

# Report on the Engineering Federated Information Systems 2003 Workshop (EFIS 2003)

C.M. Wyss  
Indiana University  
e-mail: crood@indiana.edu

W. Hasselbring  
Universität Oldenburg  
e-mail: hasselbring@informatik.uni-oldenburg.de

H. Höpfner  
Universität Magdeburg  
e-mail: hoepfner@iti.cs.uni-magdeburg.de

A. James  
Coventry University  
e-mail: a.james@coventry.ac.uk

S. Conrad  
Universität Düsseldorf  
e-mail: conrad@cs.uni-duesseldorf.de

November 7, 2003

## Abstract

*This paper summarizes the EFIS 2003 workshop, held in Coventry, U.K. in July, as part of Coventry University's Data Horizons Week. Major research issues discussed include metadata/ontologies, integration frameworks, data quality and evolution, and mobile interfaces. Topics for future work include evolution, expressiveness, maintenance, and dissemination of FIS.*

## Overview

The Engineering Federated Information Systems (EFIS) workshops bring together researchers in databases and software engineering to address issues in designing and building interoperation solutions for heterogeneous information systems. In July 2003, the fifth EFIS workshop was held in Coventry, United Kingdom, bringing together a diverse group of multidisciplinary and multinational researchers. The broad focus of the workshop was on metadata, ontologies, and standards for facilitating information exchange and interoperation. Several application domains were addressed, as well as specific architectures and related projects. Researchers from both academia and industry contributed.

EFIS 2003 was held as part of Coventry University's *Data Horizons Week*, which also included the twentieth British National Conference on Databases and the LTCS-ICS Teaching, Learning, and Assessment in Databases workshop.

Previous EFIS workshops have been held in Barcelona (1997), Kühlungsborn (1999), Dublin (2000), and Berlin (2001). Reports on these workshops are available in *ACM SIGMOD Record*, **26**, December 1997; *ACM SIGMOD Record*, **28**, September 1999; *ACM SIGMOD Record*, **29**, December 2000; and *The Computer Journal*, **45:2**, March 2002. The 2003 EFIS homepage is

<http://www.mis.coventry.ac.uk/EFIS2003/>.

The proceedings are available from IOS Press [1].

## Invited Presentations

In keeping with the tradition of bringing together researchers from both industry and academia, keynote addresses were presented by:

- Keith Jeffery, CLRC-Rutherford Appleton Laboratory, Oxfordshire: GRIDs and Ambient Computing: The Next Generation, and
- Peter Buneman, University of Edinburgh: What is Interesting about Scientific Databases?

The first keynote presentation discussed a previously proposed three-layer, agent-based GRID architecture relying on standardized metadata. GRID architectures propose to revolutionize the accuracy and efficiency of global data exchange. Major research issues identified in Dr. Jeffery's talk include

- **Metadata:** Metadata is required to represent the users, sources, and resources comprising the GRID. Expressive, formal, and machine-understandable metadata representation languages are required to underpin GRID advances.
- **Agents and Brokers:** Formal communication standards (as well as description languages) need to be produced that allow the generality and dynamicity required to interact within the GRID as well as maintain performance, reliability, and security requirements.
- **Security, Privacy, and Trust:** GRID computing requires varying levels of security protocols (including encryption and authentication) as well as availability and continuity guarantees. It is recognized that providing secure and effective GRID services necessitates representation of legally binding business documents in a virtual setting, an area of active study. Furthermore, data protection must be measured against system intelligence.

The second keynote presentation focussed on annotation, provenance, and archiving of data, especially with regard to

scientific data. The need for new data models and query languages to address these requirements was identified. Research issues in this area center around perennial problems in identifying data origination and evolution, yet these problems have grown in importance with the emerging, vast corpora of electronic data. Data quality depends on an understanding both of the reliability of the source, as well as knowledge of the *phylogeny* of participating databases. Other terms such as “data pedigree”, “data parentage”, and “data lineage” describe related issues. The central idea is that data should carry with it some notion of its origination and historical modification. A novel model for recording and querying data history that is based on semi-structured data was presented during the keynote.

## Technical Papers

Twelve technical papers were selected for presentation. These are:

1. *E. Altareva, S. Conrad (Düsseldorf, Germany):* Analyzing Uncertainties in the Database Integration Process by Means of Latent Class Analysis.
2. *D. Che, W-C. Hou (Carbondale, IL/USA):* Storing and Querying Semistructured Data using Relational Databases.
3. *H. Höpfner, K-U Sattler (Magdeburg, Germany):* Semantic Replication in Mobile Federated Information Systems.
4. *R. Iqbal, A. James, R. Gatward (Coventry, UK):* A Collaborative Platform for Interoperability of CSCW Applications.
5. *H. Kozankiewicz, J. Leszczyłowski, K. Subieta (Warsaw, Poland):* Implementing Mediators through Virtual Updateable Views.
6. *V. Morocho, F. Saltor, L. Perez-Vidal (Barcelona/Catalunya, Spain):* Schema Integration on Federated Spatial DB across Ontologies.
7. *S. Pedersen, W. Hasselbring (Oldenburg, Germany):* Structuring and Combining domain-specific Standards for Interoperability in Health Care.
8. *N. Rossiter, M. Heather (Northumbria, UK):* Four-Level-Architecture for Closure in Interoperability.
9. *W. Sigel, S. Busse, R.-D. Kutsche, M. Klettke (Berlin/Rostock, Germany):* Designing a Metadata-Based Infrastructure for E-Learning.
10. *M.-N. Terrasse, M. Savonnet, E. Leclercq, G. Becker (Dijon, France):* Building Platforms for Information System Interoperability: a UML-based Metamodeling Approach.
11. *L. Wang, M. Roantree (Dublin, Ireland):* Designing Roles for Object-Relational Databases.
12. *C.M. Wyss, J.J. Lu, S.Y. Cheung, M.M. Akhavein (Indiana/Atlanta, USA):* Intrinsic Support for Metadata in Relational Federations.

## Main Topics

The main topics addressed in EFIS 2003 largely fell into four broad categories.

### Metadata and Ontologies

As with previous workshops, metadata played a prominent role in the discussions. Several contributions focussed on the role of *standardization* in creating effective FIS systems based on common terminology and metadata. Specific standards in the Geographical Information Systems, Health-Care, and E-Learning domains were presented, with a focus on integrating differing (yet functionally similar) standards for these application areas. Also presented were general extensions to SQL99 for querying and restructuring meta-level information in database systems.

### Integrating Architectures and Frameworks for FIS

The dominant architecture for FIS has become mediator-based, where middleware solutions are responsible for providing unified access to constituent information sources.

Within this context, several architectures for FIS were presented, with the focus being on architectures and frameworks capable of integrating existing models and frameworks appropriate to a domain. Most generally, a four-level closure architecture was suggested that includes formal descriptions of data, schema, constructs (meta-schema), and concepts (meta-constructs). General categorical transformations were studied within this architecture and it was shown that all relevant transformations can indeed be captured within these four levels; this may entail that further meta-levels are unnecessary (in general) in FIS architectures.

Two specific integrating frameworks were presented. The first supports interoperability in Computer Supported Collaborative Work (CSCW) and comprises an integrating framework for generic models of CSCW systems. The characteristic that distinguishes CSCW systems from other sorts of information systems is the emphasis on coordination and user interaction. The proposed framework is thus composed of three models: an ontological model, a coordination model, and a user interface model.

The second framework involved UML based *metamodeling* where data instances and corresponding models are abstracted into “metamodels” which can then be operated on uniformly to interoperate among these metamodels. This approach also has applications for the semantic web and

service-based architectures where models specify the functionality of constituent toolboxes.

## Data Quality, Views and Evolution

Finally, an emerging group of themes centered around data presentation and quality issues, including view-based data subsets, data provenance and origination, and temporal evolution of data and schemas. Relevant work included a discussion of uncertainty accumulation in FIS integration, and approaches for reversing this accumulation using latent statistical analyses. Under this theme, work was also presented on *semantic replication* for propagating relevant views and updates in ambient, resource-restricted networks. Further work included a framework for defining mediators as virtual, updatable views as well as a framework for role-based views for the Object-Relational model.

## Mobile Information Systems

EFIS 2003 also addressed new research resulting from mobile devices that provide access to FIS. The characteristics of mobile clients differ from static clients. They are typically smaller, more lightweight and tend to work offline. This necessitates additional functions that have to be provided by the FIS, such as (1) restricting views to subsets of the underlying data, (2) contextual yet computationally restricted queries, (3) scalability, and (4) offline-updates (especially with respect to consistency conflicts).

Some of these problems have already been addressed, but typically papers about mobile information system do not consider the characteristics of the underlying FIS. Thus, it is still a challenge to augment existing federated information systems in order to solve problems specific to mobile information systems.

## Identified Topics for Future Work

The main, recurrent topic for future work was wrapped up with temporal evolution of data, schemas, and queries. Along the same lines, the need to accommodate evolving standards for metadata and ontologies was identified as a crucial feature for metadata/ontology frameworks. In general (along these lines), research addressing the systematic and ongoing maintenance of existing FIS solutions is required.

Other areas for future investigation include more expressive schema and mediator definition/manipulation languages as well as the need to address a broad range of contemporary interfaces to FIS, including resource GRIDs and mobile networks.

## Outlook

The EFIS workshops are designed to bring together researchers from both academia and industry in the general

areas of software engineering and databases to address problems in developing interoperability solutions for information systems. In keeping with this goal, the sixth EFIS workshop may be associated with the 2004 International Conference on Software Engineering to be held in Edinburgh in May.

## References

- [1] A. James and S. Conrad and W. Hasselbring (eds.). *Engineering Federated Information Systems: Proceedings of the 5th Workshop EFIS 2003*. July 17-18, 2003 (Coventry, UK). IOS Press, Amsterdam.