



From Data to Knowledge : the CODATA- O3D approach

Research data beyond 2020

J.P. Caliste, A.F. Cutting-Decelle, R. Vazille

-
- ▶ Introduction : CODATA and CODATA France
 - ▶ from Data ... to Knowledge
the transition process proposed by CODATA-O3D → ontologies
 - ▶ Some examples of ontologies
 - ▶ Usages and perspectives

Introduction

Who are we ?

CODATA is a non governmental organization

21 Countries

16 International Scientific Unions

2 Co-opted Scientific Organizations

13 Supporting Organizations (Industry, Government and Academia)

- ▶ Australia
- ▶ Brazil
- ▶ Canada
- ▶ Chinese Academy of Sciences
- ▶ Academy located in Taipei
- ▶ France *
- ▶ Georgia
- ▶ Germany *
- ▶ India
- ▶ Indonesia
- ▶ Ireland
- ▶ Israel
- ▶ Japan
- ▶ Korea
- ▶ Nigeria
- ▶ Poland
- ▶ Russia
- ▶ South Africa
- ▶ Ukraine
- ▶ United Kingdom
- ▶ USA

* Associate Member

Scientific Unions

- ▶ **International Astronomical Union (IAU)**
- ▶ **International Union of Pure and Applied Chemistry (IUPAC)**
- ▶ **International Union of Pure and Applied Physics (IUPAP)**
- ▶ **International Union of Biological Sciences (IUBS)**
- ▶ **International Geographical Union (IGU)**
- ▶ **International Union of Crystallography (IUCr)**
- ▶ **International Union of Biochemistry and Molecular Biology (IUBMB)**
- ▶ **International Union of Geological Sciences (IUGS)**
- ▶ **International Union of Psychological Science (IUPSyS)**
- ▶ **International Union of Pure and Applied Biophysics (IUPAB)**
- ▶ **International Union of Nutritional Sciences (IUNS)**
- ▶ **International Union of Pharmacology (IUPHAR)**
- ▶ **International Union of Immunological Societies (IUIS)**
- ▶ **International Union of Microbiological Societies (IUMS)**
- ▶ **International Union of Soil Scientists (IUSS)**
- ▶ **International Union of Geodesy and Geophysics (IUGG)**

Co-opted Organizations

- ▶ **International Council for Scientific and Technical Information**
- ▶ **World Federation for Culture Collections**

Supporting Organizations

- ▶ Defense Technical Information
- ▶ Springer-Verlag
- ▶ All Russian Institute of Scientific and Technical Information
- ▶ Chemical Abstracts Services
- ▶ Fachinformationszentrum Chemie, Gmbh
- ▶ Japan association for International Chemical Information
- ▶ National Library of Medicine
- ▶ Design institute for Physical Property Data (DIPPR) American Institute for Chemical Engineers
- ▶ Japan Society of Information and Knowledge
- ▶ Protein Data Bank
- ▶ Protein Information Resources
- ▶ Russian Research Centre for Standardization, Information and Certification of Materials (VNITS SMV)

History

- Established in 1966 by the International Council for Science ICSU

<http://www.icsu.org>

- Moved to France 1975
- Independent legal organization, working in close collaboration with ICSU and the ICSU family

Mission :

To strengthen international science for the benefit of society by promoting improved scientific and technical data management and use

How do we Achieve our Mission?

- ▶ Task Groups and Working Groups
- ▶ National Member Activities
- ▶ Participation in Important International Initiatives, such as : IPY (International Polar Year), eGY (Electronic Geophysical Year),
- ▶ International Conferences
- ▶ CODATA Electronic Journal
- ▶ Publications
- ▶ Workshops
- ▶ Studies and Reports
- ▶ Co-operation and liaison with other International Scientific Organizations (UNESCO, OECD, GEO, ...)

Task Groups

- ▶ Anthropometric Data and Engineering
- ▶ Data at Risk
- ▶ Data Citation Standards and Practices
- ▶ Data Sources for Sustainable Development in SADC Countries
- ▶ Earth and Space Science Data Interoperability
- ▶ Exchangeable Materials Data Representation to Support Scientific Research and Education
- ▶ Fundamental Physical Constants
- ▶ Global Information Commons for Science Initiative-EU Activities
- ▶ Global Roads Data Development
- ▶ Polar Year Data Policy and Management
- ▶ Preservation of and Access to Scientific and Technical Data in Developing Countries (PASTD)

CODATA Data Science Journal

- ▶ A peer-reviewed electronic journal
- ▶ Publishing papers on the management of data and databases in Science and Technology
- ▶ The scope of the journal includes descriptions of data systems, their publication on the internet, applications and legal issues

All of the Sciences are covered, including the Physical Sciences, Engineering, the Geosciences and the Biosciences, along with Agriculture and the Medical Science.

For more details see: <http://dsj.codataweb.org/>



Biannual International Conferences

❑ Last Conferences :

- ❑ Kiev 2008 (300 participants)
- ❑ South Africa 2010 (300 participants)
 - ❑ 22nd CODATA International Conference, 24-27th October 2010, Cape Town - South Africa, "Scientific Information for Society: Scientific Data and Sustainable Development"
- ❑ **Forthcoming Conference:** 2012 ,Taipei

➔ Joint paper written both by CODATA F and euroCRIS partners for the next Taipei Conference



CODATA 23
Taipei 2012

International Conference OPEN DATA & INFORMATION FOR A CHANGING PLANET

28 - 31 October 2012, Taipei



[Home](#)

[Call for Scientific Presentations](#)

[Practical Info](#)

[Registration](#)

[About CODATA](#)

[Call for Scientific
Presentations](#)

[Scientific Program
Committee](#)

[Program](#)

[Early Career Scientists](#)

[CODATA Prize](#)

[Sponsorship](#)

[Local Organization](#)

[Program Overview now online](#)



As President of CODATA I am delighted to welcome you to the official website of the 23rd International CODATA Conference ***“Open Data and Information for a Changing Planet”***.

Never before have scientific data and information played such a major role in addressing the problems and challenges facing our planet. The calls for freely available, publicly accessible data are gaining a stronger voice. The June 2011 editorial in *Nature* highlighted the value of open access to global nuclear-monitoring data, and the benefits of such data for a variety of applications, “in times of both calm and crisis.”

The theme ***“Open Data and Information for a Changing Planet”***

Why should a Country become a Member of CODATA?

- ▶ Develop collaboration between its national scientists and international scientists/experts : scientific networking
- ▶ Raise the profile of its national data science activities
- ▶ Attract foreign scientists and facilitate exchanges of scientists
- ▶ Take the lead on international data science activities and trends
- ▶ Improve dissemination of knowledge and know-how
Influence policy and decision makers both at national and international levels
- ▶ Economic benefits

CODATA France



CODATA's Associated Member

http://ticri.inpl-nancy.fr/ticri.fr/index.php/CODATA_France

Several actions on-going :

- ▶ Data Observatory for Sustainable Development (O3D initiative) related to a WIKI repository, promoting an International Data Kiosk for Sustainable Development
- ▶ From data to knowledge engineering : methods and tools
- ▶ Health data



All topics with a direct link to Sustainable Development are welcome in O3D.

This includes:

usual topics like climate change, demography, renewable energy, fossil resources, nuclear industry, nuclear weapons, waste, water, food, health, world agriculture, agro based chemistry, batteries, nano technologies, transport, housing...

but also human sciences, management, national and international law, norms and standards, geopolitics...

From data ... to knowledge

The transition process proposed by CODATA-03D :

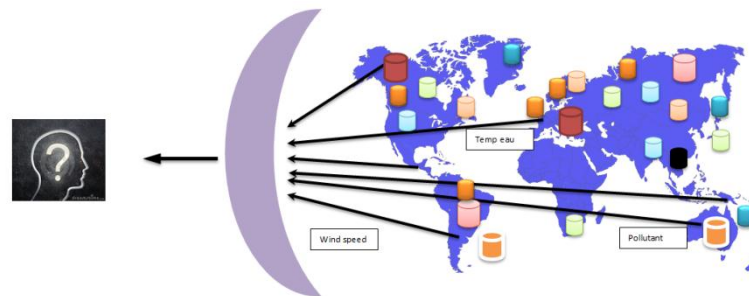
CODATA-O3D : *who are we ?*

- A « **community of interest** » in the domain of sustainable development sciences
- A **section of CODATA France**
- Our members can find in O3D :
 - An international network
 - An organisation through « clubs »
 - A methodology
 - A software support
 - Very friendly !
- **Objective : show the BRIDGE between knowledge and action, through the CODATA France approach**
 - applied to sustainable development data,
 - as an example of research data beyond 2020 ...



CODATA-O3D : working principles

- ▶ **To bridge knowledge and action**
- ▶ **To develop links** between research actors and actors of the socio-economic world
- ▶ **To suggest adapted methods and tools**
- ▶ **To encourage exchanges between members** through the network and supporting tools

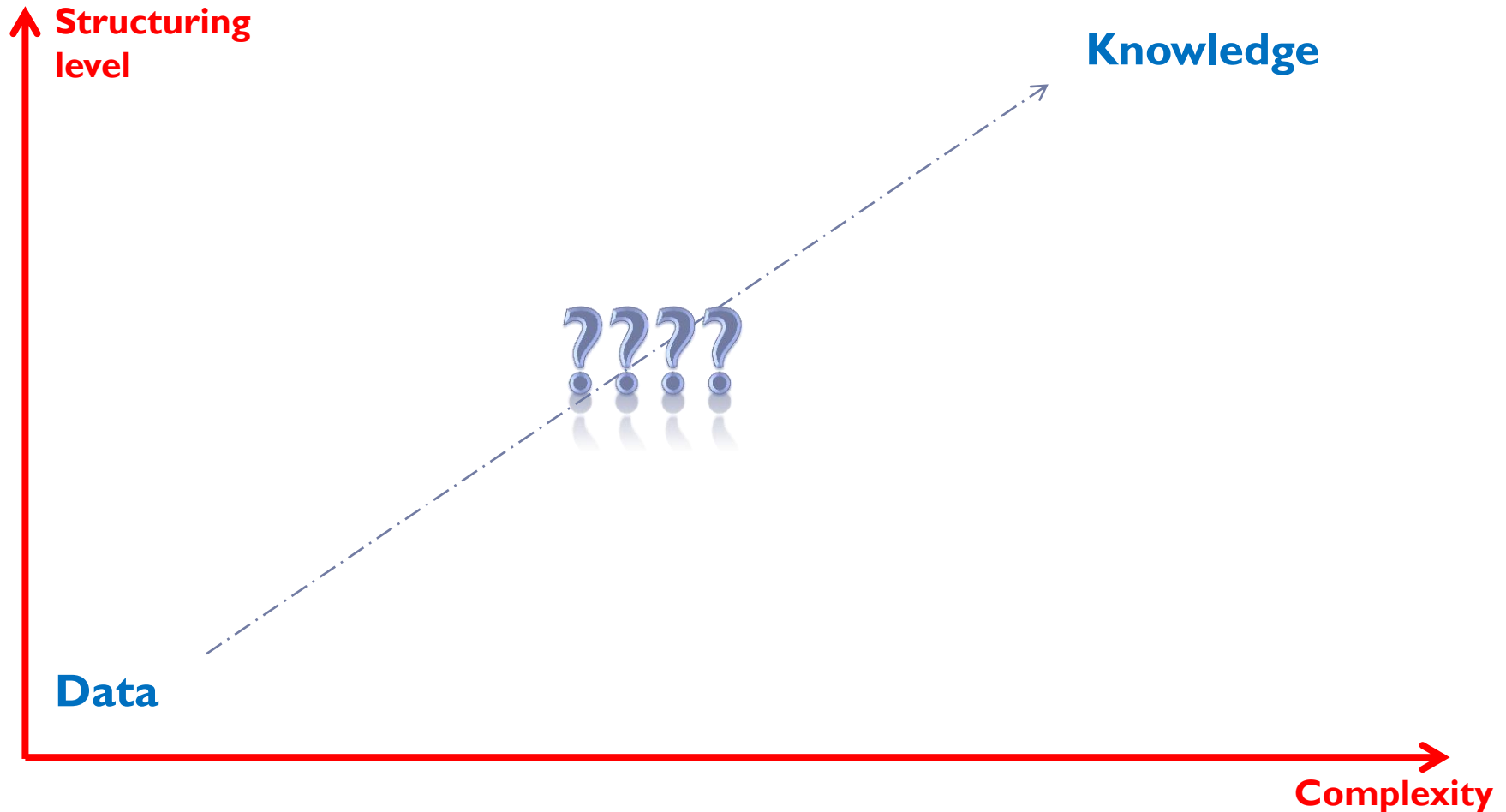


Challenge :

Data



knowledge



From data



LA BANQUE MONDIALE

Œuvrer pour un monde sans pauvreté

English | Español | Français | Русский | العربية | 中文



OK



À PROPOS

DONNÉES

RECHERCHE

APPRENTISSAGE

ACTUALITÉS

PROJETS

PUBLICATIONS

Données

PAYS | THÈMES

Par pays

Par thème

Indicateurs

Catalogue de données

Microdata

World Development Indicators

WDI



The primary World Bank collection of development indicators, compiled from officially-recognized international sources. It presents the most current and accurate global development data available, and includes national, regional and global estimates.
[more](#)

Available in the API

[DATABANK](#)

WDI & GDF (Excel)-ZIP (60 MB)

WDI & GDF (CSV)-ZIP (33 MB)

Global Development Finance

GDF



Focuses on financial flows, trends in external debt, and other major financial indicators for developing countries. Includes over 200 time series indicators from 1970 to 2010, for most reporting countries.
[more](#)

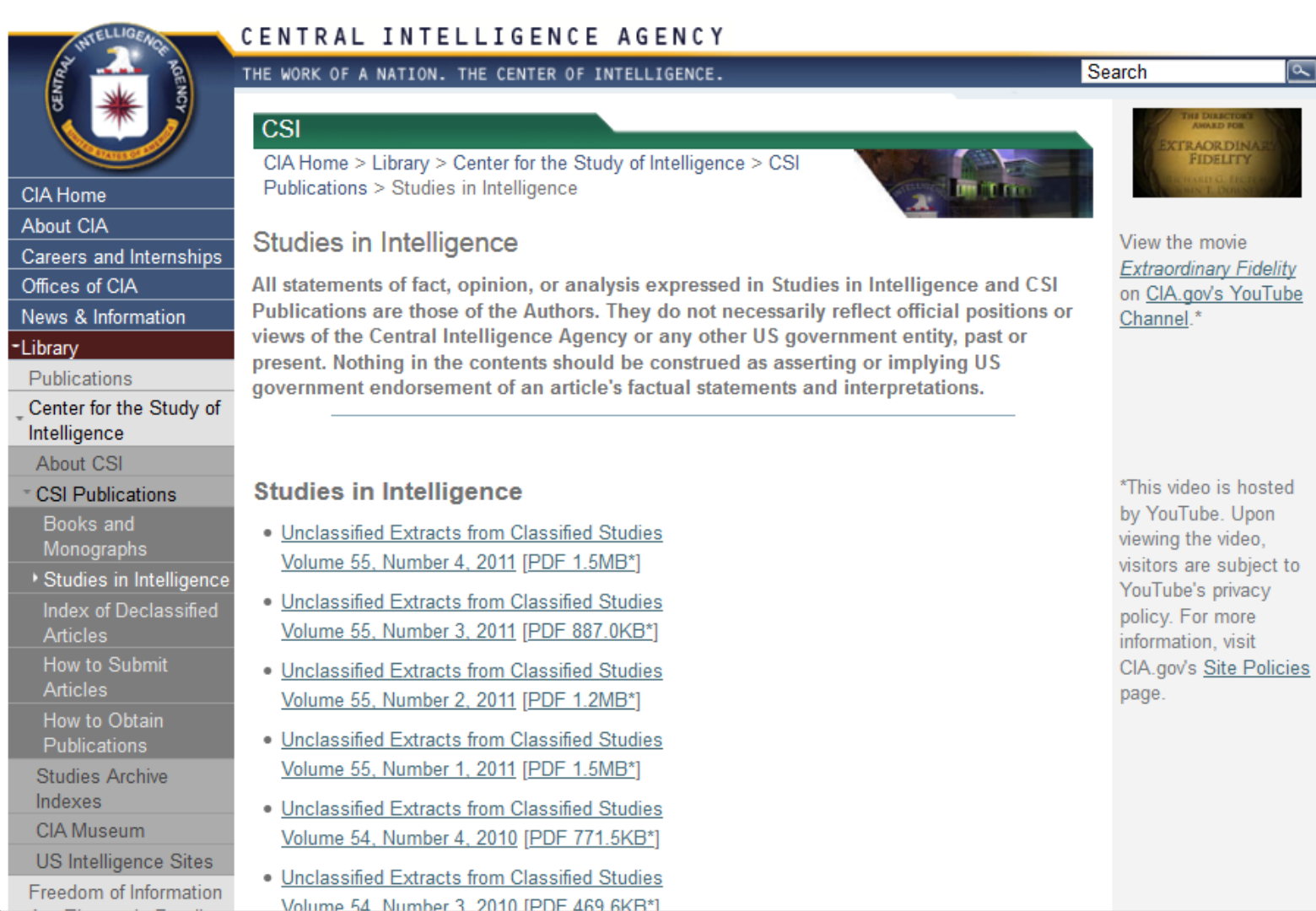
Available in the API

[DATABANK](#)

WDI & GDF (Excel) - ZIP (60.3 MB)

WDI & GDF (CSV) - ZIP (32.8 MB)

From data (cont'd)



CENTRAL INTELLIGENCE AGENCY
THE WORK OF A NATION. THE CENTER OF INTELLIGENCE.

CSI
CIA Home > Library > Center for the Study of Intelligence > CSI Publications > Studies in Intelligence

Studies in Intelligence
All statements of fact, opinion, or analysis expressed in Studies in Intelligence and CSI Publications are those of the Authors. They do not necessarily reflect official positions or views of the Central Intelligence Agency or any other US government entity, past or present. Nothing in the contents should be construed as asserting or implying US government endorsement of an article's factual statements and interpretations.

Studies in Intelligence

- [Unclassified Extracts from Classified Studies Volume 55, Number 4, 2011](#) [PDF 1.5MB*]
- [Unclassified Extracts from Classified Studies Volume 55, Number 3, 2011](#) [PDF 887.0KB*]
- [Unclassified Extracts from Classified Studies Volume 55, Number 2, 2011](#) [PDF 1.2MB*]
- [Unclassified Extracts from Classified Studies Volume 55, Number 1, 2011](#) [PDF 1.5MB*]
- [Unclassified Extracts from Classified Studies Volume 54, Number 4, 2010](#) [PDF 771.5KB*]
- [Unclassified Extracts from Classified Studies Volume 54, Number 3, 2010](#) [PDF 469.6KB*]

View the movie
[Extraordinary Fidelity](#)
on [CIA.gov's YouTube Channel](#).*

*This video is hosted by YouTube. Upon viewing the video, visitors are subject to YouTube's privacy policy. For more information, visit CIA.gov's [Site Policies](#) page.

Data :

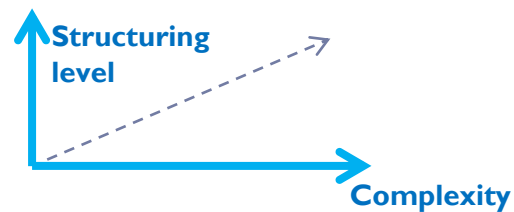
► Problems :

- **Time needed** for reading ...
- **Choice, relevance** criteria ?
- **Consistency / inconsistency** of the results ?
- Data definition **context** ?
- Difficulties to make **comparisons**
- Time validity : **long term sustainability**, ...
- Understanding of the **languages used in the documents**

..... *discouraging !!!*

From data to knowledge, an example :

- ▶ 40 = ?
 - ▶ Number of pages → **data**
 - ▶ 40 °C : patient's temperature → **information**
 - ▶ Patient with fever : suspected disease, possible contagion ?
 - ▶ Nature of the disease ? → **knowledge**
 - ▶ Fact mentioned on official monitoring networks
 - ▶ If epidemic → **organisation** as a consequence

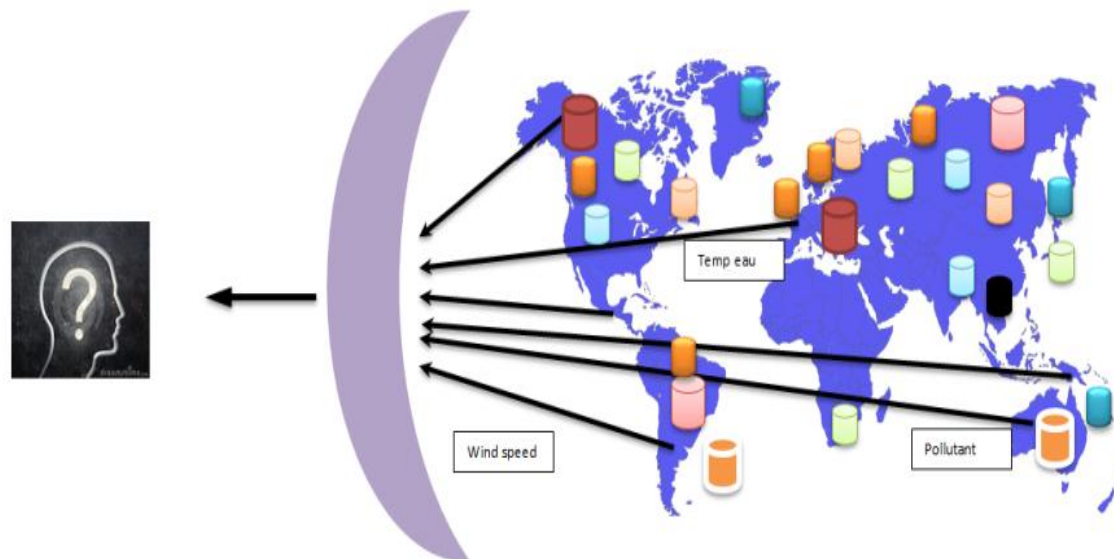


What can we do ?

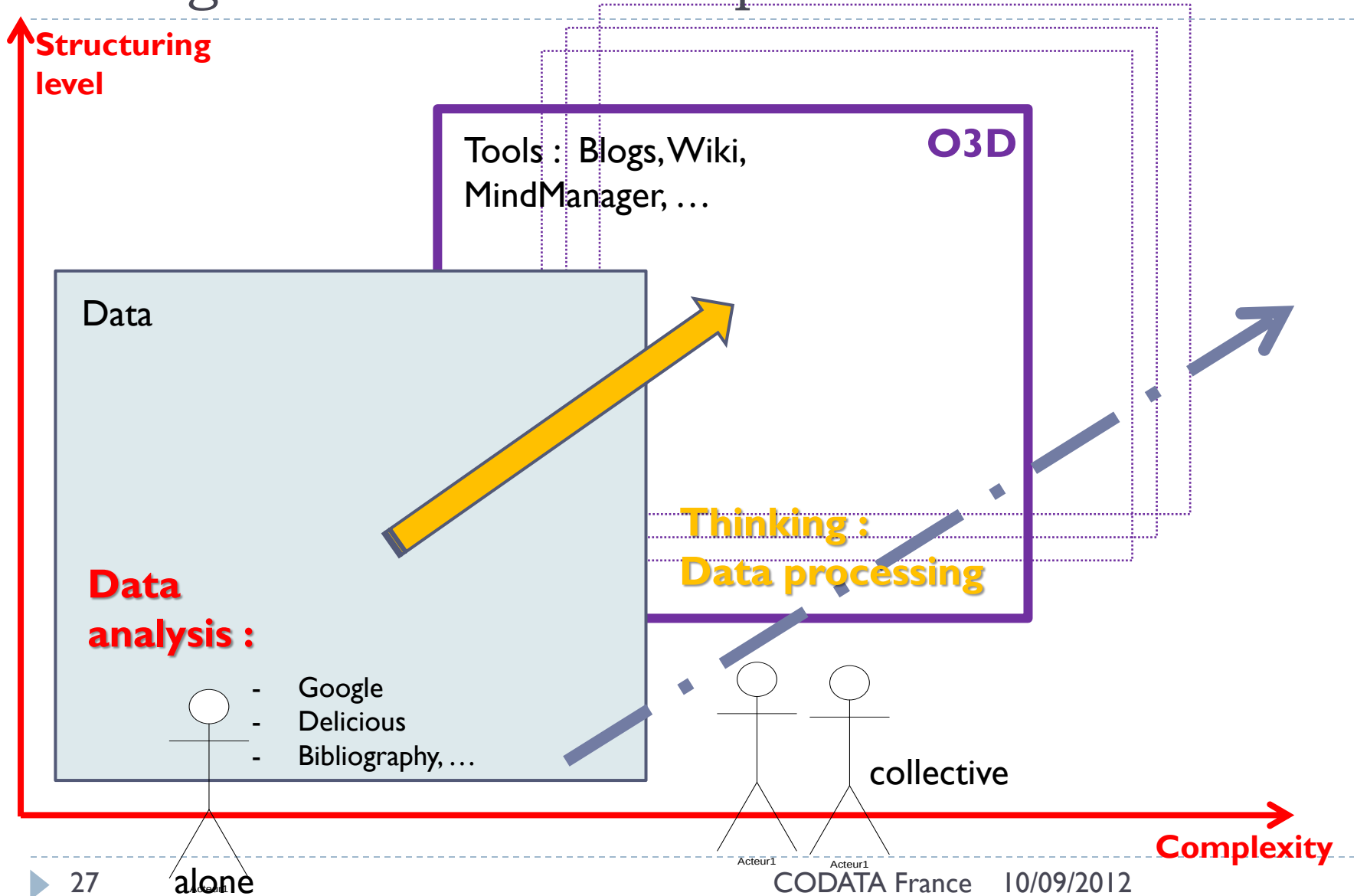
- ▶ Are there **tools** able to help to sort out those data ? Do those tools exist ?
- ▶ Are there **other persons** interested in the domain and with whom we could exchange, share information or points of view on those problems ?
- ▶ Is it possible to build up a **synthesis of the information** available ?
- ▶ How to build up **innovative ideas** on top of such an amount of information ?



The transition proposed by CODATA-O3D



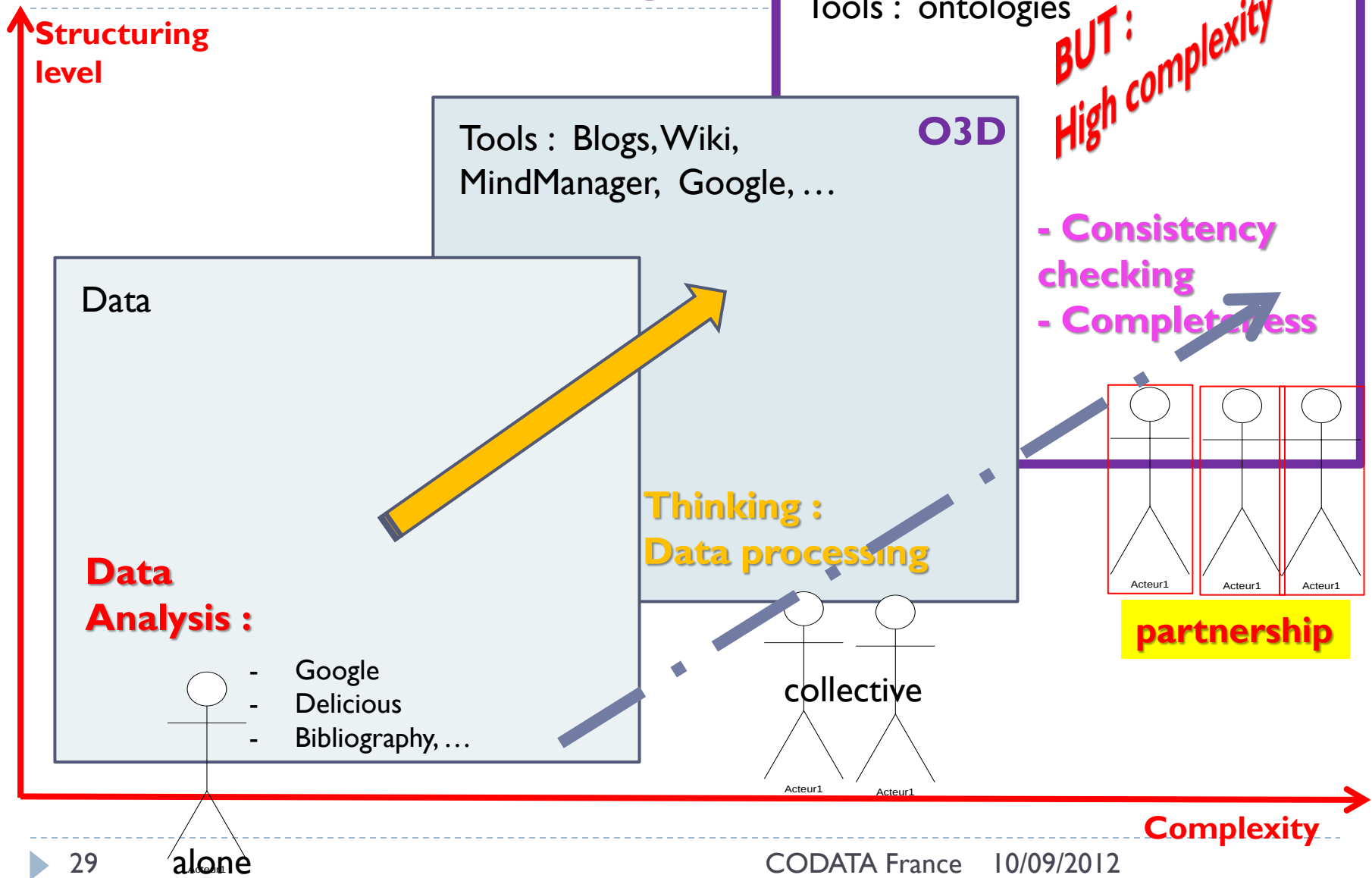
A stage of the transition process



But :

- ▶ How **to check the consistency** of my results ?
- ▶ How can I identify **redundancies, omissions** ?
- ▶ Is possible to **make transposition of methods and tools** coming from other domains, or other industrial sectors ?
- ▶ Is it possible to **deduce new information**, new rules, new methods from existing data and information (**innovation**) ?
- ▶ How **to understand, then adopt the technical language** of other professionals ?
- ▶ How to improve **the interoperability** ?
- ▶ How **to adapt to globalization** ?

... towards knowledge

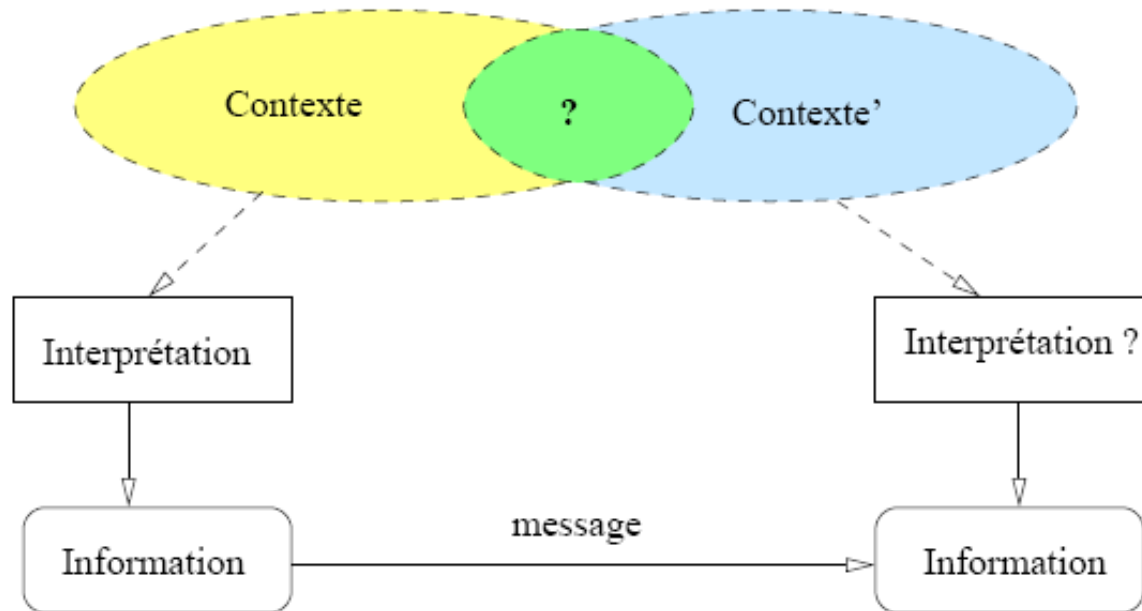


... towards knowledge

About ontologies

- ▶ Origin of the term : philosophy
 - ▶ Ontology : study of being or existence
 - ▶ « *Ontos* » (being) and « *logos* » (science, language)
- ▶ Objective : attempt to answer questions :
 - ▶ *How to specify what is ?*
 - ▶ *What does being mean ?*

Benefit of ontologies : context sharing



Why ?

- ▶ To agree on the meaning of terms used within an organisation, a community, a profession (example : terms used in steelwork construction and in timber construction)
- ▶ To enable a *mutual understanding* between people and software tools

How to enlarge the shared context ?

Two ways :

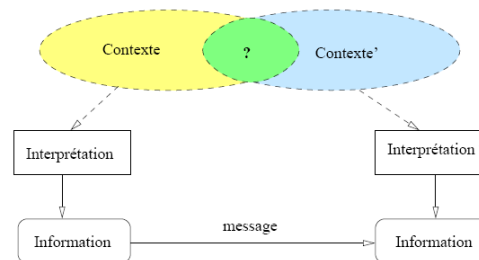
► **Communication :**

- Informal : discussion, email
- formal : protocols, training

► **Common semantic references :**

- Informal : textual description, dictionary, ...
- Formal : meta-data, ontologies, thesaurus, ...

Ontology = formal model used to represent shared knowledge



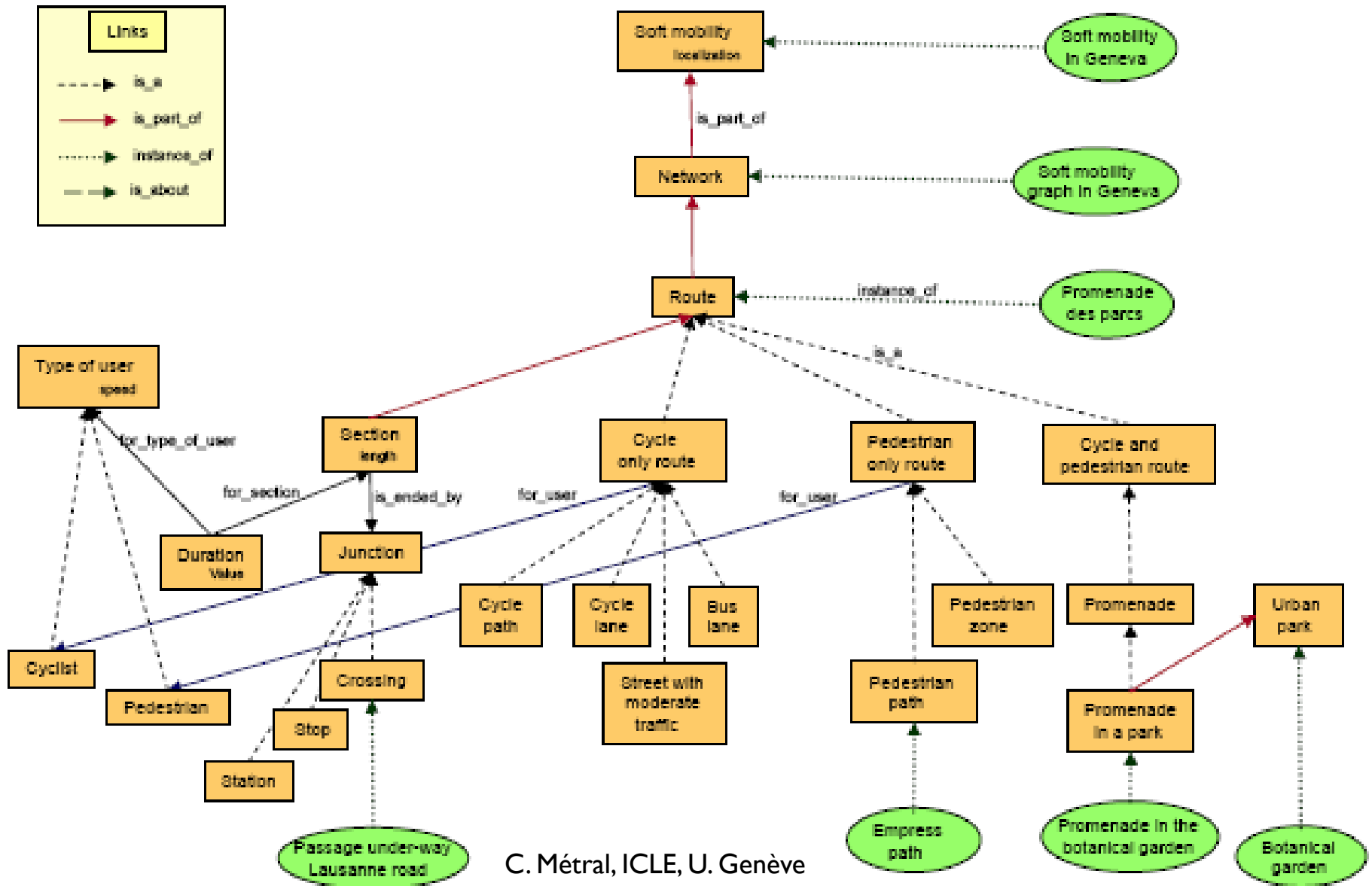
What is an ontology ?

- ▶ **An explicit description of a domain :**
 - ▶ Concepts, properties and concept attributes
 - ▶ Constraints on properties and attributes
 - ▶ Individuals (*often, but not always*)
- ▶ **An ontology defines :**
 - ▶ A common vocabulary
 - ▶ A shared understanding (grammar) : rules, axioms
- ▶ **Use : ambiguity solving of similar terms used with different meanings**

Marie-Aude Aufaure

Some examples of ontologies

Ontology of soft mobility



C. Métral, ICLE, U. Genève

An example : OTN (ontology of transport network) – REWERSE Project

OTN (<http://www.pms.ifl.uni-muenchen.de/OTN>) - [C:\Users\afcd\Desktop\cours-websem\docs-af-cours\cours\OTN.owl]

File Edit View Reasoner Tools Refactor Window Help

OTN (<http://www.pms.ifl.uni-muenchen.de/OTN>)

Active Ontology Entities Classes Object Properties Data Properties Individuals OWLViz OntoGraf

Class hierarchy: Thing

- Thing
 - Composite_Attributes
 - Accident
 - Average_Vehicle_Speed
 - Blocked_Passage
 - Construction_Status
 - Direction_of_Traffic_Flow
 - Divider
 - Emergency_Vehicle_Lane
 - Form_Of_Way
 - House_Number_Range
 - Location_Reference
 - Maximum_Height_Allowed
 - Maximum_Length_Allowed
 - Maximum_Total_Weight_Allowed
 - Maximum_Width_Allowed
 - Number_of_Lanes
 - Road_Surface
 - Special_Restriction
 - Speed_Restriction
 - Timetable
 - Toll
 - Traffic_Jam_Sensitivity
 - Travel_Time
 - Validity_Period
 - Feature
 - Land_Cover_And_Use
 - Meteorology
 - Public_Transport

OntoGraf:

Search: [] contains [] Search [] Clear []

To use the reasoner click Reasoner->Start reasoner ☒ Show Inferences

Creation of an ontology

Methodology :

- ▶ 1) **Specify** : identification of the domain and the aim of the ontology
→ important : for possible re-use
- ▶ 2) **Acquire knowledge** : expertise from experts, text mining, meta-data from databases, ...
→ list of competency questions
- ▶ 3) **Conceptualise** : identification of key-concepts of the domain, properties and relationships ; identification of terms from natural language, structuring of the knowledge of the domain
- ▶ 4) **Integration** : use or specialise an existing ontology
- ▶ 5) **Coding** : use of a formal language : CL, Ontolingua, Description logics, Conceptual Graph, RDF Schema, DAML OIL
- ▶ 6) **Document** : formal or informal definitions, complete definitions to specify the meaning of the terms of the ontology, examples of use
- ▶ 7) **Evaluation** : suitability of the ontology to the application targeted, pragmatic evaluation
→ **criteria** : consistency, completeness, concision (avoid redundancy, with a high level of granularity, ...)

Usages and perspectives

Why would we use ontologies ?

▶ **Scenario 1: Ontology as a specification**

- ▶ To facilitate the development of systems and their maintenance
- ▶ To enable knowledge re-use for wide number of applications

▶ **Scenario 2 : Ontology for data sharing**

- ▶ To reduce development costs and ensure interoperability
- ▶ To enable the access to heterogeneous data through a unique format

▶ **Scenario 3 : Ontology for information retrieval**

- ▶ To provide a generic access method to specific resources
- ▶ To provide annotation capabilities to information in documents
- ▶ To help users formulate queries
- ▶ To provide reasoning and deducing capabilities in order to check validity, correctness, consistency of the information, and also to complete missing information

Marie-Aude Aufaure

Usage : semantic integration

- ➔ Gives the feeling to use an homogeneous and centralised system
- ➔ However the real system (on the right) can be **VERY complex** !



Example of a semantic web application, developed by C. Frison et J. Pharès, EC de Lille

Structuring of **data available on the web site of the station de ski de Saint-Gervais (French Alps)**,

Aim of the application : to suggest different possible itineraries once the user has selected : the period of the stay, the type of activities, the level of skiing practice, ...

Form to be completed by the user on the web :

Activité pratiquée

- ☐ Ski de fond
- ☒ Ski de piste
- ☐ ballade à pied / raquette

Niveau en ski de piste

- ☐ Débutant - pistes vertes, bleues pour le ski de piste
- ☒ Intermédiaire - pistes vertes, bleues, rouges pour le ski de piste
- ☐ Confirmé - pistes vertes, bleues, rouges et noires pour le ski de piste

Choix de restauration souhaités

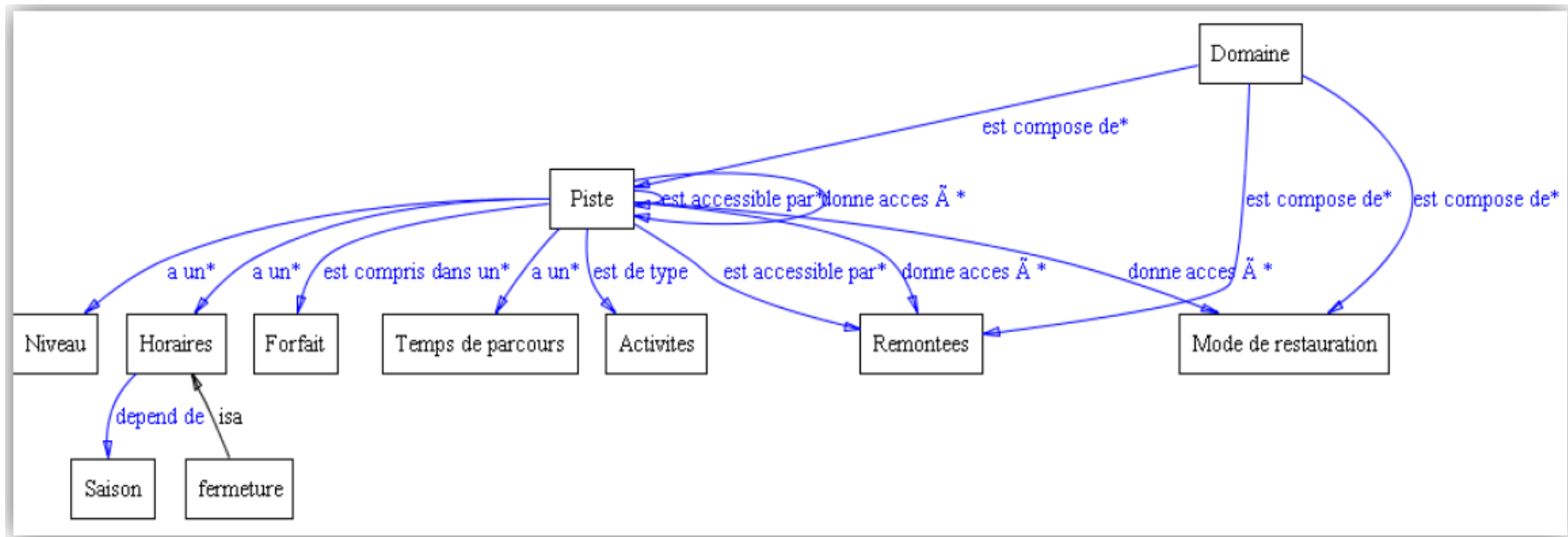
- ☐ Restaurant
- ☒ Snack
- ☒ Pique-nique sur les pistes

Centre d'interet

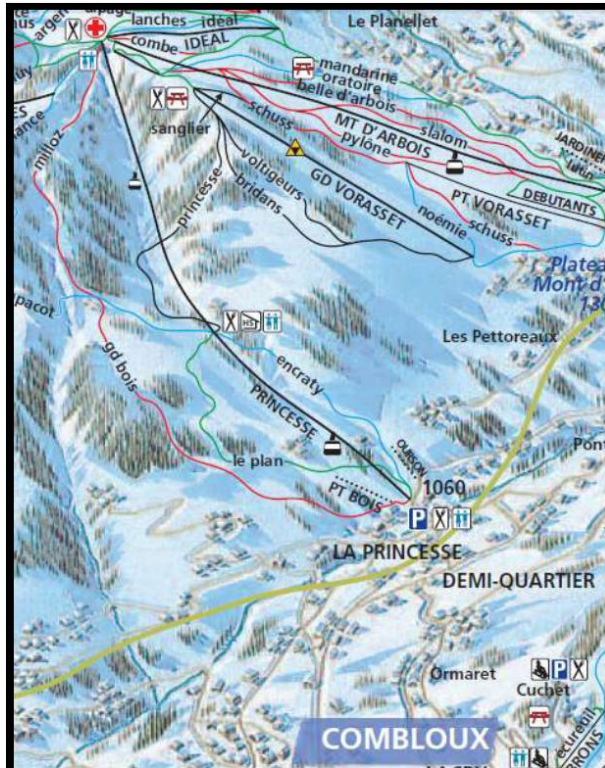
Nombre de personnes:

Example of a semantic web application, developed by C. Frison et J. Pharès, EC de Lille (2)

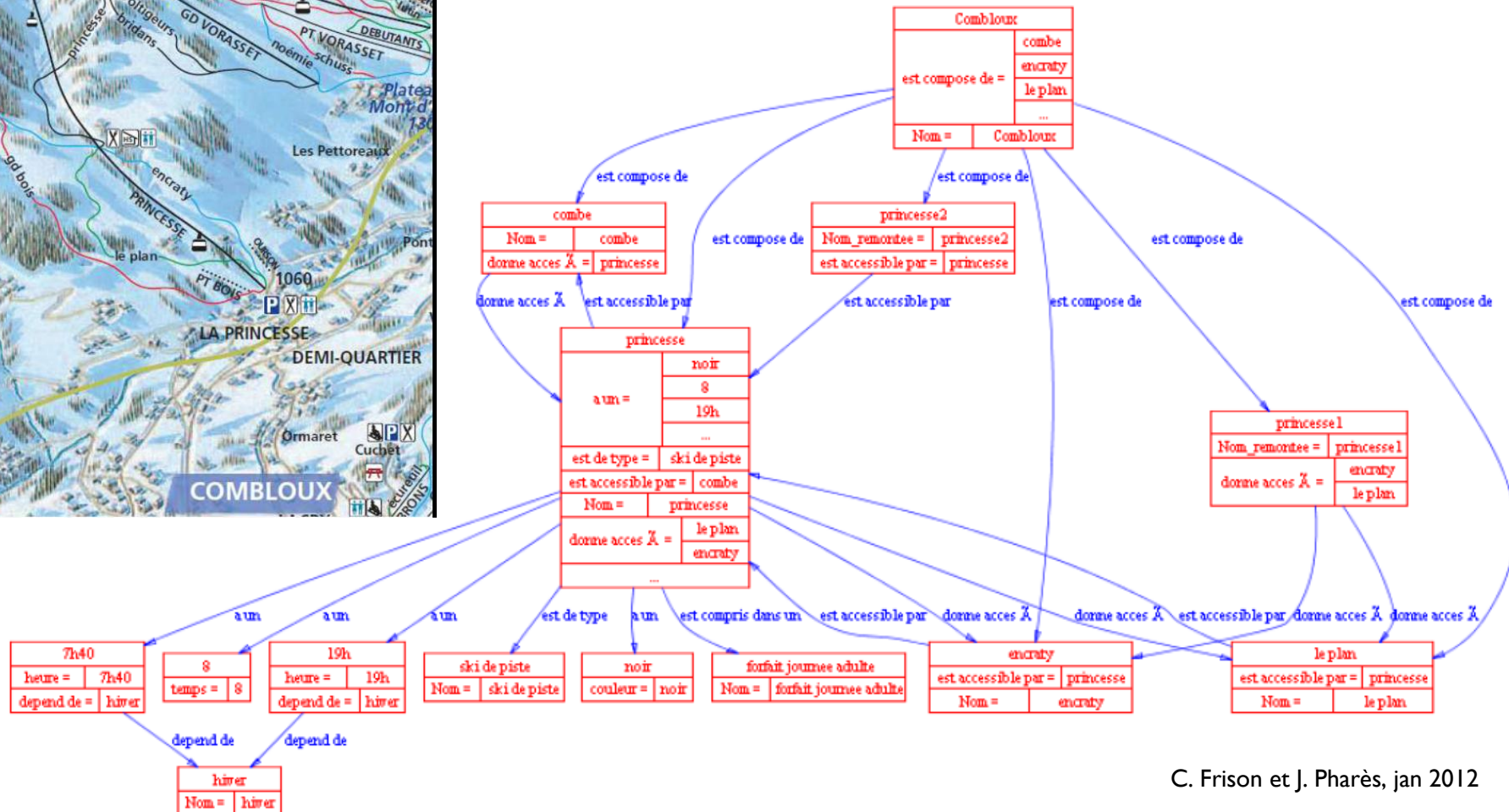
Partial visualisation of the ontology developed :



Example of a semantic web application, developed by C. Frison et J. Pharès, EC de Lille (3)



Visualisation of individuals (instances), for the « Princesse » trail :



Usage : interoperability

terms

"voiture", "caisse", "auto",
"bagnole", "tire", "tacot",
"automobile", ...

icon



comprehension / intension

Automotive transport vehicle
designed to carry a small number of
persons

concept #x23

extension

- la 206 cc 3124 SH 45
- la clio 16S de Tom
- le Galaxy bleu garé
en bas à l'entrée
- ...

NB : libellé formel du concept souvent issu des
termes pour des raisons pratiques (débugage,
prototypage ...)

Benefits of the approach :



- ▶ To provide a powerful tool for **the control of the information**, associated with tools for **automatic reasoning and information consistency checking**
- ▶ This type of analysis is already possible with numerical data (e.g. statistics)
- ▶ **Here, reasoning applies to « concepts »** (information, knowledge, organisations) and no longer on data

→ **Innovative approach**

Benefits of CODATA France – O3D : to provide methods and tools allowing the work on those concepts

→ **thus enabling to build his/her own experience**

- ▶ **Sustainable development data** : provide a « sand box » enabling the validation of the approach

--- VERY IMPORTANT FROM A POINT OF VIEW OF RESEARCH DATA IN THE FUTURE ---

REFERENCES

* Blogs O3D :

<http://codatao3dclubsebastopol.blogspot.com/>

<http://codatao3dclublionsregionnorddd1.blogspot.com/>

<http://clubo3dlionsregionnorddd2.blogspot.com/>

<http://codatao3dclubsebastopol.blogspot.com/p/codata-france.html>

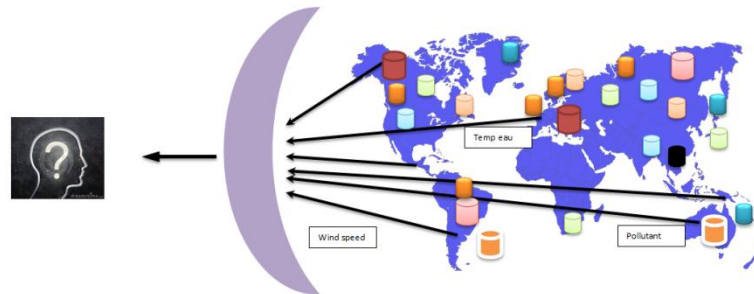
* CODATA France :

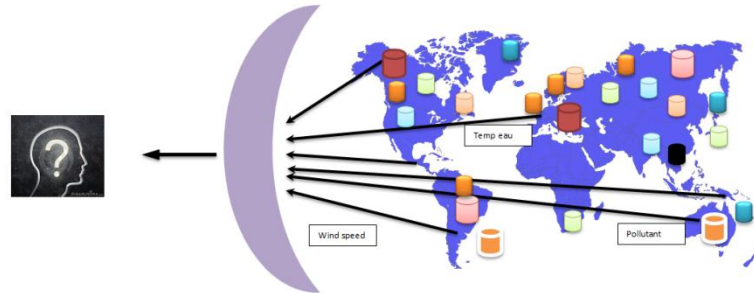
<http://ticri.inpl-nancy.fr/wicri->

incub.priv/index.php/Observatoire_des_donn%C3%A9es_du_d%C3%A9veloppement_durable

* CODATA International :

<http://www.codata.org/>





Questions ?

AF Cutting-Decelle (afcd@skynet.be)
JP Caliste (jean-pierre.caliste@utc.fr)
R Vazille (vazille.robert@wanadoo.fr)