infer new information resulting of the computations achieved (such as provenance information or domain-specific information dictated by domain-dependent rules semantically described). Traditional Semantic Web methodologies will be adapted to the distributed environment targeted and performance issues arising with large-scale system will be addressed. The domain of neurosciences in which we have already invested significant effort to extract domain-specific information and develop appropriate data structures will be particularly considered.

3.3.2 Multi-Modeling Composition

Main contributors *M. Blay-Fornarino, P. Crescenzo, Ph. Lahire, S. Lavirotte, J. Montagnat, A. Occello, A-M. Pinna-Dery, P. Renevier-Gonin, J.-P. Rigault, J-Y. Tigli*

Multi-view modeling allows a developer to describe a software system from multiple points of view, e.g. *structural and behavioral, using different modeling notations* [22].

In the last few years, tremendous progress has been made in the field of software engineering. However the complexity of the systems has increased considerably [12]. To tackle this complexity, one approach consists in modeling the various aspects of a system from different points of views. This approach allows to level the boundaries between system developers and users. So every system element is viewed differently depending on the contexts where the project members moves. This development way increases productivity and quality by raising the overall level of abstraction. However, models may overlap and have to be integrated to produce a comprehensive and consistent system. So, it does introduce a new problem of coordinating multiple different meta-models in a single system.

In the pôle we are interested in a development led by the composition of concerns. Addressing this issue appears then as a logical continuation of our research. Composition deals with code, models and can be performed at compile time or runtime. Composition may impact the entire development process from modeling to deployment. The composition addresses also various types of models like structural or behavioral ones.

In the report on the achievements of the pôle we propose several approaches and we apply them [2], mainly to a single type of models. In the real world it is necessary to handle several types of models and then to perform Multi-Modeling composition. Our objective in the near future is to address Multi-Modeling composition coupling one approach with the context of one or two others.

In particular it could be interesting to introduce variability issues into the development of the composition of services, Data/control flows, Human Computer Interface or ambient computing. This means to make the coupling of the two approaches.

Another issue which focuses more on multi-modeling is to combine the approach of composition

2. i) Software Product Lines, ii) Composition of Orchestrations, iii) Data Flows and Control Flows, iv) Ambient Computing and v) Human Computer Interface.

of services addressing the model behavior (ADORE), with the composition approach related to SmartAdapters which focuses on structural model. Indeed recent experiments have shown that structural and behavioral designs can interact to complete and improve models [22].

Works mixing the composition of control and data flows have been initiated in [28]. Works on integration of non-functional properties in data flows transparently for the user are under developments [37]. Is it possible to apply and adapt these results to support flexible and reliable adaptations of large scale workflows? Can we expect a gain of expressiveness without losing effectiveness of treatments? These issues are going to be deeply addressed.

Finally, within the pôle GLC, research is conducted on goal-driven software development [27]. In a context of software factory, we are interested in applying this approach to composing orchestrations. A first study is underway as part of the application Seduite. In this work, therefore we apply the multi-modeling in an iterative development process at two stages of the life cycle.

3.3.3 Human Machine Interaction

Main contributors *A. Occello, A-M. Pinna-Dery, P. Renevier-Gonin, M. Riveill*

Our primary objective still remains to compose all the aspects of an application from the functional core to its presentation. We have already explored the case where the composition of functional core (FC) (services, components) impacts the User Interface (UI) elements and interactions. To complete the coverage of this objective, we explore the reverse case where the composition of interactions and UI impacts the composition of FC. The aim of this approach is to generate a new application by manipulating the former UI and their interactions. Composition mechanisms are based on ontologies to connect the different concerns: functional, interactions (tasks) and UI.

Another objective is to study and deal with the multiplicity of models used in the engineering of interactive systems. Such models are involved in processes such as composition or plasticity management at design time as well as runtime. As these models share information and evolve rapidly today, there is a need to maintain collaboration links between them. We plan to explore the relationships that exist between these models and make explicit such “collaboration” in order to manage model evolution and keep them mutually consistent.

Much work has already dealt with composition at the software engineering level in services and component architectures as well as at the User Interface (UI) level such as for Mashups. The two research communities have focused on their predilection domain and its respective level (software vs UI). However an interactive system is composed of building blocks both software and interactive. Hence, information about local compositions needs to be diffused in order to infer a sound global composition. We plan to study how to combine different composition approaches and how to deal with composition driven by different entry points simultaneously.

3.3.4 Semantic Web

Main contributors *M. Buffa, C. Faron-Zucker, N. Le Than, I. Mirbel, P. Sander*

Researches in the I3S GLC Semantic Web axis aim at offering a theoretical framework for graph-based knowledge management as well as models and tools to put in practice our knowledge management supporting means. In the continuation of our contributions to support Semantic Web application development, we will more specifically focus on (i) graph-based knowledge representation and reasoning models and (ii) Semantic Web user programming.

As a proof of concept on our theoretical proposals and still in continuation of our previous work, we will work on (i) supporting online communities, (ii) Modeling technical and regulatory knowledge, (iii) Knowledge-based adaptive learning systems and (iv) Knowledge modeling for web services discovery.

In the following we detail each of these perspectives.

Graph-based Knowledge Representation and Reasoning Models. In the continuation of our work on KRR, we intend to contribute in the conception and development of the KGRAM abstract machine and to concentrate on three main objectives:

- the introduction of usability features in the query language and their handling in the machine, especially to ask explanations and traces of reasoning; our aim is to develop an Eclipse plug-in for the SPARQL-like language of KGRAM which would integrate these features;
- the handling of entailment with graph rules; this requires to study the mapping of the query language with RIF expressions; our aim is to integrate in KGRAM a RIF parser and compiler
- the mash up of heterogeneous and distributed data and the expression of pipelines of queries; this should impact KGRAM's API and algorithms

Our work on constraint checking on regulatory knowledge in the building industry should provide application scenarios.

Semantic Web User Programming. The aim of this work will be to make easier the development of Semantic Web applications by providing a JavaEE like framework for developing Semantic Web applications. In this context, we will also work on proposing a semantic wiki that combines the advantages of semantic wikis and application wikis. Existing approaches in the field of semantic wikis have focused on two areas: use wikis to create small ontologies on one hand and use wikis to annotate documents, inserting metadata in pages (instances pages) on the other hand. *A*pplication Wikis let you "encode" small user applications or integrate web services interfaces directly in wiki pages. These small applications are then cloned by users. We propose to see the future semantic wikis as tools for creating services as well as tools for integrated services.

Supporting Online Communities. In the continuation of our current work on social tagging and social network analysis, the focus of our future work will be on sampling networks and reducing complexity of network processing. One of the biggest problems of current approaches is related to the gigantic size of real social networks, and to deal with their dynamic, ever changing structure. People join these networks, but often become inactive and do not represent a pertinent part of the network. Identifying precisely the characteristic of these networks, in order to apply some sampling theory is one of our key topic for the future.

So far the semantic models that exist for describing the relations between social network actors are closely related to the web 2.0 popular networks like facebook, etc. In our own work, in particular in the work related to the ANR project ISICIL, we focused on enterprise networks and the way we could improve them using a mixture of web 2.0 and semantic web approaches. However, we are currently conducting researches on modeling emotional states of actors in order to contribute to support systems for seniors. In this context, a mean to detect emotional states would be through the analysis of interactions in the social network of the senior (i.e. analyze chats, emails, comments, emotions, etc.). The idea is to add a layer to the social network models that would represent the affectiveness of the relationship, and help define the profile of the targeted person over time.

Modeling Technical and Regulatory Knowledge. In the continuation of our work on constraint checking on regulatory knowledge, we are now interested in the problem upstream, which is the composition of these technical and regulatory documents. Our aim is to assist domain expert in this process. This implies the continuation of our work on the modelisation of both the regulatory knowledge and the process for composing the documents capturing this knowledge, the starting point of this work being the capitalization of queries and rules and their scheduling. Through this particular application, we are also interested in the general problem of explaining the results of automatic inferences to the users — which are in our case experts in the building industry.

Knowledge-based Adaptive Learning Systems. In the continuation of our work on knowledge-based learning system, we envision to concentrate on the modeling of collaborative activities in such systems. This will be done within a 3 years collaboration project with the university of Annaba in Algeria, supported by the cooperation program between CNRS and DPGRF.

Knowledge modeling for web services discovery. In the continuation of our previous work applied

in the neurosciences domain, we will concentrate on knowledge modeling for web services discovery both at design time and runtime. With regards to design time, we will focus on proving means to select a subset of the web services available in the semantic repository by reasoning on the neuroscientists high-level image pipeline specifications. In other words, we will provide means to bridge the gap between end-users requirements and web service specifications. With regards to runtime, we will concentrate on inferring semantic knowledge at runtime in order to enrich the semantic repository.

We also plan to work on web service discovery in the ambient computing fields. In this context, we are particularly interested in providing (i) means to semantically annotate services for devices and (ii) semantic extensions of the aspect of assembly model in order to improve dynamic service composition in the framework of the WComp middleware.

3.3.5 Security

Main contributors *K. Boudaoud, M. Buffa, C. Faron-Zucker, A. Gaignard, J. Montagnat, M. Riveill, J.-Y.Tigli*

Keywords: user-centric security, security of mobile environments, cloud computing security, Web 2.0 security, secure mash-ups,

Concerning future work, we are continuing to evolve towards a user-centric security that is more and more user-centric (depending on the kind of user i.e. end user or developer) and green-oriented (i.e. energy saving). In this context, we started activities that are described below.

Low energy consuming and component-based security architectures for mobiles.

Keywords: Security components, security policies, adaptable secure protocols, web services

The aim of this work is to propose another utilization of the ADEPT framework proposed by Nicolas Nobelis in the context of his PhD thesis. More specifically to use and extend some elements of ADEPT for mobile environments. Nowadays, different types of electronic data transfer are supported by applications due to the expansion of Internet. These applications may run on desktop machines (PCs, servers, etc.) or mobiles devices (mobiles phones, PDAs, smartphones, etc.). A prerequisite and a critical issue for these transfers is the assurance of security functionalities. Each communication may necessitate a number of security properties such as confidentiality, integrity, authenticity, etc. While these security properties may be offered by secure protocols such as SSL, a constraint persists: SSL is high-energy consuming and the properties are provided as a block. Energy consumption has an important impact on the battery life of mobile devices and solutions are therefore needed. We aim at proposing a security management architecture for mobile devices based on adaptable security components and externalizing security functionalities from mobile devices to save their energy resources while responding also to economical and environmental constraints. Within our work, we will treat two critical issues: 1) how we can inform the user about the applied security properties and 2) what could be done to let other parties trust our architecture. The proposed architecture has been described [20].

Secure communicative context mobile mash-up (mosh-up).

Keywords: Web 2.0 security, mash-ups, mobile environments, security policies

We started a preliminary research work about Web 2.0 security and more specifically on security of mash-ups (a mash-up is an aggregation of widgets) that has been published in [3]. Now we focus on security of mobile mash-ups (mosh-up). The main objective of our work is to design a mosh-up grouping several widgets on a single interface and to allow them to communicate securely, through the phone API or to external resources, typically the web. The innovation aspect of this work concerns the design of a fully connected platform to enable a widget to talk with the phone, the other widgets and the web. The main challenge of this work is to: (i) make all these widgets communications secure to prevent any risks encountered in such platform; (ii) take into account the possible lake of resources on mobile equipments and possible evolution and adaptation of the proposed platform to face market reality; (iii) be user-oriented (i.e. non security expert end-user and developer).

A new security model for cloud computing security.

Keywords: cloud computing, Database security, e-commerce, security properties, web services

This work started with the thesis of Christian Delettre in May 2010. The aim of this thesis is to propose a new security model for Cloud Computing and more specifically for e-commerce environments deployed in clouds platforms. The work of N. Nobelis on security components will be reused in particular to secure communications in e-commerce clouds according to the type of data exchanged. Furthermore, within an e-commerce environment many providers have the same structures of database easily available on Internet. Therefore, it is possible to realize statistics to obtain rough information even on encrypted data. To mitigate the problems of this practice, we shall develop two additional security components that will be added to those of N. Nobelis.

Impact of security on human user interfaces.

Keywords: human user interface, security composition

Starting from the security components and security properties-based security policies proposed by Nicolas Nobelis, the goal of this work is to study the impact of security at the user interface (UI) level, from three different points of views:

- study the composition of UIs security components starting from their assembly specified through a security properties-based policy;
- study the composition of UIs security properties-based security policies;
- study the effects of applying security on a UI, by allowing a non-expert end user specifies her/his security wishes to add security properties through a user interface.

Adding security and privacy to a non-secure application.

Keywords:privacy, security

The goal of this work is to facilitate the integration of security (using and extending the security components proposed by Nicolas Nobelis) and privacy into an application even after the development phase by proposing a dynamic approach for code compilation and orchestration. To achieve this goal, we plan to design a component-based and user-centric (i.e. end-user and developer) architecture that will allow the: (i) designer to add security and privacy components in an application and (ii) user to tune and configure the privacy or security she/he needs for the application she/he uses.

Security evolution from a user point of view.

Keywords:application security, security policies

When developing in a "spirale" manner, a designer can realize that she/he needs to secure her/his processes and data. Thus, the question is how to build an application by adding/removing security policies ? How to facilitate the construction of more secure applications in a flexible and “transparent” manner for the designer and end user?

Managing security at data flow level.

Keywords: authentication, autorisation

The aim of this work is to allow a control of securisation processes and flow analysis. To do this, we propose to use two approaches: (i) an improvement of the data flow allowing a composition of non-functional properties at workflow level; (ii) a simulation of the securisation processes managed effectively at the platform level to validate what is possible before the deployment.

Social networks and trust for security of collaborative web sites.

Keywords: social networks, trust, access rights

The aim of this work is to continue the work done to secure collaborative web sites from an ontology point of view by taking into account social networks and trust management notion for managing access rights.

Security visualisation.

Keywords: visualisation, administrator-oriented

The goal of this work is to study and evaluate existing visualisation approaches used in the security area by comparing them to visualisation approaches used in other domains and taking into account the administrator point of view. Thus, again our study will user-oriented in order to propose an improved approach of existing security visualisation tools or a new one.

3.3.6 Data Mining

Main contributors *C. Dartigues, D. Pallez, N. Pasquier, C. Pasquier, C. Pereira da Costa*

During the coming years, we plan to continue the activity conducted on the development of semantics, techniques and algorithms for the analysis of very large sets of heterogeneous structured and unstructured data. More precisely, our work will concern the development of Algorithms for Data Mining, Bioinformatics and Evolutionary Computation. We also plan to address the domain of User Interfaces & Visualisation that is a major domain for data mining technique applications for decision support.

The first topic we propose to address is the development of algorithms, both heuristics and meta-heuristics, for knowledge patterns extraction. In the context of heuristics and deterministic algorithms, we plan to extend the use of the concept of lattice framework, used to extract condensed representations and minimal covers of association rules, to the construction of classification models and hierarchical conceptual clustering approaches. In the context of meta-heuristics and evolutionary computation, we plan to extend approaches developed in the workgroup to the optimization of the association rule extraction using multi-objective optimization and Interactive Evolutionary Computations. We also plan to continue our study on the use of data mining techniques in order to design optimized stochastic algorithms. The PhD thesis of M. Kartick Chandra Mondal, started in 2009 in the framework of the EMMA (Erasmus Mundus Mobility for Asia) program, will be performed in the context of this activity.

We also plan to continue working on semantical aspects of data mining and expert background knowledge integration in the data mining process. Works conducted on this topics concern the development of semantical theoretical frameworks and algorithms for the integration of domain knowledge and the visualisation of extracted patterns. The objective is to improve the relevance of extracted patterns and the efficiency of the extraction by integrating in the process as much known information on the subject of the application as possible. The activity conducted on this topic, referred to as Domain Driven Data Mining in the literature, can be divided into two categories according to the type of knowledge integrated: formalized or non-formalized. Formalized knowledge is stored in heterogeneous data structures (graphs, ontologies, taxonomies, databases and information systems), knowledge bases, bibliographic bases, annotated corpus and structured or semi-structured textual files. Non-formalized knowledge, that corresponds to methodologies, how-to and procedures for example, cannot be easily described and stored in knowledge bases, due to the complexity of its formalization and the inadequacy of actual techniques to represent it. If the main problems for integrating the first are algorithmic, integrating the second requires specific techniques that are the most often interactive, visual and user-centered. These research topics are referred to as Interactive Data Mining, Visual Data Mining et User-Centered Data Mining in the literature.

In complement to the previous work, we plan to address the problem of user interfaces for knowledge patterns visualisation and management. This topic concerns the development of approaches and algorithms for the management and exploitation of extracted knowledge patterns (filtering, selection, generalization or specialization, ordering and ranking, etc.) and their visualisation by the expert of the domain. These approaches will integrate available formalized background knowledge on the domain in order to optimize and simplify as much as possible the usage of extracted patterns. We are particularly interested in the development of optimization algorithms involving one or more users during the post-processing phasis of the data mining process to interactively identify the most relevant patterns from the users' viewpoint, that is those corresponding the most to their wishes without having to formalize them.

This activity will be conducted in the context of academic and industrial collaborations on projects in the domains of bioinformatics and seed industry. In bioinformatics, recent progress in biomedical research has led to an explosive growth of biomedical data. In addition to these experimental data, researchers in life sciences have access to numerous databases dedicated to specific aspects of biological knowledge. However, the efficient use of these data is, more than ever, an important issue. One needs, indeed, to combine data from various sources and to mine them in order to extract relevant information. However, because of some specificities of biological data (uncertainty, incompleteness, high heterogeneity), the development of new algorithms or at least an adaptation of existing algorithms is often necessary. Since 2004, research aiming to the successful mining of biological data are conducted in collaboration with the Institute of Developmental Biology and Cancer (UMR6543 UNS-CNRS) and the Institute of Molecular & Cellular Pharmacology (UMR6097 UNS-CNRS). This research led to several national and international publications in bioinformatics. We plan to continue this activity and to extend our research to textual data mining, to develop new algorithms to convert information expressed in natural language into structured data. Our aim is to integrate these new data in data mining processes in order to infer new knowledge.

We also plan to continue the LABS project started in 2010 in the context of an industrial collaboration with Doriane and ActiveEon, to be submitted to the January 2011 OSEO/PACA APRF call for projects. This project concerns the domain of the seed industry, in which developing new varieties of plants is a major issue. The essence of plant breeding is the selection from within a genetic variation of plants with desirable traits that can be inherited in a stable fashion. The creation of a new variety involves plants reproduction on 4 to 6 successive generations and usually takes more than 10 years. During the selection process several important choices must be made and wrong choices at critical stages can generate significant costs. Being given that each choice is based on the analysis of a large amount of data (genotypic and phenotypic data, environmental data), the use of data mining methods seems appropriate. However, while many data mining methods have been developed in recent years, choosing which method to use in a given context remains an open issue. The objective of this project is to define an intelligent data mining system in which the choice of the algorithm to use and its settings would be based on input data and expected results.