

Research and Practice in Federated Information Systems

Report of the EFIS '2000 International Workshop

W. Hasselbring* W.-J. van den Heuvel[†] G.J. Houben[‡] R.-D. Kutsche[§] B. Rieger[¶]
M. Roantree^{||} K. Subieta^{**}

Abstract

EFIS 2000 was held at Dublin City University in June 2000. The principal aim of this third workshop was to bring together new insights from academic research with industry-driven developments and perspectives in the area of federated information systems. This report describes the observations of the workshop together with the outcome and future research possibilities.

1 Introduction

The Engineering Federated Information Systems (EFIS) series of workshops was started in 1997 and has now completed its third meeting. Reports of previous meetings are available in previous issues of this journal [CEH⁺97, CHH⁺99]. Its main goal is to bring together active researchers and industrial users of federated information systems, with submissions carefully reviewed by an international program committee. The engineering of FIS is a research area which addresses the systematic development of interoperable solutions for autonomous, heterogeneous systems covering both database sources and also non-database information sources, providing (structured or semi-structured) files, multimedia data, or other proprietary systems' data. One of the main foci is the semantic integration of heterogeneous information sources, applied in various application domains.

*University of Oldenburg, Dept. Computer Science, Software Engineering Group, Germany, hasselbring@acm.org.

[†]Tilburg University, Dept. Information Mangement, Infolab, Netherlands, W.J.A.M.vdnHeuvel@kub.nl.

[‡]Eindhoven University of Technology, Dept. Computing Science, Netherlands, G.J.Houben@tue.nl.

[§]Technical University Berlin, Dept. Computer Science, Germany, rkutsche@cs.tu-berlin.de.

[¶]University of Osnabrück, Informations Systems Group 2, Germany, brieger@uos.de.

^{||}Dublin City University, School of Computer Applications, Ireland, mark.roantree@compapp.dcu.ie.

^{**}Polish-Japanese Institute of IT & Institute of Computer Science PAS, Warsaw, Poland, subieta@wars.ipipan.waw.pl.

2 Workshop Contents

The first invited speaker, Kazimierz Subieta, spoke about mapping among heterogeneous ontologies with a special emphasis on object views. This was followed by three academic sessions under the themes of *Architectural and Design Issues*, *Federated Databases*, and *Files and User Interfaces*. In the session on architectural and design issues, the session papers were based on the topics of mixed top-down/bottom-up strategies for the conceptual design of FIS, and an architecture for secure data warehousing. In the federated databases session, the papers were based on adding conflict resolution features to integration languages for query integration, citation linking in digital libraries, and multimedia federations. Finally, the files and user interfaces session featured papers on plugging files into database federations, the automated generation of hypermedia front-ends, and the semantic integration of heterogeneous information sources.

The second day of the workshop contained two industrial sessions and the workshop review. The second invited speaker, Sean Baker, discussed the different forms of middleware which may be employed in the federation of software systems. The four industry papers discussed the use of IONA's iPortal Suite, a modular approach for federations, CASE support for federated system integration, and the use of meta-directories in building federations.

Details of the workshop review are presented in the next section, and the proceedings are available as [RHC00].

3 Workshop Observations

As is the tradition with EFIS, the final session is a workshop review. The following observations are not suggested to be a definitive comment on FIS, but reflect the issues highlighted by the workshop attendees.

- Research into formal ontologies for FIS was seen as problematic. In practice, it often turns out that formal semantics can only take place at a

programming level, with an informal approach used at conceptual levels. It was generally agreed that formal semantics typically is rather difficult to apply and often infeasible for larger systems. Thus, an immediate need for informal (or at least less formal) approaches exists. A major problem for formalization on the higher conceptual levels is to find the right abstractions that are formalizable, and at the same time, useful.

Especially a graceful translation between the higher, conceptual and the programming level is regarded as extremely important and (unfortunately) at the same time difficult. Complexity grows exponentially when one tries to map semantic descriptions at the conceptual level (often loosely defined in terms of data dictionary) to the level of programming code.

If the UML [UML] can be regarded as a semi-formal approach to modeling FIS, then this may also be used at higher levels. The reason for this viewpoint is that many FIS specification approaches are too narrow, focusing only on issues such as schema mapping. Instead, the information required to construct adequate federations of IS is often much larger.

- Object-oriented view mechanisms are often supposed to be used as the platform for creating federated schemata. However, most of these systems were considered to be too simplistic and should take more from programming languages, facilitating more increased algorithmic power. Further observations can be found in [Sub00].
- Commercial middleware is increasingly more available, and is being deployed by more and more research institutions and enterprises for tasks such as building Internet portals. Currently, there are some efforts to merge the traditional message-oriented middleware (such as MQseries) with modern object-oriented middleware (such as CORBA). However, it was observed that a gap exists between the functionality offered by these middleware solutions, and the functionality needed at the application level.
- The trend towards the use of XML and XML/EDI as a standard for the exchange of information was again highlighted. This standard helps the process of interchange of data, but it is only part of the solution. The importance of persistence and any possible role for XML should be emphasized. The often unspecified role assigned to XML (such as ‘better HTML,’ ‘standard interchange format,’ and ‘database model’) sometimes cause doubts if this technology is enough as a basis for integrating heterogeneous databases into a federated system.

With the availability of XML we are not finished, we need more in order to have a complete basis.

- Meta-directories are federated information systems also. The existence of adequate standards such as LDAP greatly alleviates interoperability in that domain.
- Understanding a FIS as a typical kind of *evolutionary* system, i.e., a system in which both existing components (legacy systems) and newly constructed ones continuously change over time, was addressed as a major topic of interest (see also Section 4).
- Federated security issues have been covered for the first time at an EFIS workshop. Within this topic, it is important to identify issues of privacy, safety, security and transactions.
- Increasingly we see different types of relationships between the data sources in the FIS, and this should be reflected in the way in which we design the system. In the context of Data Warehousing the growing need for the integration of qualitative and quantitative data can be seen as an issue for FIS. Multiple warehouses and Web services arise to support different types of information and levels of distributed decision making. The integration and management of metadata is emerging, not only for warehouse maintenance but also for end-user access and retrieval.
- FIS-concepts and methodologies seem to be well-suited for key issues of knowledge management. Two topics were addressed:
 1. integrating new kinds of information systems like knowledge based ones into FIS and
 2. extending (more static and retrieval-oriented) FIS by dynamic aspects of knowledge management (organizational learning, evolution of knowledge bases).
- In FIS, changes to application areas trigger changes to managed data objects. New issues of federated multimedia and hypermedia objects were covered for the first time.
- A short terminology discussion on the differences between FIS and mediators led to no final agreement, and ended quickly. Still a lot of terminology discussion is going on (in particular in this area). As typical in emerging disciplines, a more standardized use of terminology could better facilitate the exchange of research results.

4 Future Research Issues

A number of topics were highlighted as areas where future research would be welcomed.

- **Evolution in FIS.** This should address the dynamics of FIS during usage over time. Issues such as: how can FIS-methods (such as semantic conflict resolution and linking) be applied to the contents' maintenance on different FIS-levels (FIS-bases, FIS-cooperation), how is it possible to extract new information from FIS-usage, and how and can we store this information? There is some overlap between this topic, and knowledge management as discussed in the previous section and below.
- **Object views as mediators.** The complexity of ontologies requires object views (mediators) having full algorithmic and data processing power, stored in a database as named, protected, and removable units (similarly to SQL views), and defined through a high-level query/programming language as complex modules with own state, functions, and classes. Important aspects concern access to metadata, access to various resources outside a database and good performance.
- **Integration of Behavior.** This has been a topic for which a generalized solution has evaded researchers for a long time. Integrating components in FIS architectures means not only defining a mapping between their data structures and methods, but also their dynamic behavior. The success of new component technologies within industrial settings is dependent on usable solutions for this problem.
- **Knowledge Management.** There is a strong requirement for a separation of concerns. Data with knowledge and functional derivation on one hand, and the technical platform to distribute this data on the other hand. It should be possible to exploit concepts and methods developed for FIS to support Knowledge Management in practice (by semantically linking various information sources). How can concepts and methods of Knowledge Management be adopted and applied to extend or improve the scope of FIS?

Realizing a FIS requires a knowledge from different sources, including schemata and other documentation of local databases, but also including tacit, undocumented knowledge of people who know some important peculiarities. Thus knowledge management means some methodology of recording information about people's knowledge and some methodology which will allow us to do that step-by-step.
- **Quality of Service.** So far, mainly functional concerns of FIS were addressed. Solutions for non-

functional concerns such as fault-tolerance, security, performance, etc. are essential for successful FIS.

- **Integration of Business Rules.** The distribution (or movement) of the business process must take place in an elegant way. After transformation of data into a federated context, it is then necessary to decompose the process that was associated with the data. This type of problem may require a process-oriented design method solution.

Inter-organizational processes need to cope with (conflicting) business policies of the interacting companies. Research into the area of modeling business policies, attaching them to (virtual) enterprise goals and business processes, describing their semantics, and negotiating about them is major issue.

5 Conclusions

All of the participants agreed that the workshop was a very interesting and fruitful forum for the discussion on current research and practice in engineering federated architectures and information systems, and in highlighting industrial efforts in the same area. Plans for next year's workshop have already begun and Berlin has been agreed as a venue for the EFIS 2001 event. The fourth workshop should attempt to attract papers which target the issues raised at the EFIS 2000 workshop. Further information on the EFIS 2001 event will be available on the EFIS '2000 home page: www.compapp.dcu.ie/~efis2000

Acknowledgements: The organizers of EFIS 2000 wish to extend thanks to the two major sponsors: Materna Systems (Germany) and Orbism (Ireland).

References

- [CEH⁺97] S. Conrad, B. Eaglestone, W. Hasselbring, M. Roantree, F. Saltor, M. Schönhoff, M. Strässler, and M. Vermeer. Research Issues in Federated Database Systems (Report of EFDBS '97 Workshop). *SIGMOD Record*, 26(4):54–56, December 1997.
- [CHH⁺99] S. Conrad, W. Hasselbring, U. Hohenstein, R. Kutsche, M. Roantree, G. Saale, and F. Saltor. Engineering Federated Information Systems: Report of EFIS '99 Workshop. *SIGMOD Record*, 28(3):9–11, September 1999.
- [RHC00] M. Roantree, W. Hasselbring, and S. Conrad, editors. *Engineering Federated Information Systems (Proc. EFIS'2000)*. IOS Press (ISBN 1-58603-075-2), 2000.
- [Sub00] K. Subieta. Mapping heterogenous ontologies through object views. In Roantree et al. [RHC00], pages 1–10.
- [UML] *The Unified Modeling Language*. www.rational.com/uml.