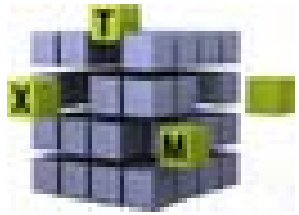
La Textométrie devient Open Source

Serge Heiden, ENS Lyon
slh@ens-lyon.fr

Séminaire CÉDITEC, Créteil, 19 Mars 2010

Projet Textométrie 2007 - 2010



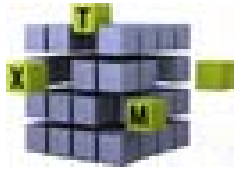
http://textometry.ens-lsh.fr

- Partenaires contractants initiaux :
  - Université de Nice Sophia-Antipolis - BCL
  - Université de Besançon - LASELDI
  - Université de Paris 3 - ILPGA
  - ENS-LSH Lyon – ICAR : hébergeant U. of Oxford – OUCS and U. de Montréal - ATO
- Participants au titre de :
  - Concepteurs de logiciels de Lexicométrie / Textométrie
  - Utilisateurs experts en Textométrie (politologie, linguistique, littérature, histoire…)

Concepteurs de logiciels

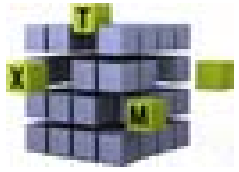
- **DTM** : **Ludovic Lebart** (développements logiciels innovants dans la suite de composants réalisés pour SPAD et SPAD-T, important logiciel de statistique et d'analyse des données diffusé par la société SPADsoft, Paris : http://www.spadsoft.com)
- **HYPERBASE** : **Etienne Brunet** (diffusé en CDROM par l'U. de Nice : http://ancilla.unice.fr/~brunet/pub/hyperbase.html)
- **LEXICO** : **André Salem** (diffusion par le web, en shareware : http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW)
- **SATO** : **François Daoust** (diffusé par l'UQAM, en CDROM et par le web : http://www.ling.uqam.ca/sato/outils/sato.htm)
- **WEBLEX** : **Serge Heiden** (diffusé par l'ENS-LSH : application web à usage académique par projets de recherche : http://weblex.ens-lsh.fr/doc/weblex.pdf)
- **XAIRA** : **Lou Burnard** (diffusé par l'OUCS : par le web, en open source : http://www.xaira.org)
- **ASTARTEX** : **Jean-Marie Viprey** (outil académique, support d'expérimentation et d'illustration de fonctionnalités textométriques innovantes :
- http://laseldi.univ-fcomte.fr/document/viprey/page_JMV.htm)
- **TRAMEUR** : **Serge Fleury :** http://tal.univ-paris3.fr/trameur

TXM - Objectifs



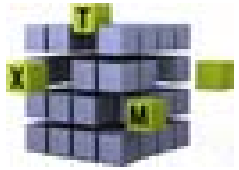
- Construire une nouvelle plateforme logicielle commune ouverte :
  - une boîte à outils pour les services fondamentaux
  - des applications prototypes :
    - Locale : Windows, Linux, *Mac OS X*
    - Distante : Navigateur Web
- Déploiement aisé pour l'utilisateur final
- Efficace

TXM – Effort principal



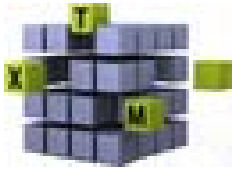
- Construire un **modèle conceptuel** commun
  - Unités de base
  - Unités de structures
  - Focus & Sélection
  - Regroupement raisonné de fonctions
- Stabiliser une **terminologie** commune
- S’accorder sur un **niveau de qualité des données d’entrée**

TXM – Modèle courant



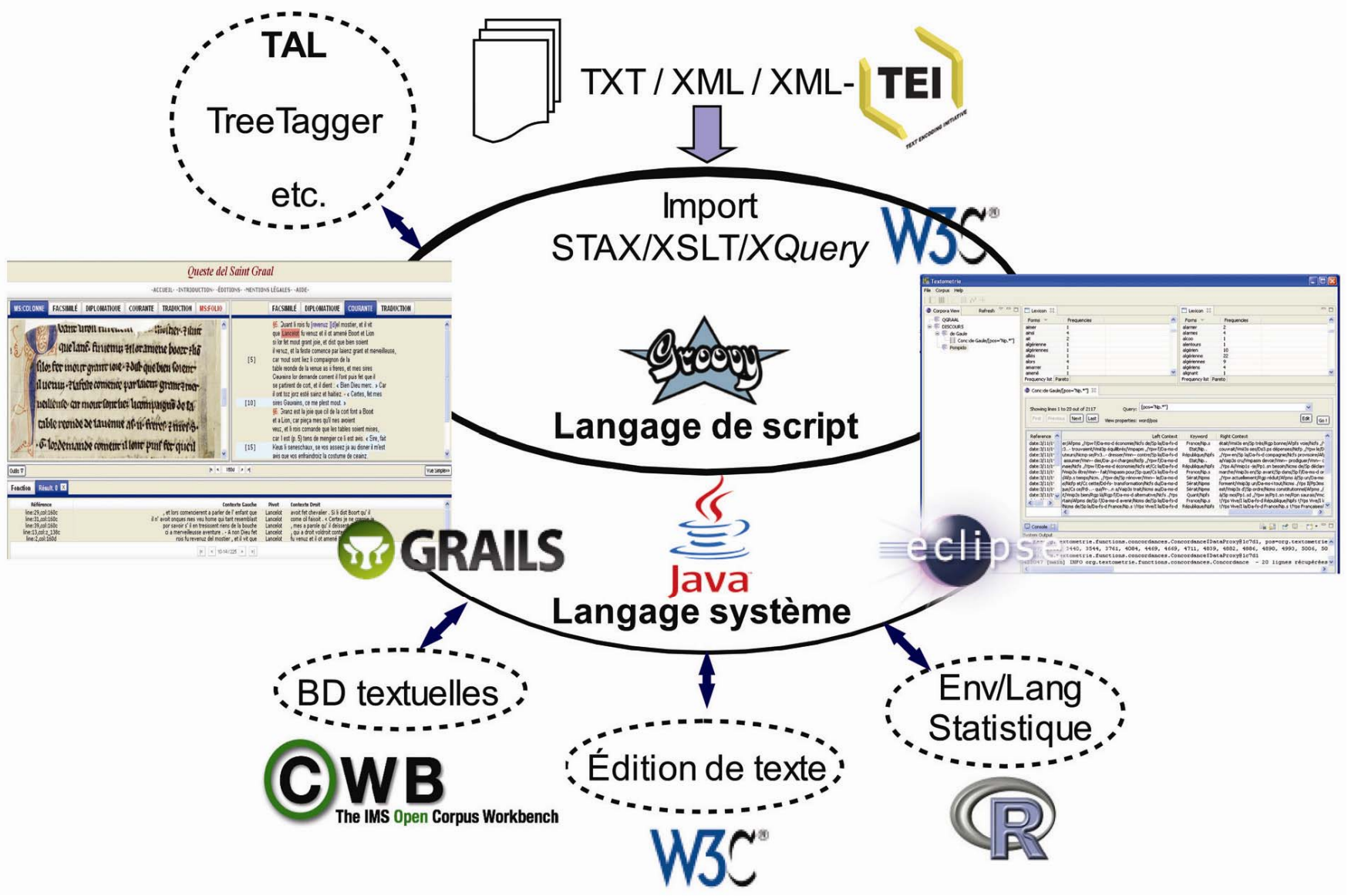
- Hiérarchie d’unités ayant d’importe quelle propriétés
- Mécanisme général de sélection d’unités
  - sélection de corpus
  - sélection de focus
- Composant d’entrée scriptable pour la construction aisée de plugins d’importation de tout format : CVS-TXT-XML-XML/TEI…

TXM - Choix



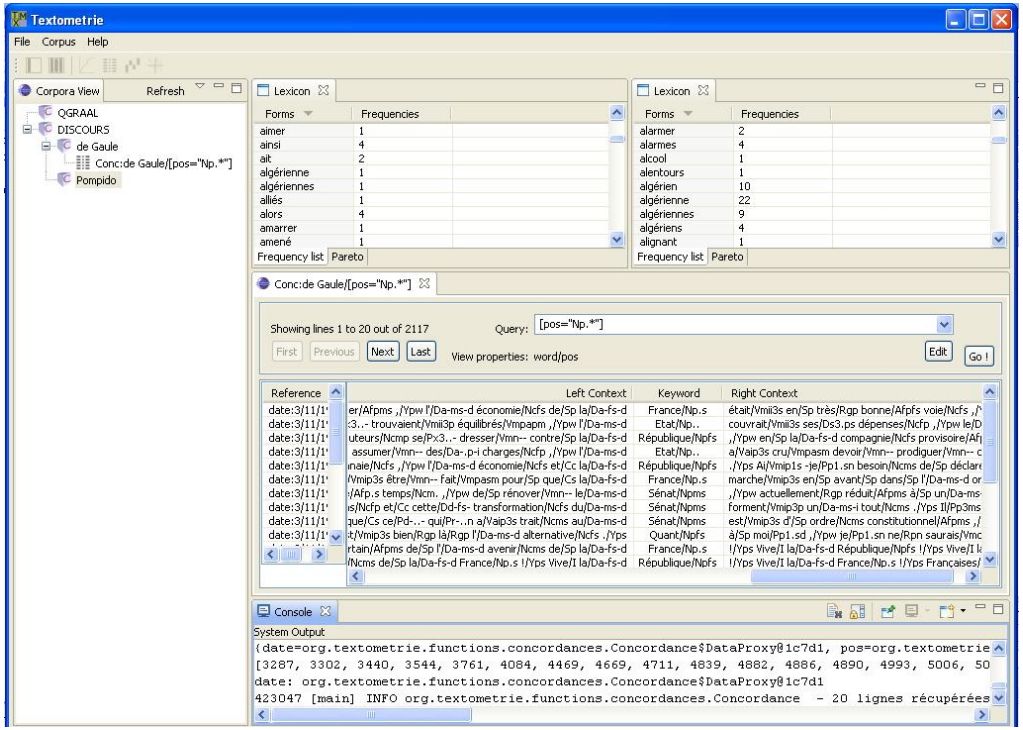
- Plateforme basée sur des standards ouverts : **Logiciel Open Source**
  - Interopérabilité
  - Extensibilité
  - Capitalisation
  - Utiliser des composants Open Source
    - IMS CWB
    - Environnement statistique R
    - TAL ouvert…
- Compatible XML-TEI : **Corpus Open Source**
  - Interopérabilité
  - Extensibilité
  - Capitalisation
  - Valider sur des corpus existants

TXM - architecture

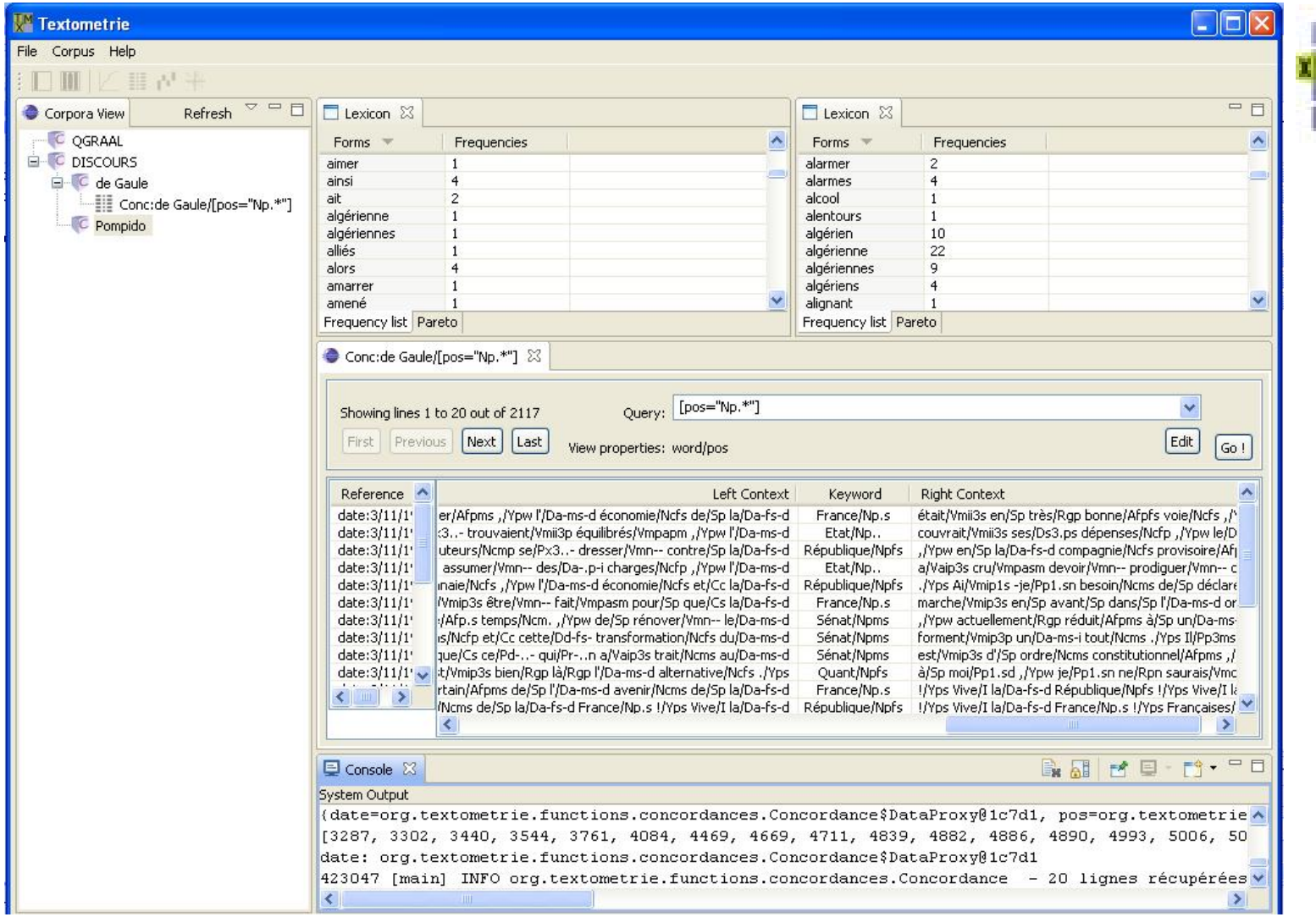


TXM - application Eclipse RCP

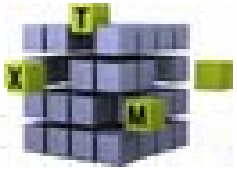
- Windows XP : Setup_0.4.0.exe
- Linux Ubuntu : TXM-0.4.0.tgz

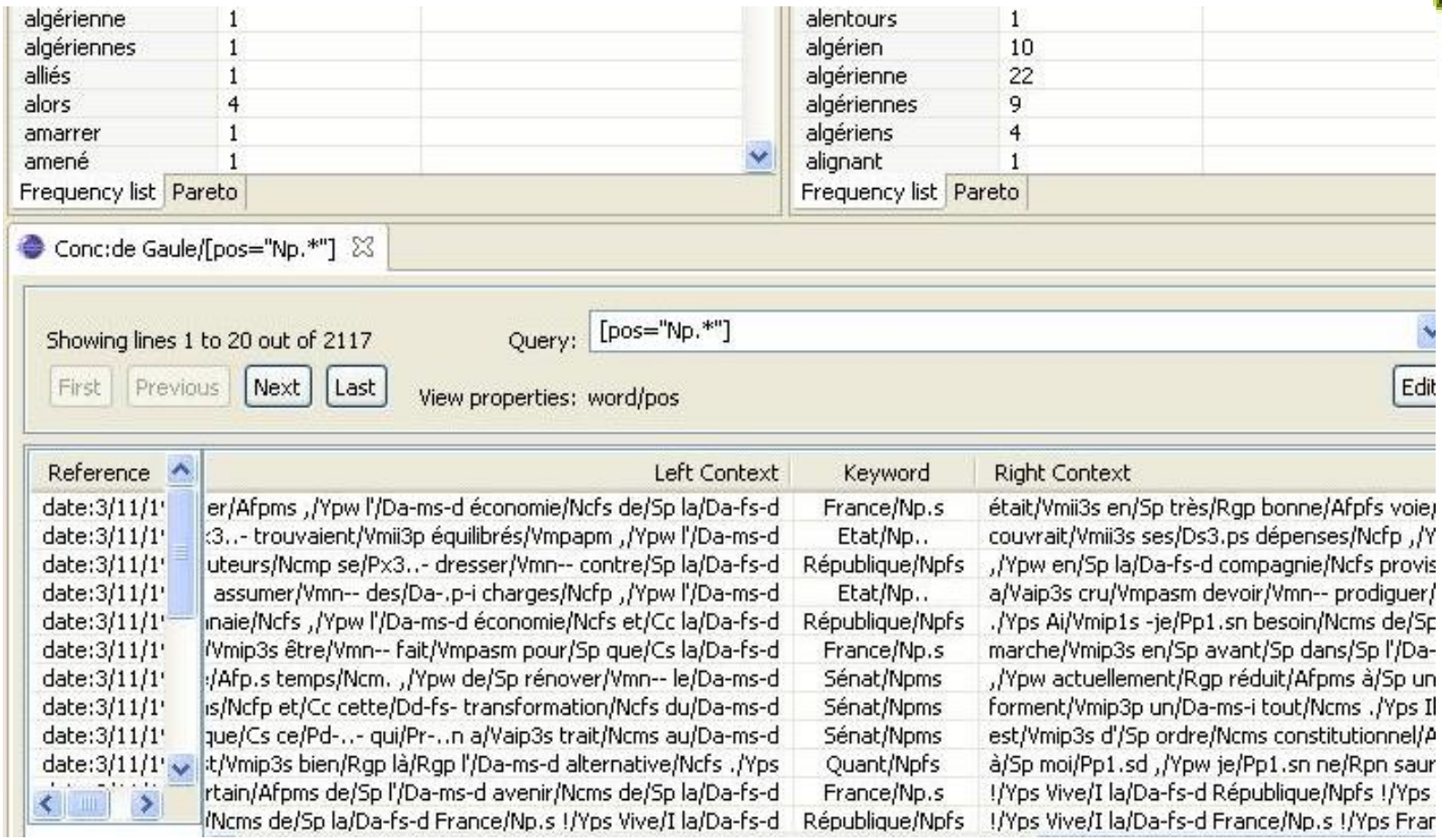


TXM Eclipse RCP application

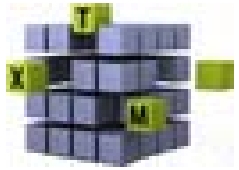


TXM Eclipse RCP application



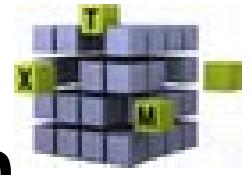


TXM Grails web application

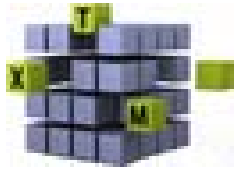


- Groovy / J2EE based + Java TXM toolbox
- YUI & Prototype AJAX javascript toolkits

Queste del Saint Graal web edition TXM Grails application



TXM - documentation ouverte & sites web & wiki



- Projet Scientifique : http://textometry.ens-lsh.fr
- Développeurs informatique : http://sourceforge.net/apps/mediawiki/textometrie/index.php?title=Main_Page
- Site logiciel principal SourceForge : http://sourceforge.net/projects/textometrie

Thank you to partners and sponsors

- **Partners**
  - Contracting : Nice University, Besançon University, Paris 3 University, Oxford University (UK), Montréal University (CA)
  - Chicago University (US) – Philologic/Philomine
  - UPR IRHT, UMR CESR, UMS RISC (FR) - corpus
  - Cultural Heritage Consortium - Izhevsk (RU) ? - corpus
- **Sponsors**
  - National French Research Agency (ANR)
  - CNRS
  - Rhône-Alps region Research Cluster 13 Agency