

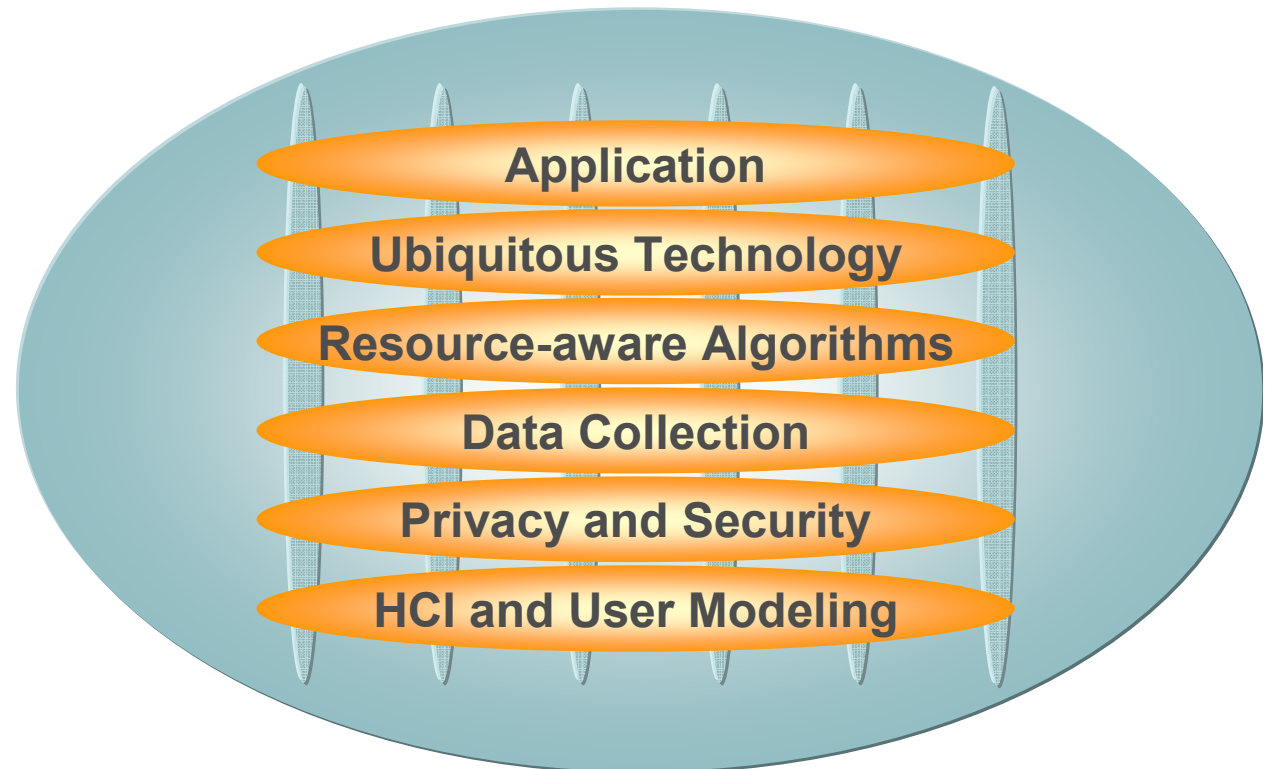
Research Challenges in Ubiquitous Knowledge Discovery



KDubiQ network

Michael May
Fraunhofer Institute for
Intelligent Analysis and
Information Systems

NGDM'07, 12.10.07
Baltimore



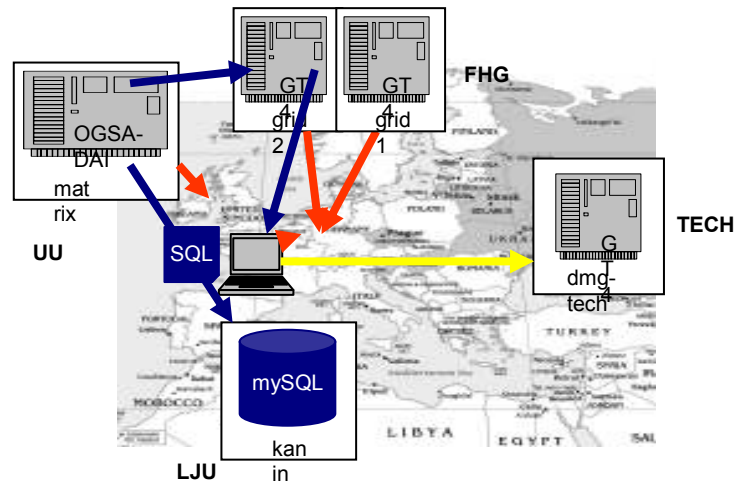
Next Generation Data Mining –

a collection of hot topics or an integrated perspective?

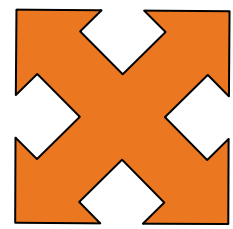


How can we get people to collaborate with each other that work in different fields?

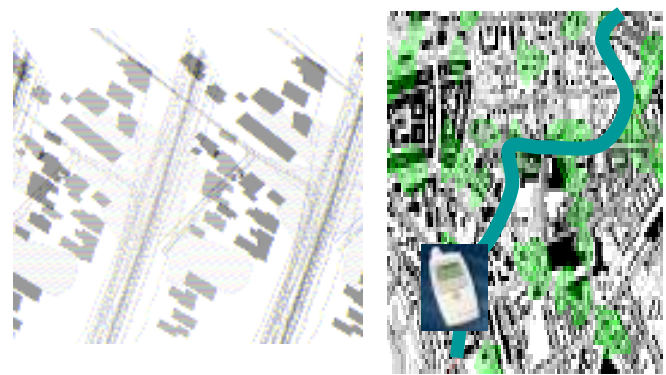
Some Challenges ...



Grid-based Data Mining & Data Mining Based Grid Monitoring (Technion, Fraunhofer, Daimler)



Data Stream Mining (Univ. Porto)

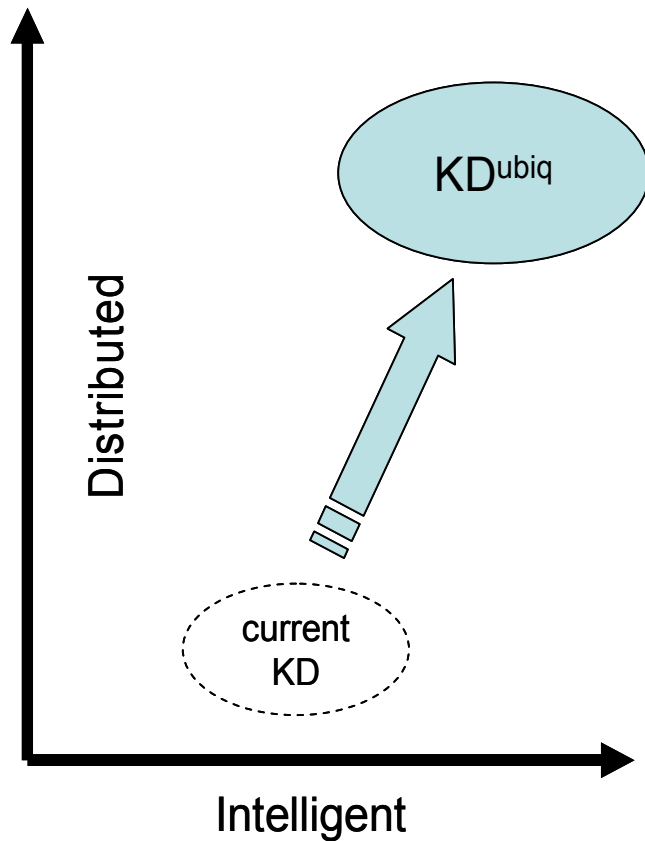


Mobility Mining from GPS-Tracks (Fraunhofer, Univ. Pisa, Univ. Sabanci)



P2P Music Mining (Univ. Dortmund)

Ubiquitous Knowledge Discovery



Knowledge discovery process in mobile, distributed, dynamic environments, in presence of massive amounts of data

= **Ubiquitous Knowledge Discovery**

Key characteristics

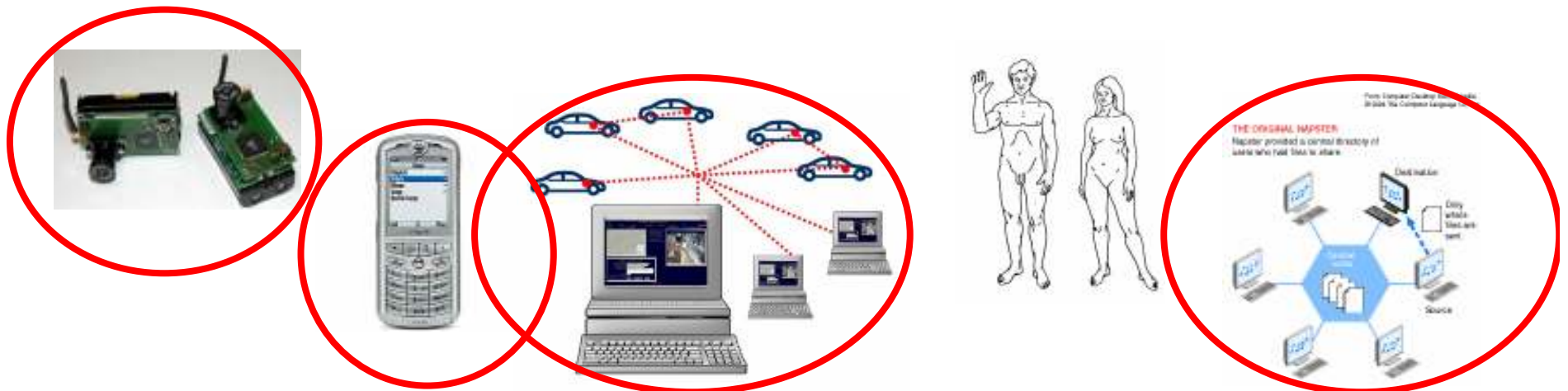


1. **Time and space.** The objects of analysis exist in time and space. Often they are able to move.
2. **Dynamic environment.** These objects might not be stable over the lifetime of an application. Instead they might appear or disappear. They exist in a dynamic and unstable environment, evolving incrementally over time.
3. **Information processing capability.** The objects are endowed with information processing capabilities
4. **Locality.** The objects never see the global picture - they know only their local spatio-temporal environment.
5. **Real-Time.** They often have to take decisions or act upon their environment - analysis and inference has to be done in real-time. The models have to evolve incrementally in correspondence with the evolving environment.
6. **Distributed.** In many cases the object will be able to exchange information with other objects, thus forming a truly distributed environment

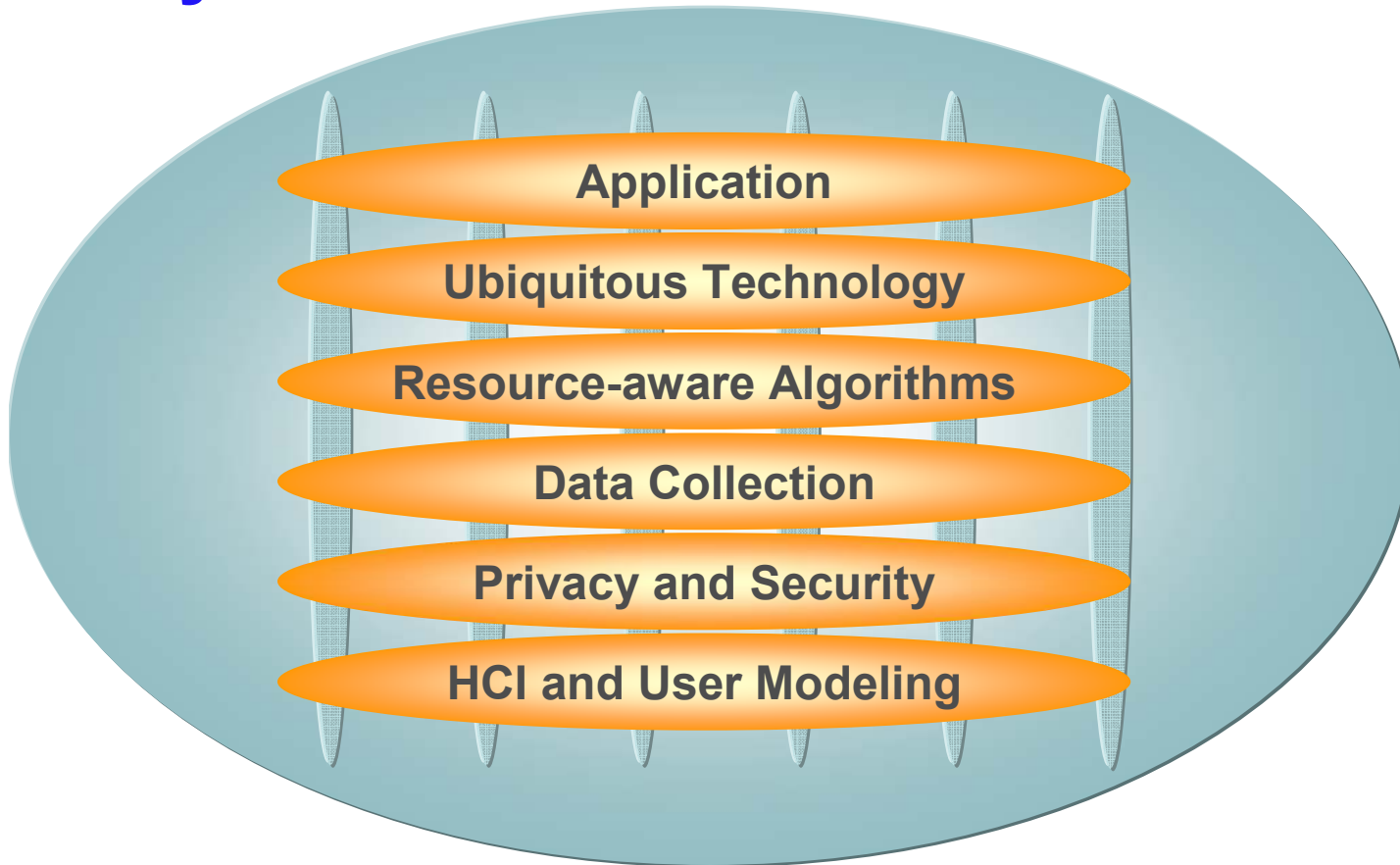
Objects of Study



- Systems that have these properties are humans, animals, and increasingly, computing devices
- KDUbiqu investigates artificial systems
 - The machine learning or data mining is not applied to data *about* the system,
 - it is rather *part* of the information processing capabilities of the system
- *This is a large departure from the current mainstream in machine learning and data mining!*



Dimensions of Ubiquitous Knowledge Discovery Systems



When designing a ubiquitous knowledge discovery system, major design decisions in each of these six dimensions have to be taken.

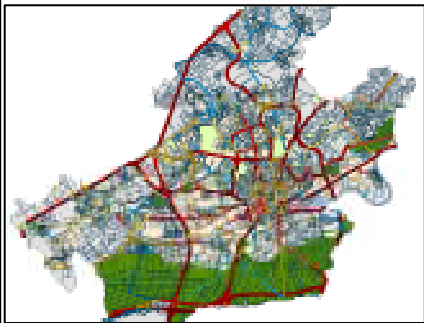
Choices are mutually constraining each other

The best way to understand a system is to build one!

KDUbiq thus adopts a *systems* view on how to build next generation knowledge discovery systems

The path to ubiquity

?



Ubiquitous Knowledge Discovery

Activity Mining inside Mobile Devices

GPS Track Mining

Traffic Prediction using video measurements

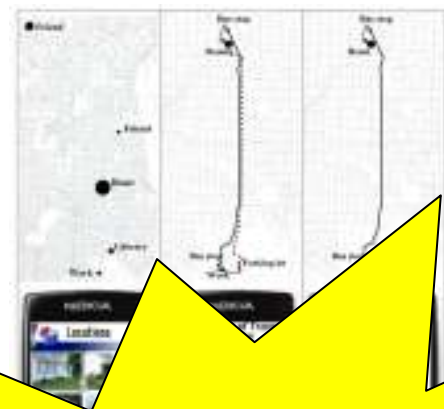
A blue background with a glowing path of binary code (0s and 1s) leading to a bright light at the top, symbolizing data discovery. A globe is visible in the background.

ubiq
KD

Some more challenges ...

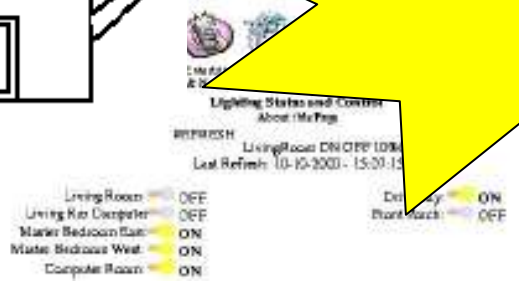


Probabilistic Robotics - Self Driving Car

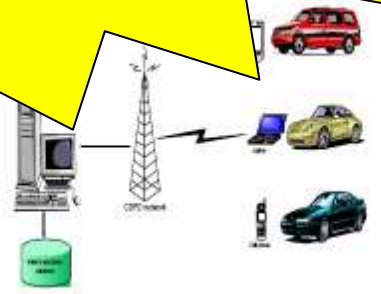


Machine Learning – Collaborative P2P Mining

Machine Learning makes the difference!



Intelligent Agents – Smart Homes



Data Mining - Vehicle Monitoring

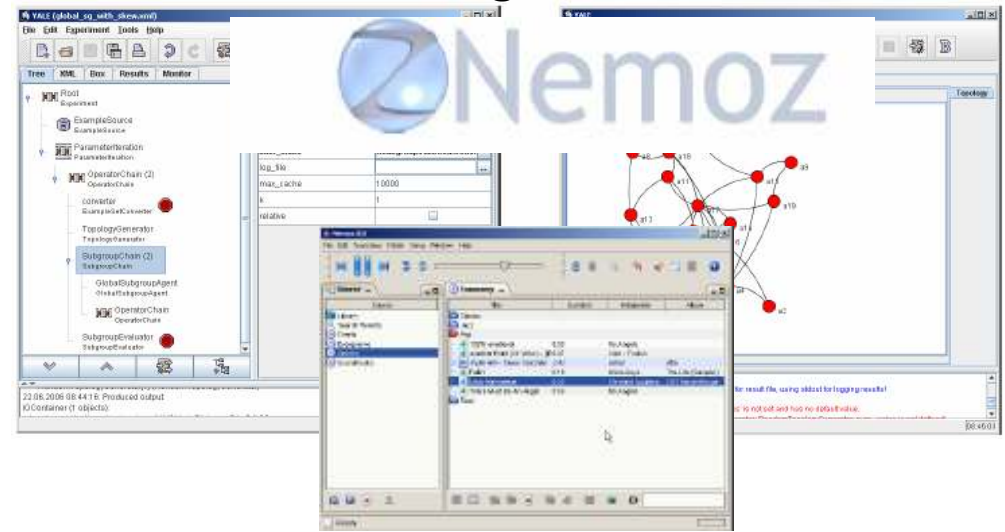




A self-driving car

What do these two systems have in common?

A Web 2.0 application for distributed, collaborative clustering of music collections



The path to ubiquity



? Communicating cars, respecting privacy

? Operating under city traffic conditions



Ubiquitous Knowledge Discovery

Autonomous Communicating Cars

Darpa 2007 Challenge

Stanley Darpa 2005

ubiq KD

What are the biggest research challenges?



- Coming up with algorithms and a theoretical framework for problems that violate the iid assumption
- Integrating results from
 - distributed data mining
 - privacy preserving data mining
 - spatio-temporal data mining
 - data stream mining
 - collaborative data miningin Ubiquitous Knowledge Discovery systems

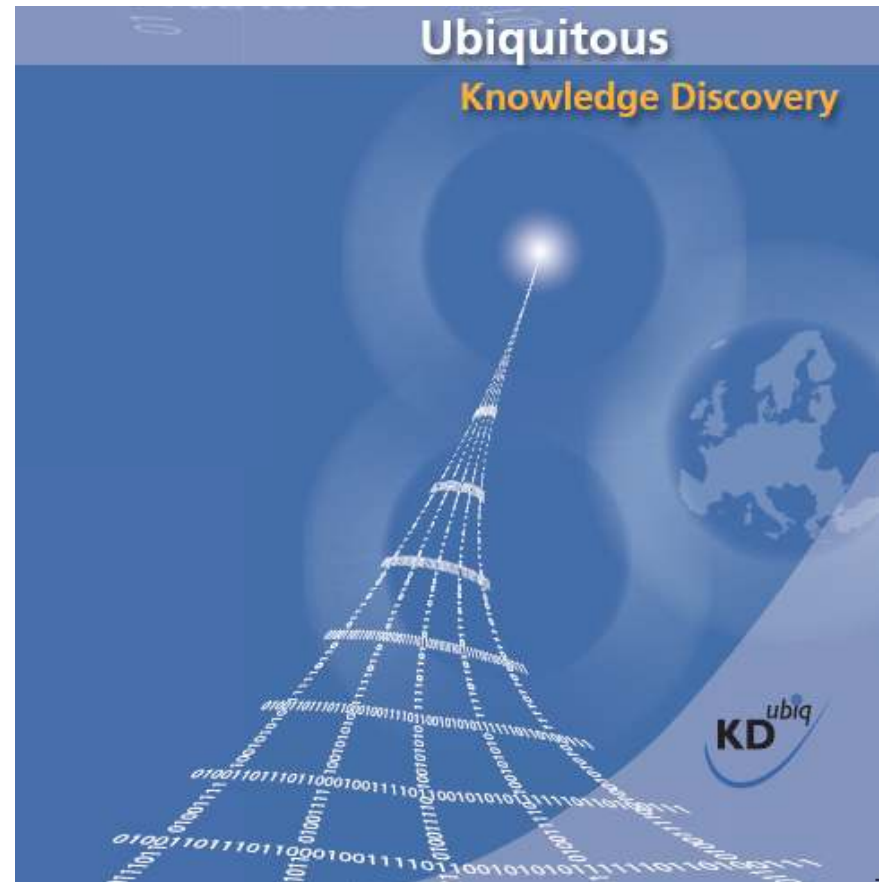


Towards KDUbiq

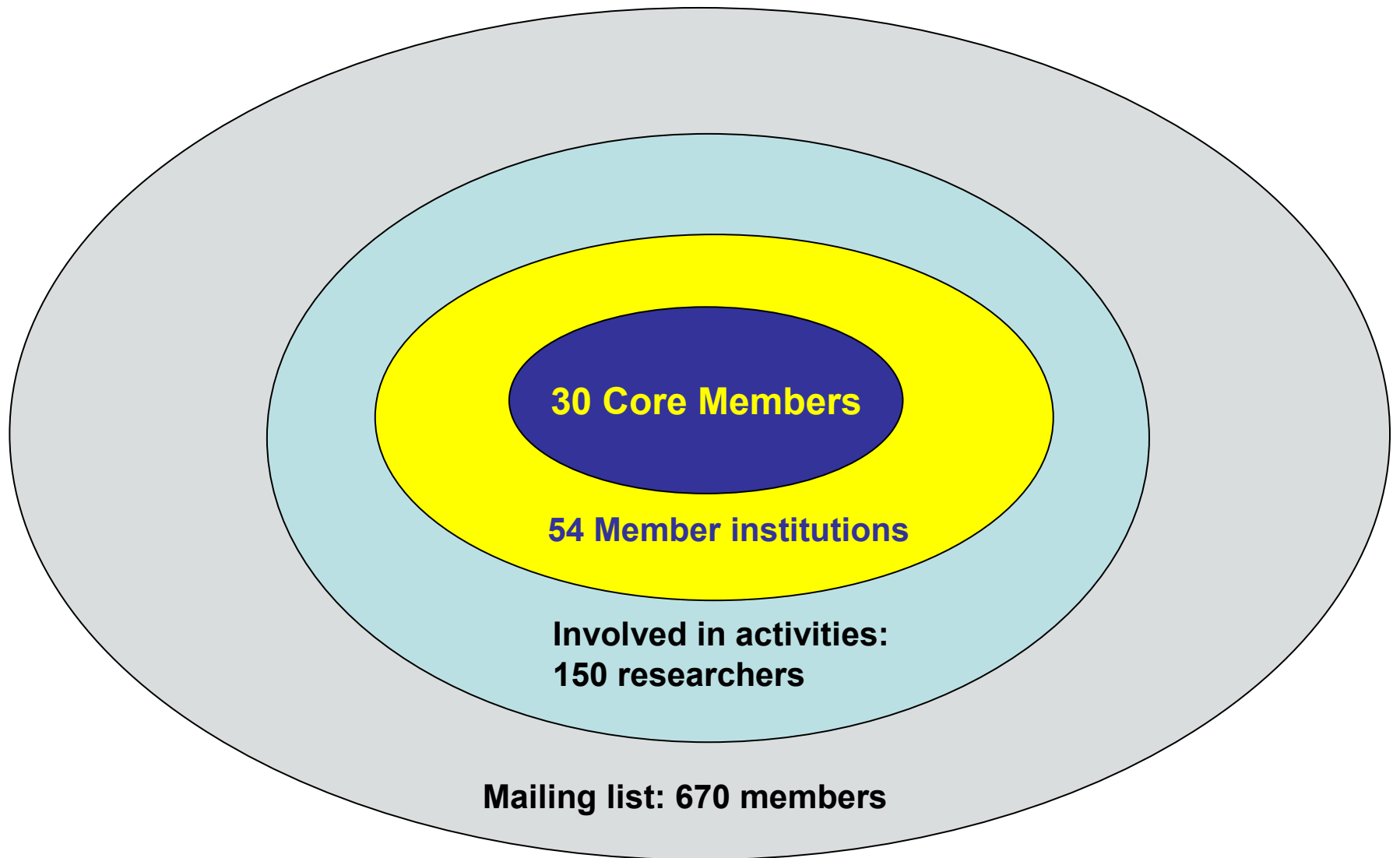
KD^{ubiq} Coordination Action



- To stimulate research, to define the field, and to shape the community in Europe, the KD^{ubiq} research network was launched in 2006.
- It is funded by the European Commission
- Currently it has more than 50 members from research and industry
- Not a research project, it's about shaping a community
- Budget 1.2 Mio \$, 2006-2008
- www.kdubiq.org



KDubiq Community



Objectives



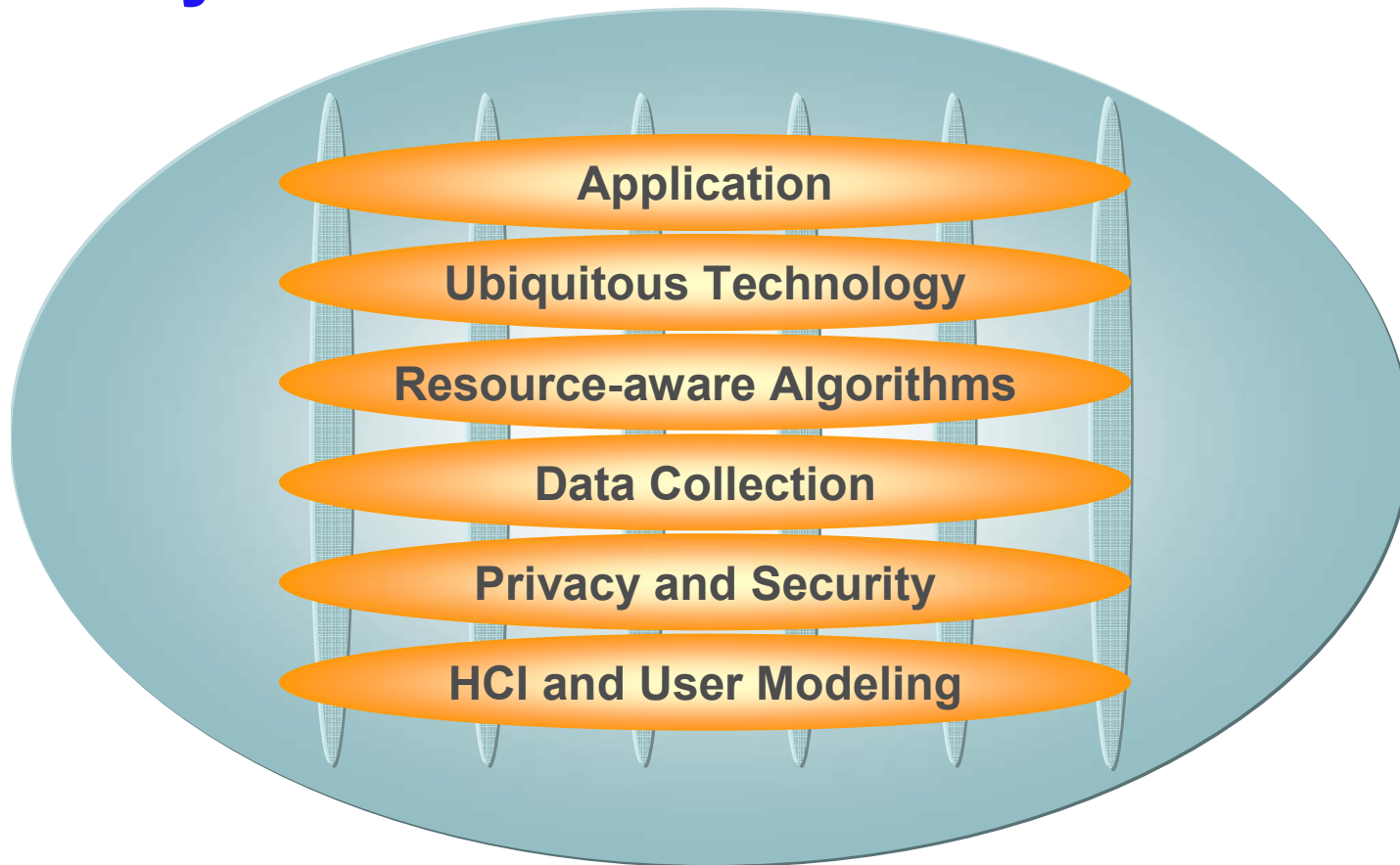
Today

- Relevant research is scattered around several communities
- Highly interdisciplinary skills are needed
- Projects usually fall short of realizing the full potential of ubiquitous knowledge discovery

Goal

- bring researchers from relevant fields together
- form critical mass
- establish community
- define a joint research agenda

Dimensions of Ubiquitous Knowledge Discovery Systems



When designing a ubiquitous knowledge discovery system, major design decisions in each of these six dimensions have to be taken.

Choices are mutually constraining each other

The best way to understand a system is to build one!

KD Ubiq thus adopts a *systems* view on how to build next generation knowledge discovery systems

Working Group Chairs



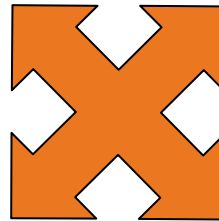
Working Group	WG Chairs	Country
Application Environments	Koon Vanhoof, Hasselt Univ.	Belgium
Ubiquitous Technologies	Assaf Schuster, Technion Ran Wolff, Univ. Haifa	Israel
Resource aware, distributed algorithms	Joao Gama, Univ. Porto Antoine Cornuéjols, Univ. Paris Sud	France Portugal
Ubiq. Interaction and Data Collection	Rasmus Pedersen, Univ Copenhagen Andreas Hotho, Univ. Kassel	Germany Denmark
Security & Privacy	Fosca Giannotti, Univ. Pisa Yücel Saygin, Sabanci Univ.	Italy Turkey
HCI & Cognitive Modelling	Bettina Berendt, Humboldt Univ, Berlin Ernestina Menasalvas, Univ. Madrid	Germany Spain

KDubiq Activities



**Integrating results
Collaborative book
editing**

**Integrating people
workshops,
working group
meetings**



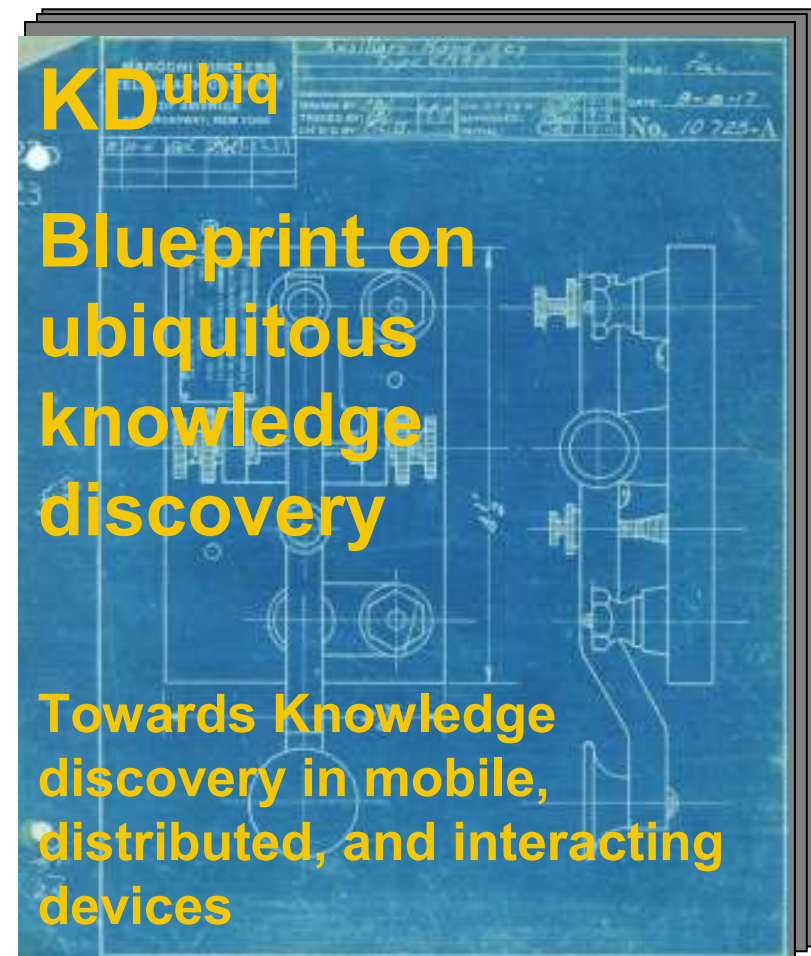
**Training:
Tutorials, Summer
School**

**Dissemination:
Portal**

Blueprint – collaborative book editing



- A collaborative effort to define the research challenges
- Six working groups corresponding to six main chapters
- 30 partners actively contributing
- Will result in a joint book in 2008



Workshops '07



International Workshop on Privacy, Security, and Trust in KDD

Home | Call for Papers | Invited Talks | Program | Committee & Chairs | Credits & Awards | Sponsors | Contact

Welcome to PrivKDD'07!
The 10th SIGMOD International Workshop on Privacy, Security, and Trust in KDD

Given the vast amount of data that is collected by service providers, system administrators, and available in public information systems, data mining practitioners are faced with the need to maintain security and confidentiality of their data. However, the application of technology to better security, while often serious systems regarding data confidentiality and personal privacy rights. The problems are global and fairly pervasive as we struggle to get national and international policies on privacy and security for data mining technology. Finding privacy and security, as well as extending to other systems of information and knowledge based systems, such as e-commerce, e-government, and health assessment. Discussion often focuses on ethical and policy aspects of the problem, emerging technology, and the evolution of a socially informed, e.g., an accessible user interface. System designers of personal information for data mining purposes in (2) can read.

For mobile, computer systems, and handheld devices that are today, have recognized that technology can be considered to support mobile devices to enable data mining capabilities without violating the privacy and security of the individual to whom the data corresponds. To meet privacy requirements and confidence data mining projects, it is necessary to understand the goals of the user. It is the goal of this workshop to create a friendly research on privacy, security, and how to apply this research within a data mining framework. We address technical and social aspects. We hope to attract interest from a

10 July 2007
End of Talk
Deadline: Extended Abstracts
Extended Abstracts (1000-1500 characters)
Keynote Speeches

10 July 2007
Accepted Papers
The list of PrivKDD'07 accepted papers is available in the "Accepted" page.

2 Aug 2007
Important Dates
Accepted papers announced
Abstract deadline (11:59 AM 2007) at www.kdd.org/standard/abstracts.html

2 Aug 2007
Workshop Website
http://www.kdd.org/standard/abstracts.html

Third International Workshop on Privacy Data Management

April 11, 2007, Istanbul, Turkey
Sponsored by
Third International Conference on Data Engineering (ICDE'07)

Scope:
The workshop is concerned with the privacy and security of data. The workshop is a part of the ICDE'07 conference. The workshop is a part of the ICDE'07 conference. The workshop is a part of the ICDE'07 conference.

Topics:
The workshop is a part of the ICDE'07 conference. The workshop is a part of the ICDE'07 conference. The workshop is a part of the ICDE'07 conference.

ICDE'07 Istanbul

- Privacy: Security, Algorithms, and Applications
- Privacy: Policy, Legislation, and Architecture
- Privacy: System, Model, Control, Language
- Privacy: Data Mining
- Privacy: Risk, Assessment, and Analysis
- Privacy: Security
- Privacy: Policy, Analysis

Monday, September 17

Workshop on Knowledge Discovery from Ubiquitous Data Streams
Chair: João Gama, Mohamed Gaber, Jesus Aguilar
Room 106

This workshop is sponsored by the Institute "State-of-the-Art in Data Stream Mining".

Session 1 - 14:00-15:30

Network-based Scientific Data Visualization: Lessons Learned and Challenges Ahead
by R. Borrajo (Invited talk)

Learning an Outlier-Robust Kernel Filter
by J. Tang, C. Theodorou, S. Schaal

A Model for Quality Guaranteed Anomaly-Aware Stream Mining
by R. Kuzuno, C. Franks, M. Gaber

Efficient Secure Query Processing in AM, Data-Stream
by Dong-Chan Ah, Sang Park

Enhanced Anomaly Algorithm for Intrusion of 100-Year Glasses
by Mark Last, A. Savakis

Query-Adapted Deviation Classifier
by R. Yoshida, E. Hruschka

15:30-16:00 Coffee break

UM 2007
11th international conference on user modeling
Corfu, Greece, 25-29 June, 2007

UBIQUITOUS KNOWLEDGE DISCOVERY FOR USER MODELING (K-UMUM'07)

Monday, 25 June 2007 - part of the WORKSHOP ON DATA MINING FOR UBIQUITOUS COMPUTING

Featuring an invited talk by **GOAD MOCHE**: The Biological Approach to User Modeling at **JAMER BOURGAIN** at **Human Pedagogical Goals Assistant**

1000 Abstracts (papers, posters, slides, and tutorials)

Mini-tutorial abstracts, knowledge discovery and data mining are of increasing importance for user modeling. Due to the increasing volume of data available and the need for sophisticated user-adaptive, knowledge discovery beyond a variety of user modeling, the focus expands to the rapidly growing and ubiquitous computing, by "ubiquitous", we mean ubiquitous devices, ubiquitous data (with spatial, temporal), and ubiquitous (distributed) processing. In addition, we also refer to the new heterogeneous in the environments and other context potentials of usage, the increasing diversity of global autonomous (space, objects), and to the goal of information systems, the effective (retrieval and accessibility of information, facing the challenges of integrating these aspects into the model representation and improved, privacy and security play a central role.

This workshop is part of the activities of the HCI and Cognitive Modeling working group of the EU FP6 project UBIQUITOUS. From that workshop, we want to start towards a reference system of ubiquitous knowledge discovery for user and user modeling. We start from Corbett et al.'s two-dimensional reference system of "Digital personas" (ICDM 2004). The authors propose to localize application domains with respect to five dimensions: by whom is the information recorded and the application controlled (self or others), and by whom is the information used (self or others). For example, a diary is a self/self application, a tutor is an others/others application, and an observer is an others/others application. We propose to add a third dimension: by whom is the information processed (people or machines). The aim of abstracting this third dimension is to investigate how and to what extent semi-automatic or fully automatic KDD can contribute to better applications.

ubicomp 07
10th International Conference on Ubiquitous Computing
16 - 18 September 2007, Istanbul - Aesthetics - Europe

Call for Workshops

As UbiComp 2007 Workshop Co-Chairs, we would like to invite proposals for the workshop program of the 10th International Conference on Ubiquitous Computing. The workshops will be held on Sunday, September 16th, prior to the main conference and will be located at the University of Istanbul.

Workshops provide an opportunity to discuss and explore specific areas of ubiquitous computing research with a group of interested researchers and practitioners. Workshops may focus on any aspect of ubiquitous computing, emerging topics, established concerns, or new ideas. The goal of the workshops is to share understandings and experiences, to foster research collaborations, to learn from each other and to envision future directions. It is our intention to provide an UbiComp Workshops Proceedings with papers from all workshops.

Important Dates

- March 15, 2007 (Friday) - Submission Deadline (workshops proposals)
- April 18, 2007 - Acceptance Notification (workshop proposals)
- May 7, 2007 - Distribution of Accepted Workshop CPDs
- June 21, 2007 - Proposed Submission Deadline (workshops, position papers)
- July 21, 2007 - Proposed Acceptance Notification (workshops, position papers)

AmI-07
European Conference on Artificial Intelligence
November 11-15, 2007, Istanbul, Turkey

Workshop on AmI-07
Artificial Intelligence and Ubiquitous Computing

Workshop on AmI-07 is a part of the AmI-07 conference. The workshop is a part of the AmI-07 conference.

Knowledge Discovery from Sensor Data (Sensor-KDD '07)

To be held in conjunction with

August 12, 2007
San Jose, California, USA

Call for Papers

With new sensor distributions, events networks, and wireless sensor networks yield massive streams of data. In this workshop, we focus on the management of this data. As sensor networks become ubiquitous, a lot of research is beginning to emerge on the high-level operations involving sensor networks and management, especially to detect change, reduce or maintain security, and the management of critical infrastructure. The new data from sensors need to be efficiently managed and transformed to enable information through

KD^{ubiq} Summer School

- Covers various areas relevant for ubiquitous knowledge discovery
 - sensor networks
 - embedded systems
 - distributed data mining
 - distributed computing and grid
 - collaborative environments
 - privacy and security
- Addresses both foundations and state of the art
- puts emphasis on exercises, will allow participants to get hands on experience in interdisciplinary work.



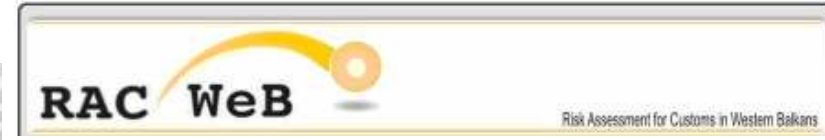
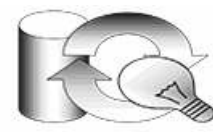
Second European Summer School on Knowledge Discovery for Ubiquitous Computing

Prof. Dr. Katharina Morik
Dr. João Gama

March 3 – 9, 2008
University of Porto, Portugal
www.kdubiq.org



Project Clustering



- KDubiq Project Presentation Day at ECML/PKDD 2006 and 2007 program
- The goal is networking and exchanging of results/experience.
- In total 24 projects presented
- each time ~ 350 participants

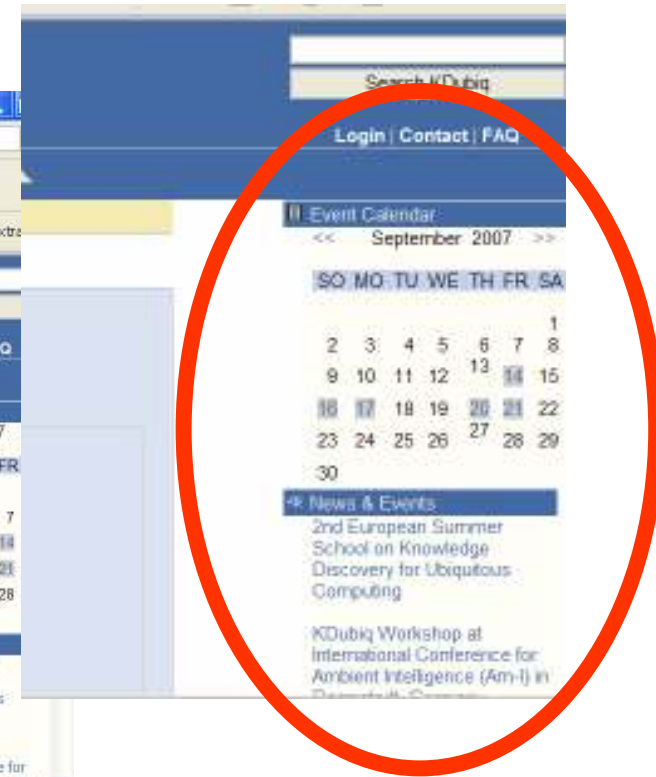
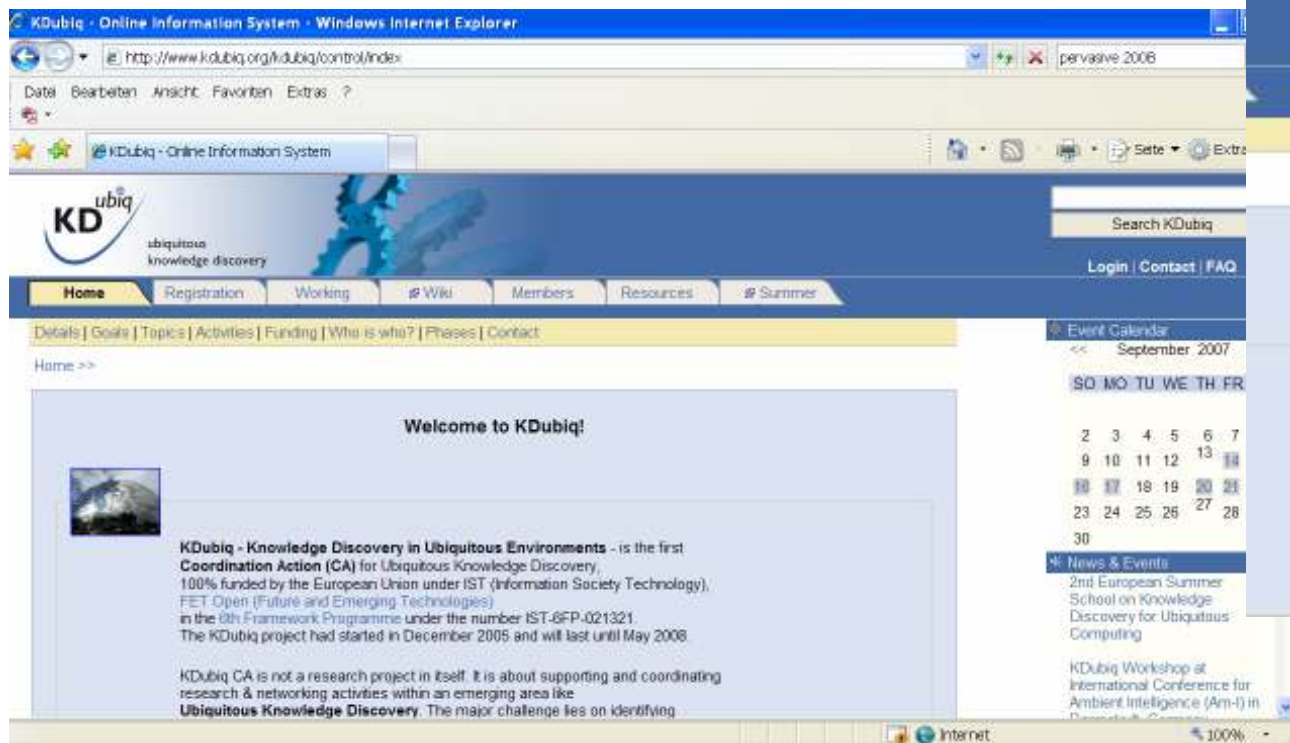


Rotes Rathaus, Berlin 2006

University Warsaw, 2007



KDubiq portal



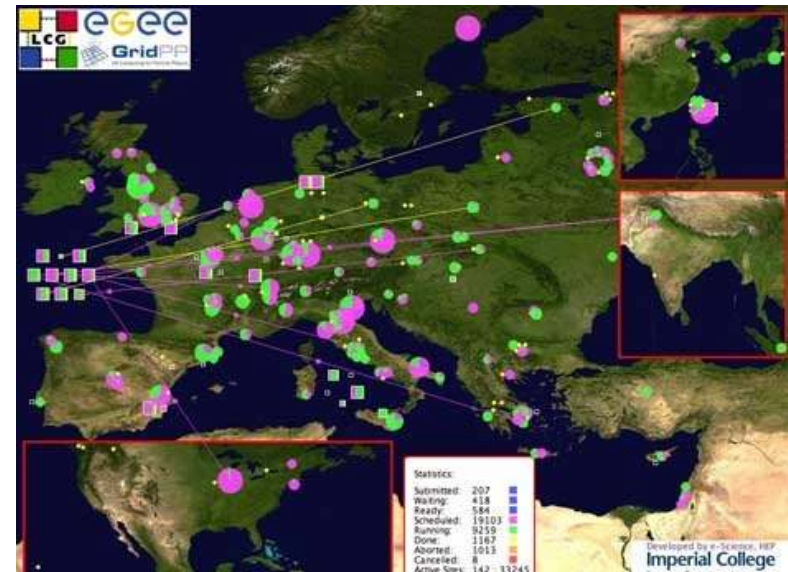
www.kdubiq.org

News Ticker
& KNet &
KDubiq
mailing lists

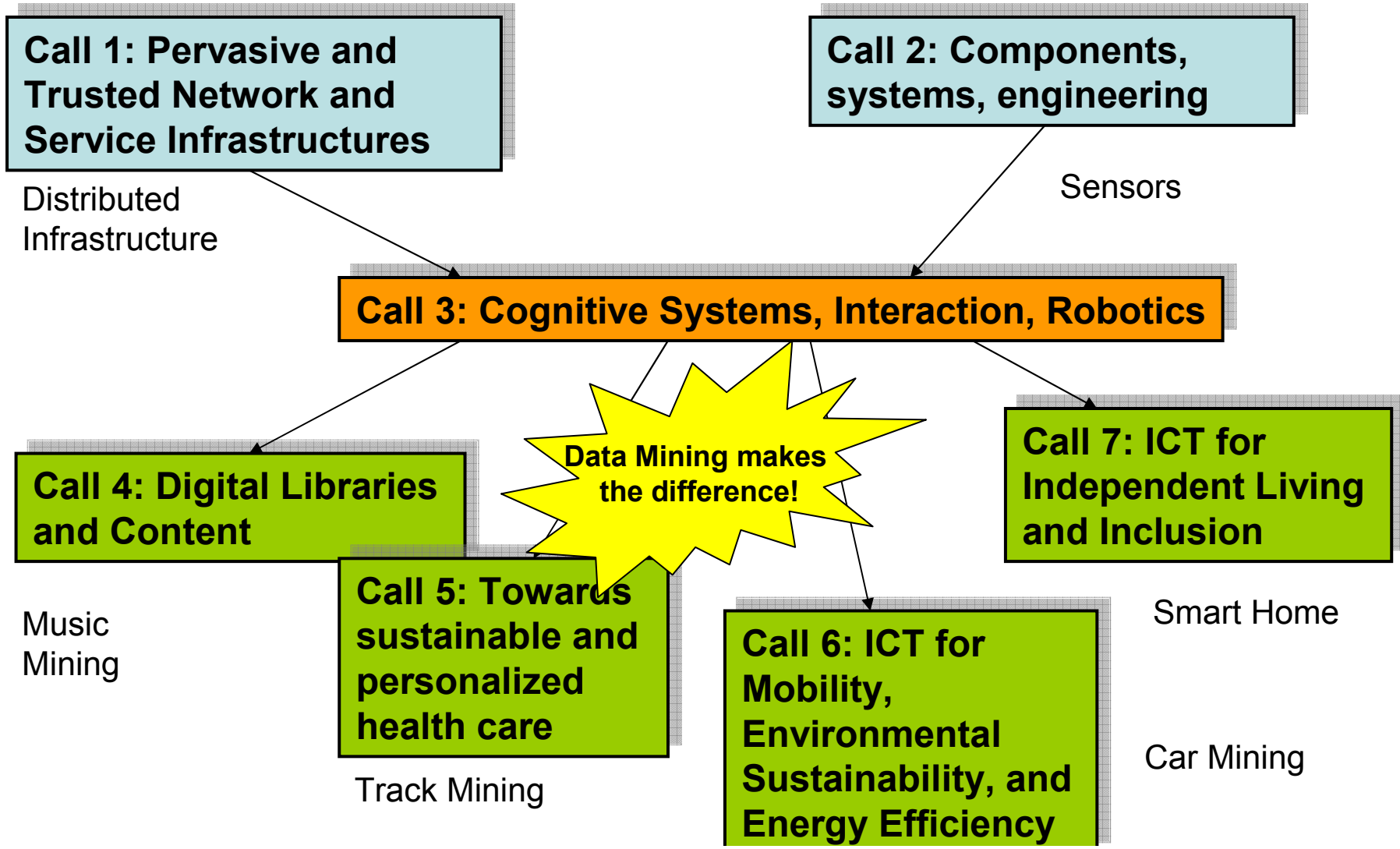
Benchmarking



- The availability of benchmarking data sets can have big impact on shaping a field
- The idea is to have some relevant and large-scale application in order to benchmark KDubiq algorithms,
- There has been work on identifying benchmarking data sets.
- a first data set from the grid computing area from the EGEE project.
- the data set will be used during the Summer School in Porto 2008.



A KDUBiq perspective on the European ICT Research Programme



Summary 1



- For many disciplines, further progress crucially depends on advances in data mining and machine learning
- Ubiquitous Knowledge Discovery provides an integrated perspective for studying data mining challenges in
 - distributed data mining
 - privacy preserving data mining
 - spatio-temporal data mining
 - data stream mining
 - collaborative data mining
- Taking a systems point view demonstrates the need for work that *combines* several (or all) of these topics
- Biggest algorithmic challenge is to study learning settings that violate the iid assumption

Summary 2



Results on KDubiq networking activities

- Data Mining is a fragmented discipline
- people that never cooperated before and did not pay attention to the other's research work started collaborating
- Everyone in the network has a much broader view on emerging issues and interrelations
- integrating research from data mining sub-communities in research projects

Resources



- CDI should
 - allow for integrating work across data mining sub-communities
 - provide benchmarking data sets
 - provide support for networking
 - establish international links