

Webdam – Kick-off meeting

Serge Abiteboul, INRIA Saclay

ENS Cachan, 01/07/2009

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

EXINRIA Centre de recherche SACLAY - ÎLE-DE-FRANCE

Webdam

Topic:	Foundations of Web Data Management
	European Research Council Advanced Grant
Founding:	2.4 millions Euros
Timing:	2009 – 2013
Institution:	INRIA Saclay – Île-de-France
Hosts:	Univ. Paris Sud (Gemo) and ENS Cachan (Dahu)

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE



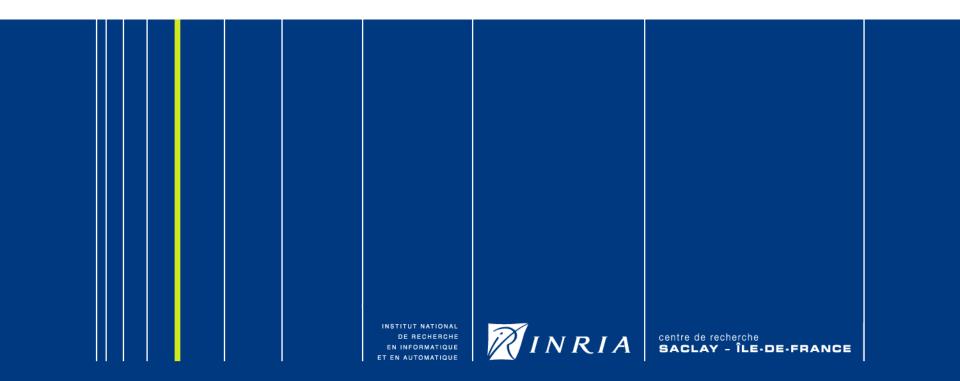
Organization

- **Technical content**
- The people
- Meetings
- Editing
- Website
- Discussion

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE T EN AUTOMATIQUE



Technical content



Webdam thesis in brief

Management of local data

- Well-accepted model: relational model
- Well-accepted theory: FO & concurrency control & etc.
- Development of systems and theory in parallel
- Management of distributed/Web data not the case
 - Rapid software development & no well-accepted model/theory

Illustration

- Relational database course: clean and neat course
- Distributed databases: a series of recipes

The Web is too complex to continue hacking systems with no unifying underlying model

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE



A first shift: distribution in autonomous systems

Web/distributed/peer-to-peer data management

Focus is on *information* residing on (possibly very many) *autonomous systems*

Information

- Structured data (relations)
- Semistructured (XML, graphs)
- Unstructured (text)
- Metadata
- Knowledge
- Physical data: indices, views, cache...

Focus is more on data management than on semantic Web but border is unclear

INSTITUT NATIO DE RECHERG EN INFORMATIO ET EN AUTOMATIO



A model for Web data management

What should be expected for a such a model?

- Describe query processing, changes, monitoring, communication, service choreography, data integration/exchange, distributed applications
- Road map

What are the issues ?

First attempt – ActiveXML: XML with embedded service calls

- With "local" logic (tree pattern queries): a calculus for Web data
- An algebra
- Optimization
- Verification

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE T EN AUTOMATIQUE



A second shift: deduction

Relational system: FO on everyone's desk

Hardwired logic

- Concurrency control with 2PL
- Integrity constraint enforcement
- Access control enforcement

On the Web or in distributed systems: this is not available

Verification of distributed applications

- Work on business processes
- Verification of ActiveXML applications

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE



What is the target?

Improve understanding of Web applications

Facilitate teaching Web data management technology

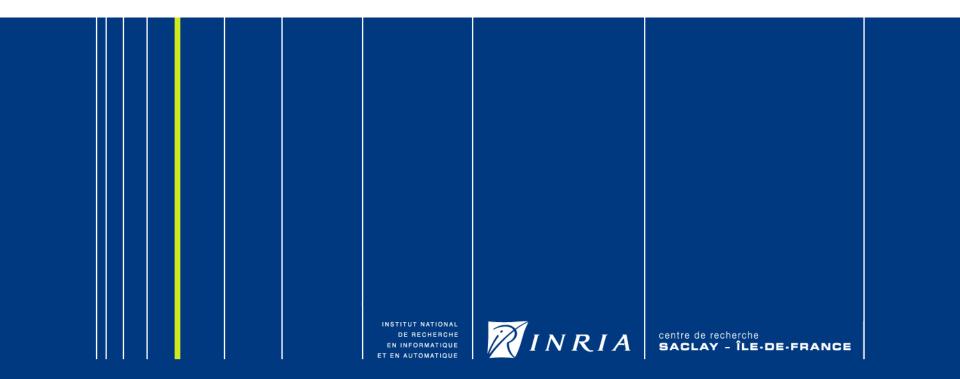
Improve Web applications: better performance, more reliable, better access control, etc.

Facilitate development of Web applications – programmers productivity

INSTITUT NATION DE RECHERCI EN INFORMATIQU ET EN AUTOMATIQU



The people



Webdam members

Already in: Abiteboul, Segoufin, Vianu, Senellart, Galland, Marinoiu, Bourhis, Kharlamov

Potentially: The members of Gemo and Dahu who are interested

Soon:

- 1. Balder ten Cate starting April 2009 for 6 months or more
- 2. Philippe Rigaux seconding "détachement" september 2009
- **3.** Marie-Christine Rousset $-\frac{1}{2}$ seconding "délégation" september 2009

Short visits: Amélie Marian, Tova Milo, Alkis Polizotis Other ideas?

> INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE



Researchers who are being considered

discussion

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE T EN AUTOMATIQUE



Hiring season

Advertise there are positions

Do it primarily by contacting friends/good groups

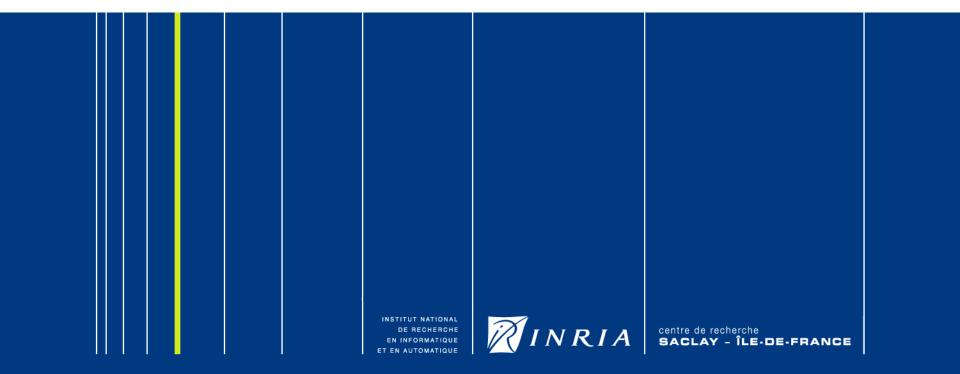
Priority for Webdam:

researchers with some experience (more than PhD or Posdoc)

INSTITUT NATIONA DE RECHERCH EN INFORMATIQU ET EN AUTOMATIQU



Events



Workshops with other projects - candidates

Local

- STREP Fox (Luc Segoufin): Foundations of XML
- ANR DataRing (Patrick Valduriez): Distributed data management
- ANR Docflow (Anca Muscholl): Verification of Active XML

ERC

- Stefano Ceri: New paradigms for Web search
- Anne-Marie Kermarrec: Web search in P2P
- Carlo Ghezzi: Self-Managing Situated Computing

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE



Workshops with other projects

Como Lake

June 19-20, 2009

Joint with ERCs of Stefano Ceri and Carlo Ghezzi

Workshop in corsica or lumini: joint with Fox in 2010?

INSTITUT NATIONA DE RECHERCH EN INFORMATIQU ET EN AUTOMATIQU



Prospective meeting

Core: Advisory board of Webdam + Webdam + guests next to VLDB - end August 2009

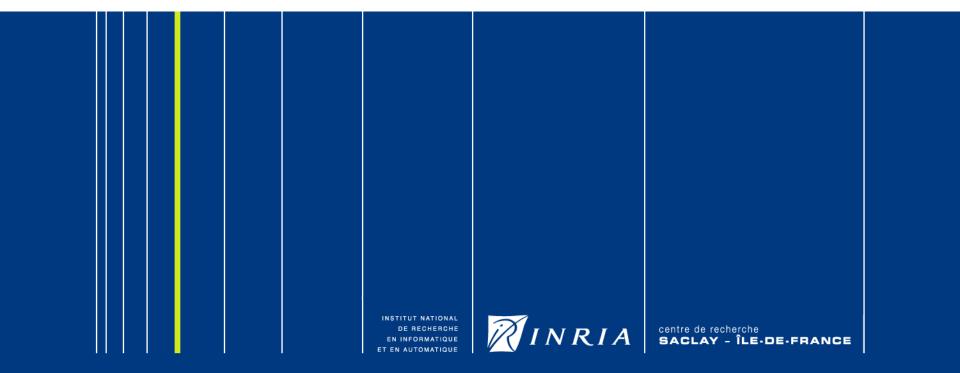
- In paris
- Deliverable:
 - Manifesto on Theory of Web data management
 - Roadmap

Other ideas?

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE T EN AUTOMATIQUE



Editing



Web Distributed Management of Data

Text book

Used in courses at Orsay, Dauphine & Telecom ParisTech

- Pushed and completed
- Rigaux, Abiteboul, Rousset, Senellart





Second Foundations

Continuation of Foundations of databases AHV

Foundations of Web data management

Web publication but high editorial standard

Topics: XML, tree automata, data integration/exchange, query processing...

Lead? Participants?

INSTITUT NATIONA DE RECHERCH EN INFORMATIQU ET EN AUTOMATIQU



Other ideas

Tutorial and text on Active XML

• To be proposed to PODS or ICDT?

Book – long paper on ActiveXML

Manifesto in favor of theory of Web data management

• With Fox?

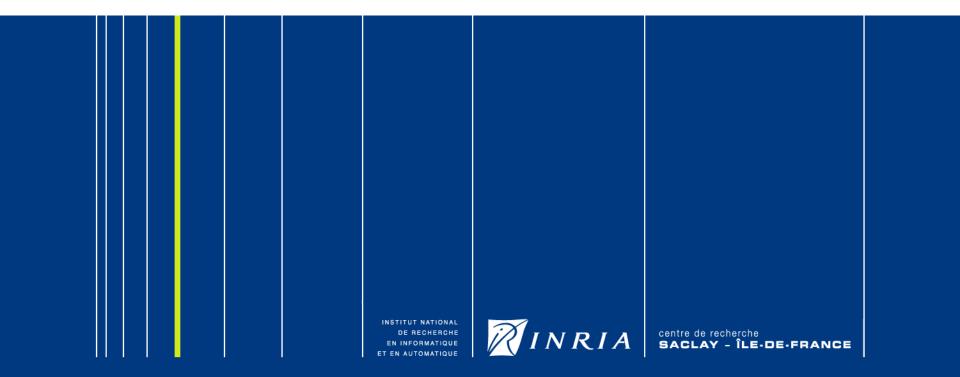
What else?

INSTITUT NATIONA DE RECHERCH EN INFORMATIQU ET EN AUTOMATIQU



Website : <u>http://webdam.inria.fr</u>

to do: intranet & mailing list



Merci

Discussion

