

# Vorlesung

Szientometrie, Wissenschaftsinformationssysteme

Hans Uszkoreit & Brigitte Jörg

Hans Uszkoreit Vorlesung  
Informationswissenschaft  
und Informationssysteme



## Scientometrics

**“Scientometrics is the science of measuring and analysing science.**

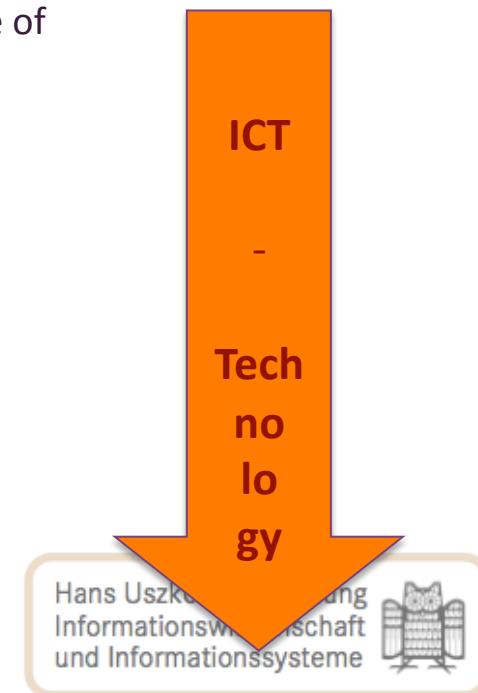
**In practice, scientometrics is often done using bibliometrics that is measurement of (scientific) publications.”**

(Wikipedia, July 2009)



## Scientometrics – some relevant dates

- **1665: First Journal:** „Philosophical Transactions of the Royal Society“ in London
- **1665: French** „Journal Scavans“ (rather scientific journalism)
- **18th Century:** largely based on personal communications
- **Late 18th Century:** emergence of citations from the specific organisation of the sciences
- **19th Century:** further institutionalization of Science; references to authors, opposed to identifiable and dated texts (later)
- **1910:** emergence of citations as currently understood
- **Early 20th Century:** substantial growth of science activities
- **1940s+1950s:** institutional phase of science policy, regular appearance of funding acknowledgements and Institutional addresses in publications
- **1960:** Foundation of ISI (Institute for Scientific Information) establishing the **Science Citation Index (SCI)** and with it, the **Impact Factor** as a measurement of Science
  
- **... the Impact Factor (based on ISI’s database) is still the dominant measure for performance of Science**
- ... but this is about to change with the ‚Open Access‘ era
- ... other Indexes appear, Hirsch-Index, download metrics, ...
  
- <http://www.arl.org/resources/pubs/mmproceedings/138guedon.shtml>
- <http://eprints.ecs.soton.ac.uk/17142/4/scientometproofs.pdf>



## Scientometrics – more recent themes

Two Decades of „Sciencometrics“ – An interdisciplinary field represented by its leading journal \*

Category	1980	1989	1997
1 Bibliometric Theory, Mathematical Models and Formalisation of Bibliometric Laws	1	6	3
2 Case Studies and empirical Papers	5	21	35
3 Methodological Papers including applications	6	24	25
4 Indicator Engineering and Data Presentation	2	6	1
5 Sociological Approach to Bibliometrics, Sociology of Science	5	9	6
6 Science Policy, Science Management and General or Technical Issues	12	1	5
<b>Total</b>	<b>31</b>	<b>67</b>	<b>75</b>

Table: Classification of Papers published in the journal „Scientometrics“ by categories

\* From an article in the „Scientometrics“ Journal. Vol 50, No. 2 (2001) 301-312:  
[http://www.steunpuntoos.be/WG\\_Papers/Scientometrics\\_50\\_3\\_301.pdf](http://www.steunpuntoos.be/WG_Papers/Scientometrics_50_3_301.pdf)



## Scientometrics, Research Information Systems

# Scientometrics – more recent themes

Topics of interest include, but are not limited to:

### GENERAL WORKS ON INFORMETRICS

- Structure of the field and relations to other fields
- History of bibliometrics
- Analysis of acknowledgements

### THEORY

- Informetric laws
- distributions and concentration
- Time and growth
- Other works on theory

### METHODS AND TECHNIQUES

- Mapping and visualization
- Text and data mining and other linguistic techniques
- Network analysis
- Classification, Other methods

### CITATION ANALYSIS

- Characterizing publications and fields
- Theoretical and methodological aspects
- Self-citation
- Ego-centered citation studies
- Quality assessment
- Errors in citation analysis
- Other Citation analysis

### INDICATORS

- H-Index
- Impact factors
- Other bibliometric indicators

### WEBOMETRICS

- Web Impact Factors
- Measures and methodological issues
- The 'academic' web
- Commercial sites
- Linking motivations and/or content analysis
- Journals and web citations

### JOURNALS

- The 'impact factor' debate
- Journal quality
- Journal coverage and structure
- Journal profiles
- Non-English journals and internationalization
- Journal editors and editorial boards

### OPEN ACCESS AND ELECTRONIC PUBLICATIONS

- Studies on Open access and/or Electronic publications
- Debate on Open access and/or Electronic publications

### PRODUCTIVITY AND PUBLICATIONS

- Country level studies
- Field specific studies
- Media coverage of science
- Interdisciplinarity
- Gender and/or other Sociological issues
- Other studies on Productivity

### COLLABORATION

- Studies in specific fields
- Studies on national, group or author level

### RESEARCH POLICY

- Research evaluation
- University rankings
- Scientometrics indicators in S&T policy
- Other studies and/or debate on Research policy

### PATENT ANALYSIS

- Patent citations and science and technology links
- Other patent indicators

### BIBLIOGRAPHIC DATABASES

- comparisons between databases, coverage, retrieval techniques and data isolation

## Scientometrics – The Journal Impact Factor

- A measure of citations to science and social science journals
- Applied to journals only (Garfield), **not individual articles**
- In charge of ISI (now part of Thomson Reuters)

**Calculation** ([http://en.wikipedia.org/wiki/Impact\\_factor#Calculation](http://en.wikipedia.org/wiki/Impact_factor#Calculation)):

- two-year period
- Average number of citations in a given year to those papers in a journal that were published during the two preceding years

**Example:** 2003 impact factor of a journal would be calculated as follows:

A = number of times articles published in 2001-2002  
were cited in indexed journals during 2003  
B = number of „citable items“  
(usually articles, reviews, proceedings or notes, not editorials  
and letters-to-the-Editor) published in 2001-2002

**2003 Impact Factor = A/B**

- The 2003 impact factor is effectively published in 2004, because it cannot be calculated until all of the 2003 publications had been received.



## Scientometrics – Problems with the Impact Factor

- originally intended as an objective measure of the reputability of a journal (Garfield), it is now being increasingly applied to measure the productivity of scientists.
- often misused to predict the importance of individual publications
- not relevant for literature, where the most important publications are books citing other books  
-> therefore, Thomson Scientific does not publish a Journal Citation Report (JCR) for the humanities
- not relevant for many areas of computer science, where the majority of the important publications appear in refereed conference proceedings and cite other conference proceedings
- cannot correctly be the only thing to be considered by libraries in selecting journals even though in practice they are applied this way
- comparison of IFs between different fields is invalid
- generally, i.e., medical journals have higher impact factors than mathematical journals and engineering journals.
- ...



## Scientometrics – Problems with the Impact Factor

- journals may change editorial policies
- journals may sometimes publish a larger percentage of review articles
- while many research articles remain uncited after 3 years, nearly all review articles receive at least one citation
- journals may change the fraction of "citable items" compared to front-matter in the denominator of the IF equation
- the types of articles considered "citable" is largely a matter of negotiation between journals and Thomson Scientific
  - > i.e. editorials in a journal do not count as publications
  - > when editorials cite published articles, they increase the citation count for articles
  - > "Letters to the editor" may refer to either class.
- ...



## Scientometrics – The Impact Factor Validity

- The IF could not be reproduced in an independent audit
- The IF refers to the average number of citations per paper
- The temporal window for citations is short
- Thomson Scientific does not contain many international journals or other types of publications (books, ...)
- Thomson Scientific contains only a few publications from languages other than English
- many high quality journals in the applied aspects of some subjects are not included (such as marketing communications, public relations and promotion management; not peer-reviewed technical magazines)
- conference proceedings publications are not indexed, including conferences, workshops and symposia.
- ...

See also: [http://en.wikipedia.org/wiki/Impact\\_factor](http://en.wikipedia.org/wiki/Impact_factor)



## Scientometrics – Some Links

- **Journal „Scientometrics“**  
Scientometrics is concerned with the quantitative features and characteristics of science. Emphasis is placed on investigations in which the development and mechanism of science are studied by statistical mathematical methods.  
<http://www.springerlink.com/content/101080/>
- **ISI Web of Knowledge:**  
<http://www.isiwebofknowledge.com/>
- **STI Indicators for the European Research Area (STI-Area):**  
<http://cordis.europa.eu/indicators/>
- **OECD Directorate for Science, Technology and Industry:**  
[http://www.oecd.org/department/0,3355,en\\_2649\\_33703\\_1\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/department/0,3355,en_2649_33703_1_1_1_1_1,00.html)
- **International Conference on Science and Technology Indicators (2008):**  
<http://systemsresearch.ac.at/sti-conference/>



## Scientometrics – Research Information Systems

### Scientometric Studies depend

- on an information system
- on the data in the system



# Research Information Systems

Current Research Information Systems (CRIS) bring together the information that underpins the Research for Quality Research Management (<http://www.euroCRIS.org/>)

- **For Researchers:**
  - easy access to relevant information and associated software,
  - processor power, storage systems and - where necessary - detectors to collect more data to overcome incomplete or inconsistent information
- **For research Managers and Administrators:**
  - easy measurement and analysis of research activity
  - easy access to comparative information
- **For Research Councils:**
  - optimisation of the funding process
- **For entrepreneurs and technology transfer organisations:**
  - easy retrieval of novel ideas and technology in a knowledge-assisted environment
  - easy identification of competitors and previously done similar research
- **For the media and public:**
  - easy access to information, software and computer power to allow easily-assimilated presentation of research results in appropriate contexts.



## Research Information

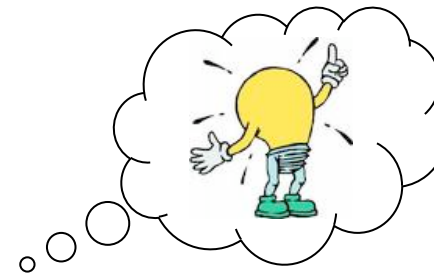
### Data/Metadata or Information about:

- Scientists
- Project Managers
- Ongoing and Completed Projects
- Research Departments
- Funding Organisations and Programmes
- Research Results
- Publications
- Equipment
- ...
- **their timely Relationships (Semantics)**

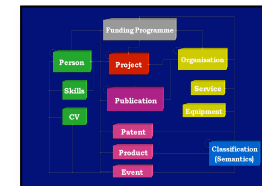


## Information Systems - Modelling

- A Concept about (Research) Entities and their Relationships  
**Specification (Conceptual Level)**



- A Description of (Research) Entities and their Relationships  
**Model (Logical Level)**



- A Formalization of (Research) Entities and their Relationships  
**Database Scripts (Physical Level)**

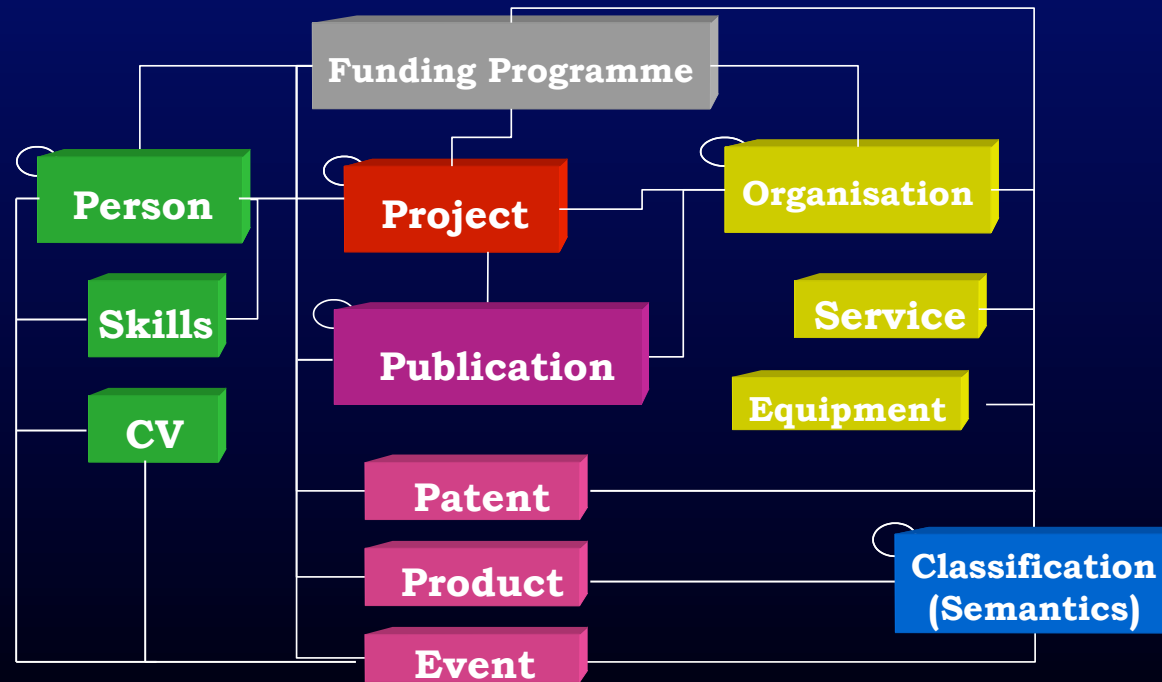
SQL Script

```
-----  
CREATE Table Person  
CREATE Table Project  
CREATE Table OrgUnit
```



Modeling: Example Data Model (CERIF)\*

Common European Research Information Format



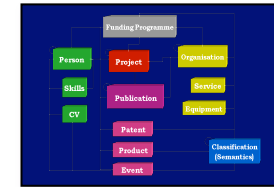
\* <http://www.eurocris.org/cerif/>

## Scientometrics, Research Information Systems

### What is CERIF?

#### Common European Research Information Format

- (1) data model
- (2) allows for a (**metadata**) representation of
  - research entities
  - their activities / interconnections (research)
  - their output (results)
- (3) offers high flexibility with (**semantic**) relationships
- (4) enables quality maintenance, archiving, access and interchange of research information
- (5) supports knowledge transfer to decision makers, for research evaluation, research managers, strategists, researchers, editors, the general public
- (6) enables a standardized setup of research information systems



„Metadata is **structured data** which describes the characteristics of a resource.“

*An Introduction to Metadata, by Chris Taylor, University of Queensland*

“Metadata is sometimes defined literally as 'data about data,' but the term is normally understood to mean **structured data about resources** that can be used to help support a wide range of operations. These might include, for example, **resource description and discovery, the management of information resources and their long-term preservation.**“

*Metadata in a Nutshell, by Michael Day, UKOLN*



## Examples of Metadata?

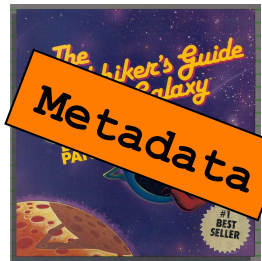
**Radio Series:**  
**Title:** The Hitchhiker's Guide to the Galaxy  
**Description:** science fiction comedy series created by Douglas Adams. Originally a radio comedy broadcast on BBC Radio 4 in 1978, [...]  
**Source:** Wikipedia  
**Date of Query:** May 30, 2008



**Book:**  
**Title:** The Hitchhiker's Guide to the Galaxy  
**Date of Publication:** 1979

**Series of five Books:**  
**Title:** The Hitchhiker's Guide to the Galaxy  
**Between:** 1979 - 1982

- Structure:**
- Type of Resource
  - Title
  - Description
  - Source
  - Date
  - Author, Creator, ...



**Game Cover Image:** The Hitchhiker's Guide to the Galaxy  
**Source:** <http://www.legotron.com/>  
**Retrieved:** May 30, 2008

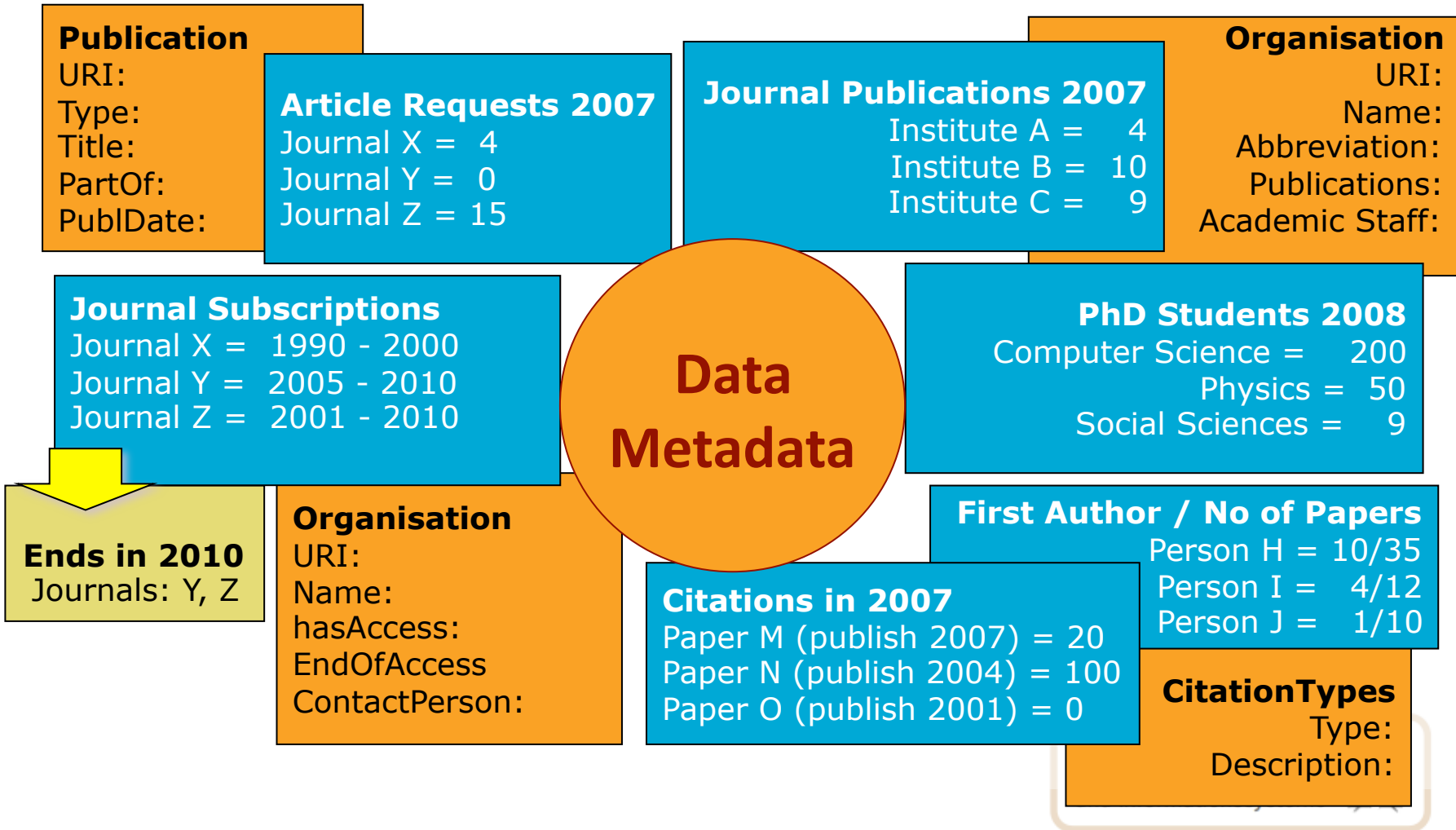
**TV Series:**  
**Title:** The Hitchhiker's Guide to the Galaxy  
**Screned:** 1981

**Computer Game:**  
**Title:** The Hitchhiker's Guide to the Galaxy  
**Released:** 1984

**Links:**  
<http://www.bbc.co.uk/cult/hitchhikers/>  
**HTML-Title:** Cult Hitchhiker's Guide to the Galaxy  
[http://en.wikipedia.org/wiki/The\\_Hitchhiker's\\_Guide\\_to\\_the\\_Galaxy](http://en.wikipedia.org/wiki/The_Hitchhiker's_Guide_to_the_Galaxy)  
**HTML-Title:** The Hitchhiker's Guide to the Galaxy

**Comic Book Adaptations:**  
**Title:** The Hitchhiker's Guide to the Galaxy  
**Between:** 1993 - 1996

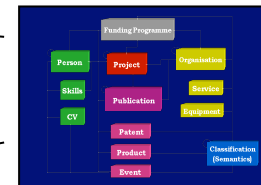
# Examples of Metadata



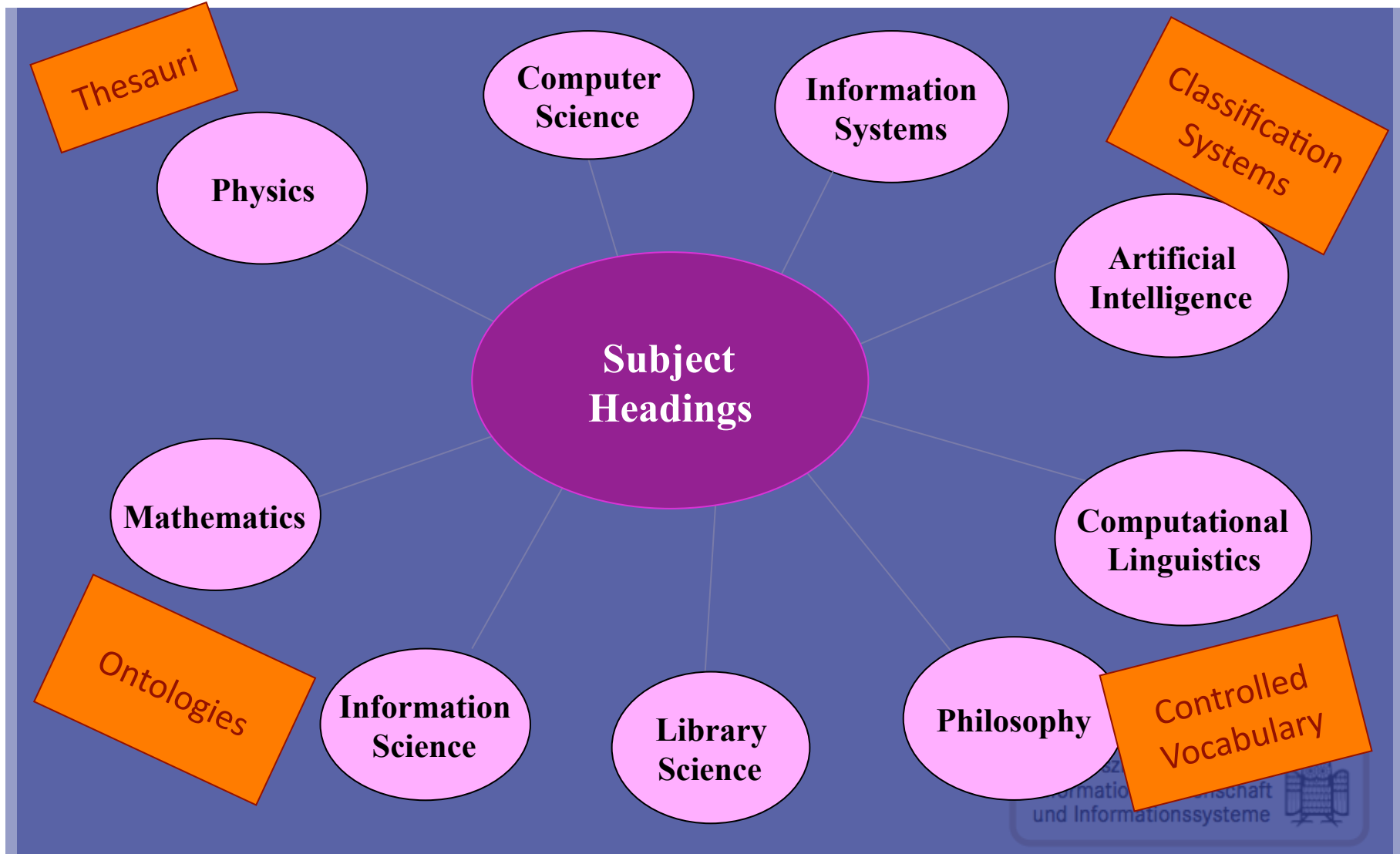
## What is Formal Metadata?

**Formal Metadata = Metadata formalized**

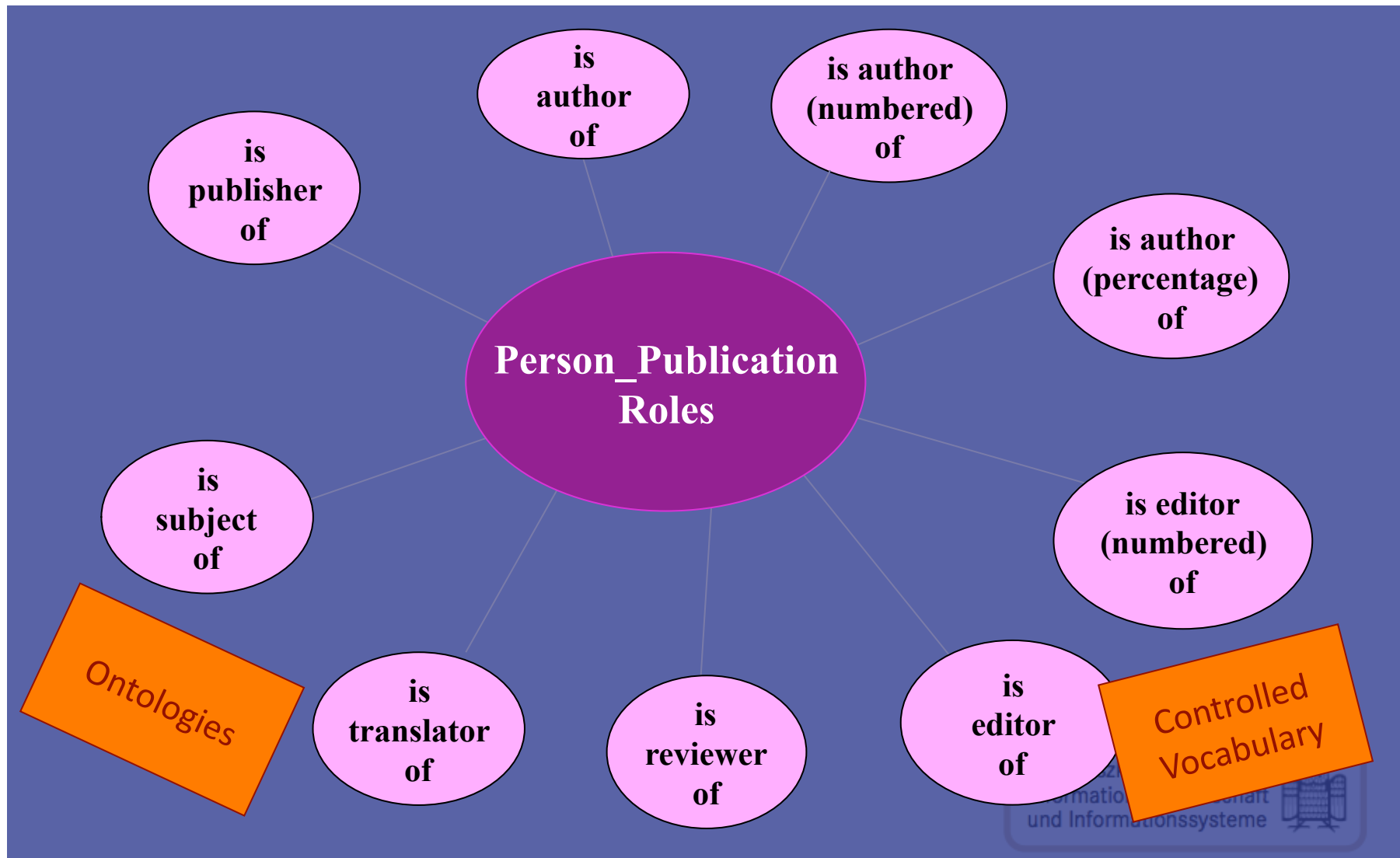
- **Typed**  
-> Based on Model Entity Types
- **Structured**  
-> Following a Formal Model Structure



## Metadata – Formal Metadata (CERIF Semantics)



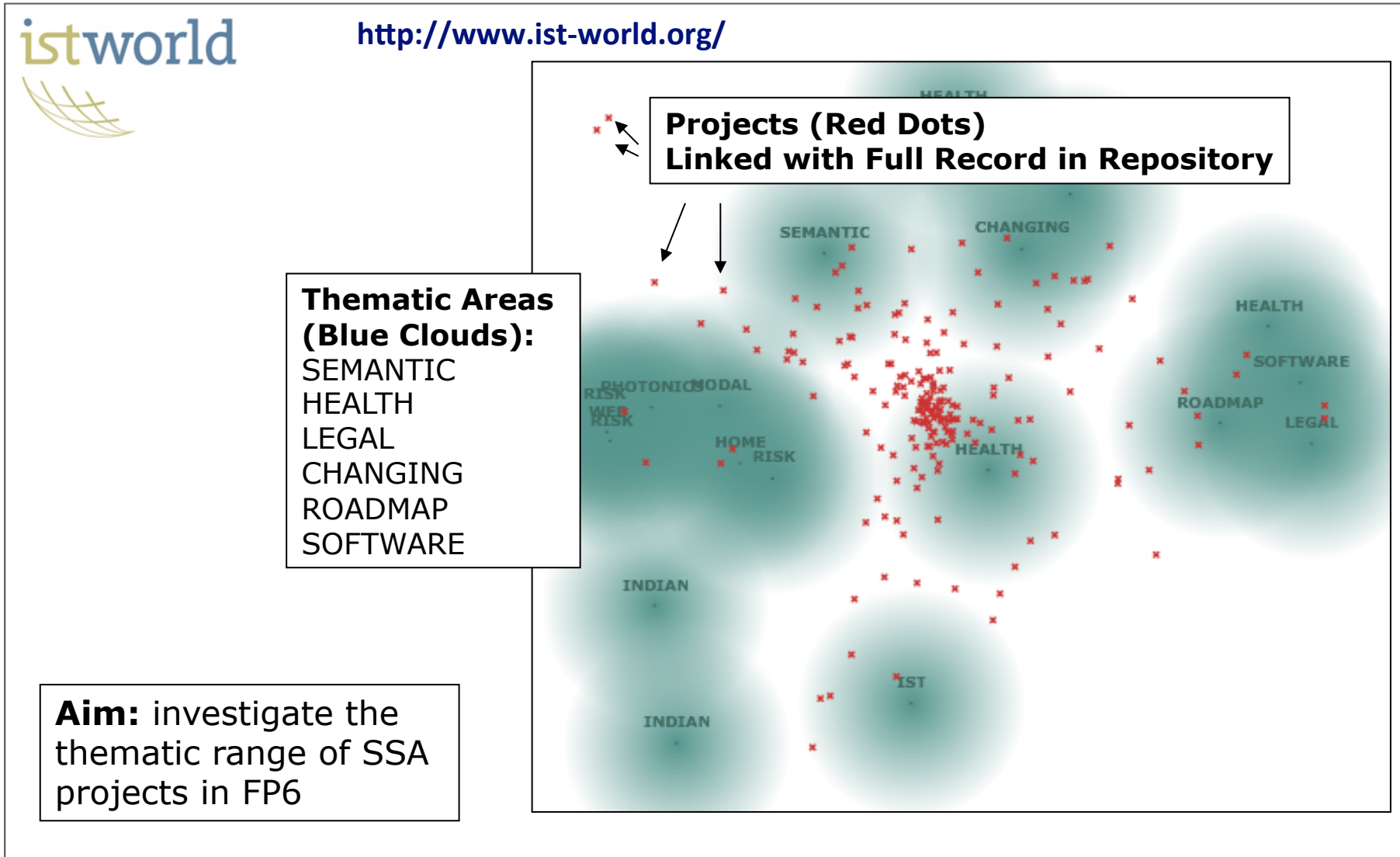
## Metadata – Formal Metadata (CERIF Semantics)



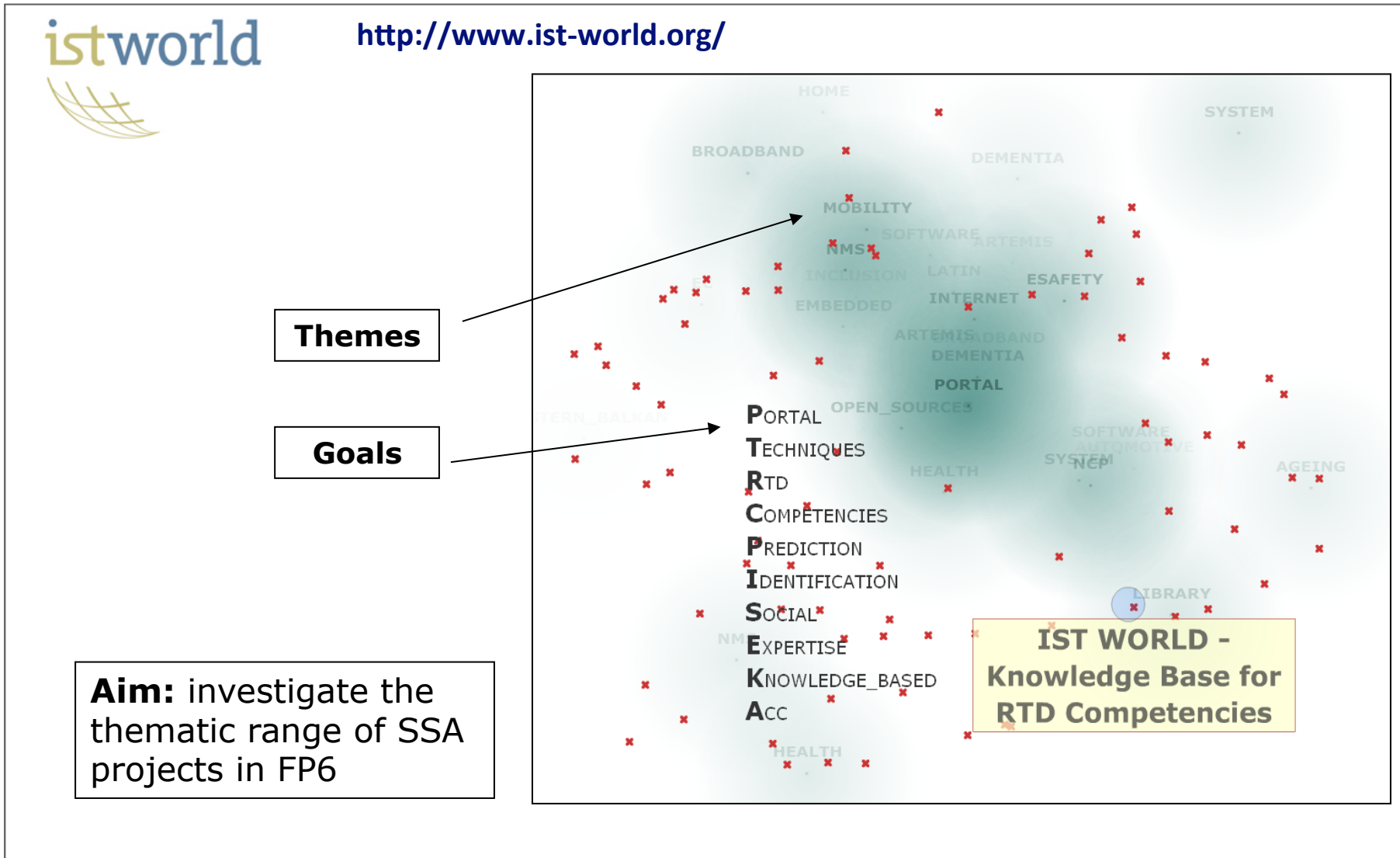
## Metadata – Formal Metadata (CERIF Semantics)



## Scientometrics – Some Examples



## Scientometrics – Some Examples



## Scientometrics – Some Examples

