IFRIS
Institute For Research and Innovation in Society

CORTEXT

INRA SenS
Sciences en Société

CorText Manager

Application Web collaborative d'analyse et de cartographie de données hétérogènes

http://manager.cortext.net/

Lionel Villard

ESIEE PARIS

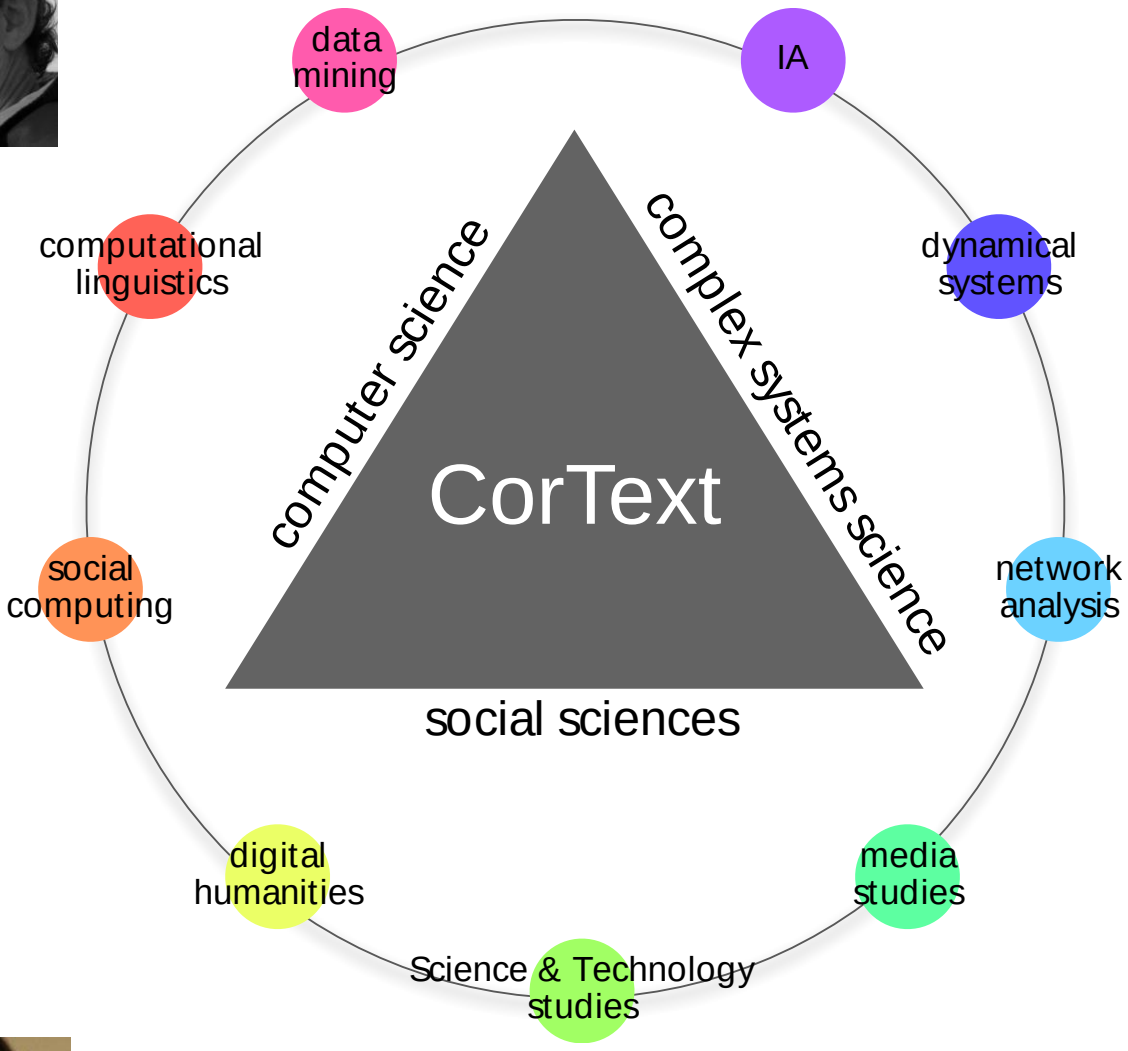
UPEM UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE



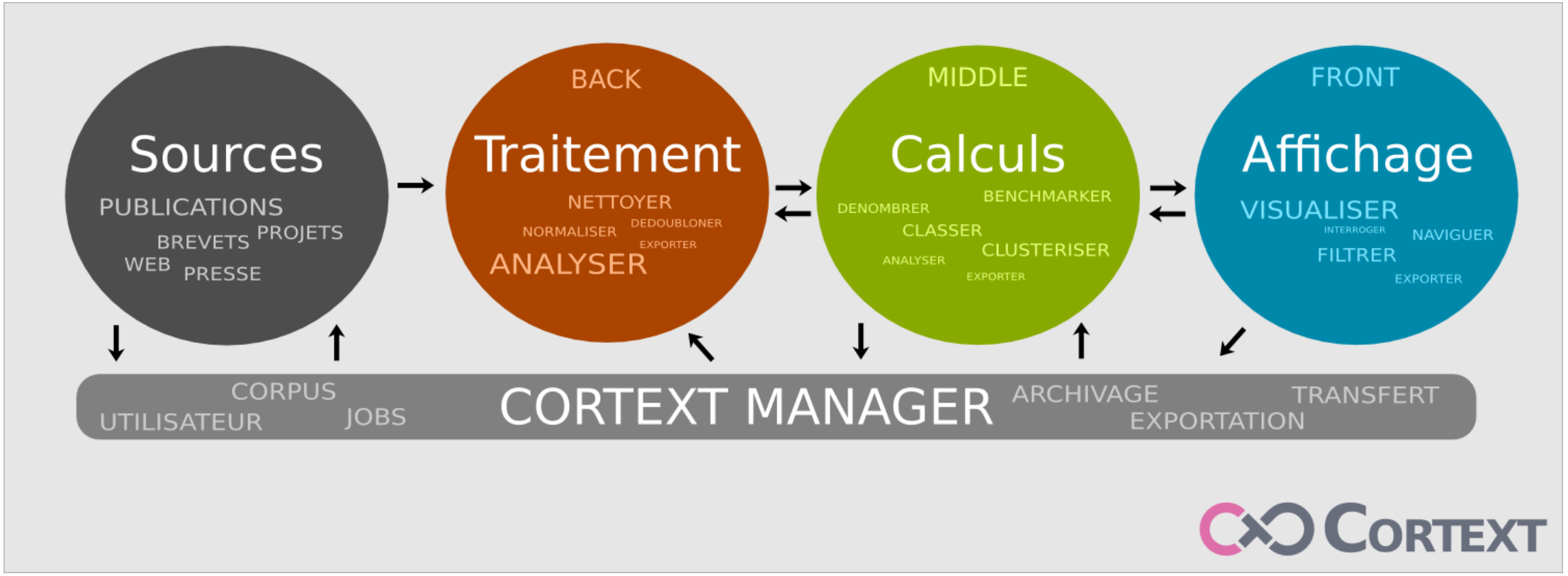

- Barbier Marc (Dir.)
- Breucker Philippe
- Cointet Jean-Philippe
- Duloquin Clhoé
- Duong Tam-Kien
- Laurens Patricia
- Martinez Cristian
- Mazières Antoine
- Mogoutov Andreï
- Schoen Antoine
- Turenne Nicolas
- Villard Lionel



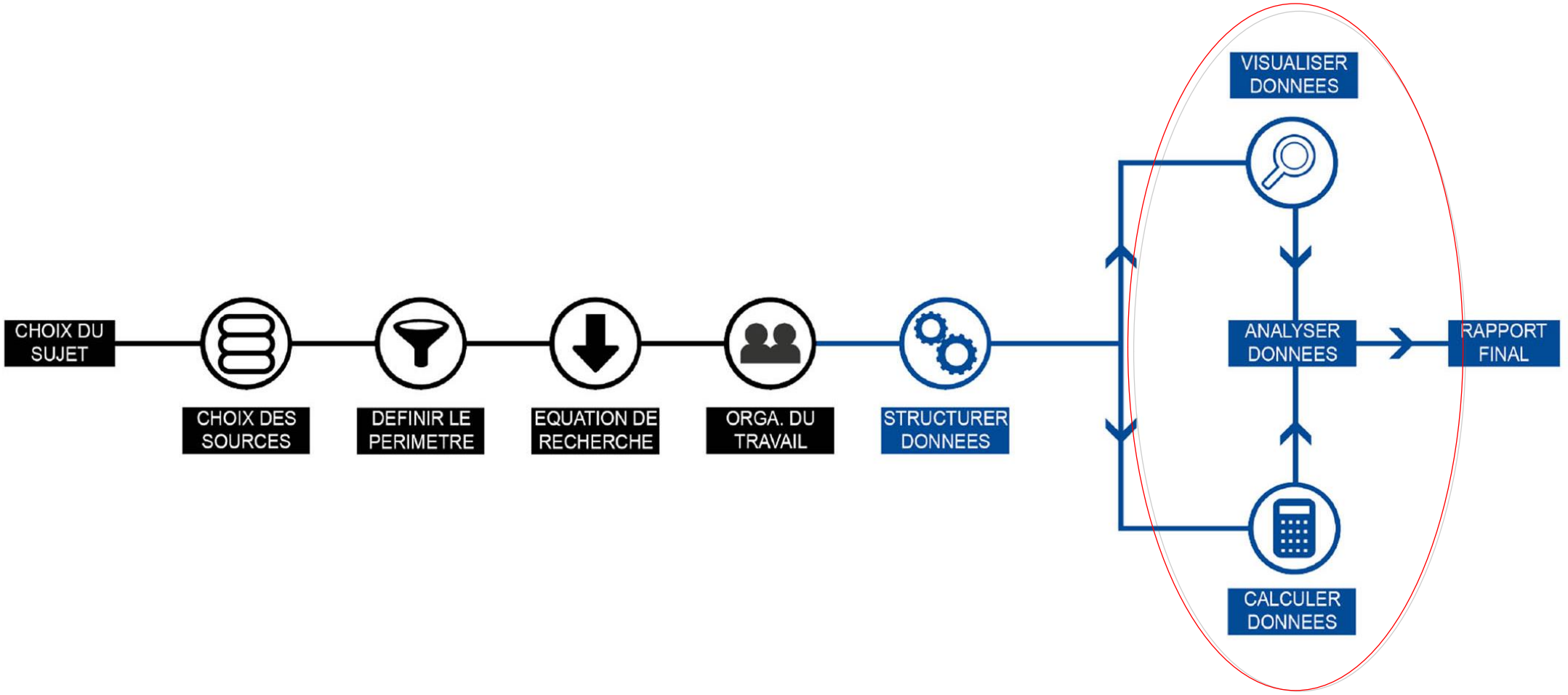




CorText Manager : principales étapes



Du **sujet** de l'étude aux résultats et **visualisations**



Des aller-retours entre données – visualisations – interprétations

scientific productions

Web Of Science ISI

specific databases

rare disease database

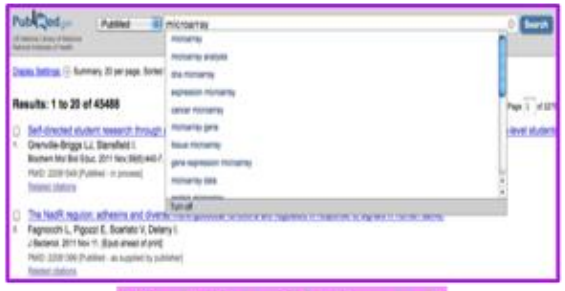
media productions (press+web)

web crawler

Microsoft Academic Search

projects database

Factiva, press articles archive



Medline Pubmed

clinical trials database

online forums

Capacity to collect, parse and handle corpora from various arenas

Intervenir sur les échelles pour **dimensionner le sujet** :

- L'échelle THEMATIQUE (les sous-disciplines et champs technologiques concernés)
- L'échelle TEMPORELLE (la période)
- L'échelle GEOGRAPHIQUE (par ex. la ville, la région, le pays, le monde)

Espaces stratégiques analysés:

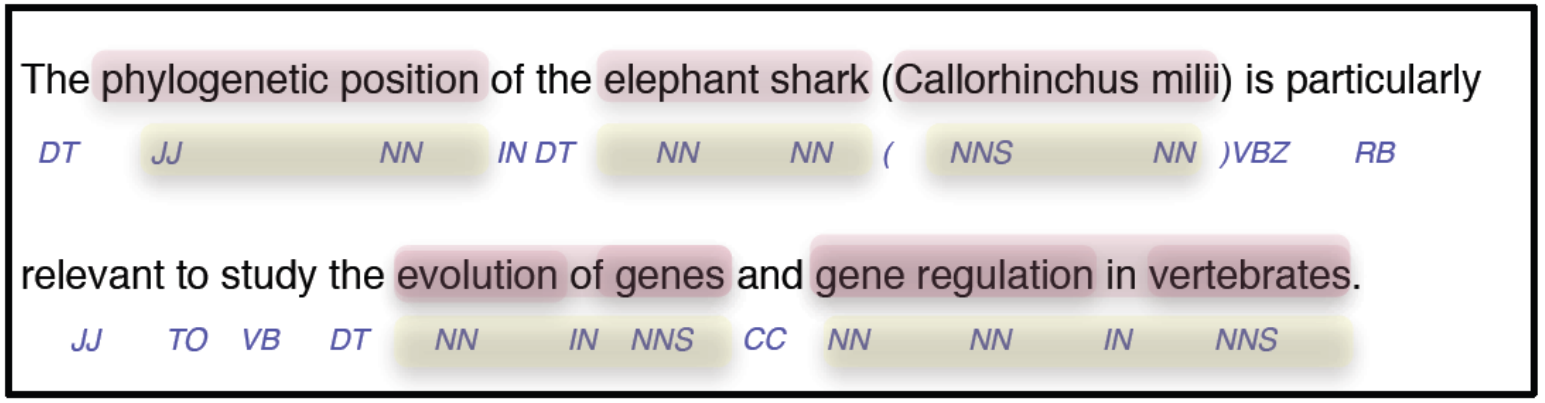
- L'espace INSTITUTIONNEL (par ex. les universités, entreprises et laboratoires publics centraux)
- La GEOGRAPHIE (par ex. les villes, les régions et les pays centraux)
- Les dynamiques SOCIO-SEMANTIQUES (association des cooccurrences de mots et des acteurs) et THEMATIQUE (disciplines, technologies…)

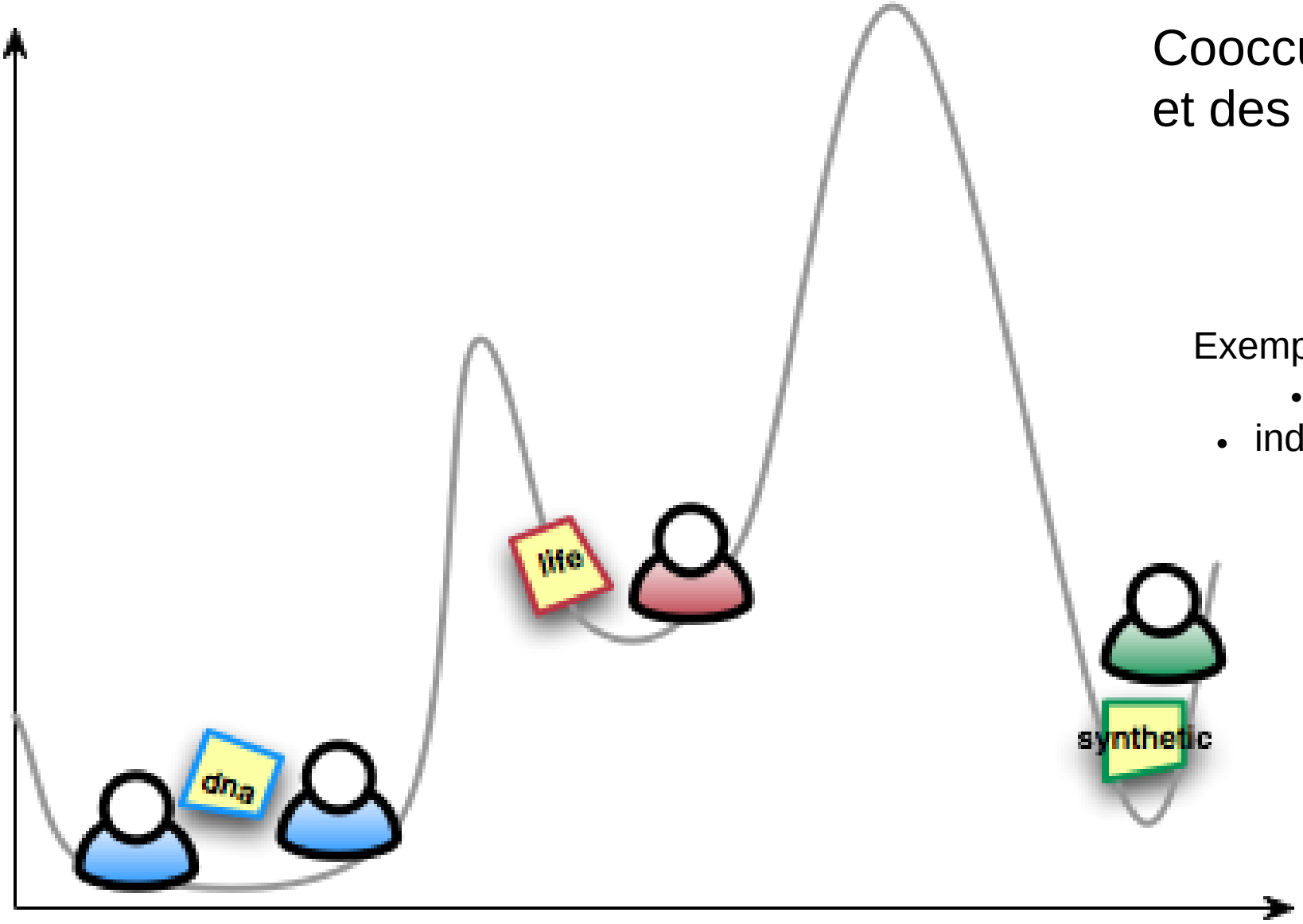
**Le choix des échelles permet notamment d'accroître la pertinence et de réduire le volume de données à extraire.*

Construire le paysage socio-sémantique

Le paysage socio-sémantique est reconstruit notamment par une EXTRACTION TERMINOLOGIQUE qui s'effectue en plusieurs étapes :

- Étiquetage morpho-syntaxique : associer aux mots des textes disponibles dans les corpus les informations grammaticales (le genre, le nombre…)
- Extraction des groupes nominaux (noms et adjectifs...)





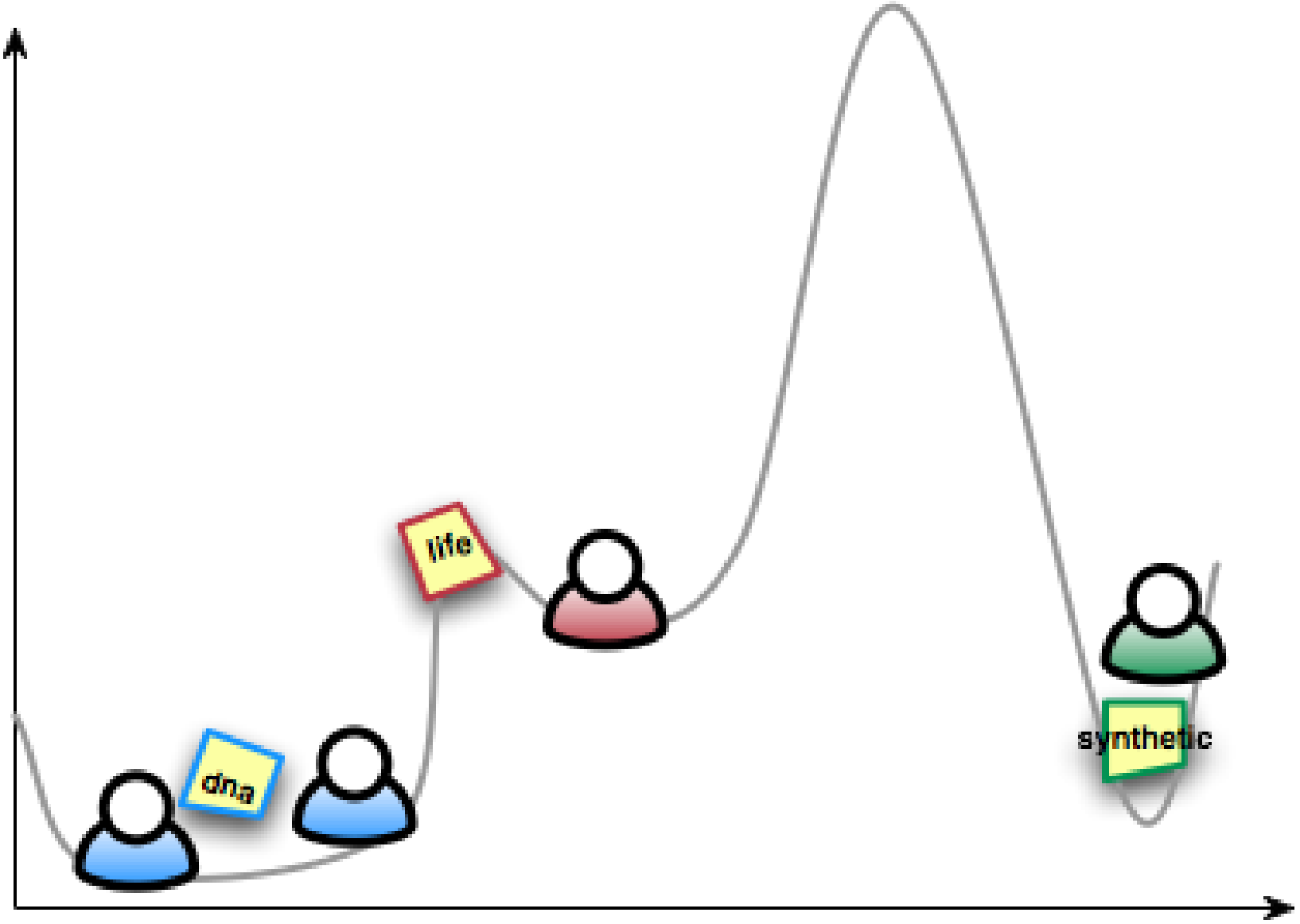
Landscape and its actors (t=1)

Cooccurrences des mots et des acteurs (individus, institutions…) par période.

Exemple : articles scientifiques

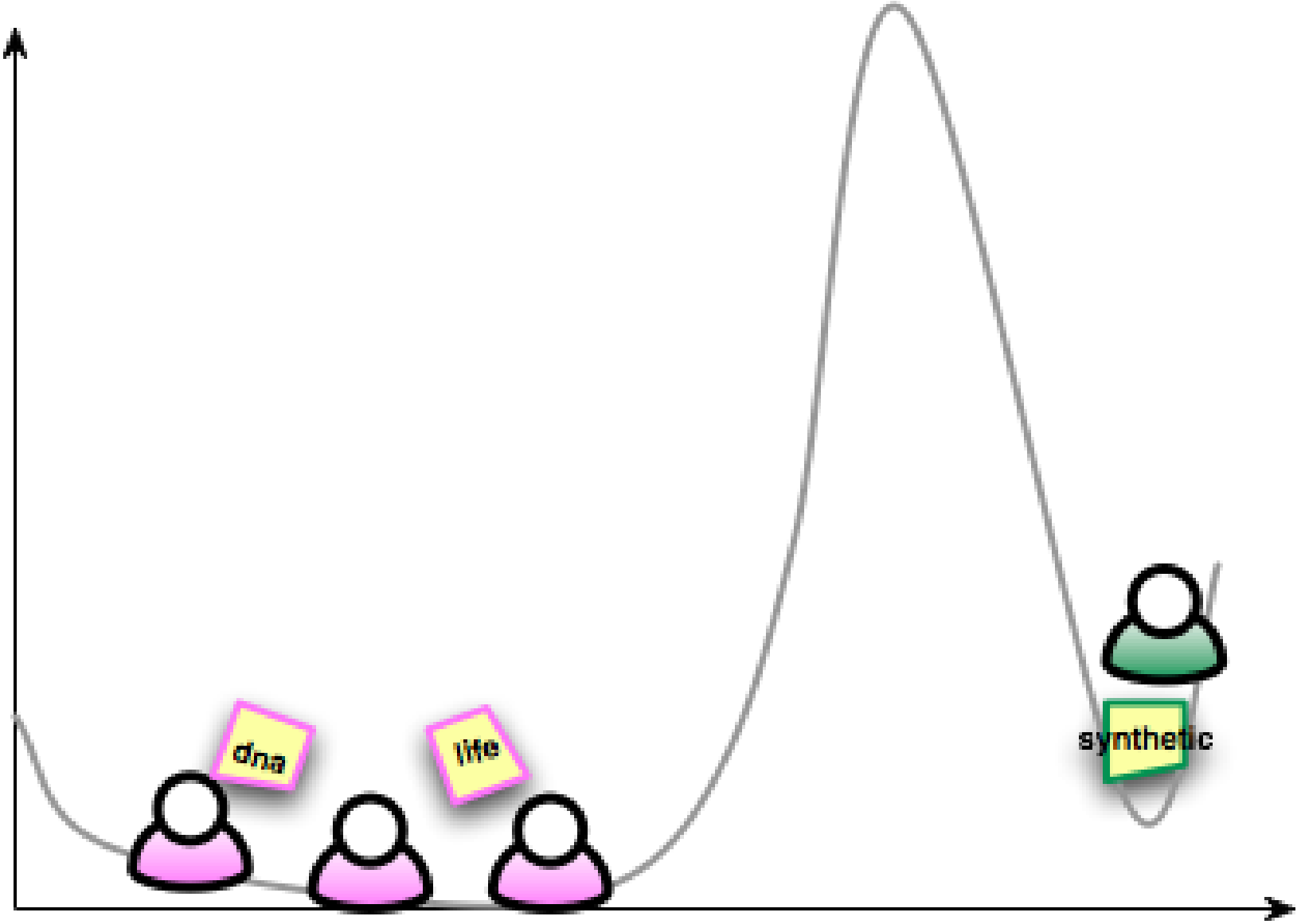
- texte : titres et abstract
- individus : auteurs dans les signatures

Déformation à t=2



Landscape and its actors (t=2)

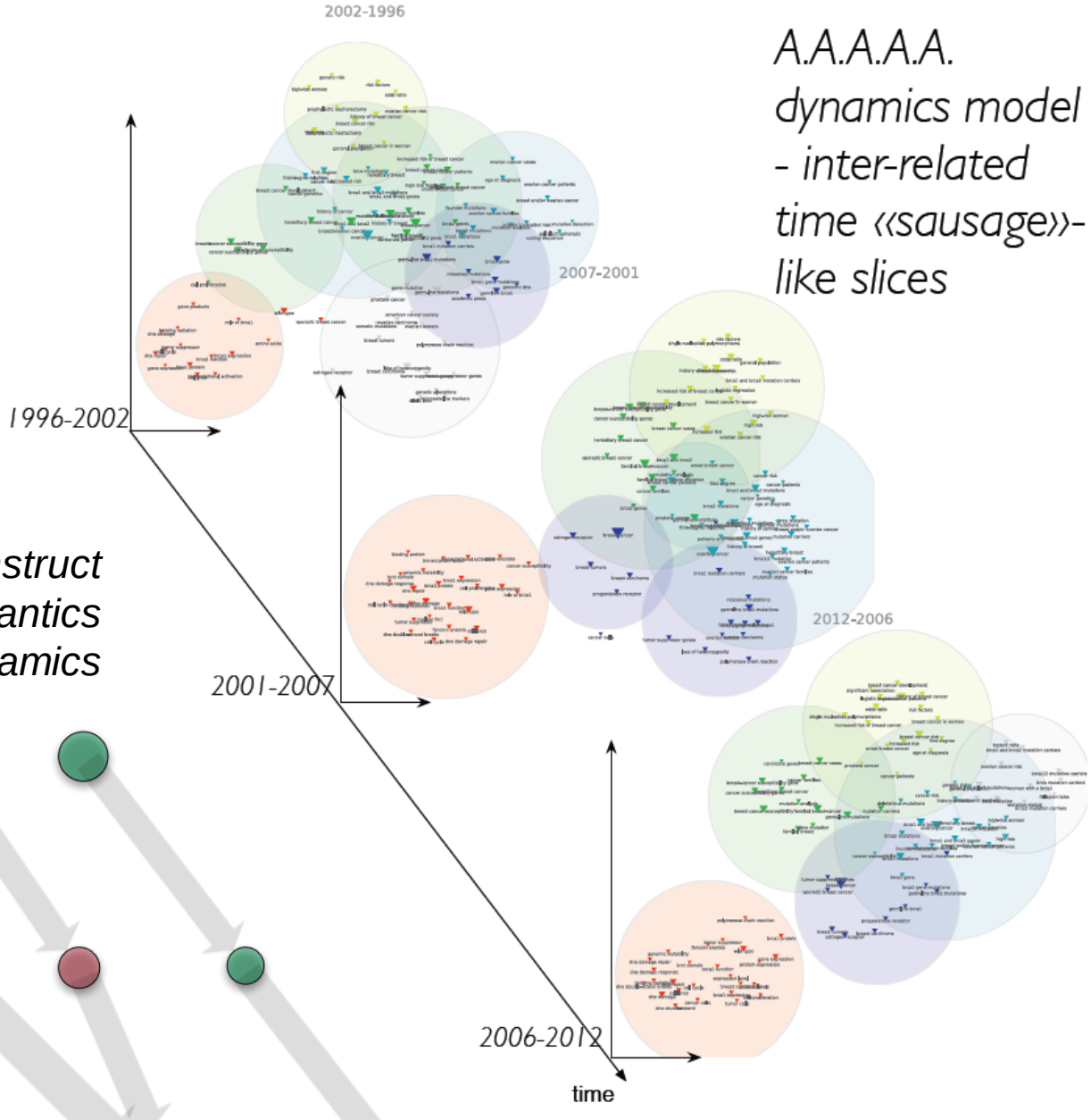
Déformation à t=3



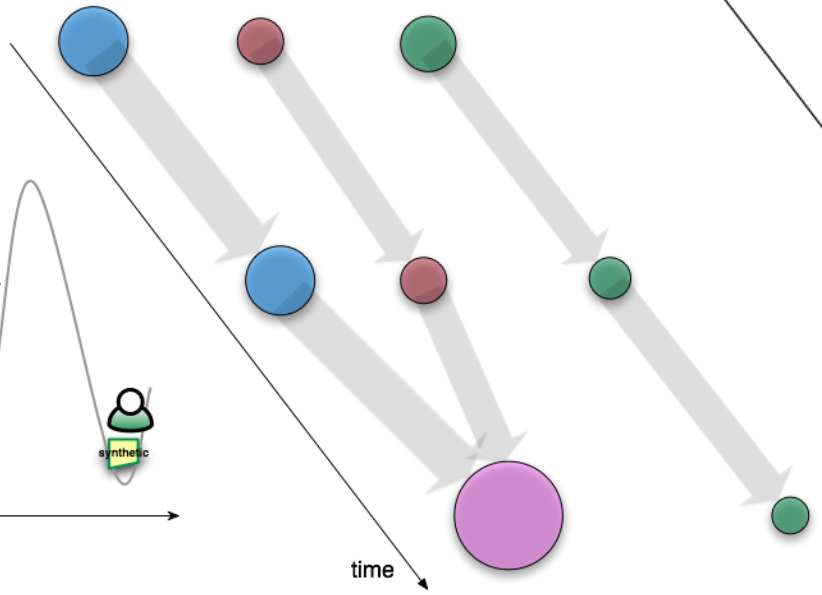
Landscape and its actors (t=3)

Reconstruction de la dynamique et des transformations

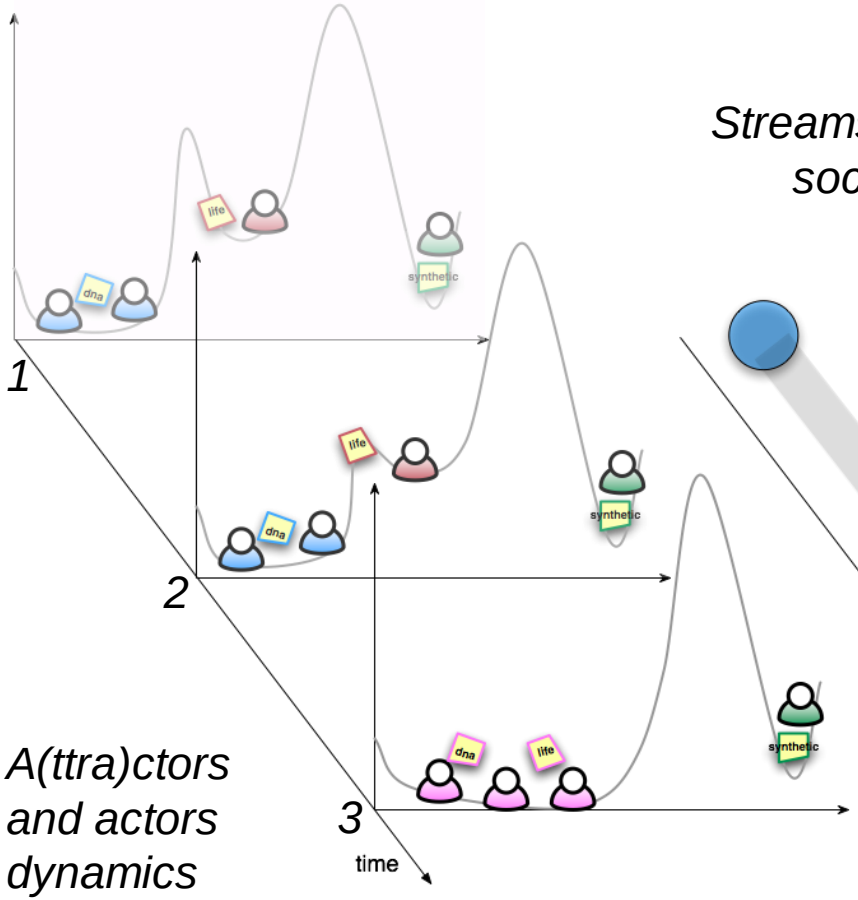
A.A.A.A.A. dynamics model - inter-related time «sausage»-like slices



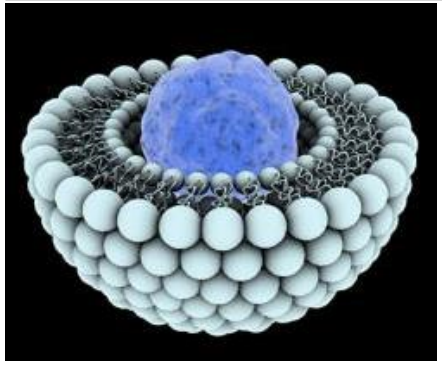
Streams reconstruct socio-semantics dynamics



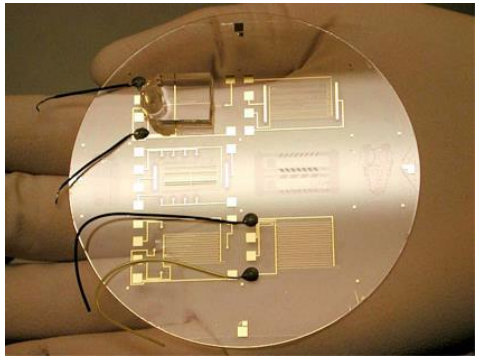
A(ttra)ctors and actors dynamics



Exemples avec un corpus scientifique concernant les nanobiotechnologies en France 2004-2013



Un liposome : les billes blanches sont les lipides phosphorescants. Ils sont liés entre eux par du cholestérol (petits fils). Le corps bleu au centre de la vésicule est le principe actif médicamenteux (doxuribicine est un ***produit toxique contre les tumeurs hépatiques****).*



Bio-puces (séquençage ADN), 2008, Philippe Houdy, La révolution des nanotechnologies, futura-sciences

Figure 8 : répartition des applications en nanobiotechnologies par sous-secteur industriel. Analyse menée à partir d'une centaine de start-up & industriels présents en France.

NanoThinking, Les nanotechnologies en France, 2013

De la recherche fondamentale aux applications

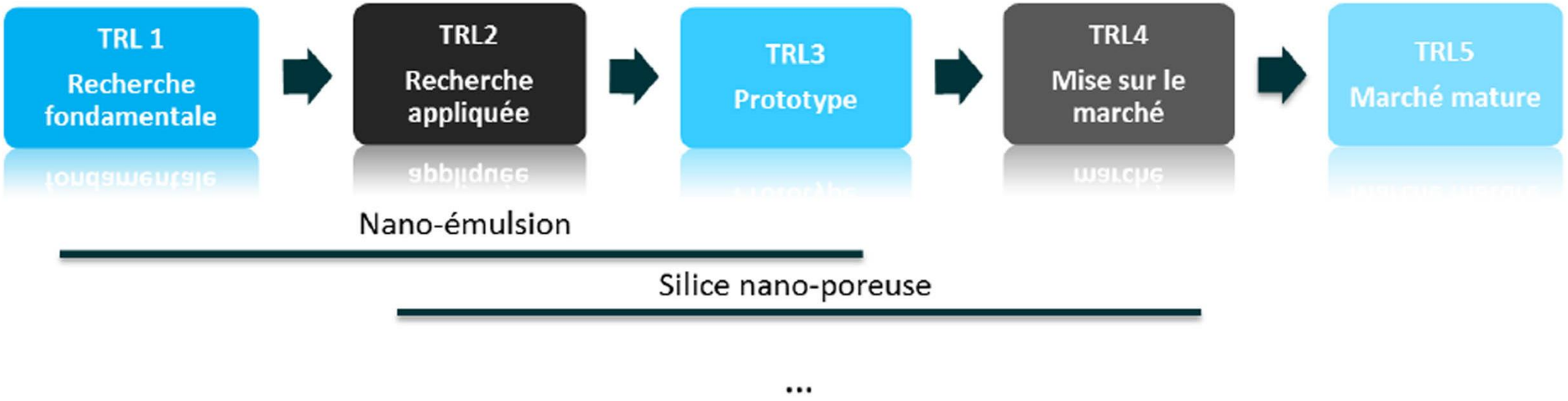


Figure 11 : niveaux de maturité pour différentes technologies.

NanoThinking, Les nanotechnologies en France, 2013

Financement du développement des nanobiotech en France

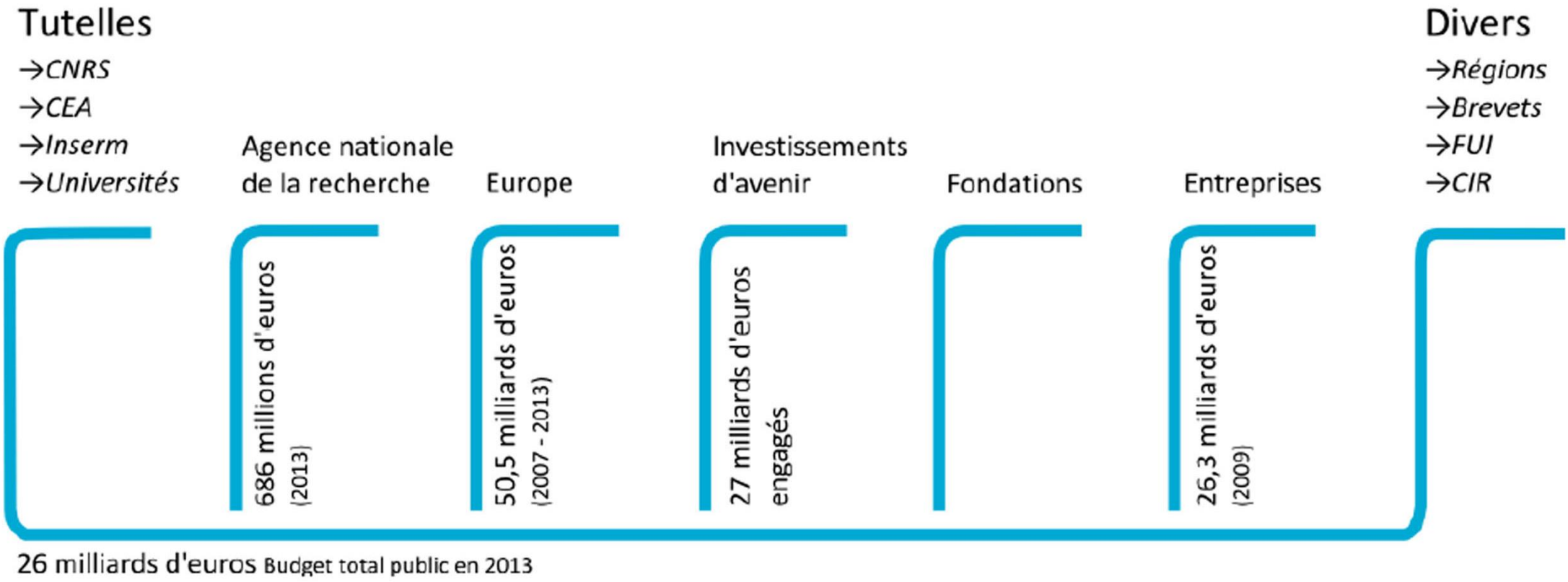


Figure 7 : mécanismes de financement de la recherche en France (source MESR).

NanoThinking, Les nanotechnologies en France, 2013

La nanobiotechnologie est donc récente et encore très orientée « recherche » !

Les nanobiotech dans les publications scientifiques

Advanced Search

Use field tags, Boolean operators, parentheses, and query sets to create your query. Results will appear in the Search History table at the bottom of the page.(Learn more about Advanced Search)

Example: TS=(nanotub* AND carbon) NOT AU=Smalley RE
#1 NOT #2 more examples | view the tutorial

```
(((TS=nano* AND (TS=bio* OR TS=molecular* OR TS="drug delivery" OR TS=protein* OR TS=cell* OR TS=biorecognition*)) OR TS=biochip*)
NOT (TS=nanosecond*)) AND PY=2004-2013 AND CU=France
```

Search

Restrict results by languages and document types:

All languages	All document types
English	Article
Afrikaans	Abstract of Published Item
Arabic	Art Exhibit Review

Booleans: AND, OR, NOT, SAME, NEAR

Field Tags:

TS= Topic	SG= Suborganization
TI= Title	SA= Street Address
AU= Author [Index]	CI= City
AI= Author Identifiers	PS= Province/State
GP= Group Author [Index]	CU= Country
ED= Editor	ZP= Zip/Postal Code
SO= Publication Name [Index]	FO= Funding Agency
DO= DOI	FG= Grant Number
PY= Year Published	FT= Funding Text
AD= Address	SU= Research Area
OG= Organization-Enhanced [Index]	WC= Web of Science Category
OO= Organization	IS= ISSN/ISBN
	UT= Accession Number

TIMESPAN

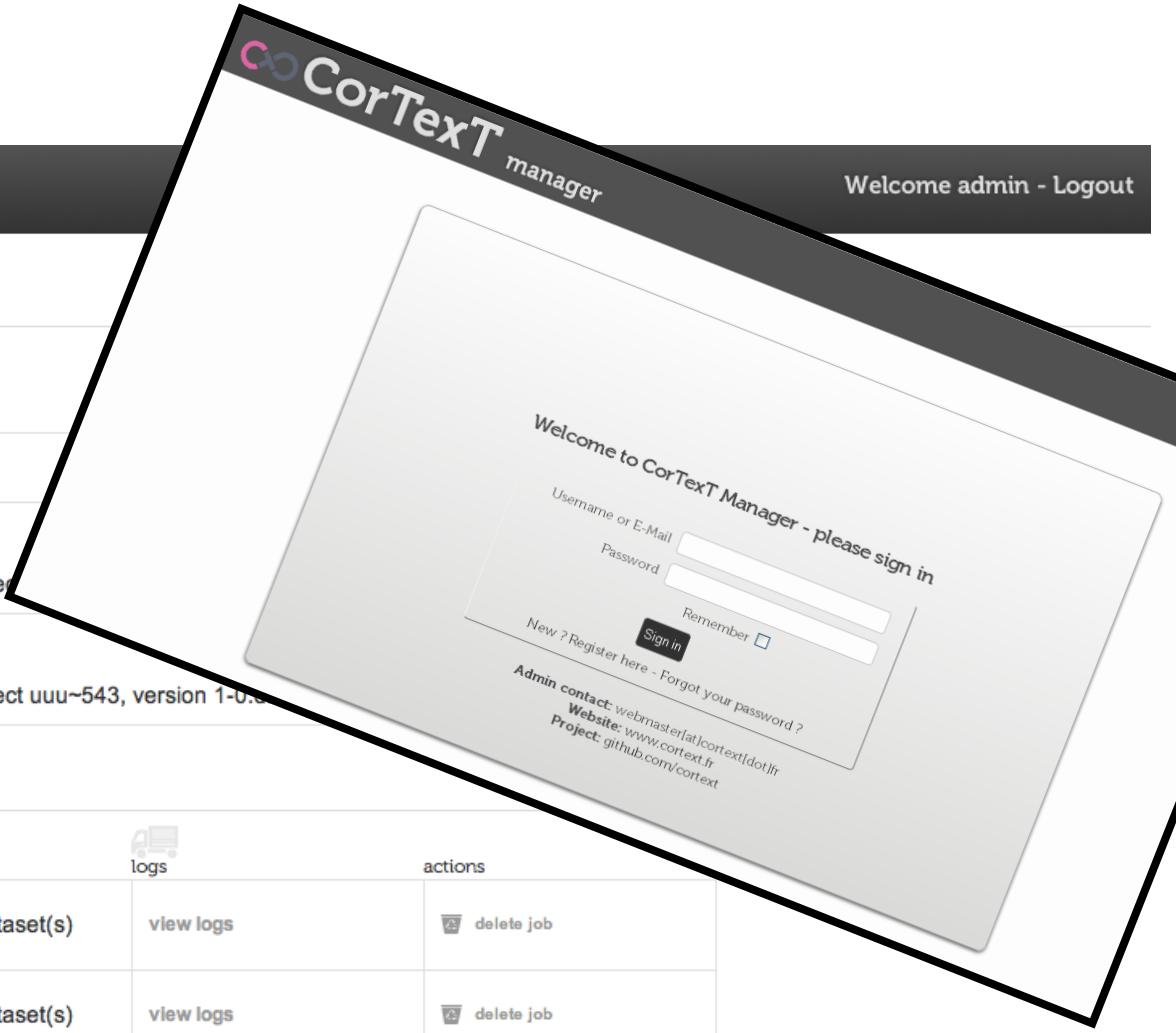
All years

From 1991 to 2014

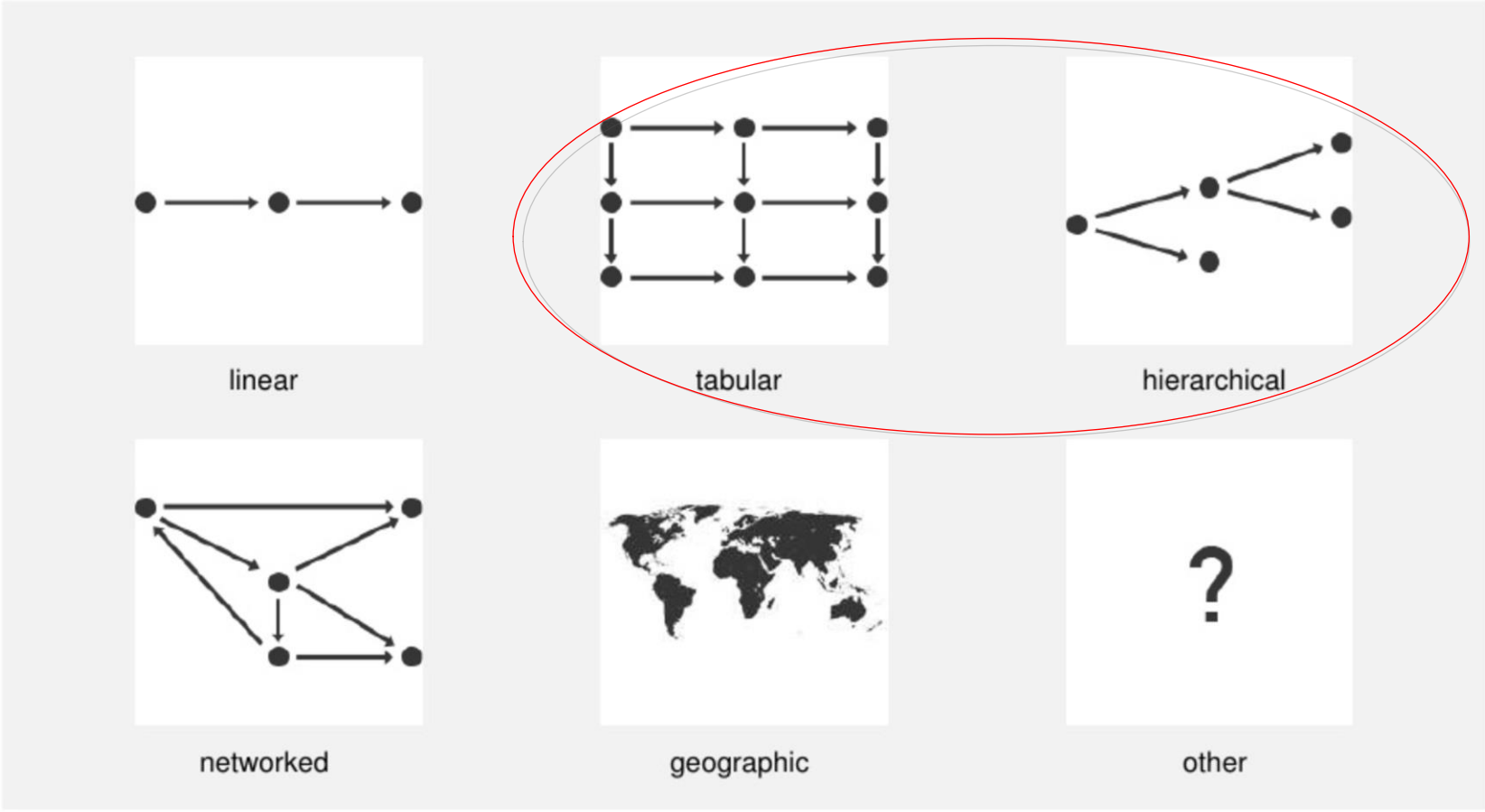
MORE SETTINGS

Search History:

Set	Results	Save History / Create Alert / Open Saved History	Edit Sets	Combine Sets (AND OR) Combine	Delete Sets Select All / Delete
# 1	10,684	(((((TS=nano* AND (TS=bio* OR TS=molecular* OR TS="drug delivery" OR TS=protein* OR TS=cell* OR TS=biorecognition*)) OR TS=biochip*) NOT (TS=nanosecond*)) AND PY=2004-2013 AND CU=France)) *AND* **LANGUAGE:** (English) *AND* **DOCUMENT TYPES:** (Article) *Indexes=SCI-EXPANDED Timespan=All years*	Edit		


CorTexT manager
Welcome admin - Logout
tox
[dashboard] [3 corpus] [3 scripts finished] [goto visualisations]
Add corpus
Start script
corpus
11/18/2011 at 02:31
uuu~651-1-uuu.db
db
105.176 Mo
06/23/2011 at 15:12
uuu~543-1-0.db
db
105.176 Mo
dataset - project uuu~543, version 1-0.
06/23/2011 at 15:11
uuu
zip
11.88 Mo
scripts [Last update : 2011-11-18 02:12:20]
Datasets
logs
actions
06/30/2011 at 16:20
651 data_parser -> uuu
View the 2 dataset(s)
view logs
delete job
06/23/2011 at 15:13
544 analysis -> uuu~543-1-0.db
View the 6 dataset(s)
view logs
delete job
06/23/2011 at 15:11
543 parser_science -> uuu
View the 1 dataset(s)
view logs
delete job
visualisations
db_viewer
uuu
db_viewer
uuu
CorTexT manager
Welcome to CorTexT Manager - please sign in
Username or E-Mail
Password
Remember
Sign in
New ? Register here - Forgot your password ?
Admin contact: webmaster[at]cortext[dot]fr
Website: www.cortext.fr
Project: github.com/cortext

Examine the Data



Trisnadi Kurniawan, *Infographics and Data Visualisation*, http://fr.slideshare.net/trisnadi/infographics-data-visualisation, 2009

Les publications

Analyse structurelle : **statistiques descriptives simples**

- **caractérisations d'ensemble;**
- indicateurs d'**activité;**
- indicateurs de **spécialisation**.

*Ces statistiques simples s'appuient sur des données au **format tabulaire** (tableaux sql et csv simples). Elles peuvent **exprimer une hiérarchie** (ex : pays / villes / discipline / Nbre de publications).*