



# RISIS

Research infrastructure for research  
and innovation policy studies



---

## Training Book CorText.Risis

SHORT COURSE TYPE A (SCA)  
6 -7 October 2015

See <http://risis.eu/events/>

Organized by CorText-Lab Team  
OF LISIS Unit in Paris-Est



UNIVERSITÉ  
PARIS-EST  
MARNE-LA-VALLÉE



---

Marne La Vallée, October 2015

## LIST OF PARTICIPANTS

Nb	Name	Surname	Institution	Mail	City/Country
1.	Gök	Abdullah	University of Manchester ; MIOIR	abdullah.gok@mbs.ac.uk	Manchester ; UK
2.	Robinson	Douglas	Université Paris-Est Marne-la-Vallée ; LISIS	douglas.k.r.robinson@googlemail.com	Paris ; France
3.	Matos	Fabio	Universidade de Aveiro	fmatos@ua.pt	Aveiro ; Portugal
4.	Akyut	Stefan	LISIS (Université Paris-Est) et Centre Marc Bloch (Berlin)	s.aykut@gmail.com	Paris ; France
5.	Ejderyan	Olivier	Transdisciplinary Lab   Department of Environmental Systems Science   ETH Zurich	olivier.ejderyan@usys.ethz.ch	Zurich ; Suisse
6.	Recauchere	Olivier	INRA - DEPE	Olivier.Recauchere@paris.inra.fr	Paris ; France
7.	Reis	Isabel	FCT	<a href="mailto:Isabel.Reis@fct.pt">Isabel.Reis@fct.pt</a>	Lisbonne ; Portugal
8.	Beldame	Diane	INRA - DEPE	diane.beldame@paris.inra.fr	Paris ; France
9.	Weeks	Madeline	LISIS, INRA	<a href="mailto:mrweeks001@gmail.com">mrweeks001@gmail.com</a>	Paris ; France

## TRAINING TEAM

Name	Mail	Function	
BARBIER Marc	Marc.barbier@grignon.inra.fr	CortextLab Director - Researcher	
COINTET Jean-Philippe	jphcoi@gmail.com	Researcher and Data Scientist	
BREUCKER Philippe	breucker@ifris.org	IT Engineer	
ORSAL Guillaume	orsal@cortext.fr	IT Engineer	
VILLARD Lionel	lionel.villard@esiee.fr	IT Engineer	
DEQUATREBARBES Constance	constance@cortext.fr	Researcher	
PROUTZOU Georgia	Georgia.Prountzou@u-pem.fr	Administrative Support	

### Address for Follow up :

In case you have connexion troubles and bugs, mail to : [breucker@ifris.org](mailto:breucker@ifris.org)

In case you intend to develop a new project mail to : [Marc.barbier@grignon.inra.fr](mailto:Marc.barbier@grignon.inra.fr)

In case you have problem with journey and refunding mail to : [Georgia.Prountzou@u-pem.fr](mailto:Georgia.Prountzou@u-pem.fr)

## **PROGRAMME OF ACTIVITIES**

### **Day 1 - October 6th, 2015**

10h00: Welcome at Ibis Hotel – Noisy Champs

10:30h-12h30: Introduction about Objectives and Lectures

12:30h-13h30: Lunch

14h-16h30: Step by Step Demonstration with Feed-Back and discussions

Coffeee Break

17h-18h: From Database to Publication: an example of using CorTexT with the presence and testimony of Researchers

19 :30 h: Diner (at Ibis Hotel)

### **Day 2 - October 7th, 2015**

9h-10h30: Work in groups on the same datasets (1): Parsing and extracting

Coffeee Break

11h-13h00: Work in groups on the same datasets (2): Datasets analysis for the characterisation of structural features

13h-14h: Lunch

14h-16h: Work in groups on the same datasets (3): Graph interpretation and temporal analysis

16h-17h: Recap, feedback, discussions and follow-up

## DESCRIPTION OF CORPUSES TO BE USED DURING THE TRAINING

The Nano S&T dynamics database (Nano), developed by CorText-Lab of LISIS Unit, collects publications and patents between 1991 and 2010 about nano S&T. One central characteristic of emerging S&T is that they do not correspond to pre-existing categorization, parsing corpuses of emerging S&T domains require thus the elaboration of semantic based queries in order to delineate relevant corpus. CorText-Lab of LISIS Unit has developed a dynamic query gathering 1.18 million publications and 735 000 priority patents. Four types of enrichments have been organized under four added extracted information:

- categorization and harmonization of institutional affiliations,
- geolocalization of all authors and inventors;
- geographical clustering of S&T activities; and
- thematic clustering of S&T activities. It offers for each publication, 14 'main units of observation' and 11 for patents.

For this training two Sub-corpuses have extracted to be handled by users

### Sub-corpus for scientific activities:

Because of the large amount of data, we choose a simple query with a large timescale to study the development of nanosciences. Thanks to the analytical tools available through the CorText.risis.eu collaborative web application, users will be able to work with all the usual attributes (title, abstract, cited references, addresses...) of **3 843 scientific selected publications** with the following specification in timescale and type of documents.

Timescale: 1995 - 2015

Type of documents: Article OR Abstract of Published Item OR Book OR Book Chapter OR Book Review OR Database Review OR Proceedings Paper OR Review In English

### Sub-corpus for R&D activities:

For this sub-corpus we have extracted all the priority patents for European Patents Office (EPO), that have been identified by a lexical query (Mogoutov A et al., 2007). First fillings (priority patents) are used: to trace the origin of an invention on one hand and to partly eliminate redundancy due to the extensions (geographical or technological extension of the initial patent) on the other hand.

With this sub-corpus of **1 415 patents**, user will be able to investigate analytical dimensions based on all the attributes of the RISIS Nano Patents database (<http://risis.eu/wp-content/uploads/2015/03/Report-Task1-Nano.pdf>), especially: geographical dynamics and institutional collaborations. with the following specification in timescale and type of documents.

Timescale: 1990 - 2007

Type of documents: EPO priority patents

## SOME SCIENTIFIC ARTICLES AND CHAPTERS THAT GROUNDS METHODS AND SCIENTIFIC USES OF CORTEXT.LAB CAPACITY

- Barbier M., Bompart M., Garandel-Batifol V., Mogoutov A. (2012). Textual analysis and scientometric mapping of the dynamic knowledge in and around the IFSA community, In: Darnhofer, I., Gibbon, D., Dedieu, B. (ed.) *Farming Systems Research into the 21st Century: The New Dynamic*, Springer.
- Cambrosio A., Keating P., Mogoutov A., (2012). "What's in a Pill? On the Informational Enrichment of Anti-Cancer Drugs." In: Volker Hess and Jean-Paul Gaudillière (Eds), *Making Drugs: Ways of Regulation in Laboratories, Factories, and Consulting Rooms*. Houndsmills: Palgrave Macmillan, 181-205.
- Chavalarias D., Cointet J.P., (2013). « Phylomemetic Patterns in Science Evolution : The Rise and Fall of Scientific Fields. », *PloS one*, 2013, vol. 8, n°2
- Chavalarias D., Cointet, J-P., Cornilleau L., Duong T. K., Mogoutov A., Villard, L. (2011). Streams of Media Issues, Monitoring World Food Security, *Poster Presented at the United Nations Conference on World Food Security*, New-York.
- Cointet J-P., Mogoutov A., Bourret P., El Abed R., Cambrosio A., (2012). The emergence and development of gene expression profiling: a key component of the 3B (bench, bedside, bytes) in translational research, *Med Sci*, 28: 7-13
- Jones D., Cambrosio A., Mogoutov A., (2011). A Detection and characterization of translational research in cancer and cardiovascular medicine, *Journal of Translational Medicine*, 9: 57
- Laredo, P. a,d Villard, L. (2015). La construction d'une Europe de la recherche et ses effets sur les relations scientifiques transatlantiques. *Cahier rouge de la fondation Jean-Monnet*, 215.
- Laurens P., Bas C. L., Schoen A., Villard L. and Laredo, P. (2014). The rate and motives of the internationalisation of large firm R&D (1994-2005): Towards a turning point?. *Research Policy*.
- Mazieres A., Trachman M., Cointet J.-P., Coulmont, B. and Prieur, C. (2014). Deep tags: toward a quantitative analysis of online pornography. *Porn Studies*, 1(1-2), 80-95.
- Parasie S., Cointet J-P., (2012). La presse en ligne au service de la démocratie locale - Une analyse morphologique de forums politiques, *Revue française de science politique*, 62 (1), 45-70.
- Schoen A., Cointet J-P., Villard L., Laurens P., Heimeriks G., Alkemade F., 2012. A global map of technology, *STI Conference - Montréal - 5-8 September*
- Tancoigne E., Barbier M., Cointet J-P., and Richard G., (2014). « The place of agricultural sciences in the literature on ecosystem services », *Ecosystem Services*.
- Turenne N., 2013. *Knowledge Needs and Information Extraction*, 288 pages, ed. Wiley.
- Venturini T., Laffite N. B., Cointet J.-P., Gray I., Zabban V., and De Pryck, K. (2014). Three maps and three misunderstandings: A digital mapping of climate diplomacy. *Big Data & Society*, 1(2).

## SAMPLE OF ARTICLES DELIVERED IN THE REPOSITORY

### **Focus: BD : Web of Science**

· Tancoigne E., Barbier M., Cointet J-P., and Richard G., (2014). « The place of agricultural sciences in the literature on ecosystem services », *Ecosystem Services*.

### **Focus: BD : Corpus of scientific communications + Web of Science**

· Barbier M., Bompert M., Garandel-Batifol V., Mogoutov A. (2012). Textual analysis and scientometric mapping of the dynamic knowledge in and around the IFSA community, In: Darnhofer, I., Gibbon, D., Dedieu, B. (ed.) *Farming Systems Research into the 21st Century: The New Dynamic*, Springer.

### **Focus: BD : Web of Science**

· Kahane B, Mogoutov A., Cointet J.P., Villard L., Larédo P., (2010). A dynamic query to delineate emergent science and technology: the case of nano science and technology. *Working Paper*.

### **Focus: BD : Web of Science**

· Chavalarias D., Cointet J.P., (2013). « Phylomemetic Patterns in Science Evolution : The Rise and Fall of Scientific Fields. », *PloS one*, 2013, vol. 8, n°2

## COURSE EVALUATION QUESTIONNAIRE



### Using the CorText-Risis Platform for Research in Science policy and Science-Technology-Studies Organized by CorText-Lab of the LISIS Unit with the support of IFRIS

Please underline the text of your selection:

**01) Why did you choose to attend this course?**

- Actual research interest
- Potential research interest
- Professional interest
- Exploration/Curiosity
- Other (specify) \_\_\_\_\_

**02) Were the course objectives clearly defined?**

- no    rather no    rather yes    yes    no opinion

**03) Were the contents of the course consistent with the course description?**

- no    rather no    rather yes    yes    no opinion

**04) Has this course stimulated your interest in the subject it dealt with?**

- no    rather no    rather yes    yes    no opinion

**05) Was the course well organized?**

- no    rather no    rather yes    yes    no opinion

**06) Was the course well structured?**

- no    rather no    rather yes    yes    no opinion

**07) Do you think that teaching materials facilitated learning?**

- no    rather no    rather yes    yes    no opinion

Please insert your comment in the

**08) What were the best features of the course?**

**09) In your opinion, are there any aspects of the course that might need to be improved?**

**10) Please rate your overall satisfaction with the course in the box**

(1-10 scale, from 1: minimum satisfaction to 10: maximum satisfaction)