Laboratoire d'Informatique et d'Automatique pour les Systèmes

Utilisation et Exploitation Données

Ladjel Bellatreche

LIAS/ISAE-ENSMA, Poitiers, France

bellatreche@ensma.fr

http://www.lias-lab.fr/members/bellatreche

Téléphone: 05 49 49 80 77

Université de Poitiers

ISAE Institut Supérieur de l'Aéronautique et de l'Espace ENSMA

Goals of this course

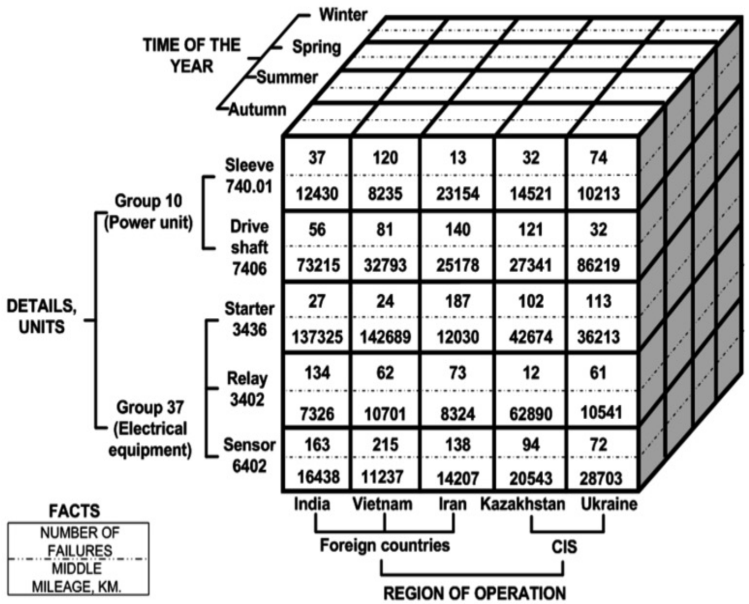
1. Student Awareness of Data Pre-processing
2. Causes of Dirty Data
3. How to Deal with Dirty Data?
4. Ontologies

OLTP Services

Characteristics	OLTP Databases
Services	Day-to-Day Applications
Model	Entity Relationship
Dependency Management (Normalization & Functional Dependencies)	Mandatory
Nature of Data	Actual, Raw
Updates	Immediate
Consolidation	Low
Query Operations	Read/Write
Size	Gigabyte
Perception	Bi-dimensional
Pre-processing	Rare

OLAP Services

Characteristics	OLAP Databases
Services	Analytical Applications: Report and Analyse Data
Model	Star Schema & Snow flake Schema
Dependency Management (Normalization & Functional Dependencies)	Rare
Nature of Data	Historical and Aggregated
Updates	Rather Deferred
Consolidation	High
Query Operations	Read/ Refresh
Size	Terabytes
Perception	Multidimensional
Pre-processing	Usual (Extract/Transform/Load)



Examples of OLAP Operations (1)

Roll up

	18	19	20
Alim.	496	520	255

Roll up

Dimension Time

	18-20
Fruits	623
Viande	648

	18	19	20
Fruits	221	263	139
Viande	275	257	116

	1S18	2S18	1S19	2S19	1S20
Fruits	100	121	111	152	139
Viande	134	141	120	137	116

Drill down

	18	19	20
Pomme	20	19	22
…	…	…	…
Bœuf	40	43	48

Drill down

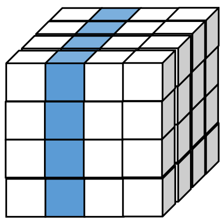
Dimension Product

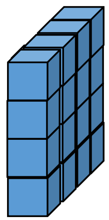
Examples of OLAP Operations (2)

- Slice ⇔ projection

		18	19	20
Œuf	Idf	220	265	284
	Ain	225	245	240
Viande	Idf	163	152	145
	Ain	187	174	184

		19
Œuf	Idf	265
	Ain	245
Viande	Idf	152
	Ain	174





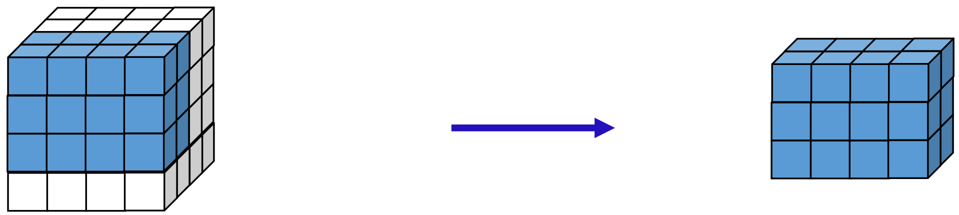
Examples of OLAP Operations (3)

- Dice ⇔ Selection

		18	19	20
Œuf	Idf	220	265	284
	Ain	225	245	240
Viande	Idf	163	152	145
	Ain	187	174	184

→

		18	19	20
Œuf	Idf	220	265	284
	Ain	225	245	240



OLTP vs. OLAP

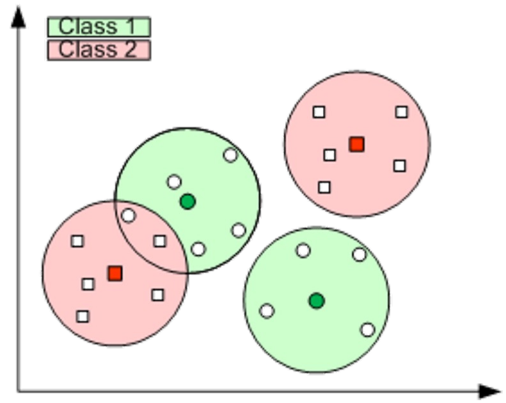
OLTP Tasks	OLAP Tasks
Find the price of a Book	Calculate Books with best profile margins
Update last Customer Transaction	Find most Loyal Customer
Keep track Employee Hours	Identify the Employee of the Month

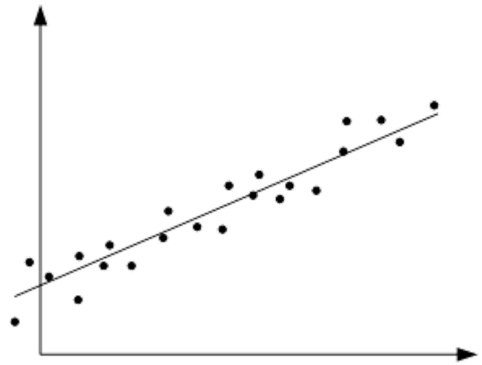
Big Data Analytics (BDA)

- **Simple:**
  - Cubes: BI, OLAP, MOLAP
    - They capture descriptive statistics (histograms, means, quantiles, plots, statistical tests)
- **Complex:**
  - Mathematical Models (ML, statistics, optimizations)
  - Graphs (Path, Triangle, Connectivity, Cliques)

Machine Learning (ML)

- A Large dataset (size n) with d *dimensions*: cannot fit in RAM, minimize I/O
- Multidimensional
  - *d*: 10-1000 of dimensions
  - Feature selection and dimensionality reduction
- Represented & computed with matrices & vectors
  - data set: mixed attributes
  - model: numeric=matrices, discrete: histograms
  - intermediate computations: matrices, histograms, equations





ML Algorithm's Performance

- Behavior with respect to dataset:
  - Few: one pass, fixed # passes (Regression, ...)
  - Most: iterative, convergence (k-means, SVM)
- Time complexity: $O(n \times d)$, $O(n \times d^2)$, $O(n \times d^3)$, $O(n^2 \times d)$

BDA: DW + ML + Text

Data Preparation

- Copy, Load
- Data Exploration, Data Quality
- Pre-processing: Data Integration, Cleaning, Transformation, …

Analytics

- Cubes
- Machine Learning, Statistics Models
- Graphs

Deployment

- Scoring
- Tuning
- Life Cycle Maintenance

Highly Iterative Process

BDA and SQL

- Preparing data takes a lot of time
- SQL helps but **it requires expertise**, statistical programming in R or SAS, spreadsheet macros, custom applications in C++/Java
- Lot of data: tabular originates in a DBMS
- Emphasis on large *n, but* many problems have many "small" files
- Data set is generally much smaller than raw data

DBA inside DBMS

- Huge data volumes: potentially better results with larger amounts of data; less processing time
- Minimizes data redundancy; Eliminate proprietary data structures; simplifies data management; security
- Caveats: SQL is not as powerful as C++, limited mathematical functionality, complex DBMS architecture to extend source code

Processing

1. Parallel processing
2. Inside DBMS: SQL, UDFs, system C++
3. Outside DBMS: MapReduce, Spark, C++/Java

BDA is not AI

- **Artificial Intelligence**

→**Goal:** build some software that can perform some tasks that are said to require *intelligence*

- **Data Analytics**

→**Goal**: discover some useful information or build some useful models from data to understand the past or predict the future

The Need for Data?

- Some traditional AI Techniques **do not require data**
  - Examples: theorem provers, path-planners, logic reasoner
- Machine learning: AI techniques that are not explicitly programmed. **They require training data**
  - Examples: Neural networks – Deep Learning

Keys to the Success of ML

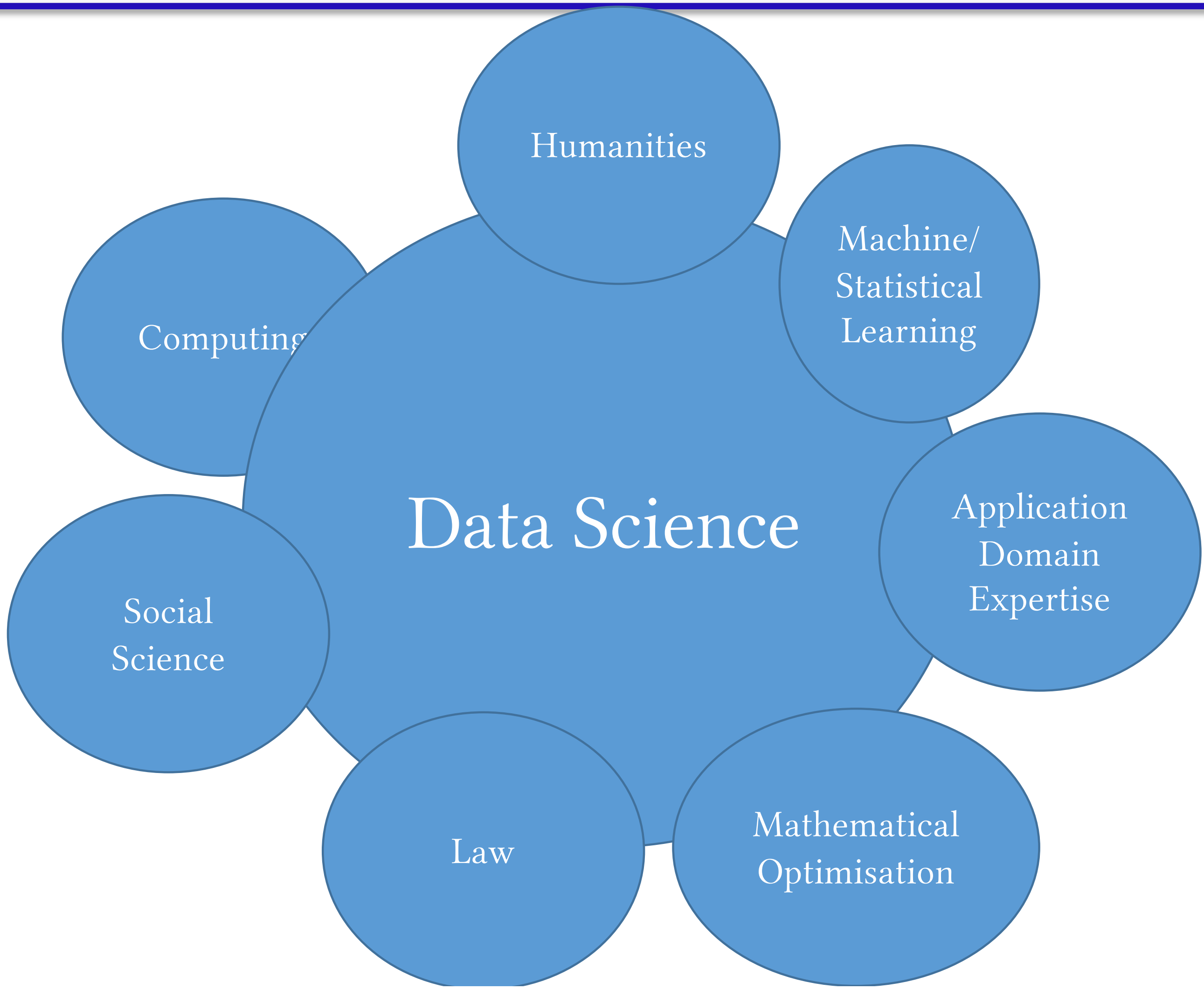
1. Clean Data
2. Better algorithms
3. Computing power (GPU, cloud...)
4. Big training data
5. Techniques to store and handle data

Data Science

- A data-based approach to problem solving by analyzing and exploring large volumes of possibly multi-modal data, extracting from it knowledge and insight that is used for better decision-making [T. Özsu, IEEE Big Data 2022]

- It involves the process of collecting, preparing, managing, analyzing, and explaining the data and analysis results.

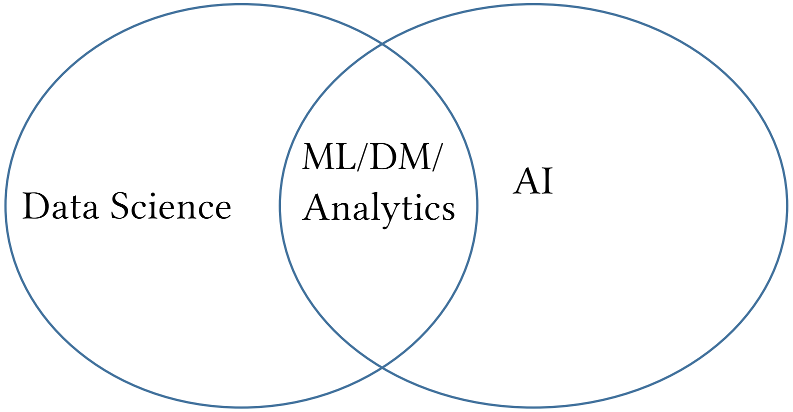
Data Science as A Unifier



Two Myths…

- Data science $\neq$ Big data
- Big data is like a raw material
- Processing it leads to data science & better understanding
- Applications are important
  - No applications →no data science

- Data science $\nsubseteq$ Machine Learning



They are related but not the same

Data Science Ecosystem

Applications

4 Pillars of Data Science

Data Engineering
- Big data management
- Data analysis
- Data understanding
- Data preparation

Data Analytics
- Explore data (data mining)
- Build models & algorithms (ML)
- Visualizations & visual analytics

Data Protection
- Security
- Privacy

Data Ethics
- Impact on individuals, organizations & society
- Ethical & normative concerns
- Bias in data
- Algorithmic bias
- Regulatory issues

Social and Policy Context

Data Engineering

Big data management

- Data enrichment, integration and storage
  - Extract/Transform/Load or Extract/Load/Transform ?
  - Data lakes
- Storage and management of big datasets
- Data processing platforms

Data understanding & analysis

- Data profiling
- Detection of anomalies
- Data changes

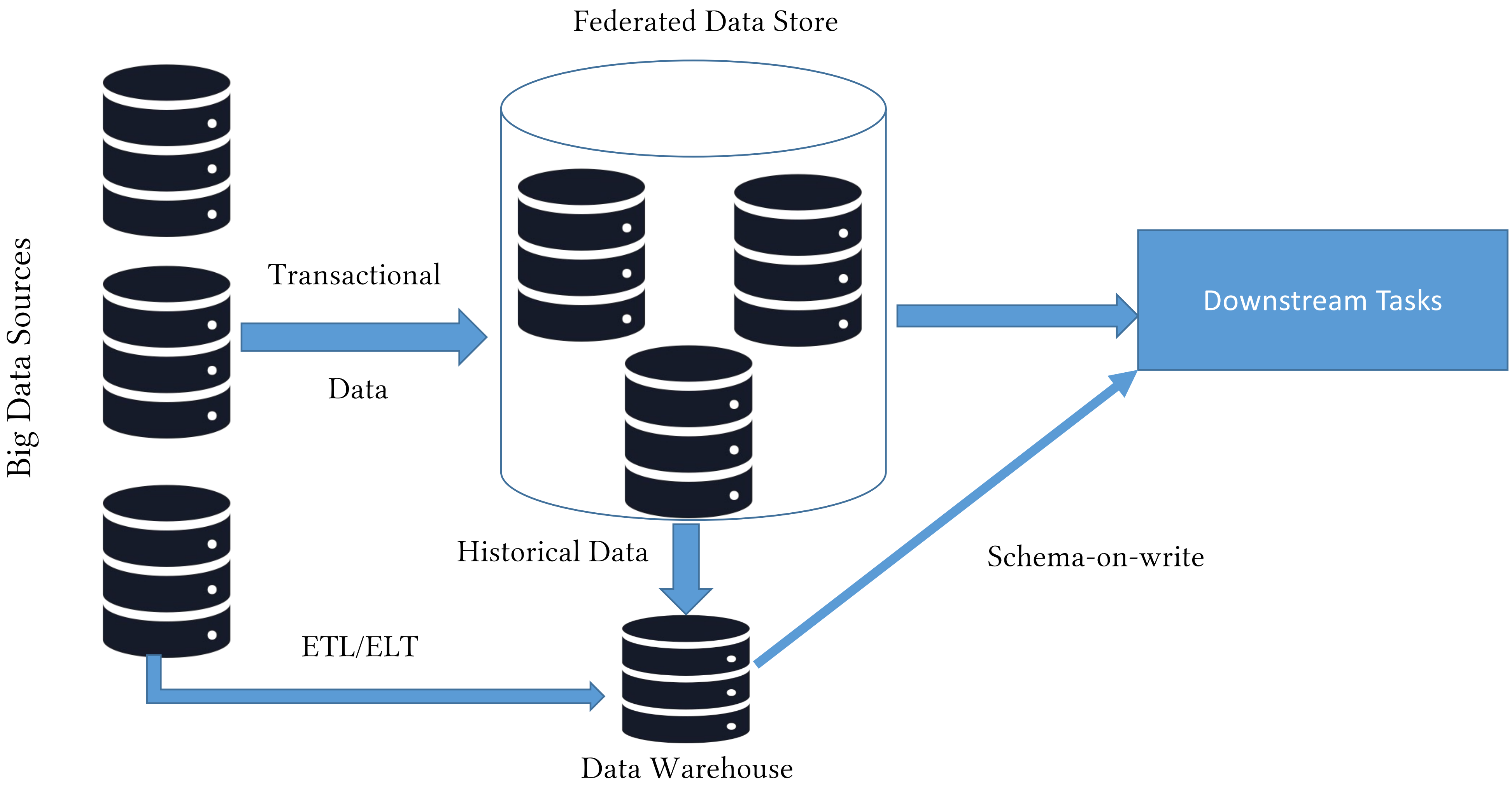
Data Preparation

- Data acquisition/gathering
- Data cleaning
- Data provenance & lineage

Data Engineering is important

Source: https://www.dataquest.io/blog/advanced-data-cleaning-r-course/

Data Integration



Data Heterogeneities

- **Schematic Heterogeneity**

S1:
- 2 classes
- 5 proprieties

S2:
- 1 class
- 4 proprieties

S1.prix*S1.TVA = S2.price

Components
- situation *(neuf ou occasion)*
- prix (*pour un composant neuf*)
- TVA

DisqueDur
- type
- capacité (*méga-octets*)

S1

HardDisk
- situation *(disponible ou indisponible)*
- price (*pour un disque dur neuf ou occasion*)
- type
- capacity *(giga-octets)*

S2

- **Semantic Heterogeneity**

Name Conflicts

synonym

homonym

Contextual Conflict

Conflict of unitz

Heterogeneity Solving

- **First Generation**
  1. Treatment of the syntactic heterogeneity problem (ex: standards of data representation: XML, EXPRESS, etc.)
  - **Semantic interoperability is manual.**
- **Second Generation**
  1. Semi-automatic integration using linguistic ontologies
  2. Ontologie linguistique : < mots, relations entre mots (SYN) > (ex: WordNet)
  - **Result of semantic integration is uncertain (e.g. homonymity problem)**
- **Third Generation**
  1. Automatic integration using conceptual ontologies

     Conceptual ontology: < classes, properties, relations between concepts and properties, relations between classes> subject to a consensus.
  - **Semantic integration easily « outillable »**
  - **Result of rigorous semantic integration**

Data Quality: By the Numbers

- Impact of poor data quality
  - In data science projects, data cleaning takes 30-80% of time/budget.
  - Erroneous data costs US businesses $600 billion/year [E02].
  - Data quality tools market is growing at 16% annually, way over 7% average for other IT segments [G07]

- How much data is erroneous.
  - Enterprise data error rates: average of 1-5%, some > 30% [R98]

- Next: **examples** to drive our intuitions, with a focus on **time** …

[1]http://download.101com.com/pub/tdwi/Files/DQReport.pdf
[2]https://www.marketsandmarkets.com/Market-Reports/data-quality-tools-market-22437870.html
[3]Divesh Srivastava Talk @ ADBIS'2021

Data Quality Dimensions

Example 1: Relational data

Misfielded Value

Representation

SIC Member Name	City	Zip	Phone
Dr. Samuel, Andres	Montpellier	34093	04 67 55 86 05
C. Pierkot	Montpellier	34093	04 67 55 86 05
Thérèse Libourel		Montpellier	99 99 99 99 99
Jean-François Desconnet	La Réunion	34093	04 67 55 86 05
Kristel Pierkot	Montpellier	34093	04 67 55 86 05
Isabelle Mougenot	Cayenne	34093	04 67 55 86 05
Dany Mitja	Montpellier	94093	04 67 55 86 05

Duplicates

Typos

Inconsistency

Incorrect Values

Missing Values

Many tuples are missing…

Data changes over time

- Everything changes over time (abstracting Heraclitus)
  - Attributes of an entity evolve over time

香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

Ladjel Bellatreche: 1999

Isae
Institut Supérieur de l'Aéronautique et de l'Espace
ENSMA

Ladjel Bellatreche: 2021

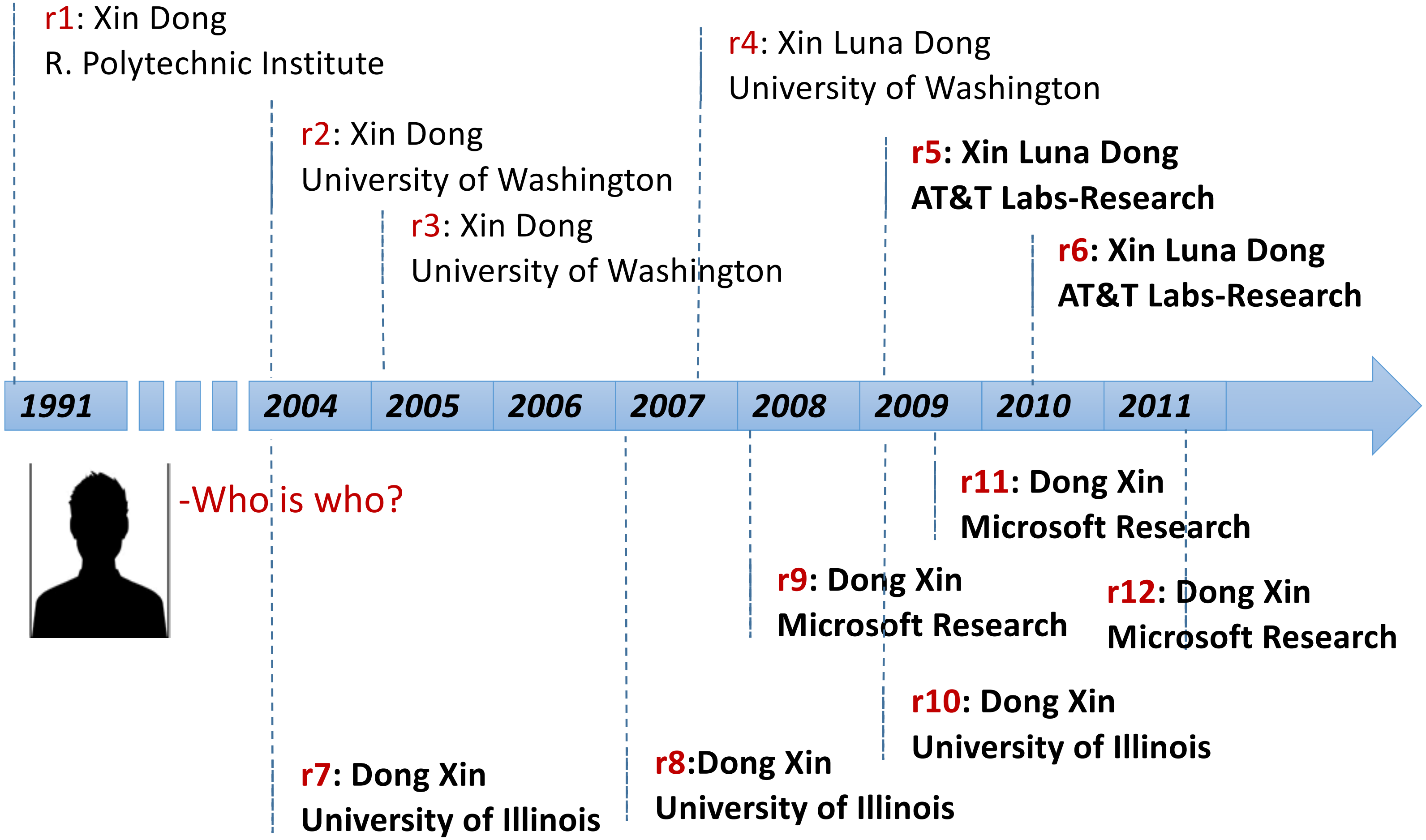
  - Different entities across time may have the same attributes

Adam Smith (1723-1790)
Philosopher & Economist

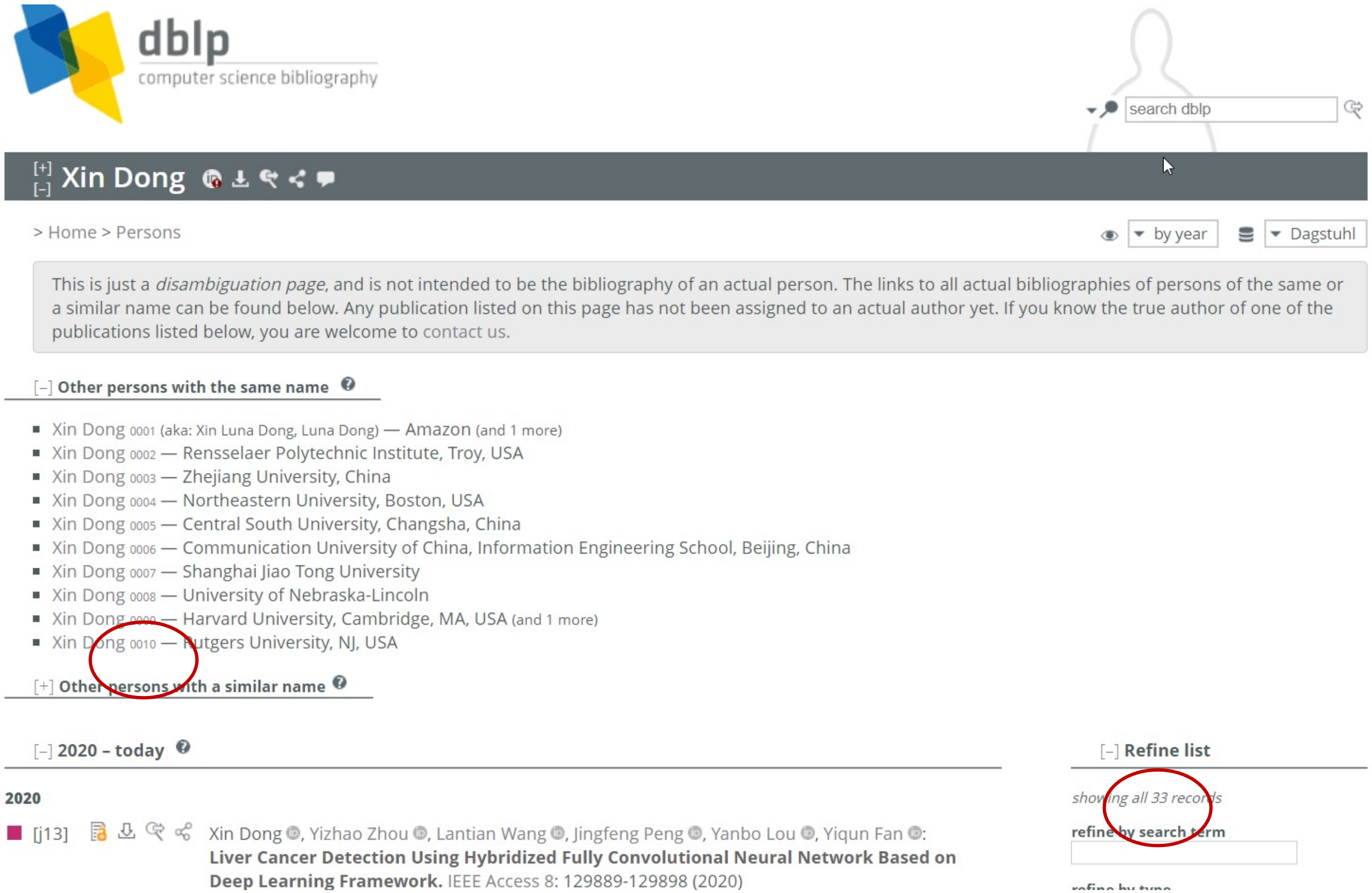
Adam Smith (1965-)
USA politician

Inspired from Divesh Srivastava Talk @ ADBIS'2021

Example: Changing Attributes Over Time



Example: Changing Attributes Over Time



Example: Changing Attributes Over Time

dblp.org/pid/35/7092.html

- Wei Wang 0266 — Iowa State University, Ames, IA, USA
- Wei Wang 0267 — Guangzhou Maritime University, GuanZhou, China
- Wei Wang 0268 — School of Resource and Environmental Sciences, Wuhan University, Wuhan, China
- Wei Wang 0269 — Beijing University of Chinese Medicine, Beijing, China
- Wei Wang 0270 — College of Information and Control Engineering, Nanjing University of Information Science And Technology, Nanjing, Jiangsu, China (and 1 more)
- Wei Wang 0271 — SER Group Ltd., Hong Kong (and 1 more)
- Wei Wang 0272 — Nanyang Technological University, Singapore
- Wei Wang 0273 — Beijing Institute of Technology, School of Information and Electronics, China

show less

[–] **Other persons with a similar name**

- Da-Wei Wang
- Liwei Wang (aka: Li-wei Wang, Li-Wei Wang) — *disambiguation page*
- Pengwei Wang (aka: PengWei Wang, Peng-Wei Wang) — *disambiguation page*
- Wei-Jen Wang
- Wei-Tsong Wang
- Wei-Yen Wang
- Jun-Wei Wang 0001 — University of Science and Technology Beijing, School of Automation and Electrical Engineering, China (and 1 more)
- Weifan Wang 0001 (aka: Wei-Fan Wang 0001) — Zhejiang Normal University, Department of Mathematics, Jinhua, China (and 2 more)
- Xingwei Wang 0001 (aka: Xing-Wei Wang 0001) — Northeastern University, College of Software, Shenyang, China
- Wang Wei — *disambiguation page*

show all similar names

[–] **2020 – today**

2021

[j560] Wei Wang, Lijuan Liu:
Complex L_p affine isoperimetric inequalities. Adv. Appl. Math. 122: 102108 (2021)

[j559] Wangli Hao, Ian Max Andolina, Wei Wang, Zhaoxiang Zhang:
Biologically inspired visual computing: the state of the art. Frontiers Comput. Sci. 15(1): 151304 (2021)

[j558] Junyang Chen, Zhiguo Gong, Wei Wang, Weiwen Liu:
HNS: Hierarchical negative sampling for network representation learning. Inf. Sci. 542: 343-356 (2021)

[–] **Refine list**

showing all 1326 records

refine by search term

refine by type

- ☑ Journal Articles (only)
- ☑ Conference and Workshop Papers (only)
- ☑ Parts in Books or Collections (only)
- ☑ Editorship (only)
- ☑ Informal Publications (only)

select all | deselect all

refine by coauthor

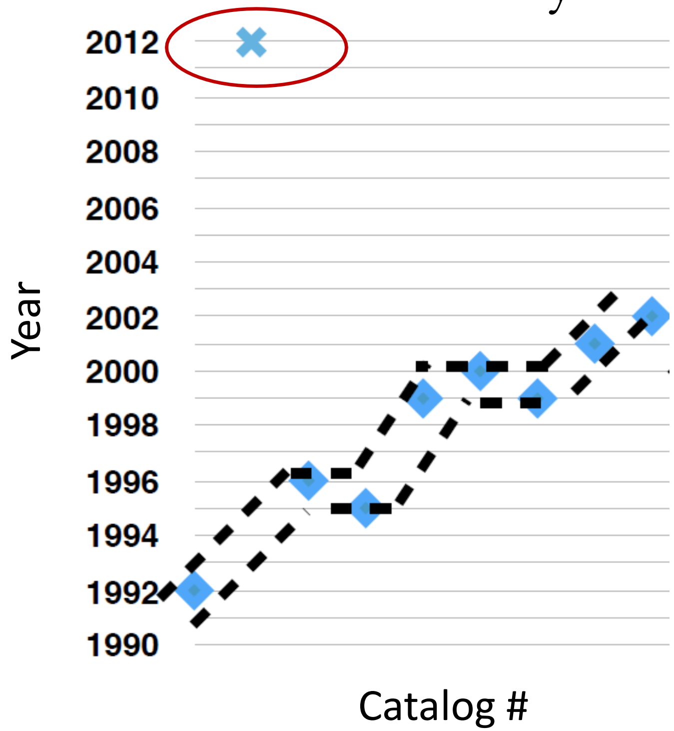
Example: Timestamps can be Erroneous

- Which record has an erroneous value of year?

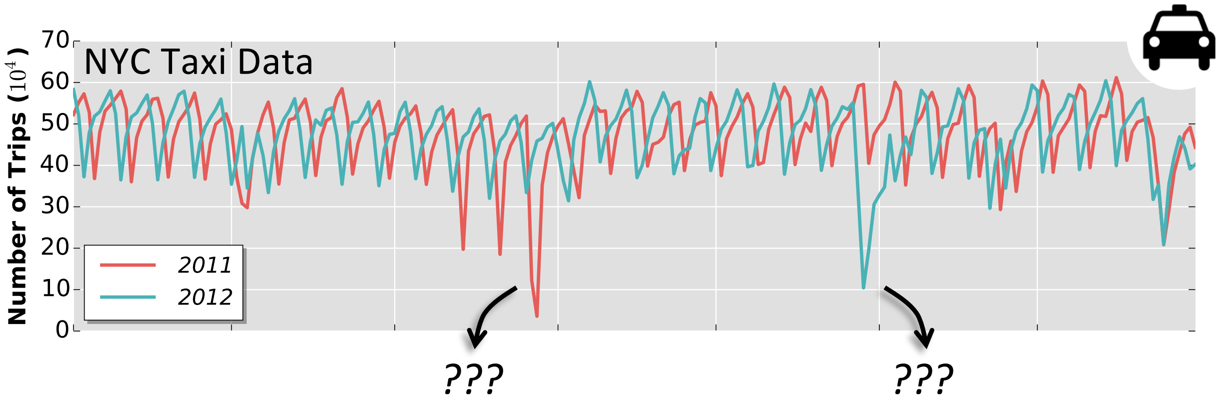
Tid	Release Title	Country	Year	Month	Catalog #
t1	Unplugged	Canada	1992	8	CDW45024
t2	Mirror Ball	Canada	2012	6	CDW45934
t3	Ether	Canada	1996	2	CDW46012
t4	Insomniac	Canada	1995	10	CDW46046
t5	Summerteeth	Canada	1999	3	CDW47282
t6	Sonic Jihad	Canada	2000	7	CDW47383
T7	Title of …	Canada	1999	7	CDW47388
t8	Reptile	Canada	2001	3	CDW47966
t9	Always …	Canada	2002	2	CDW48016

Example: Timestamps can be Erroneous

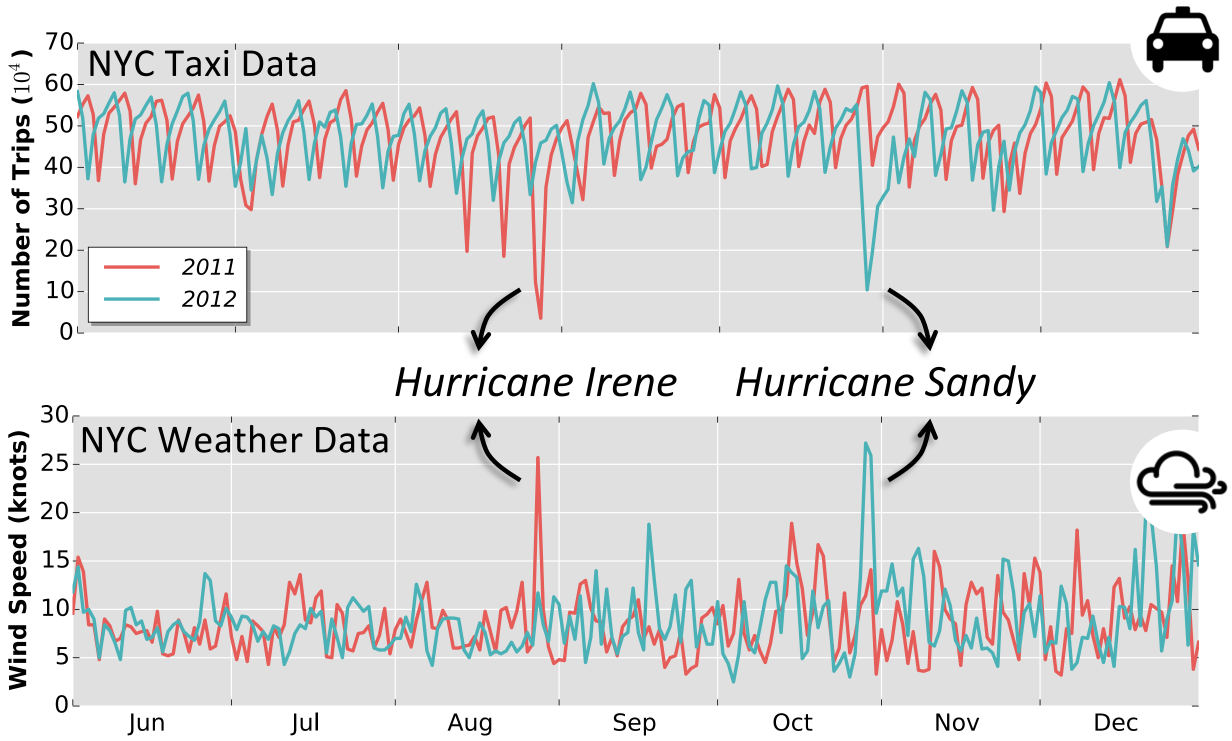
- Which record has an erroneous value of year?



Example: Time Series Anomalies



Example: Correlated Time Series Anomalies



Examples: Lessons Learned

- Big data over time (i.e., long data) can have **veracity** issues.
  - Even in domains where poor-quality data can have big impact.
  - Diversity of data quality issues involving time.

- Obtaining high-quality long data is **challenging**!
  - How soon can missing, erroneous and biased data be identified?
  - Which data can be used and when can it be used by data science?

Small Data Quality: How Was It Achieved?

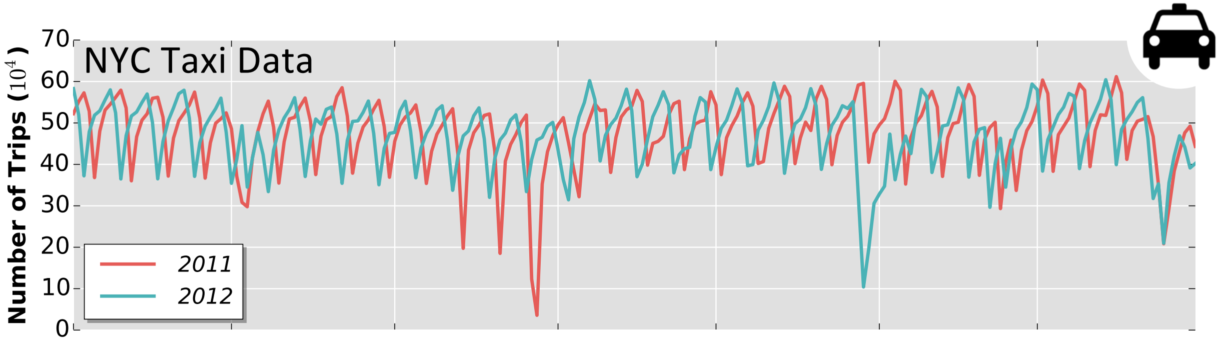
- Specify **all** domain knowledge as **integrity constraints** on data.

- Integrity constraint: formal specification that data must satisfy.
  - **Semantic** (SSN unique for person) vs **syntactic** (NNN-NN-NNNN).
  - **Qualitative** (Functional Dependencies) VS. **quantitative**

Small Data Quality: How Was It Achieved?

- Specify **all** domain knowledge as **integrity constraints** on data.

- Integrity constraint: formal specification that data must satisfy.
  - **Semantic** (SSN unique for person) vs **syntactic** (NNN-NN-NNNN).
  - **Qualitative** (FD on closing price) vs **quantitative** (# trips in given period).



Small Data Quality: How Was It Achieved?

- Specify all domain knowledge as **integrity constraints** on data.
  - **Reject updates** that do not preserve integrity constraints.
  - Works well when the domain is **very well understood** and **static**.

Big Data Quality: A Different Approach?

- Big data: integrity constraints cannot be specified a priori.
  - Data **variety, volume** → complete domain knowledge is infeasible.
  - Data **velocity, variability** → domain knowledge becomes obsolete.
  - Too much rejected data → “small” data. ☺

Big Data Quality: A Different Approach?

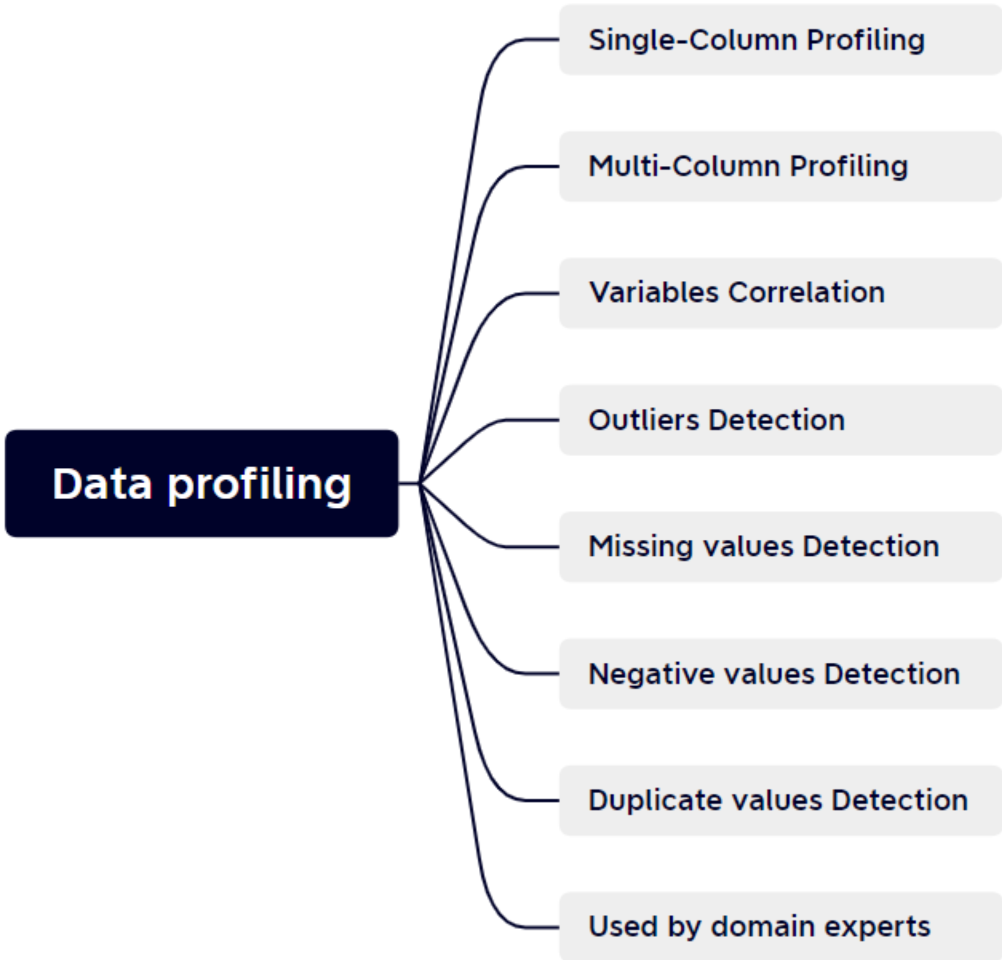
- Big data: integrity constraints cannot be specified a priori.
  - Data **variety, volume** → complete domain knowledge is infeasible.
  - Data **velocity, variability** → domain knowledge becomes obsolete.
  - Too much rejected data → "small" data. ☺

- Solution: let the data speak for itself.
  - Learn **integrity constraints / models** (semantics) from the data.
  - Identify **data glitches** as violations of the learned models.
  - Repair **data glitches and models** in a timely manner.

Data Profiling

- It is the process of statically examine and analyse the content in a data source.
- It collects information about data
- It consists
- of techniques used to analyse the data
- It helps to make a thorough assessment of data quality
- It assists the discovery of data anomalies
- It help in understanding: content, structures, relationships, ... about the data

Data Profiling



Quelques Outils de Data Profiling

- Trillium Enterprise Data Quality
- Datiris profiler
- Talend Data Profiler
- IBM Infosphere Information Analyzer
- SSIS Data Profiling Task
- Oracle Warehouse Builder

Démonstration

- Dévelopé dans le cadre du projet Européen Improvement
- Bâtiments à Zéro Energie

10-09-2018

01-02-2022

Des données météorologiques

Données horaires (1H)

- Maximum_humidity
- Average_humidity

Données Quotidiennes (1J)

- Solar_radiation
- Day_degree_hot
- Day_degree_cold
- Maximum_outdoor_temperature
- Average_outdoor_temperature
- Minimum_outdoor_temperature

Des données de consommation énergétique

- Gaz_consumption_boiler_1
- Gaz_consumption_boiler_2
- Gaz_consumption_boiler_3
- Electricity_consumption

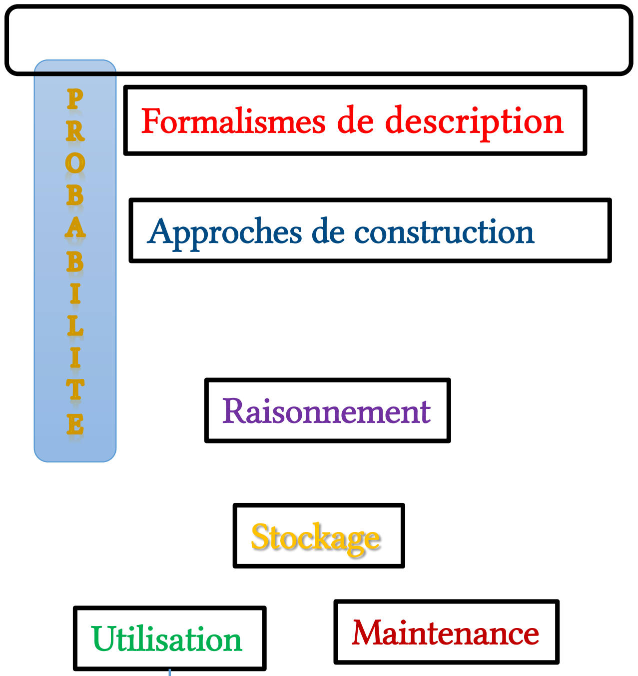
Données horaires (1H)

Lien de l'outil: https://filesender.renater.fr/?s=download&token=89ad2cc8-6f9b-42c7-afbe-5b2fed1557af

Le Monde des Ontologies

Paysage actuel

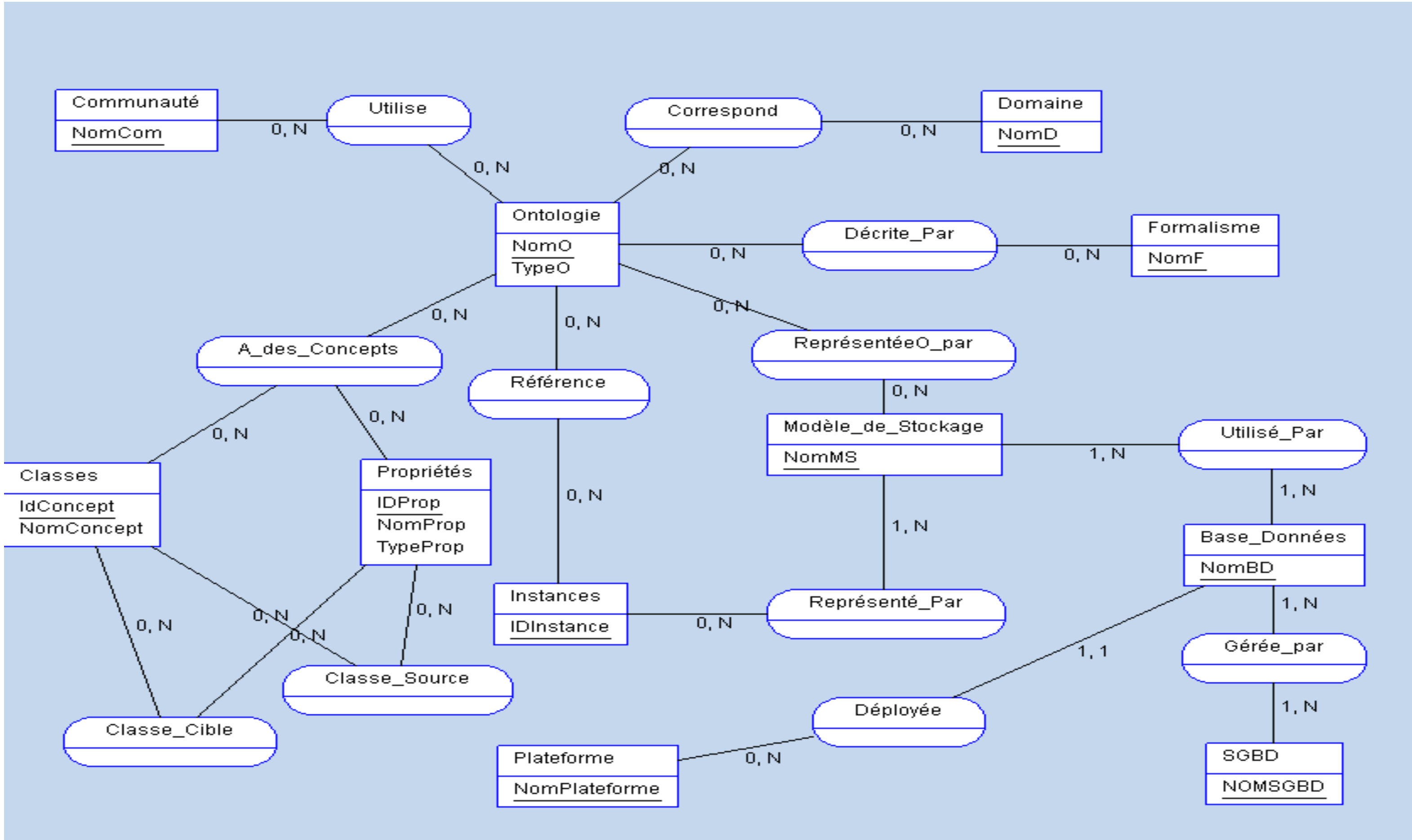
- → Workshop on Description Logics
- → Formal Ontologies meet Industry
- → Conference on Formal Ontology in Information Systems
- → Workshop on Modular Ontologies
- → Workshop on Ontologies and Conceptual Modeling
- → Web-Scale Knowledge Representation, Retrieval and Reasoning
- → Ontologies, DataBases, and Applications of Semantics (OBDASE)
- → Ontologies meet Advanced Information Systems
- → Workshop on Ontology Matching
- → Ontology-Driven Software Engineering
- → Ontology Dynamics
- → Workshop on Ontology Quality
- → Data Engineering meets the Semantic Web
- → JFO, IC
- → JoDS (Springer), Semantic Web Journal (IOS), Journal of Web Semantics (Elsevier)
- → etc.



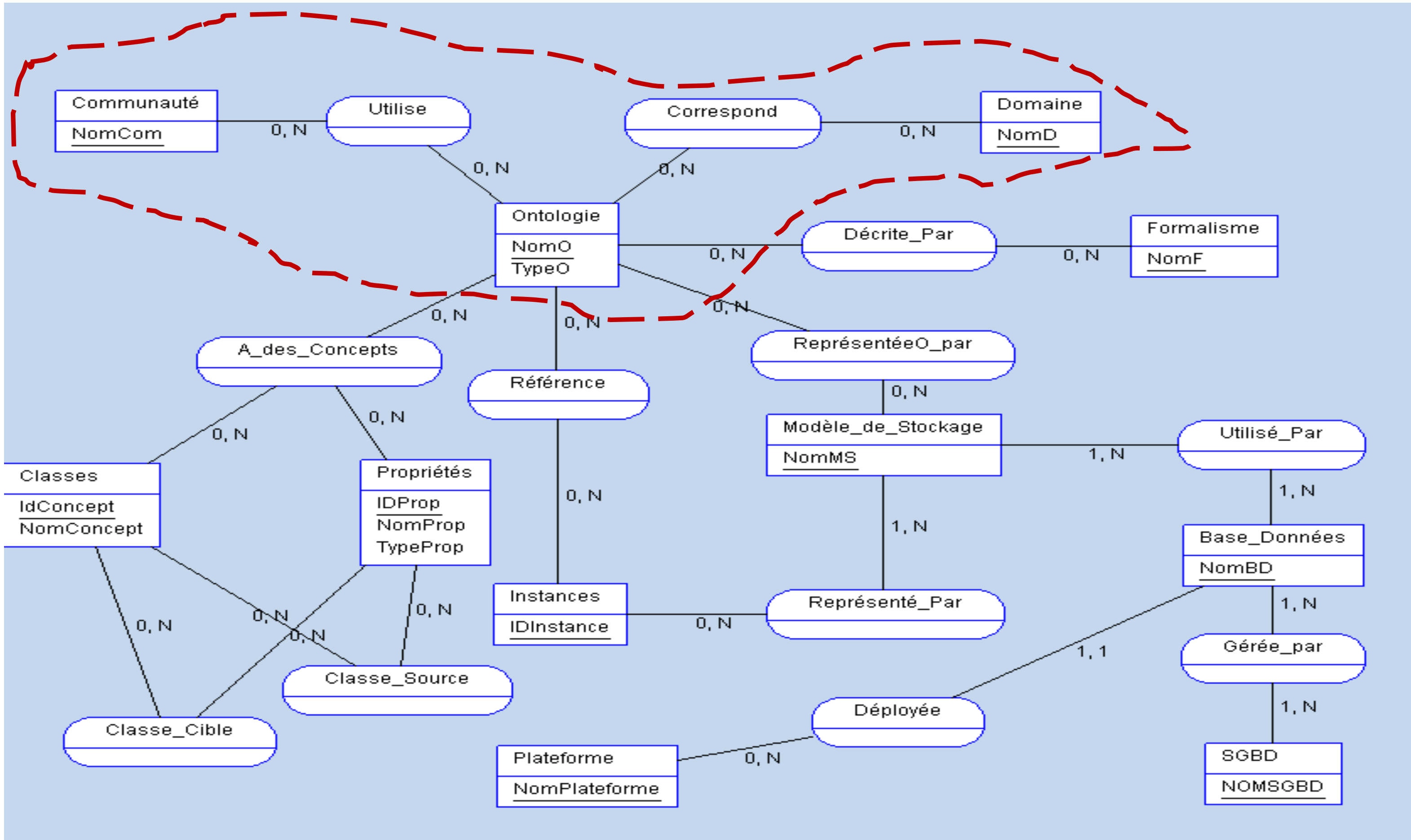
Agenda

- Monde des Ontologies
- Comment définir formellement une ontologie?
- Est-ce que les modèles de stockage traditionnels sont adaptés aux ontologies et aux instances ontologiques?
- Est-ce que l'architecture de stockage traditionnelle d'un SGBD peut être remise en question?
- Exploitation des Ontologies
  - ETL
  - Conception physique

Le Monde des Ontologies



1. Communautés



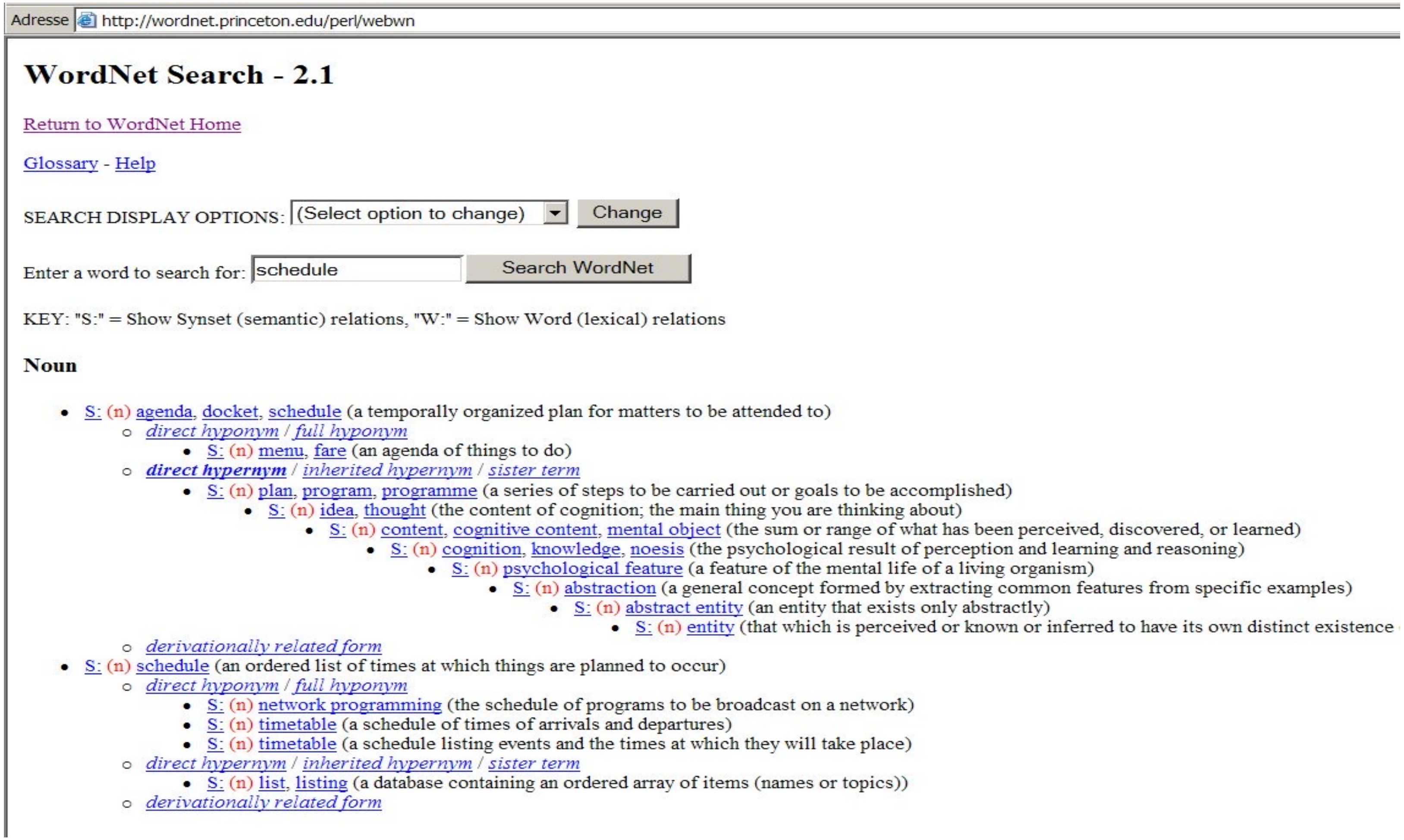
TAL

→Web : Aller de Ashford à Hastings

…Le système doit savoir que «*schedule*» et «*timetable*» sont synonymes …

Enjeu : Besoin de représenter formellement la similitude des mots

Exemple : WordNet



Adresse http://wordnet.princeton.edu/perl/webwn

WordNet Search - 2.1

Return to WordNet Home

Glossary - Help

SEARCH DISPLAY OPTIONS: (Select option to change) Change

Enter a word to search for: schedule Search WordNet

KEY: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Noun

- S: (n) agenda, docket, schedule (a temporally organized plan for matters to be attended to)
  - *direct hyponym* / *full hyponym*
    - S: (n) menu, fare (an agenda of things to do)
  - ***direct hypernym*** / *inherited hypernym* / *sister term*
    - S: (n) plan, program, programme (a series of steps to be carried out or goals to be accomplished)
      - S: (n) idea, thought (the content of cognition; the main thing you are thinking about)
        - S: (n) content, cognitive content, mental object (the sum or range of what has been perceived, discovered, or learned)
          - S: (n) cognition, knowledge, noesis (the psychological result of perception and learning and reasoning)
            - S: (n) psychological feature (a feature of the mental life of a living organism)
              - S: (n) abstraction (a general concept formed by extracting common features from specific examples)
                - S: (n) abstract entity (an entity that exists only abstractly)
                  - S: (n) entity (that which is perceived or known or inferred to have its own distinct existence
  - *derivationally related form*
- S: (n) schedule (an ordered list of times at which things are planned to occur)
  - *direct hyponym* / *full hyponym*
    - S: (n) network programming (the schedule of programs to be broadcast on a network)
    - S: (n) timetable (a schedule of times of arrivals and departures)
    - S: (n) timetable (a schedule listing events and the times at which they will take place)
  - *direct hypernym* / *inherited hypernym* / *sister term*
    - S: (n) list, listing (a database containing an ordered array of items (names or topics))
  - *derivationally related form*

Notion d’ontologie

- Qu'est-ce qu'une ontologie:
  - "An ontology is an explicit specification of a conceptualization" [Gruber 93]
- Une proposition de caractérisation [Pierra08] :

Une ontologie est une conceptualisation d'un domaine en termes de classes et de propriétés qui soit:

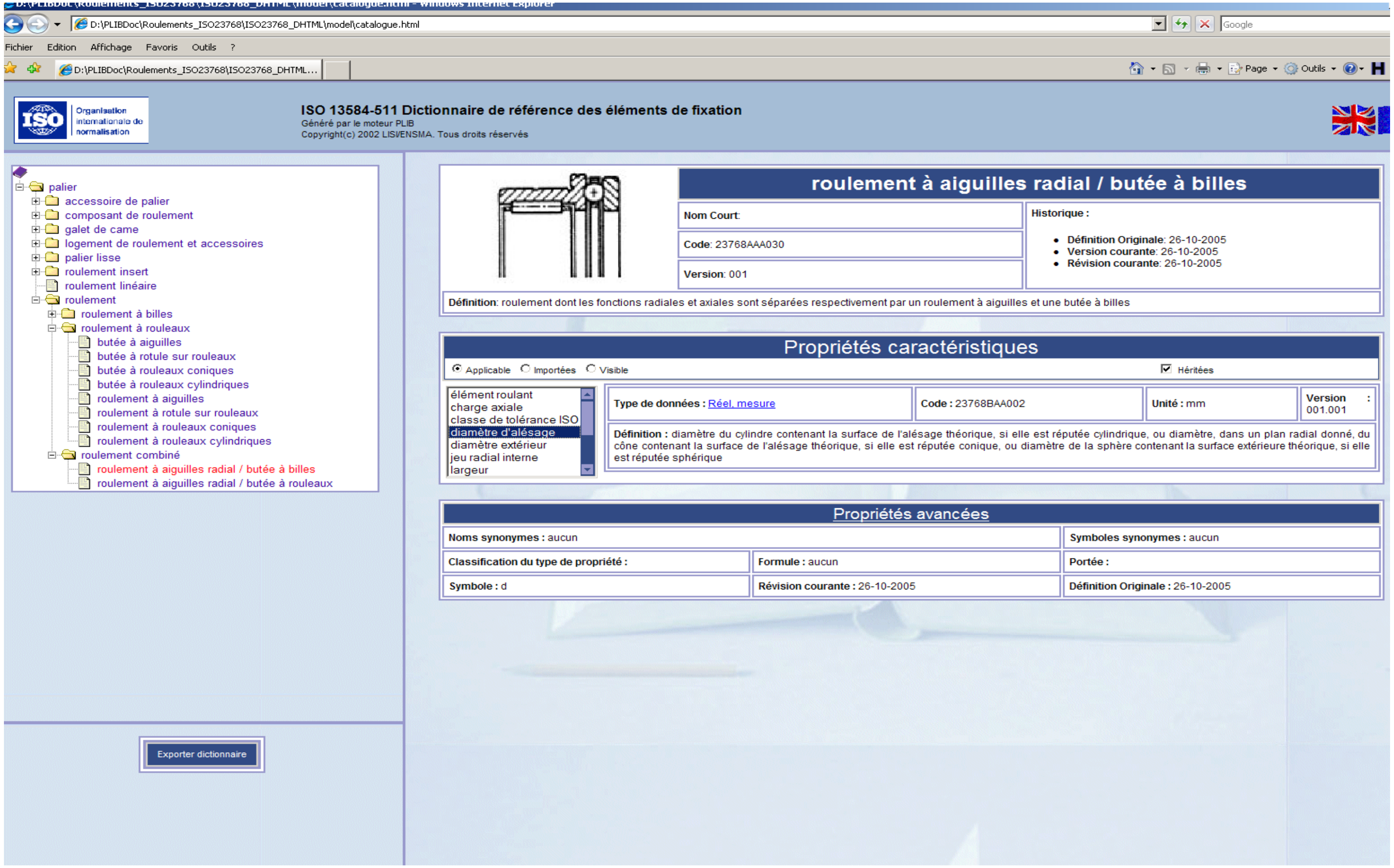
1. **Explicite;**
2. **Formelle;**
3. **Consensuelle;**
4. **Référençable par des identificateurs universels.**

- Faire l'objet de consensus…
  - chaque concept doit être associé au **contexte le plus large où il est défini pour une communauté**
  - Si ce n'est pas le cas, son **contexte doit être explicite**

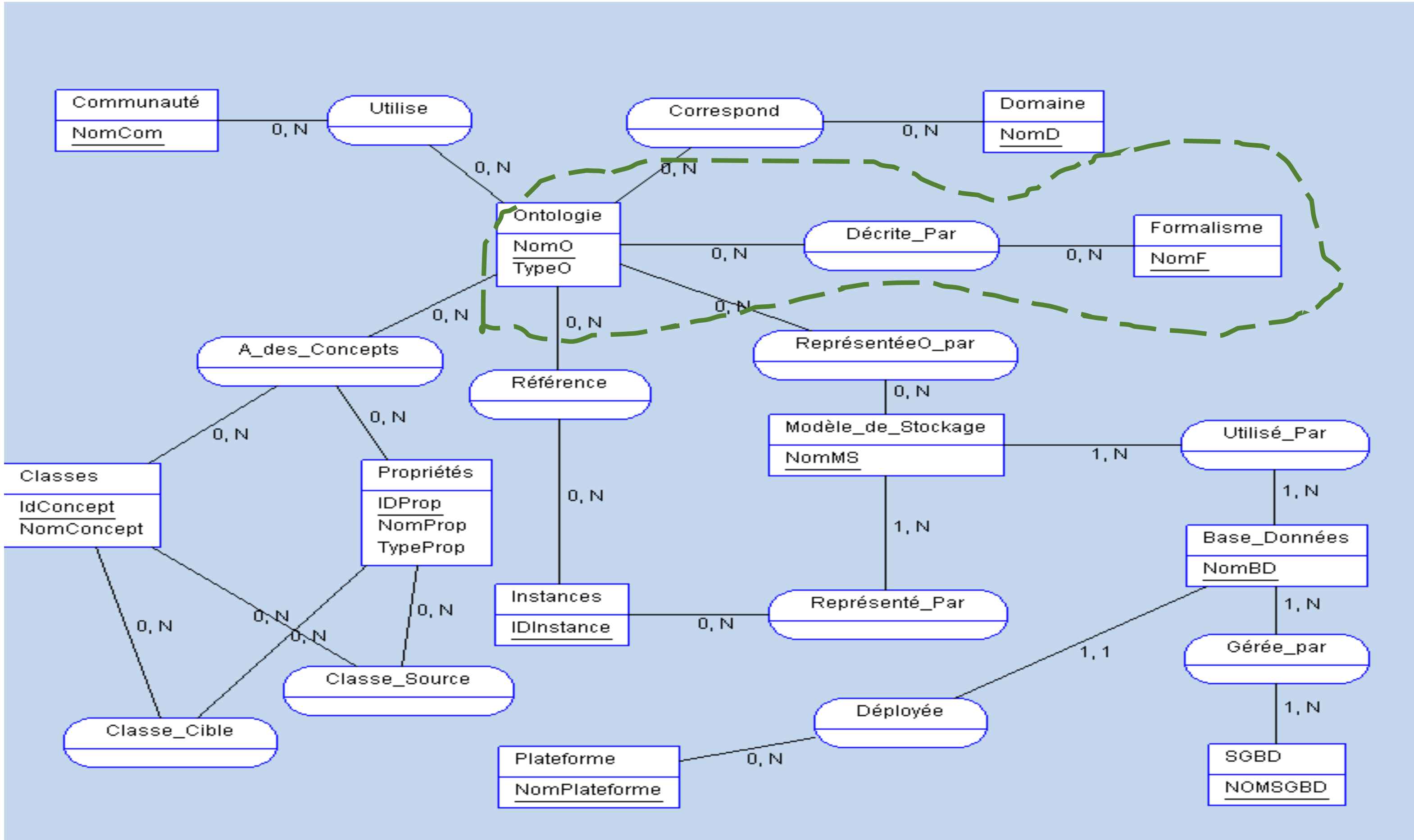
Modèles Conceptuels vs. Ontologies

	Prescription/Description	*Identification des concepts*	*Consensualité*	*Raisonnement*	*Explicitation du Contexte*
Modèles Conceptuels	Prescription	Non	Non	Non	Objectif du système cible
Ontologies Conceptuelles	Description	Oui	Oui	Oui	Explicitation des contextes

L'Ontologie PLIB



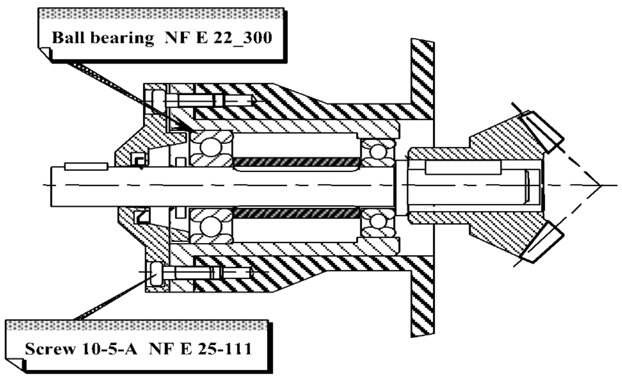
Formalismes d'Ontologies



Formalismes

- Existence de différents modèles d'ontologies : RDF-Schema, OWL, PLIB, etc.
- Capacités communes + capacités spécifiques

OWL	PLIB
Support de « synonymes » **(Origine : texte)**	**Précision** **(Origine : données structurées)**
Raisonnement logique	**Intégration automatique**



Objectifs de ISO 13584 series (PLIB)

- Modélisation
- Echange
- Intégration
- Publication

Catalogues de composants :

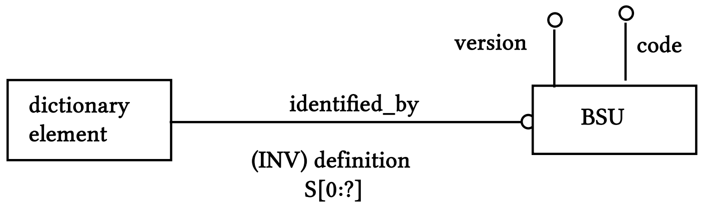
- de documents aux Données

•Un catalogue définit à la fois l'ontologie et les instances qui référencent cette ontologie

→ Besoin d'un modèle d'ontologie adapté pour les données d'ingénierie (unités de mesure, valeurs décimales, etc.)

Mécanisme d'identification (PLIB)

- BSU (Basic Semantic Unit)



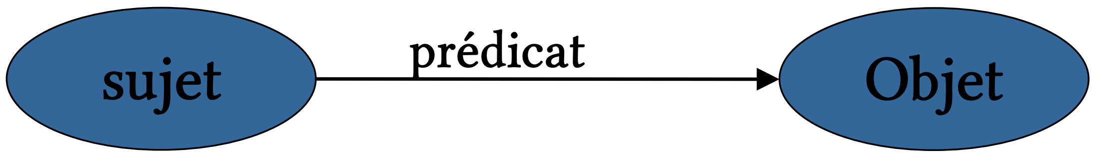
- Structure d'un identifiant
  - code_supplier_information: universal code of an organisation
  - ' - ' code_class: unique for a supplier
  - ' - ' code_property: unique for 1 class

0112/2///61630-4	CCD 124-002	AAF307-005

- BSU peut être défini, partagé, référencé, …

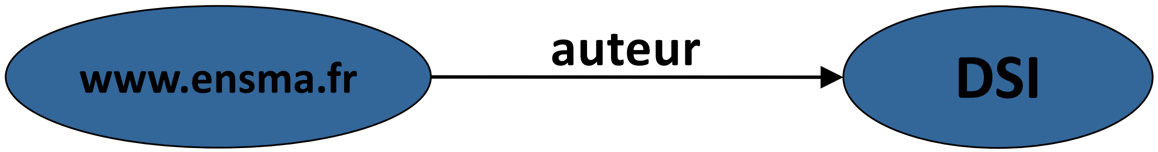
Présentation de RDF

- Description des savoirs à l'aide d'expressions
- Expression = triplet (*sujet, prédicat, objet*)



- Utilisation d'une syntaxe XML

Exemple RDF




```
(http://www.ensma.fr, schema:auteur, DSI)
```

```
<rdf:Description about="http://www.ensma.fr">
       <auteur> »DSI"</auteur>
</rdf:Description>
```

RDF Schéma

- Donne du vocabulaire à RDF
- Permet de définir des classes et types de propriétés spécifiques à une application ou un domaine.
- OWL est une extension de RDF

Exemple RDFS

```
<rdfs:Class rdf:ID='Cours'/>

<rdfs:Class rdf:ID='CoursInfo'>
        <rdfs:subClassOf rdf:resource='#Cours'/>
</rdfs:Class>

<rdfs:Class rdf:ID='CoursIA'>
       <rdfs:subClassOf rdf:resource='#CoursInfo'/>
</rdfs:Class>

<rdf:Property rdf:ID='enseignant'>
       <rdfs:domain rdf:resource='#Cours'/>
       <rdfs:range rdf:resource='#Personne'/>
</rdf:Property>
```

Web Ontology Language (OWL)

un dialecte XML basé sur une syntaxe RDF qui fournit les moyens pour définir des ontologies Web structurées

- est basé sur la recherche effectuée dans le domaine de la logique de description
- permet de décrire des ontologies, c'est-à-dire qu'il permet de définir des terminologies pour décrire des domaines concrets
- constitue une avancée importante dans la représentation et l'organisation des connaissances disponibles sur le Web
- est conçu comme une extension de Resource Description Framework (RDF) et RDF Schema (RDFS)

Avantages d'OWL

- Apporte une meilleure intégration, une évolution, un partage et une inférence plus facile des ontologies
- Ajoute les concepts de classes équivalentes, de propriété équivalente, d'égalité de deux ressources, de leurs différences, du contraire, de symétrie et de cardinalité
- Grâce à sa sémantique formelle basée sur une fondation logique largement étudiée, permet de définir des associations plus complexes des ressources ainsi que les propriétés de leurs classes respectives
- Est adéquat pour le Web sémantique, car il offre une syntaxe définie strictement, une sémantique définie strictement et selon le niveau peut permettre des raisonnements automatisés sur les inférences et conclusions des connaissances
- Le partage et l'échange dans ses formats est facile

La syntaxe de OWL, application à un exemple

Ontologie sur le vin et la nourriture

La structure des ontologies

- Les espaces de nom :

Exemple :« xmlns:vin =http://www.w3.org/TR/2004/REC-owl-guide-20040210/wine# »

- Les entêtes :

Exemple :« <owl:priorVersion rdf:resource="http://www.w3.org/TR/2003/PR-owl-guide-20031215/wine"/> »

Les éléments de base (1)

- Les classes simples

Exemples :

```
<owl:Class rdf:ID="Winery"/>

<owl:Class rdf:ID="ConsumableThing"/>
```

- L'héritage

Exemples :

```
<owl:Class rdf:ID="PotableLiquid">
        <rdfs:subClassOf rdf:resource="#ConsumableThing" />
</owl:Class>
```

Les éléments de base (2)

- Les individus

Exemples :

```
<owl:Thing rdf:ID="CentralCoastRegion" />

<owl:Thing rdf:about="#CentralCoastRegion">
<rdf:type rdf:resource="#Region"/>
</owl:Thing>
```

Ou

```
<Region rdf:ID="CentralCoastRegion" />
```

Les propriétés simples

- Caractérisent les instances des classes
- Définies par un domaine et une image

Exemples

<owl:ObjectProperty rdf:ID="madeFromGrape">

<rdfs:domain rdf:resource="#Wine"/>

<rdfs:range rdf:resource="#WineGrape"/> </owl:ObjectProperty>

Les caractéristiques de propriété

- Les différentes caractéristiques :

  TransitiveProperty, SymmetricProperty, FunctionalProperty, inverseOf, InverseFunctionalProperty.

Exemples

```
<owl:Class rdf:ID="VintageYear" />

<owl:ObjectProperty rdf:ID="hasVintageYear">
        <rdf:type rdf:resource="&owl;FunctionalProperty" />
        <rdfs:domain rdf:resource="#Vintage" />
        <rdfs:range  rdf:resource="#VintageYear" />
</owl:ObjectProperty>
```

Les restrictions de propriété

- Différentes restrictions :

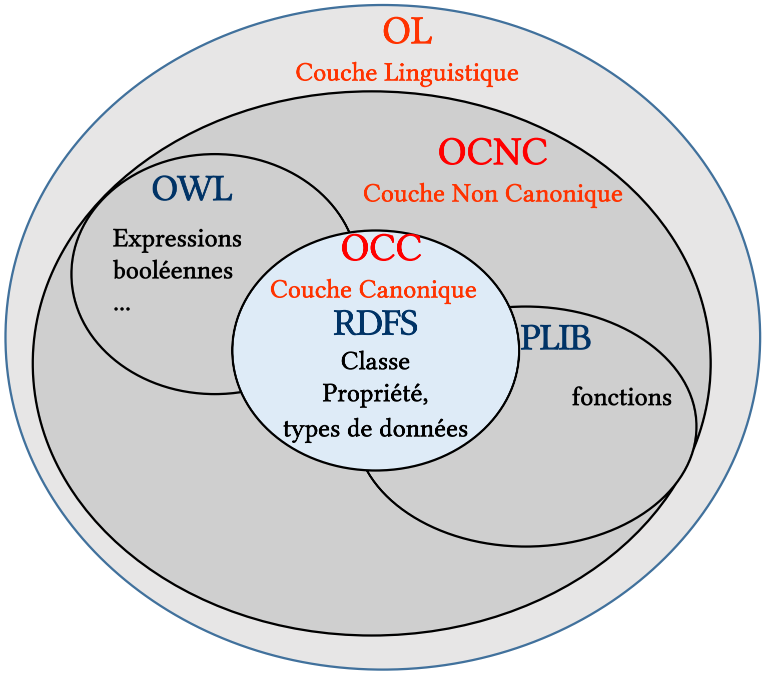
  allValuesFrom, someValuesFrom,hasValue, cardinalité

Exemples :

```
<owl:Class rdf:ID="Wine">
        <rdfs:subClassOf rdf:resource="&food;PotableLiquid" />                <rdfs:subClassOf>
                        <owl:Restriction>
                        <owl:onProperty rdf:resource="#hasMaker" />
                        <owl:someValuesFrom rdf:resource="#Winery" />
                <owl:Restriction>
        </rdfs:subClassOf>
</owl:Class>
```

Le Modèle en Oignon

- RDF-S (noyau) ≅ Classes, Propriétés, Types de données
- OWL ≅ noyau + opérateurs des Logiques de Description
- F-Logic ≅ noyau + règles logiques
- PLIB intensionnelles ≅ noyau + Fonctions de dérivation



- OCC = Constructeurs canoniques
  - Référence pour l'intégration sémantique
  - vocabulaire pour l'échange de données
- OCNC = Introduction des équivalences conceptuelles
  - Capacités d'inférence
  - Flexibilité pour l'intégration (modularité)
  - Représentation de schémas externes
- OL= Couche langagière d'accès
  - Requête en langue naturelle

Exemple:

les personnes du point de vue genre, âge, parenté