



Elena Kornyshova – Cédric du Mouza

Russian-French Workshop on Big Data Applications

2 december 2016

**Cnam-Paris
CEDRIC Lab. – ISID Team**



CNAM



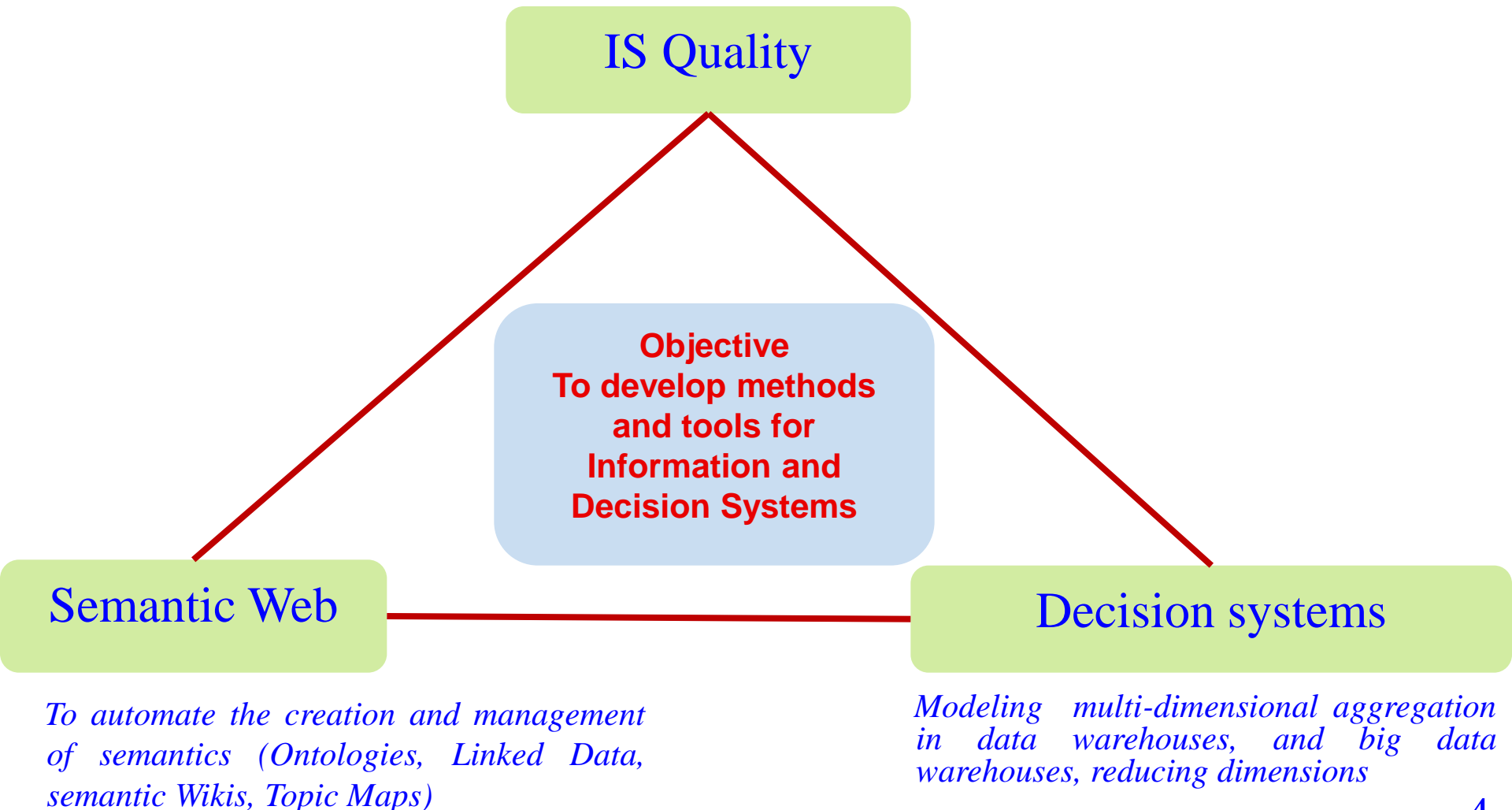
- **Founded in 1794**
- **70,000** students per year in France (mainly continuing education)
- **7,000** students around the world
- **1,000,000** alumni
- **450** diplomas and certifications offered
- **1** engineering school
- **150** sites in France
- **45** sites around the world
- **50%** of teachers belong to service and industry companies.
- **21** labs, **350** researchers

Cedric Lab.

- gathers the research activities in computer science, statistics and signal processing at [Cnam](#)
- Got a A mark from the national evaluation
- 8 teams, 80 permanent researchers, 65 PhD students
- 200-250 publications/year
- Topics:
 - Reliable systems: Certified Design and Programming
 - Information systems and Databases
 - Networks, Systems and Multimedia
 - Combinatorial Optimization
 - Statistical methods for data-mining and machine learning
 - Signal processing and electronics

ISID Team: 10 researchers

To integrate in the early phases of the development process, the non-functional requirements (quality, security, privacy, performance ...)



Contribution: Reverse engineering and Integration of Big Data

- A large experience on reverse engineering from physical to conceptual models
- How to extend this work for Big Data (especially streaming data)
- From datasets, we build a physical model with denormalized data which have to be conceptualized
- Data from various sources → integration techniques required (ontologies, topic maps, etc.)

1 PhD, 20+papers (DKE, ER, CAISE, etc.)

Contribution: Processes and Methods in Big Data

Situational engineering of processes and methods related to Big Data

- How to adapt processes and method in a particular case
- Components-based approach
- Contextual definition of method application and process execution
- Decision-based guidance through the processes and methods

Contribution: Data & Model Quality

Focusing on both data and their models

- An approach and prototypes for evaluating and improving quality in data and process models
- Handling quality in the context of multiple and various data sources with or without integration and more recently within open linked data by considering both the data and their interlinking

5 PhD students, 30+ papers (ER, ICIQ, RCIS, ADBIS, ...)

Contribution: Data Security & Privacy

- Experience on data anonymization :
 - (semi-)automated detection of sensitive data in RDBMs
 - Anonymization strategies for data publishing in the context of open data
- Current project on analysis of complex and targetted attacks from log files:
 - Very large datasets (log files)
 - Unstructured data
 - Streaming of logs
 - IS with diverse architectures, including connected components

2 PhD students, 10+ papers (IJMR, MOBID, ...)

Contribution: Data Warehouse Modeling

- Experience on data warehouse and dimensional modeling
 - MDA approach
- Current project on big data warehousing
 - Defining logical models for noSQL databases
 - Defining mappings between UML/logical/physical noSQL structures

1 PhD student, 20+ papers (DKE, ER, DSS, ESWA, ...)

Contribution: Social Web and recommendations

2 approaches:

- Content-based storing, querying, filtering and recommending Web data (Web pages, RSS items, tweets, etc)
- Topology-based storing, querying, filtering (collaborative or not) and recommending in large scale social graphs (Twitter, Facebook, etc)

6 PhD students, 20+ papers (EDBT, CIKM, WISE, JCST, RCIS, SSDBM, ...)



Thank you!