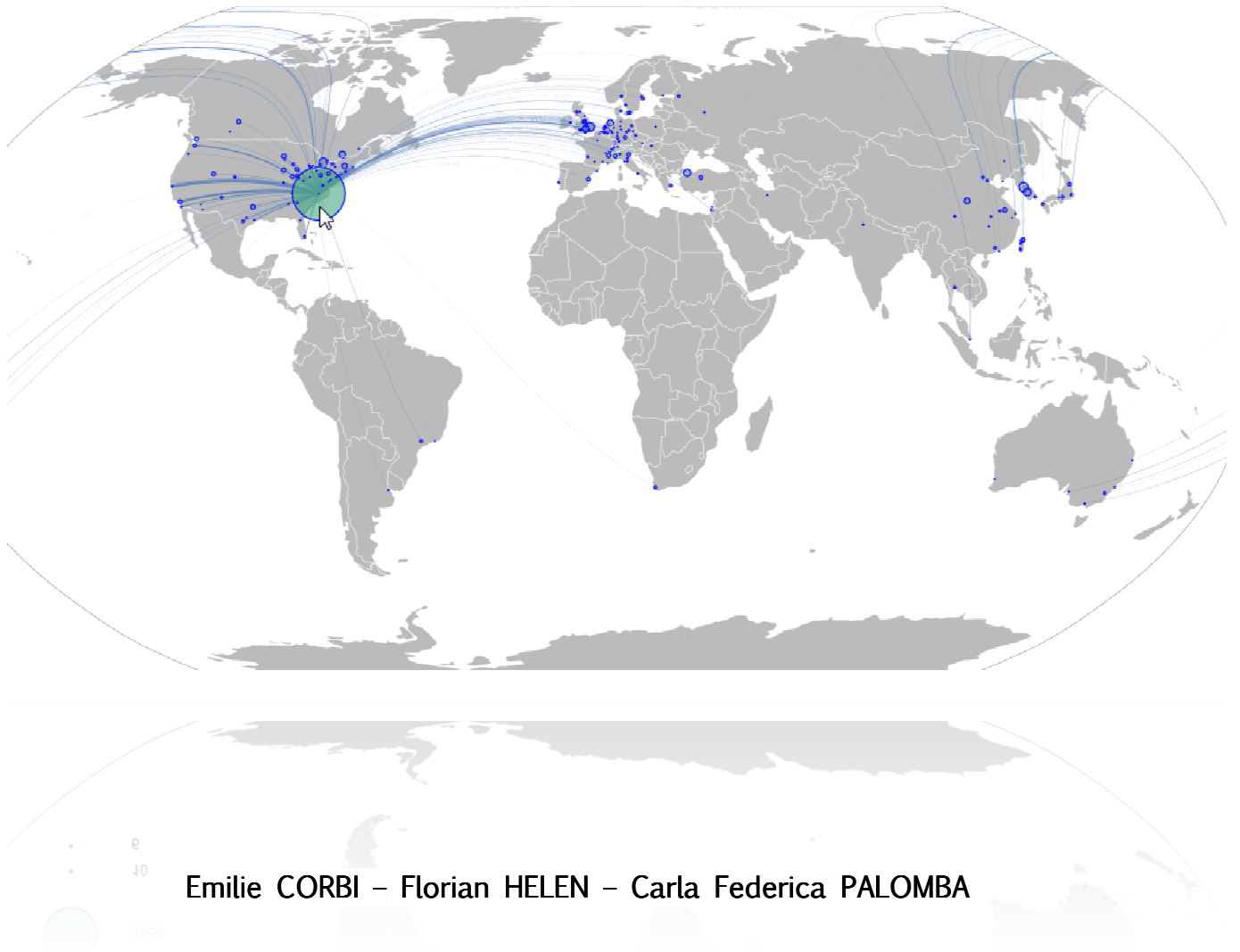


## M2 SIGMA – PROJETS GEOMATIQUES - MODULE U451

Prefiguration d'un portail de restitution du  
programme de recherche «Géosciences »



## SOMMAIRE

<b>1) INTRODUCTION</b>	<b>4</b>
<b>1.2) LA GESTION DU PROJET ET LES DEMARCHES MISES EN PLACE</b>	<b>5</b>
<b>2) CONTEXTE SCIENTIFIQUE DU PROJET: GEOSCIENCE</b>	<b>6</b>
<b>2.1) LES RESULTATS DEJA ILLUSTRÉS</b>	<b>8</b>
2.1.1) DIAGRAMME A BULLES	8
2.1.2) HISTOGRAMMES	9
2.1.3) DIAGRAMME ALLUVIAL.	10
2.1.4) CARTES	11
<b>3) CONCEPTION DU SITE</b>	<b>13</b>
<b>3.1) LE PUBLIC VISE</b>	<b>13</b>
PUBLIC 'SCIENCES ADDICTS'	14
PUBLIC CHERCHEURS	15
PUBLIC DES UTILISATEURS INTERESSES	15
PUBLIC AUTEURS/CHERCHEURS	16
PUBLIC BUTINEUR	17
PROFIL ADMINISTRATEUR	17
<b>3.2) BESOINS ET FONCTIONNALITES</b>	<b>18</b>
<b>3.3) ARCHITECTURE</b>	<b>19</b>
3.3.1) CONTENU	19
3.3.2) STRUCTURATION	19
<b>3.4) SOLUTIONS TECHNIQUES</b>	<b>22</b>
3.4.1) CMS & FRAMEWORKS	23
3.4.2) LES OUTILS DE VISUALISATION : BIBLIOTHEQUES ET WEB SERVICES	30
<b>4) PROTOTYPE ET PHASE DE TEST</b>	<b>34</b>
<b>4.1) LA STRUCTURATION DES DONNEES</b>	<b>34</b>
4.1.1) LES DONNEES	34
4.1.2) LA METHODE	35
<b>4.2) LES DIFFERENTES MODALITES DE VISUALISATION</b>	<b>37</b>
4.2.1) HIERARCHICAL BAR CHART	37
4.2.2) ZOOMABLE PACK LAYOUT	39
4.2.3) COLLAPSIBLE TREE LAYOUT	40
4.2.4) LA CARTE DES COLLABORATIONS	40
<b>4.3) PROPOSITIONS D'OPTIMISATION : LE LIEN DIRECT ENTRE LA BD ET LES VISUALISATIONS</b>	<b>43</b>

---

<b>5) CONCLUSIONS</b>	<b>44</b>
<b>BIBLIOGRAPHIE – WEBOGRAPHIE</b>	<b>47</b>
<b>ANNEXES</b>	<b>52</b>

## 1) Introduction

Le document présent illustre les démarches mises en place dans le cadre du projet « Portail Géoscience », encadré par L. Jégou et M. Maisonobe, dont l'objectif a été de préfigurer le portail Internet de restitution du programme de recherche "Géoscience".

Le portail vise à présenter les résultats du programme, à la fois méthodologiques et scientifiques. Il doit aussi prendre en compte le fait que les recherches étant encore en cours, de nouveaux résultats vont être ajoutés. Le site doit donc inclure une dimension évolutive.

Dans un premier temps, nous avons étudié le corpus d'articles issu du programme, afin de nous approprier les thématiques abordées et de bien comprendre les enjeux de présentation et de communication des résultats. Ainsi le 2<sup>ème</sup> chapitre illustre d'abord le contexte dans lequel le projet s'est déroulé. Ensuite sont analysés les cartes et graphiques illustrant les articles et le site CoSciMo. Celui-ci présente les résultats issus de l'analyse des collaborations scientifiques. Dans le deuxième paragraphe donc, nous avons détaillés les données et les idées véhiculées par les graphiques et les cartes ainsi que les possibles modifications à faire pour les rendre interactifs. Cette démarche a été effectuée dans l'optique de proposer des pistes d'amélioration, d'homogénéisation et de diffusion des résultats.

La phase suivante approche le cœur du sujet, c'est-à-dire la conception de la structure du site. Le 3<sup>ème</sup> chapitre décrit, à partir de l'analyse du public visé et des besoins qui en découlent, les possibilités de structure et les modalités de navigation que nous avons envisagées.

Enfin, la dernière partie du travail a été plus « pratique » et applicative. A partir des réflexions faites sur les modalités de visualisation, nous avons exploré les possibilités offertes par la bibliothèque JavaScript D3 (d3js.org) sur des données hiérarchisées de type père/fils. Cette structure, spécifique à la bibliothèque D3, permet en effet de créer des graphes interactifs à dimensions multiples et de présenter en une seule représentation ce qui devrait l'être par plusieurs graphiques à une dimension. Toutefois elle requiert une arborescence spécifique des données, dont l'implémentation ne peut pas être faite manuellement sur des grosses bases de données. Pour ce faire, nous avons recherché et testé des outils de structuration automatique des tables, des modalités de visualisation de ce type de données, et enfin des possibilités de connexion entre la base de données et les scripts D3.js.

## 1.2) La gestion du projet et les démarches mises en place

Le projet s'est déroulé sur 5 semaines et avait, comme mentionné précédemment, l'objectif de mettre en place les démarches nécessaires à la conception d'un site de présentation des résultats issus du programme de recherche « Géoscience ».

Dans un souci de clarté nous avons utilisé l'hexagramme de Quintilien pour recueillir et synthétiser les informations concernant le projet.

QQQOCCP	
<i>Qui</i>	Emilie CORBI, Florian HELEN, Carla Federica PALOMBA, Commanditaires : Tuteur : Laurent JEGOU (Tuteur au sein du master SIGMA) Membre de l'équipe de recherche: Marion Maisonobe (doctorante).
<i>Quoi</i>	Un cahier des charges préalable à l'implémentation d'un site internet et prototypage du site.
<i>Quand</i>	20 janvier au 21 février : période de mini-projet.
<i>Où</i>	Salle informatique du département de géographie de l'université du Mirail.
<i>Comment</i>	Moyen : machine serveur dédiée aux tests (Windows XP) disponible à l'université du Mirail; ordinateurs en salle informatique et ordinateurs personnels.
<i>Combien</i>	Aucun budget.
<i>Pourquoi</i>	Présenter les résultats du programme de recherche Géosciences.

Dans un premier temps nous avons élaboré un Diagramme de Gantt prévisionnel (cf. Annexe 4), afin de planifier et organiser le temps de travail.

On peut subdiviser le projet en trois phases principales : une première phase d'appropriation du sujet et d'analyse de l'existant. Dans une deuxième phase, représentant le cœur du travail, nous nous sommes penchés sur la conception du site, à travers la définition du public visé, des besoins à satisfaire et des solutions techniques à utiliser. En parallèle de cette phase, et pour soutenir nos choix techniques, nous avons consacré la dernière phase du projet à réaliser des prototypes de graphiques et de cartographies, afin de tester et valider nos analyses sur l'existant et nos réflexions sur la Data visualization.

L'étape finale de rédaction du rapport de projet nous a servi pour mettre à plat les concepts élaborés et définir plus concrètement les propositions finales à présenter.

Tout au long du projet, des réunions ont eu lieu, dans le but de présenter les différentes étapes du travail par nos encadrants, et prioriser les actions à mener. Les comptes-rendus des réunions sont présentés en Annexe 5.

## 2) Contexte scientifique du projet: Géoscience

Géoscience est un programme financé par l'Agence Nationale de la Recherche (2010-2013), qui s'est intéressé à l'analyse de l'évolution de la production scientifique. La recherche a été menée suivant deux approches différentes.

La première relève de l'histoire de la production scientifique et des espaces de production ; en d'autres termes les chercheurs ont reconstruit les modalités et les temps de création des centres de production scientifique dans plusieurs pays. C'est donc d'une perspective plutôt historique et sociologique du sujet.

La seconde se concentre sur l'évolution quantitative et spatiale de la production scientifique. A partir de la géolocalisation des articles scientifiques, la recherche visait à étudier et à analyser la répartition dans le temps et dans l'espace de la production scientifique. Il s'agit des informations relatives aux publications scientifiques mondiales sur deux séries temporelles : 1999-2001 et 2006-2008.

Cette partie du programme a été portée par le laboratoire LISST de Toulouse et le laboratoire OST de Montréal.

Les données ont été dans un premier temps fournies par le laboratoire CIRST de l'université de Montréal, qui travaille sur la bibliométrie, à partir des données privées de la base Web of Science, appartenant à Thompson et Reuters. Les données extraites ont été comptées et regroupées en fonction de trois niveaux géographiques : Ville, Province<sup>1</sup> et Pays. Une première série temporelle a été étudiée pour les années 1978, 1988, 1998 et 2008, afin d'apporter une analyse de long terme.

Cette première série d'études, qui avait comme but de se pencher sur la problématique de la déconcentration de l'activité scientifique, a été fondamentale pour les choix méthodologiques effectués dans la suite du projet. C'est en étudiant les trois champs géographiques que les chercheurs se sont aperçus de l'impossibilité de comparer, au niveau de la ville, des entités hétérogènes du point de vue administratif et de population. Ils ont alors conçu et mobilisé une méthode d'agrégation des villes en agglomérations, qui sera expliquée plus loin.

Dans un deuxième temps, et une fois surmontée la problématique méthodologique, les auteurs se sont penchés sur l'analyse des relations entre les villes et l'étude de l'évolution de la production de ces centres de savoir.

---

<sup>1</sup> Il s'agit ici du second niveau administratif, qui peut considérablement différer d'un pays à l'autre. Dans le cas des Etats-Unis, par exemple, ce sont les Etats, en France les départements ou les régions...

Pour ce faire, ils ont eu recours aux données fournies par l'Observatoire des Sciences et Techniques à Paris, issues là encore du Web of Science, qui établit un rapport annuel sur l'état de la recherche en France. Dans cette deuxième phase les données utilisées concernent deux périodes de trois ans : 1999-2000-2001 et 2006-2007-2008. Des moyennes mobiles ont été appliquées à ces données afin de lisser les phénomènes épisodiques et de les rendre comparables.

Aujourd'hui la base dont dispose l'équipe toulousaine dispose de 8 millions de notices de publication, associées à environ 14.5 millions d'adresses.

L'innovation du projet réside dans le fait que pour la première fois les données du WoS ont été géo-localisées à une échelle fine : celle de la ville. Cela a permis d'une part d'analyser la production du savoir et son évolution à différentes échelles géographiques, du niveau local de la ville à un niveau global. Les études ont d'autre part porté sur les collaborations entre villes à l'échelle mondiale. Afin d'homogénéiser les données relatives aux zones urbaines et les rendre comparables, une méthode de délimitation des aires urbaines a été mise au point. Il a dès lors été possible de procéder à l'agrégation des données par localités et ainsi créer de nouvelles "agglomérations scientifiques". La méthode repose sur des critères de regroupement permettant de quantifier et déterminer des zones de densité homogène de population et une cohérence scientifique, applicables au monde entier.

Grâce à cette méthodologie, les potentialités d'exploitation et d'analyse de la base ont été démultipliées. Citons, à titre d'exemple, deux nouveaux axes de recherche en cours : la production scientifique par grandes familles de disciplines scientifiques et l'analyse comparée de la visibilité de la production scientifique.

Les membres du programme ont produit des cartes et des graphiques qui illustrent leurs travaux. Nous sommes partis de l'analyse de ces représentations pour engager notre réflexion sur le site et proposer d'autres modalités de présentation des résultats.

Ensuite, nous avons choisi certaines thématiques qui n'ont pas encore été illustrées et nous avons proposé pour celles-ci des modalités de visualisation des données.

Les paragraphes suivants illustrent les démarches d'analyse effectuées. Dans le dernier chapitre, nous illustrerons les tests effectués et les graphiques produits pour certaines de ces thématiques.

## 2.1) Les résultats déjà illustrés

Étant donné que pour une partie des résultats une plateforme de visualisation à déjà été mise en place (cf. [www.CoSciMo.fr](http://www.CoSciMo.fr)), cela a constitué la base de travail pour nos futures propositions. Cette partie du projet privilégie une approche « horizontale » à travers l'analyse des interactions entre lieux pour l'étude des coopérations scientifiques.

Quatre modes d'exploration interactive des données sont proposées : une représentation matricielle, un diagramme à corde, une représentation cartographique et un graphe de réseaux, chaque modalité présentant et mettant en avant un aspect différent des résultats obtenus. Même si certaines limites peuvent être soulignées sur la plateforme, celle-ci propose déjà des représentations interactives ; nous avons donc préféré nous focaliser sur des propositions inédites et laisser aux intervenants futurs les améliorations à faire sur l'existant.

Reste donc à explorer une approche « verticale » des appartenances territoriales, afin d'étudier les entités locales intrinsèquement, et en fonction de l'évolution de la production, cela dans le but de valoriser la présentation des résultats concernant la déconcentration des centres du savoir.

Pour chaque graphique, nous sommes donc partis du message véhiculé par la représentation et nous avons exploré la façon plus adaptée de le rendre dynamique à travers l'ajout de fonctions interactives.

### 2.1.1) DIAGRAMME A BULLES

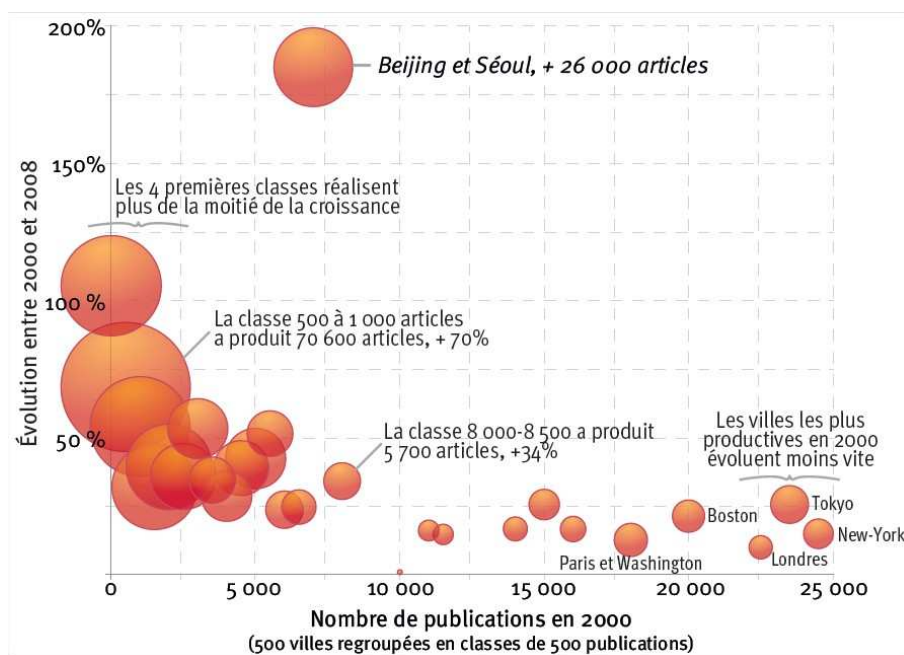


Fig. 1 : Diagramme à bulles illustrant l'évolution de la production scientifique



**Description** : Le graphique illustre en abscisses le nombre de publications en 2000, et en ordonnées le taux d'évolution de la production entre les deux périodes ; la taille des symboles est proportionnelle à la somme des publications par classe de 500 articles en 2008.

**Message** : les villes qui produisaient le moins en 2000 ont le plus évolué; on remarque notamment le rôle de Pékin et Seoul<sup>2</sup> dont la classe d'appartenance affiche un taux d'évolution de presque 200%.

**Propositions** : rendre le graphique plus clair en groupant les classes par taux d'évolution (4 niveaux : faible, moyen, important, très important) et en attribuant une couleur à chaque groupe ; ajouter des info bulles qui affichent pour chaque classe la ville qui a la plus forte/faible évolution ; ajouter des pop-up explicatifs ; prévoir des fonctions de zoom et de déplacement pour mieux explorer le graphique ; envisager un onglet de sélection par classes ou par groupes d'évolution.

### 2.1.2) HISTOGRAMMES

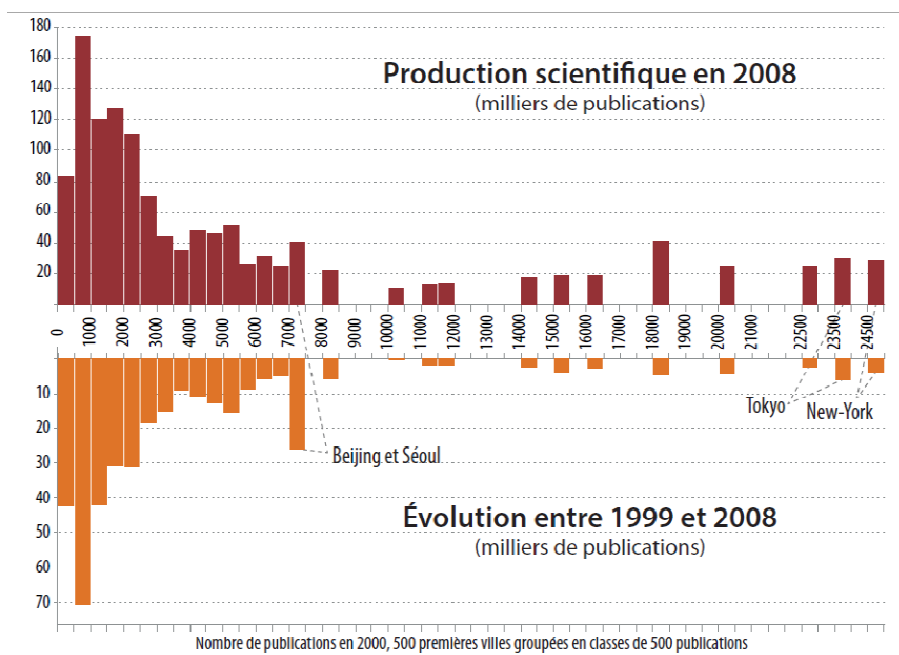


Fig. 2 : Histogramme illustrant la production scientifique et son évolution

**Description** : Le graphique illustre en abscisses le nombre de publications en 2000, en ordonnées dans la partie supérieure du graphique la somme des publications par classe de 500 articles en 2008, en ordonnées dans la partie inférieure du graphique l'évolution de la production entre les deux périodes.

<sup>2</sup> Et plus en général des villes qui produisent entre 5000 et 10000 articles dans la première période

*Message* : les villes qui produisaient le moins en 2000 ont le plus évolué ; on constate que la production des grandes villes « scientifiques » augmente moins vite que celle des petits centres.

*Propositions* : ajouter des infos bulles qui affichent pour chaque classe la ville qui a la plus forte/faible évolution ; prévoir le passage à une visualisation hiérarchisée par niveaux de visualisation, (du type « hierarchical bars<sup>3</sup> » de la bibliothèque D3) : en cliquant dans l'histogramme supérieure on pourrait accéder aux données de production de chaque classe, visualisées par ville ; sur l'histogramme inférieur aux données détaillées sur l'évolution. Une alternative pourrait être d'utiliser des histogrammes de croisement des données du type « crossfilter<sup>4</sup> » (bibliothèque js D3) pour avoir la possibilité d'appliquer des filtres par année, par taux de croissance, etc.

### 2.1.3) DIAGRAMME ALLUVIAL.

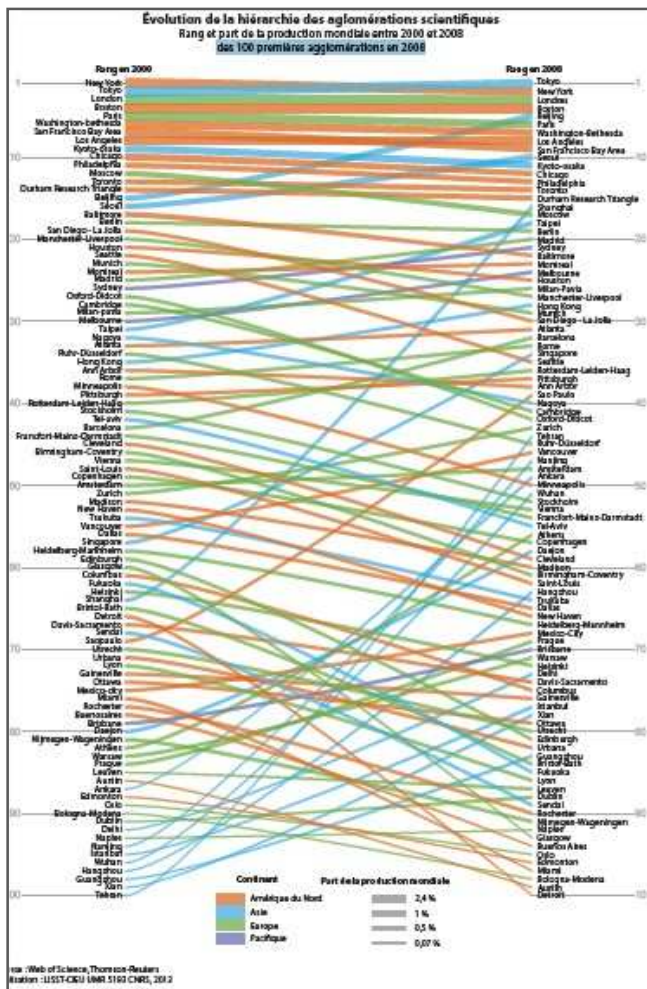


Fig. 3 : Evolution du rang des 100 premières villes dans le monde

<sup>3</sup> <http://bl.ocks.org/mbostock/1283663>

<sup>4</sup> <http://square.github.io/crossfilter/>

**Description** : Les ordonnées à gauche indiquent le rang des 100 premières agglomérations en 2000 en fonction de la production scientifique, les ordonnées à droite le rang des mêmes villes en 2007. L'épaisseur des cordes est proportionnelle à la part de production mondiale en 2000, tandis que la couleur désigne le continent d'appartenance.

**Message** : Montrer l'évolution du classement des villes par rapport à la production d'articles. On distingue en particulier la forte évolution des agglomérations qui en 2000 avaient un rang inférieur à 50.

**Propositions** : ajouter des onglets de sélection par classes d'évolution et par continent afin d'avoir deux échelles de visualisation, mondiale et nationale ; opter pour une couleur grise des cordes, dans le but de mettre en évidence avec des couleurs vives les cordes sélectionnées avec la souris (click ou survol)<sup>5</sup>.

#### 2.1.4) CARTES

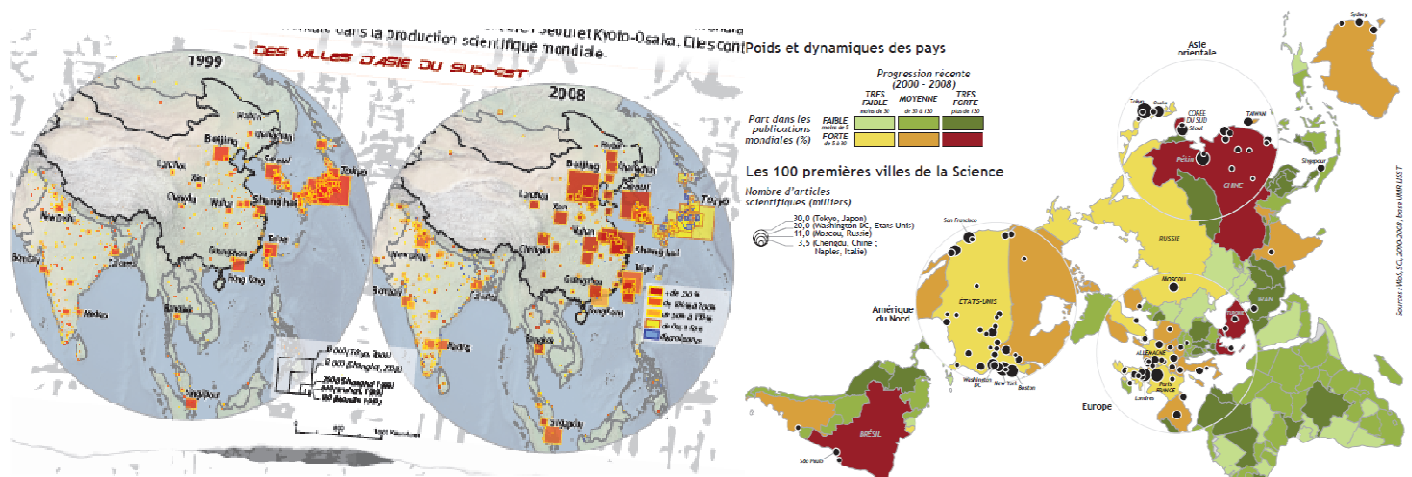


Fig. 4 : cartographies illustrant la déconcentration de la production du savoir en Asie du SE et à l'échelle mondiale

La cartographie interactive constitue le cœur du projet, soit par rapport aux sujets abordés soit par rapport aux méthodologies appliquées. En outre, les cartes constituent le moyen de visualisation le plus direct et efficace pour montrer des données géo-référencées. Nous sommes donc partis des quelques cartes produites pour des posters ou des ouvrages et nous avons essayé d'élaborer une réflexion globale sur cet outil de visualisation.

**Description** : à l'échelle internationale, les données doivent porter sur la part de chaque pays dans les publications mondiales, sur les quantités brutes et sur le taux

<sup>5</sup> Sur l'exemple du graphique à courbes du site: <http://projects.flowingdata.com/life-expectancy/>

de croissance de la production. Le niveau national doit indiquer pour chaque agglomération la valeur de production brute, ainsi que le taux d'évolution entre les deux périodes analysées. Il doit être possible d'opérer une comparaison à l'échelle globale des 100 agglomérations qui produisent le plus. D'autre part, il faut envisager des focus sur des pays spécifiques en visualisant les différentes disciplines et les typologies des centres de production (universités, écoles, IUT, etc. ).

*Message* : visualiser les pays qui évoluent le plus dans le monde et ceux qui voient leur production diminuer ; visualiser les villes qui évoluent le plus dans une zone spécifique ou au niveau global ; identifier les centres d'excellence des différentes disciplines ; analyser la croissance d'un pays en fonction d'une discipline spécifique ; comparer l'évolution des divers centres de production à l'intérieur d'un pays.

*Propositions* : deux propositions sont possibles. Créer une carte unique qui réunisse l'ensemble des données, visualisables en fonction du niveau de zoom et/ou d'une sélection par onglets. Cette solution se rapprocherait de l'option choisie par le site « Orbis »<sup>6</sup>, qui présente le résultat d'une recherche historique concernant les réseaux de transport dans l'Antiquité sur une carte unique. Toutefois, cette proposition semble problématique compte tenu de la grande quantité de données à gérer et à visualiser.

L'alternative consiste à créer plusieurs cartes qui illustrent moins d'informations et diviser par thématique ou selon une approche géographique Pays/Villes. Prévoir des onglets de sélection par années de production et, éventuellement, par disciplines. Ajouter un effet zoom/loupe animé<sup>7</sup>, et des infos bulles d'explication (exemple : nom d'une ville=nombre de publications). Cette dernière est l'approche adoptée par le site « Mapping the republic of letters<sup>8</sup> » qui présente les cartes en fonction des articles qu'elles illustrent.

---

<sup>6</sup> <http://orbis.stanford.edu/>

<sup>7</sup> <http://bl.ocks.org/mbostock/3828981> - <http://bl.ocks.org/mbostock/3795048>

<sup>8</sup> <http://republicofletters.stanford.edu>

### 3) Conception du site

Après avoir pris connaissance du contenu du programme, nous avons entamé une réflexion sur les fonctions, les besoins et la structuration du site au regard du message à véhiculer et des outils techniques disponibles.

Le contenu du site veut refléter un esprit novateur et mettre à profit les outils de data visualization actuels. La data visualization permet de transmettre de l'information statistique et géographique de manière interactive et dynamique, pour que l'utilisateur s'approprie aisément le sujet. Le but est donc de simplifier la complexité des données et de synthétiser un contenu de recherche ardu de prime abord.

Cette problématique a constitué le fil conducteur de toute la démarche de conception du site, depuis le public visé jusqu'à la phase de test.

#### 3.1) Le public visé

Afin de pouvoir cerner au mieux les besoins en termes de navigation, nous avons réfléchi aux différents types de personnes qui seraient amenées à visiter le site. Définir leurs buts, leurs besoins, leurs comportements est une première étape essentielle pour réfléchir à la conception du site. Nous avons donc déterminé des cas d'utilisation et des histoires de navigation.

Plusieurs types de publics ont ainsi été identifiés :

- le grand public intéressé par les thèmes scientifiques en général,
- les scientifiques intéressés par ces questions,
- le public plus large intéressé par les politiques de développement scientifique (journalistes, étudiants et chercheurs en sciences politiques, en économie, en sociologie...),
- auteurs/chercheurs qui font partie du projet et qui pourraient alimenter le site,
- administrateur(s) du site scientifiques et techniques.

En résumé, on peut définir trois grandes catégories d'utilisateurs. D'une part, un public technique auquel il faut fournir un accès aux données détaillées ainsi qu'à la démarche méthodologique et qui doit avoir la possibilité d'approfondir certaines thématiques en fonction d'intérêts spécifiques. D'autre part, on considère un public généraliste, n'ayant pas une maîtrise approfondie du sujet, qui doit pouvoir rapidement saisir la problématique et accéder aux informations précises recherchées.

On peut enfin définir une catégorie comprenant les administrateurs et chercheurs/auteurs qui devront gérer et/ou alimenter le site.

### *Public 'sciences addicts'*

Qui : Un public intéressé par les questions ayant trait à la science de manière générale.

Raison de la présence sur le site : Il visite le site suite à une recherche sur les thématiques qui touchent au sujet du site ou par un article sur ces recherches. Il a donc déjà une idée de ce que contient le site mais vient pour le découvrir, sans but spécifique. D'un naturel curieux, il va passer un peu de temps à naviguer s'il trouve un potentiel intérêt au contenu, il doit donc rapidement pouvoir se faire une idée des différents contenus.

Modalité de navigation : Il doit saisir la problématique du site et donc avoir une idée globale du contenu du site dès la page d'accueil. Le site doit lui paraître riche mais clair. La difficulté du propos ne lui fait pas peur. En fonction *d'intérêts scientifiques spécifiques*, il doit pouvoir effectuer une recherche dans le site. Il faut aussi lui offrir une *entrée géographique* pour accéder rapidement à un pays ou une ville. Comme tout un chacun, il veut avoir une idée du phénomène étudié sur son territoire... Curieux, il pratique une navigation par ricochets, il utilise donc beaucoup les *liens entre les pages* (tags en bas d'articles) ou les tags éventuels pour explorer les possibilités du site.

Exemple : Alex est curieux de sciences. Dans la vie, il est documentaliste et sa passion c'est les sciences. Il a lu un article sur l'évolution de la production scientifique dans le monde qui relayait des résultats récents prouvant qu'on observait un mouvement de déconcentration à l'échelle du monde. Il se faisait une toute autre idée de la question et veut aller plus loin. Il va donc consulter le site internet à la source de cet article. Sur le site, il s'intéresse à la démarche globale du programme de recherche et va ensuite naviguer vers des thématiques qui lui semblent intéressantes et se laisse guider notamment par les liens entre les contenus... Il peut passer beaucoup de temps sur le site, du moment que sa soif de connaissance est satisfaite d'autant plus si elle est alimentée de manière ludique et interactive. Il apprécie l'originalité et l'innovation mais aussi la clarté du site.

### *Public chercheurs*

Qui : chercheur ou étudiants dans différents domaines.

Raison de la présence sur le site : Il peut venir sur le site par curiosité ou pour une veille de l'état de l'art. Il peut aussi fréquenter le site par besoin bibliographique, pour utiliser les résultats du programme dans ses propres recherches.

Modalité de navigation nécessaire : Ce type de public doit pouvoir évaluer le corpus de données et la méthodologie sur lesquels reposent les résultats. Il doit disposer de fonctionnalités de *recherches spécifiques* en fonction de ses besoins et/ou discipline. Il souhaitera accéder aux *articles dans leur intégralité*. Il doit pouvoir aussi aller consulter les profils et les parcours des chercheurs impliqués dans le projet.

Exemple : Tim est chercheur en sociologie. Son sujet de recherche porte sur l'évolution des structures universitaires en France et en Allemagne. Il vient sur le site pour accéder aux résultats de recherche concernant la déconcentration et les relations entre les pôles de recherches allemands et français. Il voudrait évaluer la *méthodologie* adoptée pour vérifier la validité de celle-ci au regard de sa problématique. Il a besoin d'avoir accès à la *bibliographie* et à la *webographie*. Il souhaitera aussi éventuellement contacter les chercheurs publiant sur le site pour avoir des informations complémentaires.

### *Public des utilisateurs intéressés*

Qui : enseignants du secondaires, journalistes scientifiques, étudiants de premier et second cycle.

Raison de sa présence sur le site : Il a un besoin précis soit pour alimenter un article, un cours, un exposé...

Modalité de navigation nécessaire : Ce public doit rapidement pouvoir accéder à ce qu'il cherche. Il faut donc prévoir une possibilité de *recherche générique* dans le contenu du site, par exemple par *une zone de recherche thématique* mais aussi *géographique*. Il doit avoir un aperçu globale du cadre dans lequel s'inscrivent les résultats qu'il vient chercher. Il doit pouvoir éventuellement *recupérer les résultats* présentés sous forme de PDF.

Exemple : Chloé est journaliste pour le journal du conseil général d'Ille et Vilaine. Elle prépare un papier sur l'activité scientifique des universités rennaises, notamment à l'internationale. Elle a donc besoin de trouver des informations correspondant à son sujet. Elle veut être rapidement en mesure d'accéder aux résultats dont elle a besoin pour illustrer le propos de son article. Elle a aussi besoin pour introduire le



sujet de remettre ses résultats spécifiques dans le contexte plus large du programme de recherche. Elle souhaiterait contacter un chercheur pour un petit entretien. Elle consulte donc les profils des chercheurs et envoie un mail à deux d'entre eux pour organiser une interview.

### *Public auteurs/chercheurs*

Qui : chercheurs ou doctorants participant au programme de recherche dont le niveau et la familiarité avec l'informatique sont très variables.

Raison de son utilisation du site : Il a produit de nouveaux résultats et/ou un article qu'il souhaite communiquer par l'intermédiaire du site.

Modalité de navigation nécessaire : Pour coller à l'esprit du site, ce public doit transformer ses données en mettant l'accent sur la visualisation graphique des données et l'interactivité avec l'utilisateur alors qu'il a souvent plus l'habitude de mettre en avant le texte.

Sa capacité à alimenter le site dépend du choix de fonctionnement effectué lors de la création du site et surtout des compétences informatiques des chercheurs. On peut distinguer deux catégories.

Une partie du public chercheurs-auteurs peut être *informatico-néophyte*, en dehors peut être de Word et d'Excel. Il n'est pas question qu'il développe ses graphiques directement en JS. Il faut donc de le guider dans la création de la visualisation de ses données avec des outils spécifiques afin de faciliter le travail de l'administrateur.

L'autre partie de ce public de chercheurs-auteurs peut être en mesure de créer des visualisations de données par l'intermédiaire de bibliothèques JavaScript type D3 ou Highcharts, soit qu'il utilise des codes déjà développés par d'autre, soit qu'il utilise les fonctions pour créer de nouvelles visualisations. Les résultats sont alors plus facilement adaptables et enrichis.

Le public des chercheurs-auteurs doit donc disposer d'un *espace sécurisé* où aller chercher les informations concernant les modalités d'implémentation de résultats. Il faudra qu'il sache quels outils employer, qu'il dispose éventuellement de liens vers des web-services ou des téléchargements, des tutoriels pour aider à la création des visualisations, les normes (format de fichier, charte graphique à respecter) selon lesquelles les données doivent être transmises à l'administrateur du site ainsi que le contact de celui-ci afin de pouvoir échanger sur les éventuels problèmes rencontrés.

Exemple : Antoine est chercheur en sociologie. Il a produit une étude concernant les parcours de chercheurs en biologie moléculaire. Il souhaite intégrer ses résultats



au site du programme. Son article comprend quelques graphiques et une carte illustrant le propos. Il sait que la priorité sur le site va à la visualisation des données mais il ne sait pas trop comment restructurer son article. Il a besoin de connaître les règles à respecter pour cela (notamment pour les couleurs de ses graphiques et de ses cartes). Il n'est pas très familiarisé avec les codes informatiques. Il veut bien faire des graphiques interactifs mais il ne sait pas comment s'y prendre. Il faut qu'il ait accès à un outil qui puisse l'aider dans cette démarche. Il a quelques doutes aussi quant à la nature des fichiers à transmettre et envoie donc un mail à l'administrateur pour avoir quelques précisions. Il en profite aussi pour aller consulter les derniers articles publiés par les autres chercheurs du programme et se rendre sur le *forum de communication*.

#### *Public butineur*

Ce n'est pas le public le plus visé, mais on peut envisager des visites ponctuelles de personnes tombant par hasard sur le site.

*Qui* : n'importe quel internaute.

*Raisons de présence sur le site* : hasard de ses pérégrinations sur le net.

*Modalité de navigation* : L'utilisateur ne vient pas sur le site avec un but précis. Il ne restera sur le site que si le sujet général présente un quelconque intérêt pour lui. Il doit donc très *rapidement saisir la problématique* et le sujet du site. Une *navigation par tags* peut éventuellement lui permettre d'aborder globalement le sujet. D'autre part, son intérêt peut aussi être attiré par la possibilité d'une *recherche géographique* ou l'originalité de la présentation.

*Exemple* : Gertrude navigue sur le net à ses heures perdues. A l'origine, ses recherches portaient sur l'aide au développement. Par le biais des mots clés tels que collaborations, recherches..., elle est tombée par hasard sur le site. Après avoir survolé le texte d'introduction, elle regarde les différents développements du sujet. Une rapide navigation sur le site lui a permis d'en percevoir le contenu, elle le garde en favori pour pouvoir y revenir plus tard.

#### *Profil administrateur*

*Qui* : l'administrateur technique du site

*Raison de la présence* : Garant de l'organisation des contenus, il s'occupe aussi de la maintenance et de l'évolution du site en terme technique.

Modalité de navigation :

Les modalités d'utilisation du site dépendent considérablement des solutions techniques utilisées. Schématiquement, l'administrateur doit disposer d'un espace sécurisé par l'intermédiaire duquel il doit pouvoir gérer, organiser ou modifier les contenus déjà présents dans le site et en intégrer de nouveaux qui lui parviendraient des chercheurs-auteurs. Il doit être possible aussi de le contacter pour lui signaler des bugs éventuels. Le choix des outils peut être lié aussi à ses compétences informatiques. Il peut s'occuper du contenu d'un point de vue scientifique. Une autre personne peut se charger de cette dernière activité.

Cette réflexion a fait ressortir un ensemble de besoins en termes de navigation, de contenus et par conséquent en termes de fonctionnalités, illustrés dans le paragraphe suivant.

**3.2) Besoins et fonctionnalités**

Au vue des objectifs fixés, des données dont nous disposons et des publics que nous avons définis, divers besoins et fonctionnalités ont été déterminés, que nous synthétisons ci-dessous.

Besoins	Propositions
Approcher aisément le sujet de la recherche	Bandeau et texte de présentation en page d'accueil
Visualiser les résultats de manière interactive et attrayante	Outils de data visualization
Rechercher une thématique/pays spécifiques	Formulaire de recherche, tags
Lier les différents pages/articles pour permettre une navigation plus directe	Tags
Accéder aux articles intégraux	Lien hypertext
Accéder et évaluer la méthodologie	Section méthodologique - vidéo
Accéder à la bibliographie	Section bibliographie
Cerner rapidement les différents contenus du site	Bandeau et structure générale de la page d'accueil
Prévoir un espace d'échange avec une communication sécurisé entre les auteurs	Section privé avec login
Actualiser le site avec des nouveaux résultats	Tags, partie évolutive/ouverte

### 3.3) Architecture

Des besoins et des fonctionnalités identifiés, il faut envisager les différentes sections qui constitueront le contenu du site, puis les différentes manières de les agencer.

#### 3.3.1) Contenu

A partir des objectifs fixés, des publics visés et des données dont on dispose, on distingue deux grandes parties dans le site. L'une concerne les sections définies comme « classiques » qui permettent d'aborder le programme dans son ensemble. Il s'agit d'une partie méthodologie (texte exposant la démarche, vidéo d'introduction, bibliographie), d'une partie de présentation avec l'historique du projet, les biographies et contacts des chercheurs et éventuellement d'une partie sécurisée destinée à la communication entre les auteurs et l'administrateur. A l'intérieur de cette dernière partie pourraient se trouver les normes à respecter pour la création et la transmission des nouveaux résultats, les outils pour les créer, un forum entre chercheurs...

L'autre partie s'attache aux thématiques étudiées et aux analyses et résultats qui en découlent : la déconcentration des centres de production du savoir, l'augmentation des relations de collaborations et une analyse en fonction de grandes familles de disciplines scientifiques (en cours d'élaboration). Il s'agit aussi de mettre en avant les différentes échelles d'analyse et notamment de valoriser l'innovation de la méthode du géo-codage à l'échelle de la ville.

Cette deuxième partie doit être évolutive et donner la possibilité, d'une part de mettre à jour le site au fur et à mesure que le projet avance, et d'autre part aux publics « techniques » de construire en autonomie leur parcours dans le site.

#### 3.3.2) Structuration

Il y a trois façons d'envisager la structure du site : ouverte, figée, hybride.

→ La *première* est une organisation complètement ouverte (à l'image d'un blog), où le contenu est référencé uniquement par des tags et mots clés. Cette solution n'apparaît pas très pertinente et il nous semble ardu de transmettre des données complexes sans guider l'utilisateur. Cette option manque de substance, de corps.

→ La *deuxième* solution repose sur une arborescence figée ; les contenus sont tous intégrés de manière fixe et la navigation s'effectue par chemins prédéfinis.

Nous avons élaboré deux arborescences qui répondent à ce type de structure.

L'arborescence présentée ci-dessous<sup>9</sup> reflète une approche spatiale selon deux échelles d'analyse (ville /pays) chacune étant structurée selon une répartition par axes de recherche. La page d'accueil contient ainsi trois grandes sections :

- une partie "méthodologie" (vidéo de présentation, textes détaillés de description de la démarche et bibliographie),
- une partie "résultats" répartie entre analyse à l'échelle nationale et internationale (reprenant les résultats par thématiques),
- une partie "présentation" (historique du projet, présentation des chercheurs impliqués et des structures partenaires, actualité du programme).

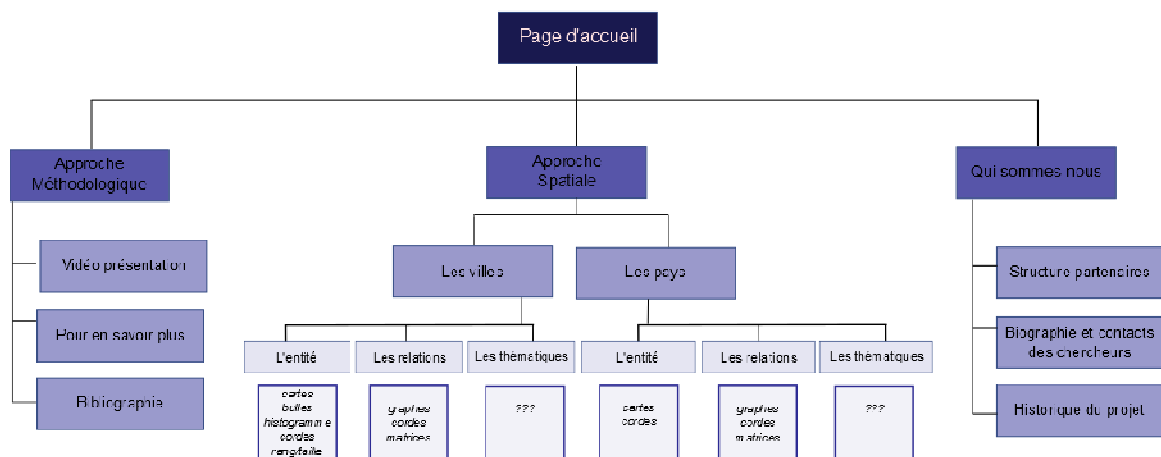


Fig. 5 : Arborescence figée, mettant en valeur la dimension spatiale

La seconde arborescence propose une structure plus souple et qui repose de façon plus évidente sur les thématiques abordées dans la recherche. Avec cette proposition l'approche spatiale est intégrée à l'intérieur de chaque section, donnant la possibilité à l'utilisateur de passer de manière plus directe d'une échelle de visualisation à l'autre.

<sup>9</sup> Pour la visualisation de détails des deux arborescences proposée voir les annexes.

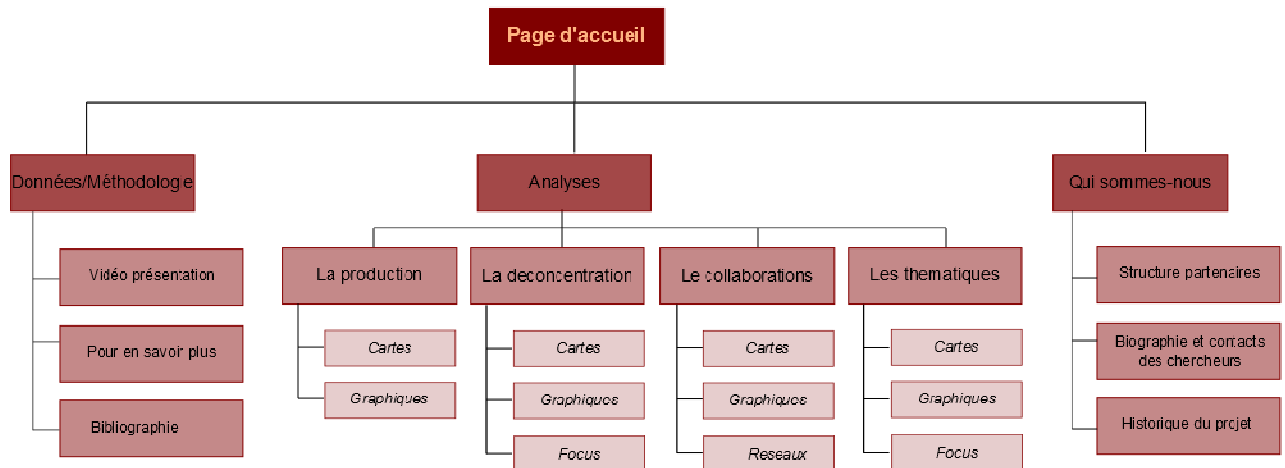


Fig. 6 : Arborescence figée, mettant en valeur les thématiques de recherche

→ La troisième solution proposée est un modèle hybride, se situant entre les deux précédentes : une arborescence structurée et fixe pour les parties déjà définies comme « classiques », et une structure par tags, permettant une organisation des données et une navigation plus ouvertes.

Un tag, en tant que marqueur ou balise est un mot clé qui permet de référencer le contenu d'un site dans sa globalité mais aussi des pages ou groupes de pages, des articles, des cartes et graphiques à l'intérieur même d'un site.

Il est possible d'accéder à un système de 'tag' par l'intermédiaire d'une simple boîte de dialogue textuelle. Il est aussi possible de passer par l'intermédiaire d'un nuage de tags, c'est à dire, un regroupement de mots-clés utilisés dans une page, comme c'est souvent le cas dans les blogs. Chaque tag renvoie à un groupe de résultats qui lui sont rattachés et qui s'affichent en fonction de la recherche effectuée.

Chaque article peut être lié à plusieurs tags qui peuvent être laissés à la discrétion de l'auteur. Cependant, il nous paraît plus pertinent que l'administrateur définisse un thésaurus de mots-clés parmi lesquels l'auteur pourrait choisir et qui permettrait de mettre en place un système d'autocomplétion pour la recherche.

Il en résulte l'arborescence schématisée ci-dessous.

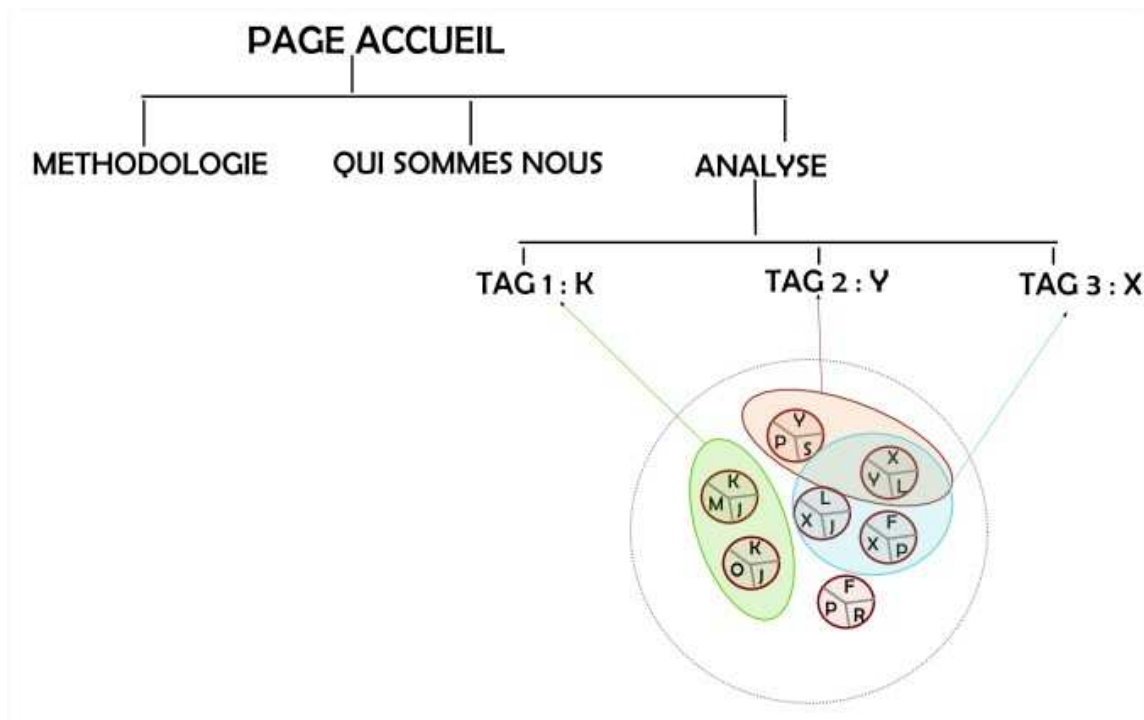


Fig. 7 : Arborescence hybrides structurée selon un système de tags

### 3.4) Solutions techniques

La création d'un site internet implique d'organiser les contenus certes mais aussi de trouver la meilleure solution pour créer la structure dans laquelle ils s'inscriront et qui déterminera aussi les modalités d'administration du site et de gestion future de ses contenus.

Il existe de très nombreuses solutions pour créer et gérer un site internet.

La première d'entre elle consiste à coder de A à Z ce site en format HTML /CSS/PHP...etc. Pour ce faire, on peut tout de même utiliser des modèles disponibles pour structurer pages, menus de navigation, styles... Cette solution envisageable pour des sites limités à un très petit nombre de pages, peut poser des problèmes de mise à jour et de compatibilité avec les navigateurs, des actions fastidieuses pour l'administration... et en bref implique un très bon niveau en développement.

Or aujourd'hui, une autre solution consiste à recourir à des framework ou CMS, Content Management System. En la matière, d'innombrables possibilités ont fleuries sur le net. Certaines de ces solutions très complètes permettent de créer et gérer un site dans son intégralité. D'autres sont spécialisées, par exemples dans les solutions CSS (comme Bootstrap), dans la gestion d'une base de données, etc. Elles peuvent parfois être combinées entre elles ou intégrées à des solutions complètes.

### 3.4.1) CMS & FRAMEWORKS

Les CMS et les frameworks sont deux types d'outils support pour le développement de site web mais leurs possibilités diffèrent en bien des points.

Le choix des outils proposés par la suite repose sur les résultats les plus fréquents lors des recherches web. Il s'avère que se sont aussi souvent ceux dont les communautés d'utilisateurs et de développeurs sont les plus importantes et les plus actives. Ce critère est un point important à plusieurs titres. Il implique souvent l'existence de nombreux tutoriaux, documentations et questions -résolues- sur les forums. Il entraîne aussi un nombre d'extensions et de plugins développés plus important. Faute de temps, nous n'avons pas installé et testé toutes ces solutions, notre évaluation repose donc sur des croisements de comparatifs et de retour d'expériences réalisés par des utilisateurs.

Les **CMS** (Content Management System) sont des outils / logiciels offrant des fonctionnalités livrées clé-en-main et relativement faciles à installer sur un serveur. Ils permettent de développer un site rapidement et sont accessibles sans nécessairement posséder de compétences avancées en programmation. En revanche, ils sont plus difficilement adaptables. Ils s'avèrent donc plus adaptés à des projets de petites et moyennes envergures développés sur du moyen ou long terme.

Dans le cadre de ce projet, le recours à un CMS peut être pertinent à condition qu'il soit possible d'intégrer des scripts ou des modules de data-visualisation. Ils permettent, pour certains au moins (Wordpress et Drupal), de gérer un système de tags en 'natif' ou par l'intermédiaire de modules extérieurs.

Les CMS les plus populaires sont : Wordpress, Drupal, Joomla et ModX. Ils sont développés en PHP.

TABLEAU DES CMS

Caractéristiques	Joomla	Wordpress	Drupal.	ModX
<i>Administrateur</i>	Très complet, bien documenté, facile à prendre en main.	Complet et bien documenté.	Complet. Interface récemment révisée et plus ergonomique.	Complet, interface claire et facile à prendre en main.
<i>Installation/ configuration</i>	Simple/	Simple et rapide à mettre en place	Pas facile à installer et configurer.	Pas facile à installer

Caractéristiques	Joomla	Wordpress	Drupal.	ModX
<i>Modules dont modules de data visualisation</i>	Nombreux modules (environ 4000) Modules de dataviz : Flashcharts, hexcharts, ChartFX, JVCharts.	Nombreux modules de dataviz : Visualiser, Maps Marker (pour la visualisation cartographique).	Très nombreux modules dont des modules de Dataviz : VIDJ, Data Visualisation Wizard et des Plugins disponibles pour adapter hightcharts et autres comme Data Visualisation API.	Moins de modules que les autres Framework et pas de modules CMS.
<i>Fonctions</i>	Gestion de base de données. Gestion des modules de publication (actualité, article,...). PNG, PDF, DOC, XLS, GIF et JPEG. Création de PDF à la volée	NC	NC	Grande liberté de mise en page.
<i>Gestion de tags.</i>	La version 3.1 de Joomla propose un système intégré de gestion des tags.	Utilisation des tags intégrés depuis Wordpress 2.3 + extensions : Smart Tag...	NC	NC
<i>Communauté /tutoriel et documentation.</i>	Communauté très active. La communauté française est assez présente en cas de besoin	Communauté très active et nombreux tutoriels disponibles en anglais et en français même pour les néophytes.	Communauté assez restreinte en France.	NC

Les **frameworks** sont des ensembles de bibliothèques, fonctionnant comme un ensemble de modules correspondant aux tâches les plus redondantes et fastidieuses du développement. Elles permettent à des développeurs de se concentrer sur les fonctionnalités spécifiques à un projet. Leur architecture est bien plus flexible que celle des CMS et leurs fonctionnalités plus avancées. Manipuler un framework implique une certaine connaissance du développement et des langages de



programmation. Ces structures permettent la mise en place de gros projets web inscrits sur du long terme.

On peut subdiviser ces outils selon le langage de programmation utilisé pour les développer. Ils permettant d'approcher divers stades de la conception du site. En d'autres termes, le type de langage et de framework utilisé peut dépendre de l'architecture du site. On distinguera alors des frameworks CSS spécialisés dans la création des styles, d'autres en PHP qui sont destinés plutôt à la gestion des données coté serveur et d'autre encore JavaScript exécutés coté Client et permettent de dynamiser l'interface utilisateur.

La majorité des frameworks est construite sur la base d'une architecture de type *Modèle-Vue-Contrôleur* : une organisation du code de façon modulaire et intelligente qui permet de scinder l'application en trois éléments distincts et quasi-autonomes. Le but est de pouvoir effectuer des modifications ou des mises à jour de certains modules indépendamment des autres. Pour résumer le fonctionnement du MVC s'articule autour de ces trois éléments :

- **Le modèle** : c'est la partie de l'application qui permet la gestion de la donnée.
- **Le Contrôleur**: c'est le moteur de l'application qui permet de réaliser des traitements sur les données.
- **La vue**: c'est la représentation de la donnée qui a subi des traitements et l'interface utilisateur.

Pour schématiser on peut dire qu'un framework de type CSS s'occupe de la « structuration » de la Vue (définition des styles, onglets de navigation, affichage..), tandis que ceux qui utilisent le langage PHP servent à l'implémentation des Modèles et Contrôleurs, et que les frameworks en JavaScript visent à faciliter la création de l'interface utilisateur et, donc, à dynamiser les relations entre Contrôleur et Vue.

Nous présentons par la suite les principaux frameworks utilisés dans les catégories évoquées précédemment.

### Frameworks CSS

→ **Bootstrap** : premier framework à connaître une grosse diffusion, très utilisé dans les réseaux sociaux et dynamisé par une grosse communauté. Il présente une grille fixe ou dynamique, des composants d'interaction avec l'utilisateur (chemins de navigation, mise en page, alertes, étiquettes..) ; met à disposition des éléments CSS fondamentaux (boutons, styles, icônes, images..) ; propose des plugins jQuery (carrousels, tableaux, accordéons, transitions animées..).

→ **Foundation** : framework CSS complet, il implique l'installation du préprocesseur CSS SASS<sup>10</sup>. Il intègre la notion de responsive design qui permet d'adapter un style à différents supports. Il propose les mêmes fonctionnalités que les autres frameworks : gestion des grilles (système à 12 colonnes), des styles, de la typographie, création et personnalisation de boutons, de formulaires et de tout autre éléments classiques de navigation (menu, tableaux, onglets...). Il offre la possibilité d'intégrer plus facilement un carrousel d'images avec le plugin Orbit, des fenêtres multiples (possibilité d'encapsuler des écrans à l'intérieur de l'écran, ou créer des fenêtres modales) avec Reveal et de créer des barres de navigation animées. Il est particulièrement simple à prendre en main.

→ **Blueprint** : framework très complet aussi. Il permet de gérer les éléments classiques d'une interface graphique : la grille mais aussi la typographie et les formulaires. Bénéficiant d'une importante communauté, il présente une très grande diversité de plugins, tutoriaux et templates. Il peut en outre facilement être intégré à des CMS comme Drupal et Wordpress.

→ **Pure** : framework développé par Yahoo. Pure présente l'avantage d'être léger en proposant une bibliothèque basique en CSS 'pure' (tout code JS a été évacué) mais fonctionnelle. Comme ses concurrents, il propose grilles, formulaires, éléments de navigation, tableaux, boutons... Surtout, il permet de facilement personnaliser son application par la création de styles propres et leur combinaison avec les styles déjà implémentés. Ses points forts résident dans sa légèreté et sa facilité de prise en main.

→ **Groundwork** : développé à l'origine pour Drupal, il repose sur le préprocesseur SASS. Il présente autant des fonctionnalités que les autres: typographie, formulaires, boutons, éléments de navigation, tableaux... Son atout principal : la flexibilité de sa grille qui permet toutes les adaptations, fractionnements ou emboîtements possibles et qui en fait l'un des frameworks 'responsive', c'est à dire capable actuellement de s'adapter le mieux à tout type d'écran. Tous ces éléments ont été parfaitement intégrés dans cette démarche. Il en découle aussi une très forte flexibilité et une très grande liberté graphique dans la création d'interface.

## Frameworks PHP

→ **Zend** : Pour réaliser applications et web services de différentes tailles grâce à des API (Google, Yahoo, Amazon). Ses caractéristiques plus importantes sont une structure orientée objet et une approche MVC, qui se reflète aussi dans la structure

---

<sup>10</sup> Un pré-processeur css est un outil permettant de transformer un langage (avec une syntaxe semblable à CSS), en CSS valide.

des dossiers d'un projet. Son point fort est la modularité de la bibliothèque, qui permet de charger seulement les composants requis. Zend est le framework utilisant le plus de mémoire. Cependant, il propose un plus grand degré de flexibilité dans le design d'applications, ce qui peut le rendre particulièrement efficace dans des environnements lourds.

→ **Yii** : offre de grosses performances et beaucoup de fonctionnalités, aussi en termes de plugins. Il combine habilement les meilleures techniques PHP et propose de bonnes solutions pour optimiser les requêtes aux bases de données. Le code source est orienté objet, moderne et propre ; la gestion des modèles est très efficace et pertinente. La documentation est claire et agréable. C'est la solution la plus flexible aujourd'hui et celle dont la cote de popularité augmente le plus.

→ **CakePHP** : plus adapté aux petits projets, facile à apprendre, avec une architecture moins flexible que les autres frameworks PHP. Son système d'extensions permet de développer de manière modulaire et de réutiliser facilement les fonctionnalités d'un projet à un autre. Malgré la gestion efficace du système MVC, Cake trouve ses limites dans la gestion d'une base de données complexe ou des gros projets, à cause des conventions très strictes de nommage et d'organisation du code. La communauté de support est réduite mais active.

→ **CodeIgniter** : framework léger bon pour les débutants qui cherchent un maximum de résultats. Malgré les nombreuses fonctionnalités proposées, il reste très léger car il se cantonne au strict minimum en termes de code et offre en option une large gamme de classes et fonctions. Bonnes performances en base de données, il est considéré comme un framework minimaliste. Bonne documentation, simple et rapide à comprendre.

→ **Symfony** : lancé en 2005 par une agence web française, il offre une série de classes écrites en PHP et de composants pour créer des applications dont la mise à jour est facile. Il intègre de nombreuses bibliothèques externes. Plus difficile d'apprentissage, même si la documentation est disponible en français.

## Frameworks JAVASCRIPT

→ **Angular** : Angular n'est pas un framework standard prêt à l'utilisation. En effet il regroupe un panel d'outils qui facilitent la création d'un framework adapté à une application spécifique. La gestion du DOM<sup>11</sup> chez Angular est particulièrement simplifiée grâce aux directives internes qui permettent de développer des attributs et

---

<sup>11</sup> Le Document Object Management est un standard du W3C permettant à des programmes informatiques et à des scripts, d'accéder ou de mettre à jour le contenu, la structure ou le style de documents XML et HTML

des éléments HTML spécifiques. Les applications deviennent alors très modulaires. Grâce à une infrastructure MVVM<sup>12</sup> et au processus de data-binding, autrement dit la fluidité de communication des données entre les différentes parties de l'architecture MVC, un élément de la page peut s'afficher à un autre endroit. Il est ainsi possible de lier directement en HTML des éléments de la vue, de sorte que tout changement sur l'un se reflète sur l'autre. Cela simplifie fortement le codage et diminue les risques de bugs.

Sa principale limite est le référencement, car Angular n'offre pas la possibilité aux moteurs de recherche d'indexer les sites.

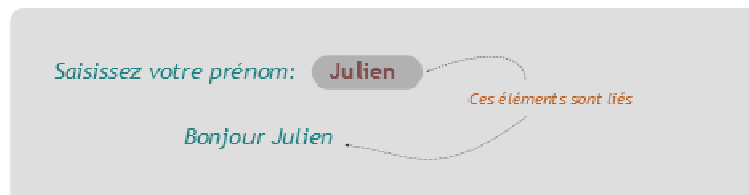


Fig. 8 : Schéma présentant le processus de data binding bidirectionnel

→ **Backbones** : framework particulièrement adapté pour développer des applications de type « single page application », en d'autres termes une page principale avec plusieurs interactions utilisateurs (exemple : formulaire). La navigation classique (par page HTML) est abandonnée et les actions sont directement gérées en JavaScript, ce par l'intermédiaire de la bibliothèque *Underscore.js* qui propose des fonctionnalités avancées de manipulation de collections d'objets et de tableaux. L'architecture MVC est approfondie et fait place au modèle MCV où le C représente les collections d'objets implémentées. La gestion du DOM est laissée à la charge de *JQuery*. Backbone est uniquement un framework client et ne permet pas la manipulation de données sur un serveur.

→ **Ember** : framework particulièrement complet dont la finalité est de créer des applications de type « single page application ». L'une de ses grandes forces est le data binding. Son mode de fonctionnement repose sur un système de composants et de templates modulables (vues) qui le rendent particulièrement flexible. Les valeurs affichées dans les vues sont directement liées à celles du modèle grâce au data-binding. Ceci permet une actualisation en continue du template au moindre changement du modèle. Sa prise en main peut cependant être un peu fastidieuse car il ne dispose pas d'une documentation très fournie et mise à jour.

---

<sup>12</sup> MVVM: Modèle-Vue-VueModèle. C'est une structure avec laquelle les données que le client saisit engendrent une mise à jour du Contrôleur, qui met à jour par ricochet la Vue.

On peut considérer un dernier point concernant les frameworks. Le site doit être amené à gérer aussi des contenus textuels. Pour cela, nous avons exploré les solutions d'édition numérique envisageables. Ces recherches se sont avérées beaucoup moins fructueuses, surtout si l'on se concentre sur les produits open-source. Lodel semble être la seule solution disponible pour cette fonctionnalité. On peut le ranger dans la famille des CMS. Il permet de charger des documents Word, OpenOffice... Il permet l'intégration et la mise en forme des contenus de manière uniformisée. Cependant, il ne semble pas être une option intéressante dans notre cas. Il reste à vérifier la possibilité de combiner Lodel avec d'autres frameworks ce qui permettrait de l'intégrer avantageusement dans la structure du site.

Ces frameworks JavaScript ont été sélectionnés car d'une part ils sont les plus importants à l'heure actuelle et disposent d'importantes communautés. Il s'agit cependant d'interfaces JavaScript côté client. De plus, ils semblent tous permettre de faciliter l'intégration de visualisation en D3 et autres bibliothèque JavaScript.

En plus des questions de compétences informatiques et du temps de développement, le choix entre Framework et CMS dépend de la complexité du projet. Il n'est pas nécessaire de s'investir dans la mise en place d'un framework pour un site requérant une organisation et des fonctionnalités peu complexes.

Les critères de choix peuvent donc être :

La *taille du site* à construire : certains outils sont adaptés à des sites de petites et moyennes tailles (comme MODx ou Wordpress) d'autres sont plus utilisés pour les sites de grande ampleur, des frameworks principalement.

Les *fonctionnalités* attendues d'un framework/CMS : gestion d'une base de données, gestion de droits d'utilisateurs, bibliothèques CSS, etc.

La *simplicité et rapidité de prise en main* : documentation disponible et dynamisme de la communauté...

### 3.4.2) Les outils de visualisation : bibliothèques et web services

Le site veut mettre en avant les résultats des travaux de recherche par le biais de la visualisation des données. D'innombrables solutions sont envisageables en la matière, des bibliothèques JavaScript aux web-services faciles d'accès. Nous avons par ailleurs concentré nos recherches sur des outils open-source.

Nous distinguons entre les solutions disponibles pour la création de graphiques et d'autre part pour la création cartographique. A l'intérieur des deux catégories de visualisation, nous présentons deux types d'outils.

Des bibliothèques, JavaScript pour la plupart, pour les auteurs possédant des compétences en développement informatique. Il pourra coder complètement sa visualisation en l'adaptant ainsi exactement à ses intentions ou utiliser un code déjà développé pour l'appliquer à son propre travail. On envisage qu'un chercheur, qui voudrait ajouter son travail au site, puisse créer son application en JavaScript et fournir à l'administrateur les éléments finalisés à intégrer dans la page web.

Concernant les chercheurs informatico-néophytes, voire informatico-réfractaires, nous avons prospecté les solutions guidées disponibles. Beaucoup de web-services permettent gratuitement de créer des représentations graphiques et cartographiques interactives. Ils fonctionnent tous de la même manière. Il suffit de charger un fichier de base, xls, csv ou shp. Puis, à la manière des tableaux croisés dynamiques d'Excel, il s'agit de sélectionner les données à représenter et la modalité de visualisation. Ils permettent pour la plupart de créer des tableaux de représentation, pour organiser sur une même page plusieurs représentations. Ces solutions permettent ensuite de charger le script qui résulte de cette création et ainsi de pouvoir l'intégrer à n'importe quelle page web. Une réserve de taille demeure cependant. Ces solutions en tant que web-service ne peuvent garantir la confidentialité des données. Il s'agit donc d'utiliser des données déjà normalisés ou non confidentielles.

#### → *Graphiques*

Il existe de nombreuses bibliothèques destinées à la construction de visualisation de données par l'intermédiaire de diagrammes, histogrammes et autres camemberts. La plupart repose sur le langage JavaScript. Les possibilités offertes par chacune d'entre elles sont détaillées en annexes, avec des liens vers les sites et tutoriels disponibles.

TABLEAU DES BIBLIOTHEQUES POUR LA DATA VISUALIZATION

Caractéristiques	D3	Highcharts	Dimple	Vega	Infoviz
<i>Visualisation disponibles</i>	+++++	+++	++	+++	+++
<i>Visualisation carto.</i>	Oui	Non	Non	Non	Oui
<i>Interactivité</i>	++++	+++	+	++	++
<i>Documentation / tutoriel</i>	+++	+++	+++	+	++
<i>Prise en main</i>	++++	++	+	+	++
<i>Le plus.</i>	La richesse des possibilités de visualisation	Les thèmes implémentés. Editeur de script.	Prise en main facile.		

A notre avis, les bibliothèques les plus intéressantes sont D3 et Highcharts car elles proposent une grande richesse en termes de visualisation et sont particulièrement bien documentées.

TABLEAU DES WEBSERVICES POUR LA DATA VISUALIZATION

Caractéristiques	Tableau public	Many Eyes	Micro Strategy	Chartle	RAW
<i>Richesse des visualisations proposées</i>	+++	+++	+++	+	++
<i>Visualisation carto.</i>	Oui	Oui	Oui	Non	Non
<i>Création de graphes.</i>	Non	Oui	Non	Non	Non
<i>Documentation disponible/ tutoriel</i>	+++	++	+++	+	++
<i>Facilité d'utilisation</i>	+++	NC	+++	+++	+++

Caractéristiques	Tableau public	Many Eyes	Micro Strategy	Chartle	RAW
<i>Le plus.</i>	Tutoriel Vidéo.	Aide au choix de visualisation en fonction de la nature des données.	Facilité de prise en main.		Des possibilités de visualisation D3 originales et inédites.

### → Cartographiques

Pour la création de cartes, il existe les mêmes alternatives : des bibliothèques de codes ou des solutions support d'aide à la création par des web-services.

De nombreuses bibliothèques JavaScript pour la création cartographique sont disponibles. Le tableau ci-dessous résume (une évaluation plus détaillée se trouve en annexe) les quelques caractéristiques des 4 bibliothèques les plus connues et les plus documentées.

TABLEAU DES BIBLIOTHEQUES JS POUR LA CREATION CARTOGRAPHIQUE

	OpenLayer	Leaflet	Polymaps	Modest maps
<i>Richesse de la bibliothèque.</i>	++++	+++	++	+
<i>Facilité de prise en main.</i>	+	+++	++	+++
<i>Travail en multicouches</i>	Oui	Oui	NC	Non
<i>Le plus.</i>	La richesse des fonctionnalités.	Facilité de prise en main.	Les templates de cartes disponibles.	La présentation des fonctions.

Il semblerait qu'OpenLayer reste indétrônable concernant la richesse des possibilités offertes. Cependant pour un projet qui ne nécessiterait que des fonctionnalités basiques, il peut avantageusement être remplacé par Leaflet.

De même que pour les représentations de données par graphiques, il existe des web-services ou des logiciels desktop qui permettent à des personnes néophytes de créer des cartes à partir de données simples. On peut émettre les mêmes réserves que précédemment quant à la sécurité des données.



Ces solutions sont globalement plus complexes à prendre en main que les solutions guidées pour la création de graphiques. Leur point fort : elles sont dédiées à la publication sur le web. Cependant, peu d'entre elles proposent une réelle interactivité. Elles peuvent donc constituer une alternative pour qui ne dispose pas de logiciel SIG ou souhaite réaliser des cartes basiques pour le web.

*TABLEAU DES WEBSERVICES POUR LA CREATION CARTOGRAPHIQUE*

<b>Caractéristiques</b>	<b>Carto DB.</b>	<b>Worldmap</b>	<b>Tilemill</b>
<i>Facilité de prise en main</i>	+++	++	++
<i>Formats acceptés</i>	+++	NC	+++
<i>Richesse des visualisations</i>	+++	++	+++
<i>Interactivité</i>	++	NC	++
<i>Le plus</i>	Possibilité d'exporter en HTML mais aussi en par une API SQL ou cartographique.	Richesse des fonds de cartes disponibles directement sur le site.	Richesse des possibilités de ... cartoCSS. Ce n'est pas un webservice.

Parmi ces solutions, CartoDB propose l'interface la plus simple d'accès et les possibilités d'enrichissement et d'interactivité la plus importante.

Après avoir cerné les différents publics susceptibles de visiter le site, nous avons déterminé les besoins et fonctions nécessaires sur le site. Cette première étape, nous a ensuite amené à rechercher les solutions techniques envisageables. Certains frameworks et CMS nous sont apparus susceptibles de constituer une option. D'autre part, des bibliothèques, web-services et logiciel desktop destiné aux chercheurs-auteurs ont été évalués et proposés pour l'alimentation du site.

## 4) Prototype et phase de test

Afin de proposer des modalités de présentation des résultats, nous nous sommes appuyés sur les possibilités offertes par la bibliothèque D3 (Data Driven Documents), créée par Mike Bostock. Développée en JavaScript et dédiée à la visualisation de données, elle s'est taillé rapidement une solide réputation. D3 permet d'illustrer de manière interactive et innovante de grandes quantités de données par des représentations graphiques et cartographiques grâce à des interfaces web. Elle fait partie des bibliothèques les plus avancées dans le domaine de la data visualization.

Cartes et graphiques illustrant les résultats obtenus ont été repensés en fonction de l'impératif d'interactivité et transformés par l'intermédiaire de la bibliothèque D3. Les tests effectués ont eu comme unique objectif la compréhension et la validation des procédures d'implémentation des graphes D3. Dès lors, la dimension sémiologique et l'aspect visuel n'a pas été prise en compte.

### 4.1) La structuration des données

#### 4.1.1) *Les données*

Nous disposons de données présentant la production scientifique à l'échelle des villes, elles-mêmes regroupées en classe en fonction du nombre total de publications. Nous disposons ainsi des quantités de publications pour 447 villes regroupées dans 8 classes. A titre d'exemple, la ville de New-York appartient à la classe plus de 20K car sa production, sur la période 1999-2001, dépasse 20 000 articles.

Afin de mettre en évidence ces données, et d'en proposer une représentation synthétique mais aussi esthétique, nous avons exploré la bibliothèque D3.js et notre choix s'est arrêté sur un histogramme dynamique. L'intérêt réside dans le fait de présenter sur un premier niveau les résultats en termes de nombre de publications à l'échelle des classes puis au niveau des villes. Ainsi, l'utilisateur a accès dans un premier temps, à la répartition du nombre de publications selon les classes de ville. Puis il peut obtenir la répartition intra-classe par un simple clic sur celle qui l'intéresse. Cette représentation interactive présente comme principal intérêt d'inviter l'utilisateur à explorer les données et ainsi mieux se les approprier.

#### 4.1.2) La méthode

Une fois récupérés les scripts et les données de l'exemple en ligne sur le site de D3.js, la principale étape du travail a été la structuration du fichier de données. Celui-ci se présente sous le format json suivant un formatage qui permet de hiérarchiser les données, afin de créer les différents niveaux de l'histogramme. Dans notre exemple, nous présentons un histogramme à deux niveaux

Les données ont été intégrées à Postgres à partir d'un fichier de type CSV. L'intégration au SGBD présente l'avantage de pouvoir interroger la base de données par le biais de requêtes SQL, et ainsi récupérer les seuls champs d'intérêt. La table qui en résulte contient 12 champs de type texte et numérique ainsi que les 447 enregistrements.

	compte numeric	dizaines numeric	id text	nom text	admintype text	af990001 numeric	af060708 numeric	gptaille_p1 text	gptaille1kp1 numeric	gpt500p1 numeric	mngp1 numeric	mngp2 numeric
1	2	0	AD15377	NEW YORK	Populated p	24809	28510	plus-20k	25	49	1	2
2	1	0	AD7725	TOKYO	Admin-0 cap	23587	29623	plus-20k	24	47	2	1
3	3	0	AD11950	LONDON	Admin-0 cap	22979	25276	plus-20k	23	45	3	3
4	4	0	AD12860	BOSTON	Admin-1 cap	20491	24864	plus-20k	20	40	4	4
5	14	1	AD1975	TORONTO	Admin-1 cap	8364	11823	5k-10k	8	16	13	14

Extrait de la table de données.

L'étape suivante est le formatage des données en JSON. Pour ce faire un script php a été écrit, qui permet l'interrogation de la base de données par requête SQL simple, mais aussi le formatage du résultat en flux JSON.

```
//Paramètres de connexion à la BDD PgSQL
require('connexion.php');

$chco = "host=".$pg_hote." dbname=".$pg_base_de_donnees." user=".$pg_utilisateur." password=".$pg_mot_de_passe;
$dbconn = pg_connect($chco) or die("Connexion impossible");

//Requête SQL pour récupérer les données qui nous interessent

$sql = "SELECT gptaille_p1 as dep, nom as name, af990001 as size from ville_histo";
$res = pg_query($sql) or die(pg_last_error());

//Génération de la chaîne GeoJSON
$data = array();

for ($x = 0; $x < pg_num_rows($res); $x++) {
    $data[] = pg_fetch_assoc($res);
}

echo json_encode($data);
```

Script php pour le formatage des données

```
[{"dep": " plus-20k ", "name": "NEW YORK", "size": "24809"}
, {"dep": " plus-20k ", "name": "TOKYO", "size": "23587"}]
```

Extrait du résultat de l'exécution du script php

Suivant ce format, le script de D3.js de l'histogramme choisi ne peut fonctionner, car la structure présentée ci-dessus n'est pas hiérarchisée. Pour proposer un graphique interactif par arborescence, la structure des données doit être définie par l'intermédiaire de nœuds et de liens. Ainsi l'arborescence prend en entrée la racine des nœuds qui peuvent se présenter sous la forme d'un tableau de nœuds fils (le père étant la racine des nœuds) caractérisés par une certaine valeur. Concrètement dans notre exemple, la classe (ex plus-20K) se présente comme le père d'un tableau de nœuds contenant le nom de la ville et le nombre de publications, les fils.

```
1 {
2   "name": "Arborescence",
3   "children": [
4     {
5       "name": " plus-20k ", ← Définition de la classe parent
6       "children": [
7         {
8           "name": "NEW YORK",
9           "size": "24809"
10        },
11        {
12          "dep": " plus-20k ", ← Classe Parent
13          "name": "TOKYO", ← Tableau de noeud:
14          "size": "23587" ← Nom de la ville et nombre de publications
15        },
16        {
17          "dep": " plus-20k ",
18          "name": "LONDON",
19          "size": "22979"
20        },
21        {
22          "dep": " plus-20k ",
23          "name": "BOSTON",
24          "size": "20491"
25        }
26      ]
27    },
28  ]
29 }
```

*Extrait du fichier de données formaté.*

Le premier niveau se présente comme le fils d'un niveau 0 qui correspond à l'ensemble de l'histogramme. Le premier « children » est constitué des classes de villes tandis que le second « children » fait référence aux villes contenues dans chaque classe.

Afin d'obtenir cette structure, nous avons utilisé un script D3.js. Celui-ci prend en entrée les données issues de la requête présentée précédemment.

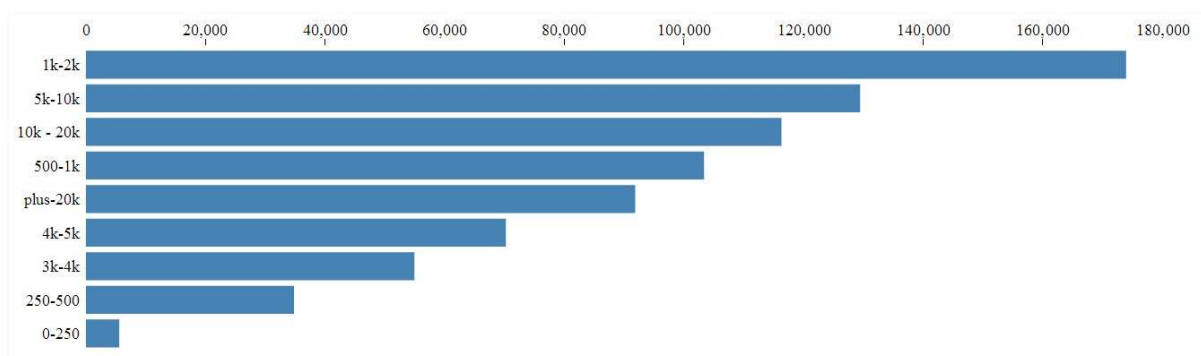
```
// initialisation d'une nouvelle variable qui recevra le résultat du formatage
var newData = {"name": "Arborescence", "children": {}};
newData.children = Object.keys(newData.children).map(function (key) { return newData.children[key]; });
// Appel du fichier json plat
d3.json("get_data.php", function(error, data) {
  //creation des noeuds et des branches
  data.forEach(function(d) {
    if(typeof newData.children[d.dep] !== 'undefined') {
      newData.children[d.dep].children.push(d)
    } else {
      newData.children[d.dep] = {"name": d.dep, "children": [{"name": d.name, "size": d.size}]}
    }
  });
  // edition au format json
  d3.select('body').append('pre')
    .text(JSON.stringify(newData, null, ' '));
});
```

*Script JS pour la structuration des données selon une architecture hiérarchisée*

A l'issue de ces traitements, nous disposons d'un fichier JSON correctement formaté pour être lu par le script du « Hierarchical bars » que nous avons copié sur le site D3.js.org.

## 4.2) Les différentes modalités de visualisation

### 4.2.1) Hierarchical bar chart



*Fig. 9 : Premier niveau - les publications par classe*

Le premier niveau représente des classes de villes regroupées en fonction de leur production sur la première période (1999-2001). En d'autres termes, toutes les villes qui ont publié moins de 250 articles entre 1999 et 2001 sont regroupées dans la classe 0-250.

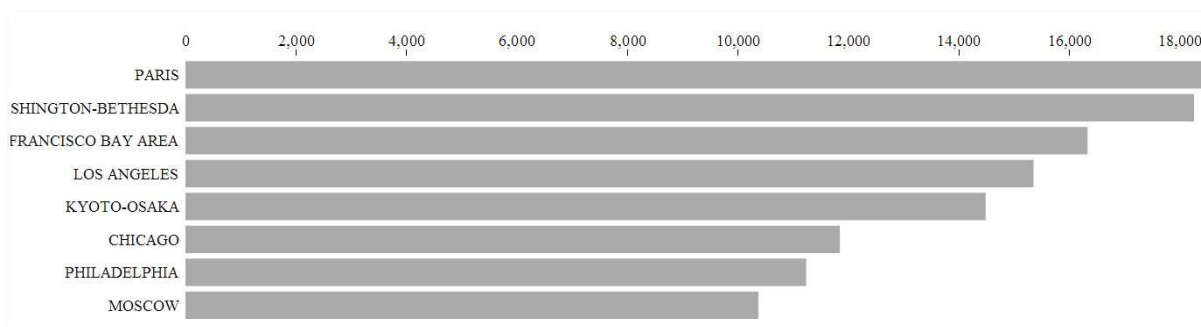


Fig. 10 : Second niveau - les publications par villes pour la classe 10-20K

Le second niveau de l'histogramme indique le nombre d'articles scientifiques publiés dans chacune des villes appartenant à une classe de production. L'utilisateur peut obtenir une vision détaillée de la production au sein des classes et ceci de manière interactive.

L'histogramme est construit de telle sorte que l'affichage des classes (ordonnées) s'effectue de manière décroissante. Concrètement, la valeur affectée à chaque « barre » de l'histogramme correspond à la somme des publications des villes incluses dans la classe. Soit pour le premier niveau, on a :  $B = \sum_{i=1}^n size_i$  avec  $n$  est le nombre de villes par classe et  $B$  la valeur de la « barre » de l'histogramme.

Nous aurions voulu intégrer un troisième niveau, présentant l'évolution des publications pour chaque ville. Cependant compte tenu de nos données, il est difficile de mettre en œuvre ce troisième niveau. En effet, pour chaque niveau, une barre de l'histogramme correspond à la somme des éléments qu'elle 'contient'. Or le nombre de publications à l'échelle des villes ne peut évidemment s'obtenir par la somme d'un taux d'évolution.

Cet outil présente l'avantage de proposer sur une même interface un grand nombre de données. L'utilisateur est invité à naviguer dans l'histogramme suivant ses intentions. Dans l'interface finale, il faudra guider son parcours dans l'histogramme par une phrase invitant le visiteur à "entrer" dans le graphique, en lui expliquant le fonctionnement. L'interactivité suscite la curiosité de l'utilisateur et ceci quel que soit son profil. Le visiteur néophyte est attiré par l'arborescence proposée. En naviguant dans l'histogramme, certaines informations peuvent captiver son attention et l'inviter à creuser le sujet. A l'opposé, pour l'utilisateur spécialiste du sujet, l'histogramme peut lui permettre d'accéder rapidement à l'information qu'il recherche.

Grâce à la structure des données sous forme de liens et de nœuds, d'autres types de représentation sont possibles. Ces possibilités, nous les avons découvertes au fil de nos explorations de la bibliothèque D3.js. Il nous semble intéressant, de

proposer plusieurs manières de présenter la même donnée en utilisant un panel de graphes différents.

#### 4.2.2) Zoomable Pack Layout

Ce graphe permet à l'utilisateur d'explorer le contenu des classes à l'aide d'un zoom. En effet par simple clic, on accède au contenu de la classe d'intérêt. De plus, les données des quantités d'articles scientifiques produits par ville sont représentées par des symboles proportionnels. Et donc comme pour l'histogramme la taille de la classe (diamètre du symbole) correspond à la somme du nombre de publications par ville.

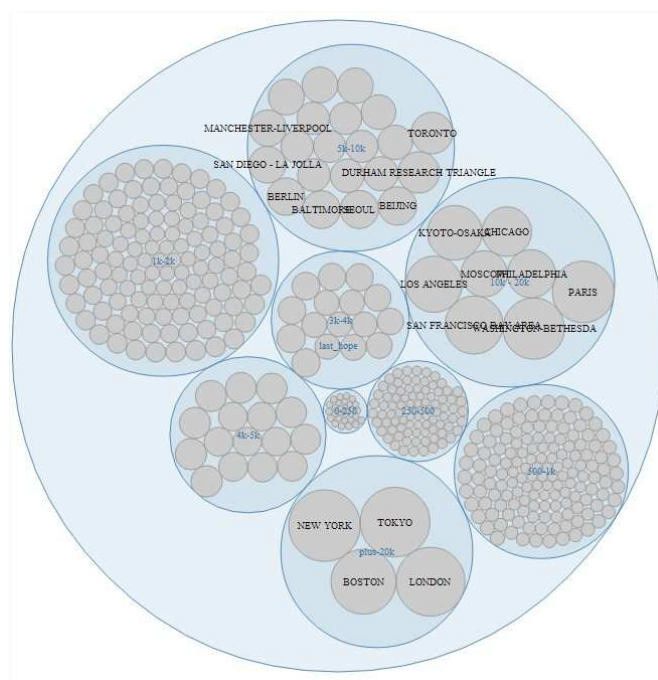


Fig. 11 : Vue d'ensemble du graphe

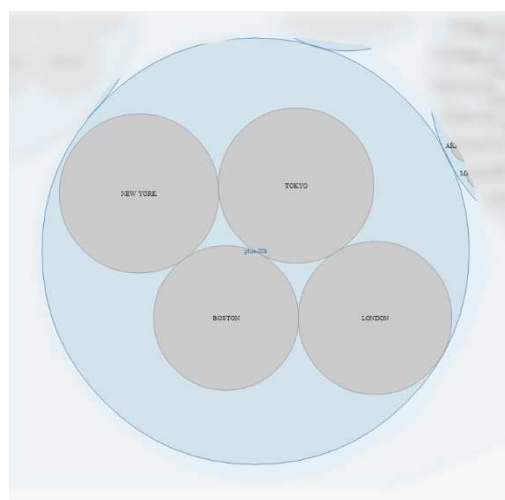


Fig. 12 : Zoom sur la classe plus-20K



### 4.2.3) Collapsible tree Layout

Ce Dendrogramme interactif offre la possibilité à l'utilisateur de déplier les « branches de l'arborescence » par un simple clic sur les nœuds. Dans notre exemple, le visiteur du site peut ainsi explorer le contenu de toutes les classes ou seulement celle(s) qui l'intéresse(nt). Contrairement à l'histogramme proposé précédemment l'utilisateur n'a pas accès à la donnée de quantité de publications. Mais le graphe propose une vue synoptique du contenu des classes toujours dans une optique interactive et esthétique.

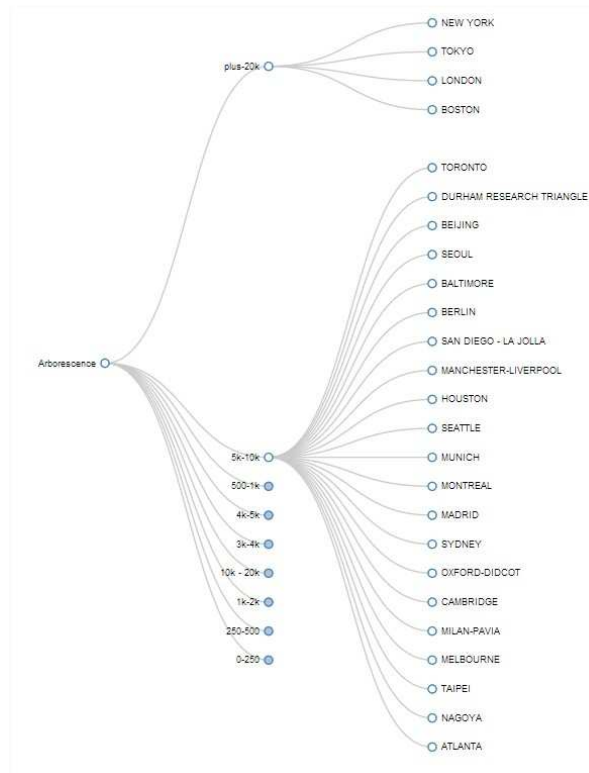


Fig. 13 : Arborescence déployée pour les classes plus-20K et 5-10K

Pour aller plus loin, il pourrait être intéressant de représenter le nombre de publications en faisant varier l'épaisseur du lien.

### 4.2.4) La carte des collaborations

A l'image de ce qui est présenté sur le site CoSciMo, il nous a paru intéressant de développer une cartographie des collaborations à l'échelle de la France pour explorer d'autres fonctionnalités de D3. La notion de déconcentration des pôles de production est largement démontée par (R. Levy et all. Mappemonde, 2013). Afin d'étayer ces propos, nous proposons une cartographie interactive mettant en relation, par des arcs, les liens qui existent entre les pôles qui publient des articles. Les symboles proportionnels ponctuels apportent une information sur la quantité



d'articles produits dans chacune des villes. Les arcs matérialisent les relations (collaborations) entre les pôles. Il pourrait être intéressant de rendre l'épaisseur des liens proportionnels à l'intensité de collaboration.

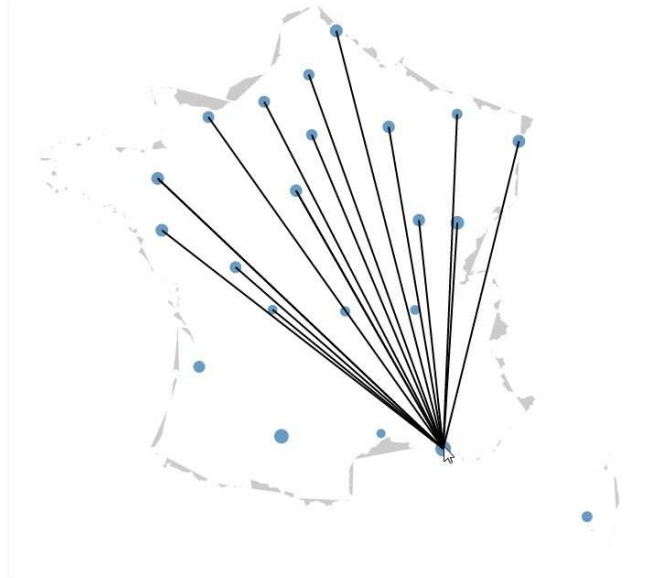


Fig. 14 : Marseille et ses collaborations

L'emprise de la cartographie est obtenue à partir d'une couche vectorielle (IGN : ROUTE 500) dont la géométrie a été formatée en Json à l'aide de Postgis. Puis la couche a été intégrée par le biais de l'appel de la classe d3.json.

```
// appel du fichier de géométrie (fond de carte) au format json
d3.json("COTE_FRONTIERE.json", function(collection) {
  states.selectAll("path")
    .data(collection.features)
    .enter().append("svg:path")
    .attr("d", path);
});
```

Extrait du script D3.js

Les pôles scientifiques sont géo référencés par des coordonnées longitude et latitude exprimées en WGS-84 (EPSG : 4326). Ces données sont obtenues à l'aide de Qgis et d'une couche vectorielle (IGN : ROUTE 500) des villes contenant les pôles scientifiques de la base de données. La géométrie et les données de la table attributaire sont alors stockées dans un fichier CSV.

```
// appel du fichier des pôles scientifiques
d3.csv("town_export.csv", function(universities) {

  // on ne considère que les pôles avec au moins une collaboration
  universities = universities.filter(function(university) {
    if (countByuniversity[university.town_ID_U]) {
      var location = [+university.longitude, +university.latitude];
      locationByuniversity[university.town_ID_U] = location;
      positions.push(projection(location));
      return true;
    }
  });
});
```

*Extrait du script D3.js*

Un fichier au format CSV indique les liens et l'intensité de production des pôles scientifiques.

```
origin,destination,count
U_LIL,U_AMI,853
U_LIL,U_CAE,1
U_LIL,U_REN,805
U_LIL,U_BES,465
U_LIL,U_BOR,247
U_LIL,U_MON,997
U_LIL,U_AJA,3
U_LIL,U_CHA,9
```

*Extrait du fichier json : collaboration.csv*

En utilisant ces données, le script D3 est en mesure de matérialiser les liens (arcs) et de paramétrer la taille des symboles proportionnels (champ : « count »).

```
// appel du fichier de collaboration
d3.csv("collaboration.csv", function(coll) {
  var linksByOrigin = {},
      countByuniversity = {},
      locationByuniversity = {},
      positions = [];
  // initialisation des arcs
  var arc = d3.geo.greatArc()
    .source(function(d) { return locationByuniversity [d.source]; })
    .target(function(d) { return locationByuniversity [d.target]; });
  //affectation des liens
  coll.forEach(function(coll) {
    var origin = coll.origin,
        destination = coll.destination,
        links = linksByOrigin[origin] || (linksByOrigin[origin] = []);
    links.push({source: origin, target: destination});
    countByuniversity[origin] = (countByuniversity[origin] || 0) + 1;
    countByuniversity[destination] = (countByuniversity[destination] || 0) + 1;
  });
});
```

*Extrait du script D3.js*

D3.js présente, par l'intermédiaire de la fonction D3.geo, un panel de projections possibles. Les données des géométries contenues dans les fichiers d'appel sont alors ré-projetées directement par le script D3.js. Pour des raisons esthétiques nous avons opté pour une projection de type gnomonic. C'est une projection cartographique azimutale transformant les grands cercles en lignes droites.

```
var projection = d3.geo.gnomonic()
  .center([2.454071, 47.279229]) // On centre la carte sur la France
  .scale(2000)
  .translate([w / 2, h / 2]);
```

*Extrait du script D3.js*

#### 4.3) Propositions d'optimisation : le lien direct entre la bd et les visualisations

Si durant notre travail, il nous a semblé intéressant d'utiliser Postgres pour formater nos données en json, nous nous sommes posé la question de la création d'une passerelle directe entre les scripts D3.js et le SGBD. Nous souhaiterions ainsi automatiser la chaîne de traitement qui permet le formatage des données (Postgres) en json hiérarchisé. Malheureusement, nous avons rencontré des difficultés techniques que nous n'avons pu surmonter dans le temps imparti.

Cependant, comme nous l'avons déjà évoqué dans le paragraphe..., l'utilisation de frameworks JS peut fluidifier la communication entre le Modèle, le Contrôleur, et la Vue permettant ainsi d'abandonner les requêtes prédéfinies.

L'utilisation de Framework de type Angular.js et Embers.js permet, en effet, la dynamisation du lien entre les données brutes (Modèle) et le résultat des traitements (Vue) opérés par le Controller. En d'autres termes, ils permettent la création d'une forme de passerelle entre les données et l'interface utilisateur. Généralement cette dernière se présente sous la forme d'une simple page HTML. L'encapsulation de D3.js intervient au niveau du Controller et l'affichage des résultats est intégré à la Vue par de simples balises HTML qui appellent les traitements du Controller. L'utilisateur n'a ainsi pas besoins de gérer le rafraîchissement de la page HTML.

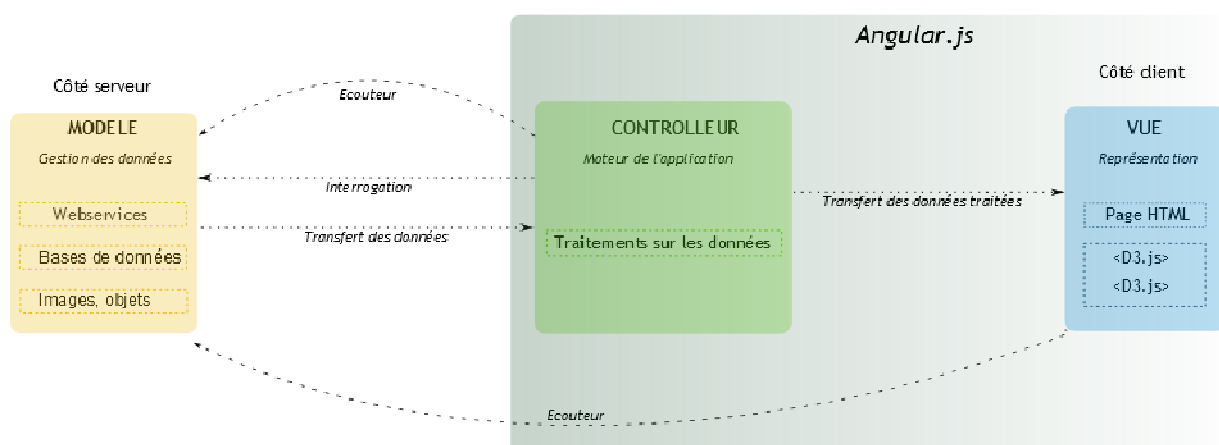


Fig. 15 : Positionnement du framework Angular au regard de la structure MVC

## 5) Conclusions

Le projet illustré dans le présent document porte sur les possibilités de visualisation des résultats issues du projet de recherche « Géoscience » et sur la conception d'un portail web de restitution. Grâce à la géolocalisation des articles scientifiques, le programme a permis d'étudier et d'analyser la répartition spatiale et temporelle de la production scientifique mondiale.

Nous avons principalement effectué un travail de recherche, afin de connaître les procédures nécessaires à la conception d'un site web. Une longue phase de réflexion a été menée sur la façon de présenter des résultats scientifiques complexes et sur les possibilités offertes par les outils de data visualisation. Pour ce faire, nous nous sommes basés sur les productions réalisées par les auteurs. Ces dernières sont constituées, d'une part des articles rédigés par les membres du projet et d'autre part par le site CoSciMo. Celui-ci approche la partie de la recherche concernant les collaborations entre les villes.

Parallèlement à la réflexion sur les modalités de visualisation, nous avons étudié les possibilités de structuration du site afin de dresser un panel de solutions techniques adaptées aux besoins. Ces derniers ont découlés de l'analyse du public touché et plus particulièrement de la définition de profils d'utilisateurs. Dès lors, nous avons réfléchi à une structure qui doit s'articuler autour de deux grandes parties. A savoir, dans une section figée la présentation du programme, et dans une section évolutive les analyses et les résultats des travaux déjà effectués et à venir.


Ainsi trois architectures ont été déclinées proposant différents types d'organisation: ouverte, figée et hybride. La *solution hybride* nous paraît être une piste de travail intéressante car elle offre une navigation ouverte et laisse à l'utilisateur le choix de construire son propre parcours dans le site à l'aide d'un corpus de mots clefs évolutif.

Après avoir déterminé l'architecture adéquate nous nous sommes concentrés sur l'étude comparative des solutions possibles. Nous avons exploré d'une part les outils de création et de gestion, notamment les CMS et les frameworks, et d'autre part les outils de visualisation, à savoir les bibliothèques java script et les web services. Face à la complexité du monde des Framework et des CSS, et à l'impossibilité de tester personnellement les solutions du marché, nos propositions reposent sur des études comparatives et des retours d'expérience. De ce fait notre étude représente le préambule nécessaire au choix définitif.

Le résultat de notre démarche est synthétisé dans la matrice besoins-propositions-solutions techniques, présentée ci-dessous.

TABLEAU DE SYNTHESE

PROPOSITIONS BESOINS	BANDEAU/TEXTE DE PRÉSENTATION	OUTILS DE DATA VISUALIZATION	FORMULAIRE DE RECHERCHE	TAGS	LIEN HYPERTEXTE	MÉTHODOLOGIE VIDÉO	SECTION BIBLIOGRAPHIE	SECTION PRIVÉ AVEC LOGIN
APPROCHER LE SUJET DE LA RECHERCHE	FW CMS							
VISUALISATION INTERACTIVE ET ATTRAYANTE		CMS B WS						
RECHERCHE UNE THÉMATIQUE			FW CMS	FW CMS				
LIASON DES PAGES/ARTICLES				FW CMS				
ACCÉDER AUX ARTICLES					FW CMS			
EVALUER LA MÉTHODOLOGIE						FW CMS		
ACCÉDER À LA BIBLIOGRAPHIE							FW CMS	
CERNER RAPIDEMENT LES CONTENUS	FW CMS B WS							
ESPACE D'ÉCHANGE ET DE COMMUNICATION SÉCURISÉ ENTRE LES AUTEURS								FW CMS
ACTUALISER LE SITE ET AJOUT DE NOUVEAUX RÉSULTATS	FW CMS B WS			FW CMS B WS				



La volonté d'apporter une touche d'interactivité aux cartes et graphiques du projet Géoscience, nous a conduits à effectuer des tests sur la bibliothèque D3.js. Cette étape de notre travail ne vise pas à construire des représentations abouties mais plutôt à montrer en quoi les outils de Data-visualisation peuvent être efficaces pour communiquer des résultats scientifiques complexes. En ce sens, nous proposons trois graphiques différents à partir d'un même jeu de données. Nous avons été confrontés au problème de la dynamisation des liens entre les données et les graphiques. Il en ressort que les frameworks JavaScript (Angular et Ember) pourraient être une solution envisageable.

Le parcours effectué durant ce projet pose les bases d'une réflexion menée en amont de l'étape de réalisation du site internet. Le travail de conception impliquant de partir de zéro, nous a conduit à trouver notre propre méthodologie et à construire un parcours spécifique. Ce fut un exercice quelque peu dépaysant que de concevoir les étapes qui nous amèneraient à la structuration du site. Néanmoins, malgré les difficultés rencontrées, le projet a été formateur et nous espérons qu'il apportera des éléments de réponse aux questions qui se poseront lors de la phase de réalisation.

## Bibliographie – Webographie

### Le projet Géoscience.

Grossetti Michel, Eckert Denis, Jégou Laurent, Maisonobe Marion, Gringas yves et Larivière Vincent, "Diversification des espaces de production du savoir", Ceriscope, consulté le 22/01/2014

<http://ceriscope.sciences-po.fr/puissance/content/part2/la-diversification-des-espaces-de-production-du-savoir>

Grossetti Michel, Eckert Denis, Jégou Laurent, Gringas yves, Larivière Vincent et Milard Béatrice, " Cities and the geographical deconcentration of scientific activity: A multilevel analysis of publications (1987–2007)", urban Studies, publié le 20/11/2013, consulté le 23/01/2014

[http://usj.sagepub.com/content/early/2013/11/20/0042098013506047.abstract%20\(pi%C3%A8ce%20jointe\)](http://usj.sagepub.com/content/early/2013/11/20/0042098013506047.abstract%20(pi%C3%A8ce%20jointe))

Eckert Denis, Baron Myriam, Jégou Laurent, "Les villes et la science : apports de la spatialisation des données bibliométrique", Mappemonde, n°110, février 2013, consulté le 23/01/2013 <http://mappemonde.mgm.fr/num38/articles/art13201.html>

Maisonobe Marion, "Diffusion et structuration spatiale d'une question de recherche en biologie moléculaire", Mappemonde, n°110, février 2013, consulté le 23/01/2013 <http://mappemonde.mgm.fr/num38/articles/art13202.html>

Levy Rachel, Sibertin-Blanc Mariette, Jégou Laurent, "La production scientifique universitaire dans les villes françaises petites et moyennes (1980-2009) ", Mappemonde, n°110, février 2013, consulté le 23/01/2013 <http://mappemonde.mgm.fr/num38/articles/art13203.html>

### Sites comparables.

Scheidel Walter, Meeks Elijah, "Stanford geo-spatial model for roman world", <http://orbis.stanford.edu/>

Standford Humanities Center, "Mapping the republic of letters", <http://republicofletters.stanford.edu/>

George Damien, Kneijens Rob, "Paperscape", mise à jour : 20/02/2014, <http://paperscape.org/>.

### Sites consacrés à la data visualization :

La fonderie, "Boîte à outils", Expoviz, mise à jour : 02/2014, <http://outils.expoviz.fr/>

Kirk Andy, "Essential collection of visualisation resources", Visualising data, <http://www.visualisingdata.com/index.php/resources/>

Gavin Kirk, "30 best tools for Data Vizualize", CodeGeez, mise à jour : 13/11/2013, <http://codegeekz.com/30-best-tools-for-data-visualization/>

"Density design", Design Density, research lab, <http://www.densitydesign.org/blog/>

Clafolie, "Dataviz", Pearltrees, [http://www.pearltrees.com/#/N-u=1\\_950135&N-p=64778335&N-s=1\\_6891768&N-f=1\\_6891768&N-fa=6881920](http://www.pearltrees.com/#/N-u=1_950135&N-p=64778335&N-s=1_6891768&N-f=1_6891768&N-fa=6881920)



## Conception du site.

Olivier Ertzscheid, Cahiers des charges de réalisation de site internet, mis à jour en septembre 2009, consulté 8/02/2013  
[http://oecoursiut.files.wordpress.com/2009/09/cahier\\_des\\_charges\\_de\\_realisation\\_de\\_site\\_internet.pdf](http://oecoursiut.files.wordpress.com/2009/09/cahier_des_charges_de_realisation_de_site_internet.pdf)

Lerigoleur Emilie, "Base de donnée Paleopyr pour les données paléo-environnement pyrénéennes : de la conception de la base à la conception d'interface cartographique", mémoire de stage de fin d'étude, 47p, septembre 2013, Toulouse.

### *Ergonomie, taxonomie et arborescences.*

Ergonomie, la base d'un site web, OpenClassRoom, mise à jour le 6/12/2013, consulté le 10/02/2014 <http://fr.openclassrooms.com/informatique/cours/le-referencement-de-son-site-web/l-ergonomie-la-base-d-un-site-web>

Panaï Enrico, La taxonomie un cagibi pour l'oubli, Journal du net, 18 janvier 2011, consulté le 10/02/2014  
<http://www.journaldunet.com/solutions/expert/49579/la-taxonomie-pour-eviter-le-cagibi-de-l-oubli.html>

Créer l'arborescence de votre site web en quatre étape, OpenCil, mise à jour : 13/11/2012, consulté le 29/01/2014  
<http://www.lecil.fr/fr/webdesign/creer-arborescence-site-web>

Structure de l'information, Ergonomie Web, mise à jour inconnue, consulté de 29/01/2014 <http://ergonomie-web.studiovitamine.com/arborescence-site-web-structure-de-l-info,350,fr.html>

### *Solutions techniques.*

#### CMS et frameworks.

Blondiau Geoffrey, "Code Igniter et l'architecture MVC", Macdeb, mise à jour : 26/05/2013, consulté le 18/02/2014  
<http://macdeb.net/blog/codeigniter-2-larchitecture-mvc/>

#### Sites de comparatifs

"Comparatif des framework PHP", Beta Social Compare, mise à jour : 07/02/2014, consulté le 18/02/2014  
<http://socialcompare.com/fr/comparison/php-frameworks-comparison>

Pousséo Guillaume, "Pourquoi développer avec un framework est plus avantageux qu'avec un CMS", RevSquare, mise à jour : 02/12/2012, consulté le 06/02/2014  
<http://www.revsquare.com/fr/framework-ou-cms/>

"Comparaison entre l'utilisation d'un framework et d'un CMS pour le développement web", Kreatys, mise à jour inconnue, consulté le 06/02/2014  
<http://www.kreatys.com/Technologies/Comparaison-framework-CMS>

"Comparatif des meilleures CMS du marché", ManiakGeek, mise à jour : 01/07/2012, consulté le 05/02/2014  
<http://maniacgeek.net/2012/06/30/comparatif-des-meilleurs-cms-sur-le-marche-drupal-joomla-spip-wordpress-prestashop-3997>

"Tableau comparatif", Comparateur CMS, mise à jour inconnue, consulté le 05/02/2014  
<http://comparatif-cms.com/tableau-comparatif/>

Bichet, Bruno, "Framework CSS + reset CSS : design from scratch", CSS4design, mise à jour : 04/06/2007, consulté le 15/02/2014  
<http://css.4design.tl/frameworks-css-reset-css-design-from-scratch/>

D'Amico Simone, "24 high quality free CSS framework", Webification, mise à jour : 07/02/2011, consulté le 15/02/2014  
<http://webification.com/24-high-quality-free-css-frameworks>

## Sites officiels des CMS/Frameworks

### *CMS*

Drupal : <https://drupal.org/drupal-7.0/fr>

Wordpress : <http://fr.wordpress.org/>

Joomla : <http://www.joomla.fr/>

MODx : <http://modx.com/>

### *Framework JS*

Angular : <http://angularjs.org/>

Ember : <http://emberjs.com/>

Backbone : <http://backbonejs.org/>

*Framework PHP*

CakePH : <http://cakephp.org/>  
 Zend : <http://framework.zend.com/>  
 Symfony : <http://symfony.com/>  
 Yii : <http://www.yiiframework.com/>  
 CodeIgniter : <http://ellislab.com/codeigniter>

*Framework CSS :*

Blueprint : <http://www.blueprintcss.org/>  
 Bootstrap : <http://getbootstrap.com/>  
 Pure : <http://purecss.io/>  
 Skeleton : <http://www.getskeleton.com/>  
 Foundation : <http://foundation.zurb.com/>  
 Groundwork : <http://groundworkcss.github.io/>

**Sites officiels des bibliothèques et solutions guidées.**

Dans les annexes, le détail des recherches effectuées concernant les solutions techniques est présenté avec des liens vers d'autres sites, tutoriels et supports divers.

*Bibliothèques de visualisation graphique :*

D3: <http://d3js.org/>  
 Highcharts : <http://www.highcharts.com/>  
 Dimple : <http://dimplejs.org/>  
 Vega  
<https://github.com/trifacta/vega/wiki/Tutorial>  
 Infoviz : <http://philogb.github.io/jit/>

*Bibliothèques de création cartographique:*

OpenLayer : <http://openlayers.org/>  
 Leaflet : <http://leafletjs.com/>  
 Polymaps : <http://polymaps.org/>  
 Modest maps : <http://modestmaps.com/>

*Webservices de création de graphiques.*

Tableau public  
<http://www.tableausoftware.com/public/communi>  
 ManyEyes :  
<http://www-958.ibm.com/software/analytics/manyeyes>  
 Chartle : <http://www.chartle.net/>  
 Raw: <http://raw.densitydesign.org/>  
 Micro Stratégie : <http://www.microstrategy.fr/>

*Webservices de création cartographiques.*

CartoDB : <http://cartodb.com/>  
 TileMill : <https://www.mapbox.com/tilemill/>  
 Worldmap : <http://worldmap.harvard.edu/>

## Annexes

<b>ANNEXES 1 : DOCUMENT DE SYNTHÈSE DES RECHERCHES EFFECTUÉS SUR LES OUTILS D'AIDE À LA CRÉATION DE GRAPHIQUES ET DE CARTES (WEB-SERVICES).....</b>	<b>53</b>
<b>ANNEXES 2 : GANTT PRÉVISIONNEL .....</b>	<b>62</b>
<b>ANNEXES 3 : COMPTES RENDUS DES RÉUNIONS .....</b>	<b>63</b>
<b>ANNEXES 4 : LISTE DES FIGURES .....</b>	<b>68</b>

## **Annexes 1 : Document de synthèse des recherches effectuées sur les outils d'aide à la création de graphiques et de cartes (web-services).**

Le résultat des nombreuses recherches effectuées concernant les outils disponibles en matière de visualisation graphique et cartographique a été résumé dans le rapport par l'intermédiaire de tableaux récapitulatifs. Le détail de ces recherches est présenté ci-dessous. Chaque outil est commenté plus précisément. Des liens vers des tutoriels et support complémentaires sont indiqués.

D'autre part, soulignons que cette recherche d'outils disponibles pour la visualisation de données se concentre sur les outils open source. Ils sont nombreux, diversifiés, stables et s'adressent à des publics diversifiés. Certains proposent une interface "clé en main" et s'adresse à des utilisateurs néophytes. Ce sont souvent des web-services. D'autres, bibliothèques JavaScript pour la plupart, peuvent nécessiter des compétences informatiques minimales (permettant de comprendre le fonctionnement d'un Framework voire le fonctionnement d'un code) voire avancées si il s'agit d'adapter le code ou de développer de nouvelles fonctions.

### **Les bibliothèques :**

Une bibliothèque est un ensemble de fonctions développées qui peuvent être assemblées, combinées pour créer des visualisations graphiques et cartes. La différence entre les diverses bibliothèques reposent sur divers critères. De la richesse des fonctions découle les possibilités de visualisation et d'interactions plus ou moins importantes. Celles ci peuvent être complétées par des extensions et plugins qu'il est possible d'ajouter à la bibliothèque.

L'importance et le dynamisme de la communauté est aussi un paramètre important puisque de ce critère découle souvent le nombre de tutoriel et des questions -résolues- sur les forum, la richesse de la documentation et des exemples.

### **Les web-services :**

Le principe est souvent le même. Il propose de charger des données en format xls ou csv voire dbf sur un serveur web. Il est ensuite possible de créer un tableau regroupant un ou plusieurs graphiques et/ou tableaux. Il est ensuite possible soit de communiquer les résultats par mail ou réseaux sociaux soit de récupérer le script et d'intégrer les résultats dans une page web. Ces solutions permettent gratuitement de facilement créer des visualisations dynamiques

Il existe d'autres possibilités d'outil d'aide à la création de graphiques mais souvent il ne propose qu'un export en pdf ou svg et ne conserve par conséquent pas d'interactivité.

## Bibliothèque pour la création de graphiques.

### D3 :

*Liens/documentation/ exemples :*

<http://d3js.org/> : site de la bibliothèque comprenant de nombreux exemples et la documentation complète.

<https://github.com/mostock/d3/wiki/Gallery> :

*C'est quoi ?*

C'est une bibliothèque JavaScript dédiée à la visualisation de données. Créée par Mike Bostock, elle s'est taillée rapidement une solide réputation. Elle propose de très nombreuses possibilités de visualisation et d'interactivité graphiques et cartographiques. Elle fait partie des bibliothèques les plus avancées dans le domaine de la data visualization. Elle possède une bonne documentation et de très nombreux plug-in qui permettent d'adapter les graphiques. A noter : la bibliothèque est gratuite pour des usages non commerciaux.

*Visualisation* : la bibliothèque la plus fournie : Area chart, Bar chart, Line chart, Column Chart, Line Chart, Scatter Plot, Bubble Chart. Pie et donuts Charts, Polar Charts, Spider Charts, Box plots...

*Interactivité* : pop-up, légende interactive (sélection des données et modifications des visualisations par sélection dans la légende), graphiques hiérarchisés.

### Highcharts :

*Liens/documentation/ exemples :*

<http://www.highcharts.com/> : site officiel.

<http://www.highcharts.com/demo/> : section démo à visiter absolument, très didactique avec la possibilité de récupérer tous les scripts js et css.

<http://www.falaf.net/2013/03/11/utilisation-de-highcharts-avec-php-et-mysql/> : tutoriel pour la mise en place d'un pont Php.

*C'est quoi ?*

C'est une bibliothèque JavaScript dédiée à la visualisation de données, très complète. Propose de très nombreuses possibilités et une interactivité tout à fait satisfaisante. Bonne documentation et de très nombreux plug-ins qui permettent d'adapter les graphiques. A noter : la bibliothèque est gratuite pour des usages non commerciaux.

Il peut être aussi associé à Highstock, bibliothèque JavaScript pour la création de timeline.

*Visualisation* : Area chart, Bar chart, Line chart, Column Chart, Line Chart, Scatter Plot, Bubble Chart. Pie et donuts Charts, Polar Charts, Spider Charts, Box plots.

*Interactivité* : pop-up, légende interactive (Sélection des données et modifications des visualisations par sélection dans la légende), graphiques hiérarchisés.

### Dimple :

*Liens/documentation/ exemples :*

<http://dimplejs.org/>

*C'est quoi :*

Bibliothèque JS qui reprend les fonctions D3 avec une documentation détaillée. Permet d'aborder D3 avec uniquement des graphiques simples et de manière très guidée.

*Visualisation :* Area chart, Bar chart, Line chart, Column Chart, Line Chart, Scatter Plot, Bubble Chart.

*Interactivité :* pop-up.

**INfovis :**

*Liens/documentation/ exemples :*

<http://philogb.github.io/jit/index.html>

*C'est quoi :*

Bibliothèque JS permettant de créer des visualisations interactives de données moins riche que D3. Propose aussi de nombreuses solutions de création de graphes.

*Visualisation :* Area chart, Bar chart, Line chart, Column Chart, Scatter Plot, Bubble Chart, Treemap, Graphes.

*Interactivité :* pop-up, graphiques hiérarchisés, sélection d'entités particulières à partir de la légende, graphes interactifs (réorganisation des entités...).

**Vega :** open source.

*Liens/documentation/ exemples :*

<http://trifacta.github.io/vega/>

<https://github.com/trifacta/vega/wiki/Vega-and-D3> : positionnement de Vega / D3.

<http://trifacta.github.io/vega/editor/index.html?spec=population> : démonstration des possibilités de Vega.

*C'est quoi ?*

Vega est une bibliothèque développée par Stanford. Elle propose les mêmes possibilités de visualisation que D3. Il se présente comme un outil plus puissant que D3, notamment en ce qui concerne le data-binding. La nature du projet est un peu obscure. Il présente que peu de documentation si ce n'est le Github.

TABLEAU DES BIBLIOTHEQUES JAVASCRIPT POUR LA CREATION DE GRAPHIQUES.

Caractéristiques	D3	Hightcharts	Dimple	Vega	Infoviz
Visualisation disponibles	+++++	+++	++	+++	+++
Visualisation carto.	Oui	Non	Non	Non	Non
Visualisation de graphes	Oui				Oui
Interactivité	++++	+++	+	++	++
Documentation disponible/ tutoriel	+++	+++	+++	+	++
Prise en main	++++	++	+	+	++
Le plus.	La richesse des possibilités de visualisation	Les thèmes implémentés. Editeur de script dans le site.	Prise en main facile.		

## Web-services d'aide à la création de graphiques et de visualisation :

### Tableau public :

*Lien/documentation/exemples :*

<http://www.tableausoftware.com/public>

<http://www.tableausoftware.com/public/how-it-works> : vidéo qui montre clairement comment le business fonctionne et ça semble très prometteur.

*C'est quoi :*

Un support web pour créer facilement des graphiques et des cartes interactives à partir de fichier csv, xls... et les intégrer à un support web. Il semble plus puissant que les précédents et propose une dimension dynamique plus importante avec la possibilité de jouer sur les paramètres et redessiner immédiatement les graphes et cartes. Certes les visualisations sont plus simples qu'avec D3 mais je trouve la simplicité et l'interactivité offerte vraiment intéressante. Documentation très détaillée, beaucoup de tutoriels vidéo très clairs. Mais tout en anglais.

*Visualisation disponibles :* scatter plots, matrices, graphes, bar charts, bloc histogram, bubble chart, line graph, stack graph, pie chart, tree map, word tree, tag cloud, outils cartographiques.

*Interactivité :* pop-up, sélection d'entités à partir de la légende, graphes interactifs (réorganisation des entités...).

### Manyeyes:

*Lien/documentation/exemples :*

<http://www-958.ibm.com/software/analytics/labs/manyeyes/#home>

*C'est quoi :*

Un support de développement de visualisation graphique a priori basé sur du D3 au vu des type de graph. Comme les précédentes solutions, l'interface permet à partir d'un fichier excell de créer es visualisations dynamiques et interactives (comme Tableaux publics). Les possibilités en termes de graphiques sont très complètes et très simples à prendre en main.

*Visualisation disponibles :* scatter plots, matrices, graphes, bar charts, bloc histogram, bubble chart, line graph, stack graph, pie chart, tree map, word tree, tag cloud, outils cartographiques.

*Interactivité :* non évaluée.

### MicroStrategy :

*Lien/documentation/exemples :*

<http://www.datavisualization.fr/blog/pilotage-reporting/>

<http://apandre.wordpress.com/2013/10/22/free-microstrategy-visualization/>

*C'est quoi :*

Un outil gratuit proposé par la société MicroStrategy, éditeur de logiciels de business intelligence, en parallèle de ces offres payantes. Elle propose un nombre considérable de possibilités et de création de tableaux.

*Visualisation disponibles :* scatter plots, matrices, graphes, bar charts, bloc histogram, bubble chart, line graph, stack graph, pie chart, tree map, word tree, tag cloud, outils cartographiques



**RAW :**

*Lien/documentation/exemples :*

<http://raw.densitydesign.org/>

<https://github.com/densitydesign/raw>.

*C'est quoi :*

C'est une solution élaborée par le laboratoire DensityDesign Research Lab de l'école polytechnique de Milan. Elle repose sur la bibliothèque D3. Contrairement aux autres solutions guidées, Raw propose des visualisations D3 inédites : cluster dendogram, sankey... C'est son principal atout.

*Visualisations disponibles :* treemap, bubble chart, circle pocking circular dendogram, hexagonal binning, alluvial diagram.

*Interactivité :* non évaluée.

**Chartle.net :**

*Lien/documentation/exemples :*

<http://www.chartle.net/gallery>

*C'est quoi :*

Un outil sur le web qui permet un peu comme Raw de créer des diagrammes, des histogrammes... à partir de données sur fichiers Excell puis de les importer dans un site web notamment. Les possibilités proposées sont assez limitées et sans interactivité. Elle ne permet pas en outre la possibilité de créer de visualisations cartographiques.

*Visualisation possible :* Line charts, Bar charts, Column charts, Area charts, Pie charts, Maps,

*Interactivité :* non évaluée.

Tableau des solutions guidées pour la création de graphiques.

Caractéristiques	Tableau public	Many Eyes	Micro Strategy	Chartle	RAW
Richesse des visualisations proposées	+++	+++	+++	+	++
Visualisation carto.	Oui	Oui	Oui	Non	Non
Création de graphes.	Non	Oui	non	Non	Non
Documentation disponible/ tutoriel	+++	++	+++	+	++
Facilité d'utilisation	+++	NC	+++	+++	+++
Le plus.	Tutoriel Vidéo.	Aide au choix de visualisation en fonction de la nature des données.			Des possibilités de visualisation D3 originales et inédites.

## Outil de développement de cartographie interactive.

### **Openlayers :**

*Lien/documentation/exemples :*

<http://openlayers.org/>

<http://geotribu.net/node/3>

*C'est quoi :*

C'est a priori l'API open source la plus complète en matière de web-mapping. Elle offre un nombre considérable de fonctions et de possibilités. Elle est particulièrement flexible. Elle est aussi l'option la plus adaptée lorsqu'il s'agit de mettre en place une liaison avec Postgis.

*Le plus:* c'est ma bibliothèque la plus complète et la plus puissante.

### **Leaflet :**

*Lien/documentation/exemples :*

<http://leafletjs.com/>

<http://giscollective.org/tutorials/web-mapping/leaflet-1/> : tutoriel de base.

<http://www.youtube.com/watch?v=1VYvjHb6KeM> : petit tutoriel vidéo de mise en place de base.

<http://www.geotribu.net/node/423> : petit tuto réalisé par Géotribu.

*C'est quoi :*

Développée par Cloud Made, cette bibliothèque de création de carte interactive, particulièrement légère, prend de plus en plus d'ampleur. Elle est notamment dédiée au développement d'applications mobiles. De nombreux plugg-in permettent d'améliorer ses performances en fonction des besoins de l'utilisateur sans alourdir inutilement le package. Quelques tutoriels sur le site permettent de prendre en main les basiques de l'application.

*Interactivité :* non évaluée.

*Le plus :* la facilité de prise en main et les nombreux tutoriels disponibles pour des premiers pas.

### **Polymaps :**

*Lien/documentation/exemples :*

<http://polymaps.org/>

*C'est quoi :* une bibliothèque JavaScript de création de cartes interactives en format svg avec des formats de données de type Json et GeoJson. Adapté plus particulièrement à des échelles régionales et urbaines.

Documentation : bibliothèque détaillée et quelques exemples avec des scripts disponibles.

Interactivité : zoom, mouse Wheel control,

*Le plus :* légèreté et simplicité de la bibliothèque, des templates originaux de visualisation des cartes.

### **Modestmaps :**

*Lien/documentation/exemples :*

<http://modestmaps.com/>

*C'est quoi* : une bibliothèque JavaScript de création de carte interactives, développée par Tom Carden certes modeste mais qui offre des possibilités basiques et souvent largement suffisante d'interactivité. Son atout principal est la légèreté de son code, argument non négligeable alors que se développent les applications mobiles. Ses points faibles : il ne supporte pas le format GeoJson et ne permet de travailler en multi-couches.

*Interactivité* : non évaluée.

*Le plus* : La bibliothèque d'exemples des différentes fonctions qui fournie pour chacune d'entre elles un index HTML, un script commenté avec parfois la structure de données et un css. Les extensions qui permettent d'améliorer les possibilités de la bibliothèque en fonction des besoins.

### **Kartograph.**

Lien/documentation/exemples :

<http://kartograph.org/> : site officiel. ,

<http://blog.bandinelli.net/index.php?post/2013/08/16/Des-cartes-avec-Kartograph> : tutoriel rapide pour la création d'une carte interactive basique.

C'est quoi :

Développé par Gregor Aisch, Kartograph est un framework léger qui réunit deux bibliothèques de fonctions. Une bibliothèque en Python permet de gérer les fonds de cartes en format .svg. La seconde bibliothèque est développée en JavaScript. Elle s'appuie sur la bibliothèque JS Raphael. Cette dernière permet de gérer n'importe quel objet vectoriel, notamment des graphiques. La bibliothèque JS s'occupe des symboles et des outils d'interactivité.

Le plus : la légèreté de la bibliothèque.

	OpenLayer	Leaflet	Polymaps	Modest maps
Richesse de la bibliothèque.	++++	+++	++	+
Facilité de prise en main.	+	+++	++	+++
Possibilité de travailler en multicouches	Oui	Oui	NC	Non
Le plus.	La richesse des fonctionnalités.	Facilité de prise en main.	Les templates de cartes disponibles.	La présentation des fonctions dans des scripts.
Richesse de la bibliothèque.	++++	+++	++	+

### **Des supports de développement cartographique.**

Ces solutions sont globalement plus complexes à prendre en main que les solutions guidées pour la création de graphiques. Leur point fort : elles sont dédiées à la publication. En effet, peu d'entre elles proposent une réelle interactivité. Elles peuvent donc constituer une alternative pour qui ne dispose pas de logiciel SIG ou souhaite réaliser des cartes basiques pour le web.

**CartoDB** : web service.

*Lien/documentation/exemples.*

<http://cartodb.com/>

<http://geotribu.net/node/486> : petit tour d'horizon de l'outil par la 'géotribu'.

<http://here.com/livingcities/> : application développée montrant au cours d'une journée le trafic routier dans certaines grandes villes du monde.

<http://vizzuality.github.io/rollingstonesmap/#/> applications qui montre au fil des ans les cartes des tournées des Rolling stones.

*C'est quoi :*

Une application développée par... qui permet de charger des données à partir de différents formats (xsl, csv, shp, tab, dump sql...) possédant bien sûr des champs latitude/longitude. Il est ensuite possible de construire une cartographie interactive de ces données. Il est ensuite possible d'intégrer ces représentations à des sites internet. Elle donne énormément de possibilités pour modifier la visualisation. Cette solution est gratuite, dans une certaine limite, 5 tables et 5mb de mémoire.

**TileMill** : logiciel.

*Lien/documentation/exemples :*

<https://www.mapbox.com/tilemill/> : site de Mapbox pour TileMill.

<http://dataforradicals.com/the-insanely-illustrated-guide-to-your-first-tile-mill-map/> : tutoriel pas à pas pour la création d'une première carte avec TileMill.

*C'est quoi :*

Un logiciel multiplateforme qui permet de réaliser des cartes destinées à la publication sur le net. Il est possible d'insérer des éléments interactifs minimaux (pop-up et mouse-over) et des graphiques. Cependant il n'est possible de travailler que sur une seule couche pour le moment. L'application accepte des formats très divers de données : shm, klm, geojson, csv... et permet une connexion avec une base Postgis. Il est possible de sauver et de réutiliser la carte réalisée. D'après les nombreux témoignages, elle nécessite un petit temps de prise en main.

**WorldMap** : web-service.

*Lien/documentation/exemples :*

<http://worldmap.harvard.edu/>

*C'est quoi :*

Une solution de cartographie guidée développée par Harvard. Elle permet de créer des cartes à partir de différents fonds. Comme CartoDB, elle offre la possibilité de charger des données et de les mettre en forme avec une pointe d'interactivité : sélection d'objet sur la carte, pop-up, sélection des couches... Toutefois, l'interface n'est pas très engageante et la prise en main manque d'explications.

*Le plus :* Cet outil dispose d'un très grand nombre de types de fonds de cartes.

Caractéristique	Carto DB.	Worldmap	Tilemill
Facilité de prise en main	+++	++	++
Formats acceptés	+++	NC	+++
Richesse des visualisations	+++	++	+++
Interactivité	++	NC	++
Le plus	Possibilité d'exporter en HTML mais aussi en par une API SQL ou cartographique.	Richesse des fonds de cartes disponibles directement sur le site.	Richesse des possibilités de ... cartoCSS. Ce n'est pas un webservice.

TABLEAU DES OUTILS CARTOGRAPHIQUES GUIDES.

Annexes 2 : GANTT prévisionnel



## Annexes 3 : Comptes rendus des réunions

### *Réunion de lancement du projet du 20/01/2013*

#### 1. Présentation du projet géoscience.

Géoscience est un programme de recherche qui a duré 3 ans entre 2010 et 2013, pour lequel ont collaboré trois laboratoires, Toulouse, Montréal et Paris. Son objectif : évaluer l'organisation spatiale des publications en fonction de la BD Thompsen et Reuters, Web of Science. Cependant à ce jour, toutes les données n'ont pas été traitées et les recherches déjà effectuées ont ouvert de nouvelles pistes de réflexion.

##### 1.1. La première étape du programme :

Evaluer l'évolution dans le temps de l'organisation spatiale de la production scientifique. Différents axes de recherches ont été tirés de ce premier objectif.

La première relève de l'histoire de la production scientifique et des espaces de production.

Le second se concentre sur l'évolution de la production scientifique quantitativement et spatialement. Les données ont été dans un premier temps fournis par l'université de Montréal CIRST, laboratoire qui travaille sur la bibliométrie, qui achète les données de la base Web of Science, qui répertorie les publications et les métadonnées concernant ces publications (auteurs et localisation). Les données extraites ont été comptées et regroupées par ville, province et pays. Les données au niveau de la ville représentent une avancée considérable puisque jusqu'à présente les données les plus fines concernaient le niveau régional. Une première série temporelle a été étudiée pour 1978, 1988, 1998 et 2008. Cette première série d'analyses a nourri la publication d'un premier article dans 'Urban studies'. Celui-ci a permis de réfuter la thèse de la concentration de l'activité scientifique défendue par Mathiessen.

##### 1.2. La deuxième étape du projet.

Après ces premiers constats, il s'est avéré nécessaire d'avoir recours à des données plus fines d'un point de vue géographique et d'autre part d'extraire des informations quant aux relations entre les villes.

Un des aspects méthodologiques important a été de définir les entités géographiques de base. Il est très vite apparu que l'adresse "communale" n'était pas la plus indiquée. En effet, en fonction de l'organisation spatiale des villes suivant les pays, cette échelle présentait de prime abord des disparités trop importantes. En effet, Ensat est référencé à Auzeville alors qu'elle appartient à la "sphère" de Toulouse. Il a fallu établir une méthodologie d'agrégation des villes en se concentrant sur les variations de densité.

On a d'autre part choisi de travailler sur des données plus précises en se focalisant d'une part sur différentes disciplines et d'autre part sur la visibilité des articles par les indices. Pour ce faire, on a eu recours aux données fournies par l'Observatoire des Sciences et Techniques à Paris, qui établit un rapport annuel sur l'état de la recherche en France. Il a fourni les données du web of science contre les méthodologies créées par le laboratoire. Les données utilisées sont focalisées sur deux périodes de trois ans : 1999-2000-2001 et 2006-2007-2008. Ces

temporalités permettent de lisser les phénomènes épisodiques et de rendre les données comparables.

Aujourd'hui il y a 8 millions de données disponibles, 14.5 millions d'adresses (qui représente un an et demi de géocodage, particulièrement sur la France). Ce qui est ressortie de ce projet : des données géocodées proprement, une méthodologie novatrice, des articles, un premier site de visualisation des relations entre les villes.

Objectifs du projet : le but final : création d'un site pour rendre le projet de recherche visible et compréhensible par le grand public.

Trouver des solutions de visualisation et d'interactivité pour rendre le projet visible et abordable par le grand public en limitant la complexité.

Etablir un cahier des charges : liste d'opérations à effectuer de manière obligatoire et/ou facultatives.

Gestion de projet.

Pistes de réflexion.

Lecture des articles fournis et fouiller sur le net pour trouver des bibliothèques et des exemples de visualisations.

Reprendre les visualisations statiques déjà créés au bureau carto.

Réfléchir à l'approche graphique et cartographique.

Mettre mieux en avant les situations nationales différenciées.

Rendre CoSciMo interactif avec la base de données.

Mettre en place un système de publication automatique.

Rechercher du côté des CMS pour la partie conception web plutôt que de passer par du Html.

Sécurité des données : elles ne doivent pas être lisible directement. Mais présentées agrégées.

Faire connaître les publications pour que les articles soient réutilisés.

### *Réunion du 28/01/2013*

#### **Générale:**

Va-t-on travailler sur les données brutes ou les données statistiques que vous avez créées?

*Sur les données brutes. L'application va aller chercher les données directement sur le serveur.*

Quel niveau protégé? Es ce que le regroupement par ville est un niveau suffisant de protection? Es ce que le nombre de publications par ville passe le seuil de sécurité? Par exemple Barcelone a produit 15251 articles en 2008.

*Il ne faut pas que les valeurs de comptage du nombre de publications par ville et par année soit accessibles. C'est le résultat de tout un ensemble de travaux et personne ne doit être en mesure de récupérer ces données et produire à partir de celles-ci. Un moyen de palier à ce problème : utiliser la normalisation des données et des indices. Ex CoSciMo données comparable entre chaque onglet mais elles sont normalisées entre 0 et 1000. Mais il ne faut pas se focaliser sur ce problème puisqu'il sera posé au moment de l'implémentation du site.*

**Direction générale à donner à la recherche.**

**Réfléchir au public et à l'utilisation du site.**



Intégrer la dimension utilisateur à notre recherche. Il faut définir le public qui sera susceptible de visiter le site, ses objectifs de navigation, ses besoins. A partir de cette base de réflexion, il faut adapter l'ergonomie et l'organisation des données en fonction des paramètres ainsi déterminer.

On peut rapidement imaginer deux types de publics. D'une part, un public "initié" probablement constitué de chercheurs d'horizons différents (géographes, sociologues...), de doctorants qui va probablement venir sur le site avec un besoins particulier et qui doit pouvoir rapidement accéder à l'information. D'autre part, on peut imaginer un public néophyte (prof de lycée, journaliste scientifique...) qui va venir sur le site sans idée préconçue et qui doit rapidement savoir quel parti tirer des données disponible et comprendre comment utiliser le site (outils, interface...). Pour ce dernier public, il pourrait être intéressant de réfléchir à des moyens de guider l'accès à l'information par l'intermédiaire d'animations, de vidéos... Le visiteur lambda doit pouvoir savoir où entrer dans le site et où sortir et ce qu'il peut tirer du site.

On peut faire le choix de créer une interface et laisser le visiteur explorer les possibilités proposées au risque que celui-ci se perde et ne comprenne pas l'intérêt et les ressources proposées par le site.

Par exemple, l'idée d'une carte mondiale regroupant beaucoup d'informations accessibles par zoom ou onglets peut s'avérer complexe pour un utilisateur néophyte qui pourrait passer à côté de beaucoup d'informations. On pourrait par exemple imaginer un accès direct par pays.

On pourra se référer à différents sites qui présentent des informations statistiques complexes de manière simple et interactive. Gapminder. Org (<http://www.gapminder.org/>), site sur l'excellence des universités (<http://www.excellencemapping.net/>), la présentation de Al Gore (<http://www.encyclovideo.net/une-verite-qui-derange.html>) ou le site de la FAO.

Il faut maintenant aborder la question dans l'autre sens : réfléchir à ce qu'on veut présenter et non plus aborder le comment.

Pour faire passer simplement des informations complexes, il est possible de 'raconter une histoire', de mettre en scène l'information ou, en d'autres termes, établir un fil directeur. Pour lier les informations les unes aux autres et amener le visiteur d'idées complexes.

#### **Réfléchir à la structure du site.**

Pour l'instant nous avons réfléchi à la question de la déconcentration à laquelle nous pourrions ajouter la dimension relation de co-publication déjà présenté par le site CoSciMo.

D'autres approche de la question peuvent être intégrer. Marion Maisonobe et Laurent Jégou s'engagent à nous fournir une liste des thématiques annexes sur lesquelles ils travaillent et dont les résultats pourraient être intégrer au site.

Le site doit refléter les principales idées et apports de l'équipe de recherche. La notion d'agglomération et de regroupement est la première innovation mise au point par l'équipe. Ensuite, c'est l'idée de déconcentration qui constitue l'avancée majeure de l'équipe. Cette idée peut être développée en divers "chapitres". On peut imaginer un focus sur la situation de la Russie, ou à des situations de villes/agglomérations spécifiques.

Le site doit présenter deux aspects de la question de recherche. D'une part, on doit pouvoir accéder à une mise à plat des résultats. Puis ouvrir vers les analyses apportées par l'équipe.

### **Quelques réflexions sur les propositions effectuées par le groupe.**

La réflexion menée était une étape incontournable de la démarche du projet. Globalement, les propositions effectuées ont été validées. Quelques observations ont été faites sur l'approche proposée pour chacune des représentations.

*Diagramme a points.*

*Histogramme.*

L'idée de graphe à plusieurs dimensions est intéressante.

L'idée d'utiliser le mode "*cross-filter*" est intéressante mais complexe à mettre en place. La notion de temporalité étant restreinte, il faudrait éventuellement remplacer cette notion par une notion géographique.

*Cordes.*

Les idées évoquées pour améliorer le graphe à cordes sont validées.

*Carte de l'atlas.*

Une carte globale à l'échelle mondiale semble un peu trop complexe pour l'utilisateur. Il faudrait penser à un moyen de ... il est envisageable de créer une animation pour aiguiller le visiteur quant aux possibilités offertes par l'outil.

Concernant l'idée de zoomer sur un pays pour obtenir des informations complémentaires. Il peut être particulièrement compliqué de créer une telle interface. Par contre il pourrait être intéressant de pouvoir accéder pour un pays à deux cartes présentant les deux états

### *Réunion du 04/02/2013*

La réunion commence avec la présentation du cahier de charges :

1. La première partie est un peu trop générique, il faut approfondir les paragraphes concernant le contenu du site et les propositions d'arborescence .
2. Il manque en outre une partie qui explicite les modalités de navigation des différents usagers : qui vient sur le site, qu'est ce qu'il cherche et comment il le trouve ? Le travail doit être fait pour les différents profils d'utilisateur. Il faut donc écrire un texte présentant pour chaque profil d'utilisateur la façon dont celui-ci utilisera le site.
3. Ajouter une autre catégorie d'usagers : l'administrateur général du site et les auteurs/chercheurs qui font partie du projet et qui alimenteront le site au fur et à mesure que la recherche avancera.

Pour ceux d'entre eux qui n'ont pas de compétence de développement il faut prévoir une manière de faciliter le travail . Exemples :

- A. alimenter la base de données en ligne par des fichiers texte ou excell, par des frameworks qui convertissent les données en tableaux directement utilisables par les bibliothèques de data visualisation.
  - B. Ou encore des outils qui convertissent automatiquement des tableaux en fichiers .csv
  - C. Des outils qui permettent d'interroger la base sans passer par des requêtes « en dur »
4. Le site doit pouvoir évoluer avec la recherche et avec la structure de la base de données. Dans cette optique les arborescences proposées sont trop rigides et doivent

prévoir une partie plus souple qui peut être modifiée au fur et à mesure. Donc, tout en gardant les deux possibilités de structuration du site, et sans avoir à choisir entre les deux, il faut prévoir une partie centrale en évolution selon la progression du projet. Cette partie ne doit pas être forcément détaillée mais il faut effectuer des recherches et prévoir des possibilités de structure, qu'on pourra approfondir, développer, tester dans le futur.

Concernant les modalités de navigation et utilisation du site, il est important d'imaginer une utilisation par sélection individuelle : à travers des rubriques ou des mot-clés on doit pouvoir accéder aux différentes thématiques sans avoir à parcourir toute l'arborescence. L'exemple donné est celui de WordPress et des blogs qui sont alimentés en continu : l'auteur écrit un texte et à partir des mot-clé celui-ci est référencé dans le site. (Certains blogs fonctionnent avec des rubriques automatiques créées à partir des tags. Pour celui qui implémente cela permet d'attacher des mots clés à un contenu qui entre alors automatiquement dans la bonne rubrique. )

5. La deuxième partie du cahier de charges sera plus technique et plus détaillé, avec la description de la charte graphique, l'analyse des besoins, les fonctionnalités utiles..

Pour cette partie il faut prévoir plusieurs propositions (tant pour les styles que pour les frameworks) mais il ne faut pas nécessairement tester les solutions ; cela dépend aussi du temps qu'on veut consacrer à la partie « réalisation ». On peut rester à un niveau assez générique, de recherche et comparaison théorique entre plusieurs outils. A titre d'exemple, explorer les possibilités offertes par Bootstrap, Raw du Density design Lab, etc.

**Annexes 4 : Liste des figures**

Fig. 1 : Diagramme à bulles illustrant l'évolution de la production scientifique.....	8
D. Eckert, M. Grossetti, L. Jégou	
Fig. 2 : Histogramme illustrant la production scientifique et son évolution.....	9
D. Eckert, M. Grossetti, L. Jégou	
Fig. 3 : Evolution du rang des 100 premières villes dans le monde.....	10
L. Jégou, M. Maisonobe	
Fig. 4 : cartographies illustrant la déconcentration de la production du savoir en Asie du SE et à l'échelle mondiale.....	11
Carte Asie SE: M. Baron, M. Maisonobe, L. Jégou	
Carte du monde: M. Baron, L. Jégou	
Fig. 5 : Arborescence figée, mettant en valeur la dimension spatiale.....	20
Fig. 6 : Arborescence figée, mettant en valeur les thématiques de recherche.....	21
Fig. 7 : Arborescence hybrides structurée selon un système de tags.....	22
Fig. 8 : Schéma présentant le processus de data binding bidirectionnel.....	28
Fig. 9 : Premier niveau - les publications par classe.....	37
Fig. 10 : Second niveau - les publications par villes pour la classe 10-20K.....	38
Fig. 11 : Vue d'ensemble du graphe.....	39
Fig. 12 : Zoom sur la classe plus-20K.....	39
Fig. 13 : Arborescence déployée pour les classes plus-20K et 5-10K.....	40
Fig. 14 : Marseille et ses collaborations.....	41
Fig. 15 : Positionnement du framework Angular au regard de la structure MVC....	43