



**HAL**  
open science

## L'EMI en partage : essai de cartographie des acteurs

Élisabeth Schneider, Alexandre Serres, Angèle Stalder

### ► To cite this version:

Élisabeth Schneider, Alexandre Serres, Angèle Stalder. L'EMI en partage : essai de cartographie des acteurs: A qui appartient l'EMI (Éducation aux Médias et à l'Information)? 10e Congrès des Enseignants Documentalistes de l'Education Nationale "Enseigner-apprendre l'information-documentation", FADBEN (Fédération des Associations de Documentalistes Bibliothécaires de l'Education Nationale), Oct 2015, Limoges, France. sic\_01217549

**HAL Id: sic\_01217549**

**[https://archivesic.ccsd.cnrs.fr/sic\\_01217549](https://archivesic.ccsd.cnrs.fr/sic_01217549)**

Submitted on 21 Oct 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

« *L'EMI en partage : essai de cartographie des acteurs* »

**Elisabeth Schneider**, maîtresse de conférences, ESPé de Caen,

**Alexandre Serres**, maître de conférences, URFIST de Rennes, Université Rennes 2

**Angèle Stalder**, doctorante, ATER, Université de Rouen

Note : ce texte s'accompagne d'un certain nombre de fichiers (ppt, pdf, xls, gexf), disponibles à partir du [Carnet du GRCDI](#) sur Hypothèses.

### **Introduction**

Le rapport de la concertation sur la future Loi de refondation, paru début 2013, inclut l'éducation aux médias et à l'information dans un paragraphe sur « apprendre le numérique ». Lorsqu'est organisée la Conférence nationale « Cultures numériques, éducation aux médias et à l'information », Jean-Louis Durpaire évoque l'éducation à l'information née d'une réflexion commune entre un groupe de travail de l'IGEN et de l'Ifé depuis mars 2012. Eric Bruillard, lui, parle d'« éducation à l'information et aux médias » dans le mot introducteur de la conférence et utilise information et informatique mais non médias dans la suite. L'annonce de la loi sur la refondation de l'école mentionne tout d'abord « une véritable éducation aux médias ». Mais cette conférence restera comme l'EMICnf. Ces éléments montrent que l'expression et le sigle qui semblent consacrés aujourd'hui, « EMI », ont oscillé entre différents domaines jusqu'à ce que l'édiction de la loi les stabilise.

Cependant, dès les premiers temps et encouragés par la demande de participation des organisateurs à l'EMICnf, des professionnels, des institutionnels ont commencé à publier, mettre en ligne, discuter ce qui semblait relever de ce nouveau domaine d'enseignement. Et rapidement des questions polémiques se sont fait jour : qui est chargé de cet enseignement ? A quel autre domaine peut-on/doit-on le rattacher ? Comment le définir, le distinguer, le rapprocher de l'EAM ? Sans pour autant que des réponses définitives soient données et acceptées au fil des mois. Pour un lecteur, professeur documentaliste ou de discipline qui chercherait à comprendre de quoi il s'agit, cela constitue un ensemble foisonnant, pluriel et insaisissable.

Nous-mêmes, à la lecture d'un certain nombre de publications, nous nous sommes demandés ce qu'il fallait en comprendre concernant le contenu mais aussi le positionnement éditorial qui se construisait. En effet, publier en ligne correspond à des enjeux politiques (au sens large) de visibilité. Qui veut donner à penser quoi sur l'EMI ?

Très vite, il nous a semblé que la méthode de recherche la mieux adaptée à la mise en visibilité des acteurs de l'EMI sur le web était une cartographie : pour les rendre visibles tout d'abord, puis pour mieux comprendre les stratégies d'acteurs, leurs différentes conceptions de l'EMI, les variations dans les définitions, leurs manières de s'impliquer, les différents

niveaux d'échelle de leurs interventions, les différentes approches pédagogiques. Dans quelle mesure les acteurs en ligne configurent le champ de l'EMI et contribuent à le définir ? Quels liens s'établissent entre des acteurs de natures diverses et proposant des ressources différenciées ? Quelles récurrences s'organisent dans cette mise en réseau ? Au-delà des liens, comment peut-on qualifier cette dernière et que nous dit-elle d'un processus de configuration et de mise en visibilité d'un domaine d'enseignement en émergence ?

Après avoir explicité notre démarche méthodologique, nous essaierons de montrer comment les diverses publications contribuent à délimiter des espaces de l'EMI en les référant à des acteurs, des enjeux, des textes scientifiques, mais aussi comment se constituent des « acteurs-réseau » de ce domaine.

En explorant rapidement le web avec pour requête « EMI », nous avons identifié des hauts-lieux de diffusion, de discussion, divers dans leur nature. D'autres moins visibles mais que l'on retrouvait régulièrement dans les résultats ou les documents. Cependant, comment aller au-delà du simple repérage et mesurer cette dimension du réseau ? C'est le détour par les travaux sur les acteurs-réseaux du sociologue Bruno Latour qui a semblé répondre à notre interrogation.

Selon Latour, est un acteur celui qui dit ou apporte « quelque chose, qui fait quelque chose qui fait la différence ». Il n'est pas obligatoirement un être humain mais une entité qui peut être un dispositif, une institution, qui contribue à configurer le réseau dans lequel il s'inscrit, qui contribue à le faire évoluer en agissant.

Associée à la notion d'acteur, il faut donc comprendre celle de réseau, nécessaire quand il s'agit d'explorer des contenus sur le web. Un réseau est constitué de nœuds que sont les acteurs, les sites par exemple, et de liens. Le réseau produit une réalité qui est plus que la juxtaposition des liens. Les nœuds ont un impact les uns sur les autres et les interrelations configurent le réseau en mode dynamique : de nouveaux liens peuvent apparaître, certains se renforcer, etc.

### **Brève présentation de la démarche**

Comment cartographier les acteurs de l'EMI et leurs relations, identifier les pôles les plus influents, les catégories d'acteurs présentes ? Le recours aux outils d'exploration et de cartographie du web s'est avéré ici indispensable, un travail purement manuel étant littéralement impossible. Ces outils sont-ils pour autant la panacée ? Évidemment non, et leur utilisation exclusive signifie par exemple un abandon de fait de tout document non accessible sur le web. De plus, ces outils ne portent pas sur l'analyse des contenus diffusés, mais sur les relations entre ceux qui les diffusent. Une fois ces deux limites méthodologiques rappelées, quelle a été notre démarche ? Schématiquement, on peut y distinguer quatre grandes étapes, assez classiques dans tout travail de ce type :

- la collecte et la constitution du corpus ;
- l'affinement, la description, le traitement de celui-ci ;
- le paramétrage et la production de cartographies ;
- l'interprétation de leurs résultats.

Trois outils ont été successivement utilisés :

- le premier est bien connu, le gestionnaire de signets Diigo, pour commencer à constituer « à la main », un premier noyau de sites et de ressources web ;

- le deuxième, beaucoup moins connu, est un outil permettant la constitution semi-automatique de corpus et l'établissement de cartographie des liens entre sites ; il s'agit de Hyphe, outil développé par le MediaLab de SciencesPo ;
- et le troisième est un outil de cartographie et de visualisation, qui connaît un certain succès dans les milieux de la recherche : Gephi, outil en open source.

Avant de présenter plus en détails cette démarche et ses résultats, deux mots sur ses contraintes et sur ses conditions de réalisation. Ce travail de recherche, visant la cartographie de corpus web, a nécessité un temps considérable (plusieurs dizaines d'heures), pour deux raisons : d'abord, tout projet d'analyse et de cartographie du web, même avec de puissants outils, nécessite toujours plus de « temps humain » qu'on ne le pense, car le traitement automatique ne remplace jamais le nécessaire travail de sélection, de description et d'interprétation, que seuls peuvent faire les pauvres humains. Et la seconde raison tient à l'inexpérience des dits-humains concernés, qui ont dû apprendre en marchant à utiliser et maîtriser ces outils, parfois très complexes comme Gephi. Ce travail est donc une première étape, un résultat provisoire d'une démarche de cartographie, appelée à se poursuivre et à s'affiner.

## **Quel corpus ?**

### **La méthode d'élaboration du corpus**

Trois étapes dans l'élaboration du corpus : la collecte, le traitement, la description.

D'abord la collecte. Étant donné les délais assez courts qui nous étaient impartis, ce recensement a été très sélectif. En effet, nous avons réduit notre sélection aux acteurs (humains et non-humains) les plus impliqués dans l'EMI. D'autres critères de sélection se sont ajoutés très vite :

- un critère géographique tout d'abord : nous avons limité la recherche aux publications d'acteurs français ; tout en prenant aussi quelques acteurs francophones importants ;
- un critère temporel : nous sommes partis de la Loi d'orientation de 2013, instituant l'EMI ;
- un autre critère temporel qui portait sur la durée, à savoir l'ancienneté de l'implication des acteurs (individus, organisations) dans l'éducation aux médias et/ou à l'information.

Notons que ce travail a été fait par trois chercheurs qui sont eux-mêmes partie prenante de la réflexion sur le domaine de l'EMI : nous participons au GRCDI, à l'ANR Translittératie, avec des centres d'intérêt scientifiques en grande partie convergents (la culture de l'information, la didactique de l'information-documentation). Notre démarche est ainsi réflexive.

La recherche des textes, ainsi que leur qualification, ont été conduites de façon collective à l'aide d'une plateforme de partage de signets, Diigo en l'occurrence. Il s'agissait de capitaliser les textes, de les indexer en vue de les exporter dans Hyphe. Cette étape de recherche, sélection, collecte et description a permis un premier filtrage de nos données, grâce à une méthode d'indexation partagée qui a porté sur les éléments du résumé. Nous avons décidé de modifier systématiquement le résumé généré automatiquement par Diigo en élaborant un cadre descriptif pour disposer d'une grille d'analyse commune, permettant une approche « latourélienne » de ces différents acteurs, i.e. une lecture neutre des acteurs, de

leurs stratégies, de leurs discours. Le choix des tags a également fait l'objet d'une méthodologie : qualifier systématiquement l'acteur, distinguer dans le contenu ce qui relève de l'EMI, de l'EAM (Éducation aux Médias) et de l'EAI (Éducation à l'Information), mais aussi de la culture de l'information et de la culture numérique.

Deux mots sur les « textes » répertoriés pour constituer un corpus des publications des acteurs de l'EMI. Ici, il faut entendre texte au sens où Yves Jeanneret le définit : « *Le texte est un objet matériel, singulier, complexe, hétérogène ; cet objet repose sur une union intime entre le support et le message ; il repose sur des codes stricts et d'autres plus flous en matière d'assemblage de signes (le texte alphabétique étant un cas particulier) ; il peut être doté de sens par la confrontation à des modèles acquis ; il propose des marques pour une relation énonciative (implication de communication) et des représentations du monde ; tout en définissant ses propres frontières, le texte est ouvert, car il entre en relation, explicite ou non, avec d'autres textes.* »<sup>1</sup>.

Quels furent les résultats de cette première collecte ? 147 références, collectées de juillet à fin août 2015 ; les textes parus depuis n'ont pas été intégrés à ce premier corpus, où nous avons constaté la prédominance de textes issus de la sphère professionnelle (54), en particulier celle des professeurs documentalistes, ce qui marque une appropriation indéniable de l'EMI par ces derniers. Nous avons également noté une faible représentation des textes institutionnels (21). Enfin, il y a encore moins de textes scientifiques (15). Pour ce qui est des autres textes, ce sont des textes issus du monde associatif.

Une analyse des occurrences de tags montre aussi une prédominance des acteurs professionnels. Quelques « figures » émergent, les chercheurs identifiés au champ de l'Éducation aux Médias (Jacques Kerneis, Divina Frau-Meigs ou Geneviève Jacquinet-Delaunay). Ce qui est également intéressant, ce sont les sujets corrélés à l'EMI : EAM, EAI, culture de l'information, mais aussi culture numérique, littératie et translittératie. Si l'on regarde les tags les plus utilisés, encore une fois, on peut noter également la fréquence du tag « éducation aux médias ».



<sup>1</sup> Yves Jeanneret, « Informatique literacy : manifestations, captations et déceptions dans le texte informatisé », *SPIRALE - Revue de Recherches en Éducation*, 2001, n° 28

Mais cette analyse est celle des textes collectés, et non celle des résultats de la mise en réseau de tous les liens contenus dans ces textes et qui vont dessiner le domaine de l'EMI sur le web. Autrement dit, après la phase de la première collecte par un questionnaire de signets, est venue la phase cruciale de la cartographie de ce domaine sur le web.

## Présentation de Hyphe et de la méthode de collecte des sites

Qu'appelle-t-on au juste la cartographie du web ? Sur quels principes repose-t-elle et que cherche-t-elle à montrer ? Il nous semble important de fournir ici quelques points de repère théoriques et méthodologiques sur ce domaine de recherche, en plein essor depuis plus d'une dizaine d'années.

Issue du croisement de plusieurs disciplines et axes de recherche (la scientométrie, l'étude des réseaux, l'étude de l'hypertexte, la théorie des graphes, l'approche de Latour-Callon sur les controverses, etc.), la cartographie du web vise à fournir une visualisation d'un corpus de sites web, en se fondant sur l'élément essentiel de la morphologie du web : les liens hypertextes. Elle repose sur un principe assez simple, comme le rappelle Guillaume Sylvestre : « *La cartographie du web ou des réseaux sociaux s'appuie sur l'idée que les liens créés sur le web entre différents acteurs (sites web, comptes twitter...) peuvent être perçus comme des liens sociaux.* »<sup>2</sup>. Faire un lien d'un site web A vers un site web B n'est pas seulement une opération technique, c'est la trace d'une interaction sociale et documentaire. Or de nombreuses études ont mis en évidence une particularité très intéressante de ce réseau de liens hypertextes, soulignée par Marta Severo : « *L'intérêt de cette technique dérive de deux régularités observées maintes fois dans la pratique de la création de liens hypertextes :*

- *les auteurs de sites web ne citent d'autres sites que s'ils partagent un intérêt thématique ou social ;*

- *les auteurs de sites web ne citent pas les sites qui ont un point de vue opposé au leur, même ceux traitant des mêmes thématiques.* »<sup>3</sup>

Autrement dit, « *Qui se ressemble se connecte* », selon la formule de Franck Ghitalla, l'un des pionniers de la cartographie du web. Le principal objectif de celle-ci est donc de « *visualiser le web de façon synoptique pour formuler des hypothèses sur le positionnement des acteurs* »<sup>4</sup>, comme le souligne Mathieu Jacomy, autre pionnier de cette discipline et co-concepteur des outils Navicrawler, Gephi et Hyphe.

Deux premières notions sont essentielles à comprendre : les domaines et les agrégats. Ghitalla et Jacomy parlent de domaine « *pour désigner l'ensemble des ressources qui parlent d'une même chose, quelle qu'elle soit.* »<sup>5</sup>. Ainsi peut-on parler du « domaine de l'EMI », pour

---

<sup>2</sup> Guillaume Sylvestre, "Utilisation du logiciel Gephi pour l'analyse cartographique", In *Master Intelligence économique et stratégies compétitives*, 2 juin 2015. Disp. sur <http://master-iesc-angers.com/utilisation-du-logiciel-gephi-pour-lanalyse-cartographique/>

<sup>3</sup> Marta Severo, "La cartographie du Web : le lien social sur le Net", mars 2012. Disp. sur : <https://halshs.archives-ouvertes.fr/halshs-00678768/document>

<sup>4</sup> Mathieu Jacomy, "La cartographie du web entre démocratisation et maturité", s.d. , p. 5. Disp. sur : <http://webatlas.fr/tempshare/pratic.pdf>

<sup>5</sup> Mathieu Jacomy, Franck Ghitalla, *Méthodologies d'analyse de corpus en Sciences Humaines à l'aide du Navicrawler*, 2007. p. 4. Disp. sur : <http://webatlas.fr/wp/share/navicrawler/Guide%20m%E9thodo%20NC%202007.pdf>

désigner l'ensemble des sites et des ressources du web qui traitent de l'EMI. Mais un domaine ne forme pas toujours un « agrégat » : « *La théorie des agrégats, issue de travaux statistiques probabilistes sur l'étendue et la structure du web, stipule que les documents qui traitent du même sujet ont une plus forte probabilité d'être connectés (par des liens hypertextes). Les sites d'un même domaine sont souvent connectés, et forment ainsi ce qu'on appelle un « agrégat », centré sur une thématique.* ».<sup>6</sup> Un agrégat recouvre donc un ensemble de sites traitant d'un même sujet, plus ou moins fortement connectés.

Autre élément à comprendre : les travaux sur la morphologie du web « *montrent que le web est loin d'être homogène et continu : il est au contraire fortement structuré et hiérarchisé ; il possède des couches très visibles et bien connectées et d'autres plus profondes et invisibles.* »<sup>7</sup>. Ghitalla parle de « modèle en couches », pour désigner la structure du web, composée de trois grandes couches :

- la « couche haute » est constituée des grands sites génériques, qui dominent le web et sont omniprésents sur de très nombreux sujets (Google, Wikipedia, Amazon, mais aussi des grands sites institutionnels, gouvernementaux, etc.) ;
- la « couche intermédiaire » est la plus intéressante, car précisément constituée de ces agrégats, des communautés en ligne, des réseaux de sites interconnectés ; c'est cette couche du web qu'il s'agit d'explorer pour cartographier un agrégat, par exemple sur l'EMI ;
- enfin la « couche profonde », constituée des bases de données, qu'on appelle aussi le web invisible.

La cartographie du web repose sur deux types d'outils : d'abord des outils de type « crawler », qui vont parcourir le web, l'indexer, le crawler et constituer éventuellement un premier réseau de nœuds et de liens. Le second est un outil de cartographie de l'information, de visualisation, capable de construire des cartographies lisibles. Nous en parlerons plus loin.

Notre outil de collecte et de crawling s'appelle Hyphe<sup>8</sup>. Il s'agit d'un outil développé au Medialab de Sciences Po, dans le cadre du projet « Hypertext Corpus Initiative », par Mathieu Jacomy et son équipe. Issu de l'équipe de chercheurs autour de Bruno Latour, il a été conçu aussi comme un outil de cartographie des controverses. Outil en open source, Hyphe est disponible en deux versions : une version à télécharger et une version de démonstration en ligne, où l'on peut faire un projet de corpus. C'est cette version que nous avons utilisée. Que fait cet outil, qui a pris la suite de l'ancien Navicrawler ? Il permet de constituer un premier corpus de ressources web, qu'il définit comme des « entités web ». Cette notion très intéressante est à la base de Hyphe, qui permet de déterminer soi-même la « longueur » de chaque « entité web » : un site entier, sa page d'accueil, telle ou telle page, telle ressource, etc. Hyphe constitue ainsi des corpus « d'entités web » et non de « sites web » au sens strict.

Voilà donc notre problématique et nos questions de recherche posées : quelle est la réalité du domaine de l'EMI ? Ce domaine constitue-t-il un agrégat ? Et comment est-il structuré ?

---

<sup>6</sup> *Ibid.*, p. 4

<sup>7</sup> Guillaume Ollivier, "Jalons pour une étude (con)textuelle du web", In *Socio-informatique et argumentation*, 16 janvier 2012. Disp. sur : [http://socioargu.hypotheses.org/3427/comment-page-1?lang=fr\\_FR](http://socioargu.hypotheses.org/3427/comment-page-1?lang=fr_FR),

<sup>8</sup> Hyphe est disponible en version de démo ici : <http://hyphe.medialab.sciences-po.fr/demo/#/login>

## **Comment constitue-t-on un corpus sur Hyphe ?**

De deux façons complémentaires : manuelle et semi-automatisée. On peut d'abord charger sur Hyphe un lot d'URL, préalablement sélectionnées : nous avons procédé ainsi à partir des signets de Diigo. A partir de ces URL, il faudra décider pour chacune la longueur de « l'entité web ». Hyphe va ensuite les parcourir, les crawler et recenser tous les liens entrants et sortants. Et il va proposer de nouvelles ressources, liées aux premières. Le corpus initial va ainsi s'enrichir rapidement de tous les sites en lien avec les premiers.

### **Traitement du corpus**

Le corpus ainsi constitué a été traité en trois temps itératifs. En effet, il ne peut s'agir d'étapes rigoureusement successives puisqu'à chaque moment, l'une des actions peut conduire à ajuster le traitement.

Un premier travail de sélection parmi les entités web a conduit à explorer les entités découvertes par Hyphe et marquées « Discovered », signifiant ainsi qu'elles pouvaient être explorées et crawlées par le robot pour que soient repérés, à partir d'elles, les liens sortants vers de possibles nouvelles entités. Cette activité est nécessaire pour ne pas fonder le seul travail d'exploration du réseau sur nos recherches manuelles. Les sites crawlés nous ont permis de faire de nouvelles découvertes d'entités, par exploration du voisinage des sites déjà trouvés.

Le deuxième travail a consisté en un nettoyage des entités identifiées. En effet, il s'est avéré que des pages repérées par le crawling appartenaient au même site. Lorsque les contenus étaient congruents, nous avons considéré qu'il s'agissait d'une seule et même entité.

Progressivement se posait la question de la clôture du corpus. En effet, quels pouvaient être les indicateurs signalant que nous avons identifié tous les acteurs qui étaient à trouver ? Nous nous sommes appuyés ici sur les travaux de Ghitalla sur la cartographie des réseaux.

Au fur et à mesure de l'exploration des sites, on repère les contenus relatifs au domaine qui nous intéresse, ici l'EMI. Ces sites sont déclarés « IN » : ils sont clairement des acteurs du réseau et restent dans les limites intérieures du corpus. En avançant de proche en proche, de site en site, on repère que l'EMI devient un sujet annexe ou que le site concerné n'a pas de proche dans le même domaine. Il semble être à la marge du réseau constitué autour de la thématique de l'EMI. Ainsi une brève de presse qui mentionnerait une action dans un EPLE : on a là l'exemple d'un élément qui traite du contenu mais ne peut être considéré comme un acteur. Il sera alors exclu du corpus. Ce travail de limitation puis de clôture du corpus est délicat pour deux raisons : parce qu'il demande d'explorer (ou de vérifier rapidement) toutes les entités pour repérer leur place et parce que la particularité de la cartographie du web nous conduit à travailler sur les dispositifs socio-techniques qui imbriquent des actions humaines et des procédures automatisées, qui produisent aussi du contenu sur le réseau. Ainsi l'activité de partage automatisé conduit à une duplication possible des contenus sur les plate-formes.

### **Le corpus constitué**

Nous avons ainsi constitué un corpus de 233 entités web et nous avons exclu environ 6500 autres ressources, découvertes par Hyphe.

Avant de passer à la présentation du travail de visualisation de ces données sous forme de cartographie, il convient de donner des précautions méthodologiques nécessaires quant à l'utilisation de ce corpus. Tout d'abord, s'il ne peut prétendre à l'exhaustivité, parce qu'il s'agit d'une exploration en milieu ouvert et dynamique, il peut cependant prétendre à la représentativité en raison de la démarche à trois brins que nous avons adoptée : exploration manuelle, automatisée et itérative. Ensuite, la notion d'entité web est très intéressante dans la perspective d'une cartographie parce qu'elle permet de repérer les nœuds du réseau ; mais déterminer ces entités peut conduire à perdre des informations hétérogènes dont peuvent être porteuses certaines de ces entités. Le dernier point concerne le caractère provisoire du corpus. Nous avons arrêté l'exploration fin août, les publications continuent, les contenus circulent, sont commentés, partagés, etc. Notre corpus constitue un état du réseau traitant de l'EMI, volontairement et arbitrairement figé pour en permettre l'analyse. Conjointement, cet état est lié à un contexte éducatif, social, technique.

### **Description, qualification des données**

Une fois le corpus d'entités web soigneusement délimité, une autre phase s'est ouverte : celle de la qualification des données. Étape très importante, car elle va déterminer directement le type et le nombre de cartographies possibles sous Gephi. En effet, le choix des catégories et des libellés dépend ici complètement de la problématique de recherche, de ce que le chercheur veut mettre en évidence sur une cartographie. C'est un aller-retour constant entre la problématique, la méthode et les résultats. Une cartographie est toujours le produit d'un choix initial.

Voulant montrer la diversité des acteurs et des « actants » impliqués dans la thématique de l'EMI sur le web, deux premières catégories nous ont semblé évidentes : une qualification des types d'acteurs et des types de sources. Les Types d'acteurs se devaient d'emprunter aux notions issues de la sociologie de la traduction et nous avons donc gardé la partition Acteurs humains et Non-humains, en prenant les labels suivants : « Individu », pour les acteurs humains, « Organisation » pour qualifier les « acteurs organisationnels », i.e. tous les types d'organisations (institutionnelles, professionnelles, associatives, etc.), « Texte » pour désigner tous les « textes-réseaux » (articles, textes de loi, textes pédagogiques, etc.), et enfin une catégorie plus incertaine, « Dispositif » pour qualifier toutes les entités socio-techniques (un cours, un portail, un site de curation ou de veille, etc.).

La catégorie des Types de sources posait moins de problèmes, et nous avons fait une distinction classique selon la nature des sources : institutionnelles, scientifiques, associatives, journalistiques...

Nous avons ajouté deux (et même trois) autres catégories : les Types de ressources (selon le type de contenu de la ressource : ressource éducative, politique, scientifique, professionnelle, de veille, etc.) ; et les Domaines d'appartenance, en distinguant quelques thématiques qui nous semblaient pertinentes : EAI, EAM, EMI, Education, Culture numérique, Presse, etc.

Une dernière catégorie a été renseignée et concernait les Types de supports (site web, blog, site de curation, revue en ligne, etc.), mais nous ne l'avons finalement pas utilisée dans les cartographies.

Quand on qualifie un acteur « non-humain », on a des incertitudes ! Il y a eu aussi des difficultés pour qualifier les types de ressources. C'est un travail long et un peu fastidieux,

qui exige de la rigueur, mais qui est facilité par le traitement dans une temporalité réduite, puisqu'une sorte d'automatisme se met en place.

Cette étape primordiale de la qualification a permis d'explorer ce second corpus, et elle a montré notamment l'importance de la place des acteurs non-humains, des organisations et aussi des dispositifs de type 2.0. De nombreuses entités web sont des plateformes collaboratives de blogging, de veille ou de curation. Les résultats ont confirmé en grande partie le travail d'analyse du premier corpus, celui collecté et traité dans Diigo.

Signalons enfin la tâche, absolument cruciale pour tout travail de cartographie, de nettoyage des données, d'homogénéisation des labels, etc. Nous ne nous attarderons pas sur la dimension technique de ces opérations, sauf pour souligner la question très délicate de l'import des données en format csv dans Gephi, tâche qui s'est avérée plus délicate que prévu.

Cette étape de la qualification des données du corpus est décisive à double titre : elle va conditionner fortement le type de cartographies obtenues (concrètement, si nous voulons pouvoir visualiser les types d'acteurs ou de sources d'un domaine, encore faut-il les identifier au préalable) et elle doit être menée également en anticipant les visualisations obtenues. C'est ainsi qu'un premier travail de qualification trop détaillé, avec de nombreuses sous-catégories (par exemple dans les Organisations, en indiquant aussi le type d'organisation), a dû être abandonné, car il induisait une visualisation trop difficile, un manque de lisibilité des cartes.

### **Premières observations**

Ce travail de qualification<sup>9</sup> nous a conduits à faire les premières constatations sur ce corpus d'entités. Pour ce qui est des Types d'acteurs tout d'abord, ils se répartissent comme suit sur les 233 entités :

<b>Types d'acteurs</b>	
Organisation	125
Individu	63
Dispositif	31
Texte	14
<b>TOTAL</b>	<b>233</b>

Il semble que dans cette phase d'émergence de l'EMI, ce sont d'abord les organisations qui se l'approprient, qu'elles soient institutionnelles ou associatives.

Pour ce qui est des ressources, on constate, de manière générale, la prédominance des ressources éducatives et des ressources professionnelles. Cela est en résonance avec l'analyse du corpus Diigo évoqué précédemment, qui laissait apparaître dans l'indexation des ressources collectées la prédominance du tag « ressource professionnelle ».

On remarque par ailleurs la très faible présence de ressources dites politiques, produites sur l'EMI ; ce n'est pas un sujet qui fait débat dans la société. Peut-on aller jusqu'à dire que l'EMI est intrinsèque au monde éducatif ?

---

<sup>9</sup> On trouvera, sur le Carnet du GRCDI, le [fichier xls](#) du corpus.

Enfin, si l'on croise les Types de ressources avec les Types d'acteurs :

Types d'acteurs / Types de ressources	Dispositif	Individu	Organisation	Texte	TOTAL
Ressource culturelle	1		1		<b>2</b>
Ressource de veille	17		1		<b>18</b>
Ressource éducative	2	5	67		<b>74</b>
Ressource journalistique			6		<b>6</b>
Ressource législative			2	2	<b>4</b>
Ressource pédagogique	6	5	11		<b>22</b>
Ressource politique			2		<b>2</b>
Ressource professionnelle	2	42	19	1	<b>64</b>
Ressource scientifique	3	11	16	11	<b>41</b>
<b>Total</b>	<b>31</b>	<b>63</b>	<b>125</b>	<b>14</b>	<b>233</b>

On voit alors :

- que les organisations, acteurs les plus représentés, sont plus diversifiées dans la production de ressources, mais on note la nette prédominance des ressources éducatives ;
- que les individus produisent surtout des ressources professionnelles, à destination des professeurs documentalistes en l'occurrence (c'est l'analyse fine des ressources une à une qui permet de le dire) ;
- que les dispositifs de types 2.0, qui sont essentiellement des curations, des revues de hashtags, des Scoop.it, mais aussi des comptes Twitter, produisent des ressources de veille ou des ressources éducatives, émises par des relais institutionnels de l'EMI (CLEMI, Canopé, sites académiques), mais aussi et plus faiblement par quelques individus. Le poids des organisations se retrouve dans cette analyse.

## Cartographies et résultats

Une fois fini ce long travail de collecte et de traitement du corpus sur Diigo et surtout sur Hyphe, nous avons exporté le fichier obtenu sur Hyphe, sous format GEXF, et nous sommes passés à la troisième et à la quatrième étape de notre démarche : celles de la cartographie et de la visualisation, et surtout de l'interprétation des résultats.<sup>10</sup>

## Rapide présentation de Gephi

Qu'est-ce que Gephi<sup>11</sup> ? Selon sa notice Wikipedia, « *Gephi est un logiciel libre d'analyse et de visualisation de réseaux, développé en Java et basé sur la plateforme NetBeans. Il a été*

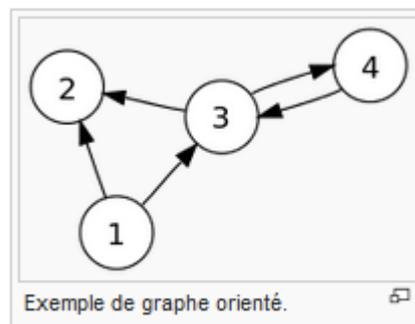
<sup>10</sup> Ce fichier est disponible [à partir du Carnet du GRCDI](#).

<sup>11</sup> Voir le site officiel : <https://gephi.github.io/>

initialement développé par des étudiants de l'Université de technologie de Compiègne (UTC) »<sup>12</sup>. On peut préciser que cet outil a été développé par une équipe, composée notamment de Sébastien Heymann et Mathieu Jacomy, sous l'impulsion de Franck Ghitalla, qu'il a été conçu et développé pour concrétiser les projets de cartographie du web de Ghitalla, et notamment pour compléter le Navicrawler, l'un des premiers outils de collecte et de cartographie du web. Depuis son lancement en 2008, Gephi est devenu l'un des plus puissants outils de cartographie et de visualisation de l'information, reconnu au plan international par de nombreuses communautés scientifiques<sup>13</sup>. Son développement est porté par un Consortium<sup>14</sup> depuis 2010, dans lequel figurent plusieurs grands acteurs de la recherche.

Que permet Gephi ? Il analyse des données, calcule des graphes, établit des cartographies de ces graphes et permet différentes visualisations. Outil très riche en fonctionnalités<sup>15</sup>, il a permis de créer de très nombreuses cartographies du web, de Twitter, de data journalisme, etc. Nous ne pouvons pas présenter ici ses nombreuses fonctionnalités et nous n'insisterons brièvement que sur quelques notions importantes à comprendre.

Tout d'abord, qu'est-ce qu'un graphe ? Selon Wikipedia, « un graphe est un ensemble de points nommés nœuds (parfois sommets ou cellules) reliés par des traits (segments) ou flèches nommées arêtes (ou liens ou arcs). L'ensemble arêtes entre nœuds forme une figure similaire à un réseau. »<sup>16</sup>.



(source : [Wikipedia](#))

Autre notion centrale : la spatialisation. C'est la première étape de visualisation d'un corpus : il s'agit de choisir un algorithme de traitement du graphe. Plusieurs algorithmes existent, qui ne sont pas tous propres à Gephi, et qui permettent de créer des graphes selon des modes de calcul différents. L'algorithme choisi généralement est *Force Atlas2*, qui fonctionne selon un principe d'attraction-répulsion entre les nœuds. C'est celui que nous avons utilisé. Le choix de l'algorithme de spatialisation va donc déterminer « l'architecture » de la cartographie, à la fois sa forme globale et le positionnement des nœuds les uns par rapport aux autres. Pour filer

<sup>12</sup> <https://fr.wikipedia.org/wiki/Gephi>

<sup>13</sup> Il a obtenu plusieurs prix scientifiques.

<sup>14</sup> Le Gephi Consortium, localisé au Medialab de Sciences Po : voir le site

<https://consortium.gephi.org/index.html>

<sup>15</sup> Voir sa fiche descriptive sur PLUME : <https://www.projet-plume.org/fiche/gephi>

<sup>16</sup> [https://fr.wikipedia.org/wiki/Th%C3%A9orie\\_des\\_graphes](https://fr.wikipedia.org/wiki/Th%C3%A9orie_des_graphes)

une métaphore architecturale, la spatialisation serait le plan d'une maison, sa forme globale : elle pourrait être ronde, carrée, rectangulaire, etc.

## Quels résultats ?

Que nous montre Gephi concernant le domaine de l'EMI sur le web ? Précisons tout d'abord que notre méthode globale d'interprétation s'est faite en quatre temps : la lecture des cartographies, leur analyse, des conclusions intermédiaires et une conclusion générale.

### Le paysage de l'EMI

Comment sont structurés les domaines et les agrégats ? Citons de nouveau Ghitalla et Jacomy pour intégrer trois notions de la cartographie du web : « *On distingue trois composantes à un domaine ou agrégat pour se repérer :*

- *Le coeur, qui contient des sites souvent très gros et très fortement connectés ;*
- *La nébuleuse, qui comprend la majeure partie des sites du domaine mais pas les plus connectés ;*
- *Les filaments, qui est constituée de sites souvent petits, qui entrent dans le champ de la thématique mais qui sont peu connectés avec le domaine lui-même. »<sup>17</sup>.*

Quelles observations pouvons-nous tirer de cette première cartographie globale<sup>18</sup>, établie sur le corpus complet, composé de 233 nœuds et 1408 liens ?

D'abord l'EMI est un domaine fortement connecté, et donc un agrégat, avec une moyenne de 6 liens par nœud (autrement dit, chacun des 233 nœuds serait, en théorie, relié à tous les autres par 6 liens). Notre première hypothèse est vérifiée : l'EMI sur le web constitue un agrégat de sites web très interconnectés.

On y distingue nettement les sites génériques, qui ne sont pas propres à l'EMI mais qui concernent le domaine : les sites du Ministère (Education.gouv, Eduscol, Education, etc.). Si l'on reprend les catégories de Ghitalla, le cœur de cet agrégat est constitué des nœuds les plus fortement connectés : les sites institutionnels (CNDP, CLEMI, sites académiques), professionnels (DocspourDocs, Café pédagogique, Trois couronnes, etc.), associatifs (FADBEN), mais aussi quelques sites de curation (Scoop.it) et scientifiques (GRCDI). La « nébuleuse » de l'EMI est représentée ici par la multitude des « petits » nœuds (nombreux blogs, comptes Twitter, sites académiques peu connectés, etc.). Quant aux « filaments », ils sont constitués de tous les nœuds de petite taille, qui entourent le corpus ; il est intéressant d'observer que ces « filaments » sont souvent des sites scientifiques (comptes Twitter de chercheurs, d'équipes scientifiques, de revues, d'événements), individuels (nombreux comptes Twitter), parfois institutionnels, ou des sites de presse.

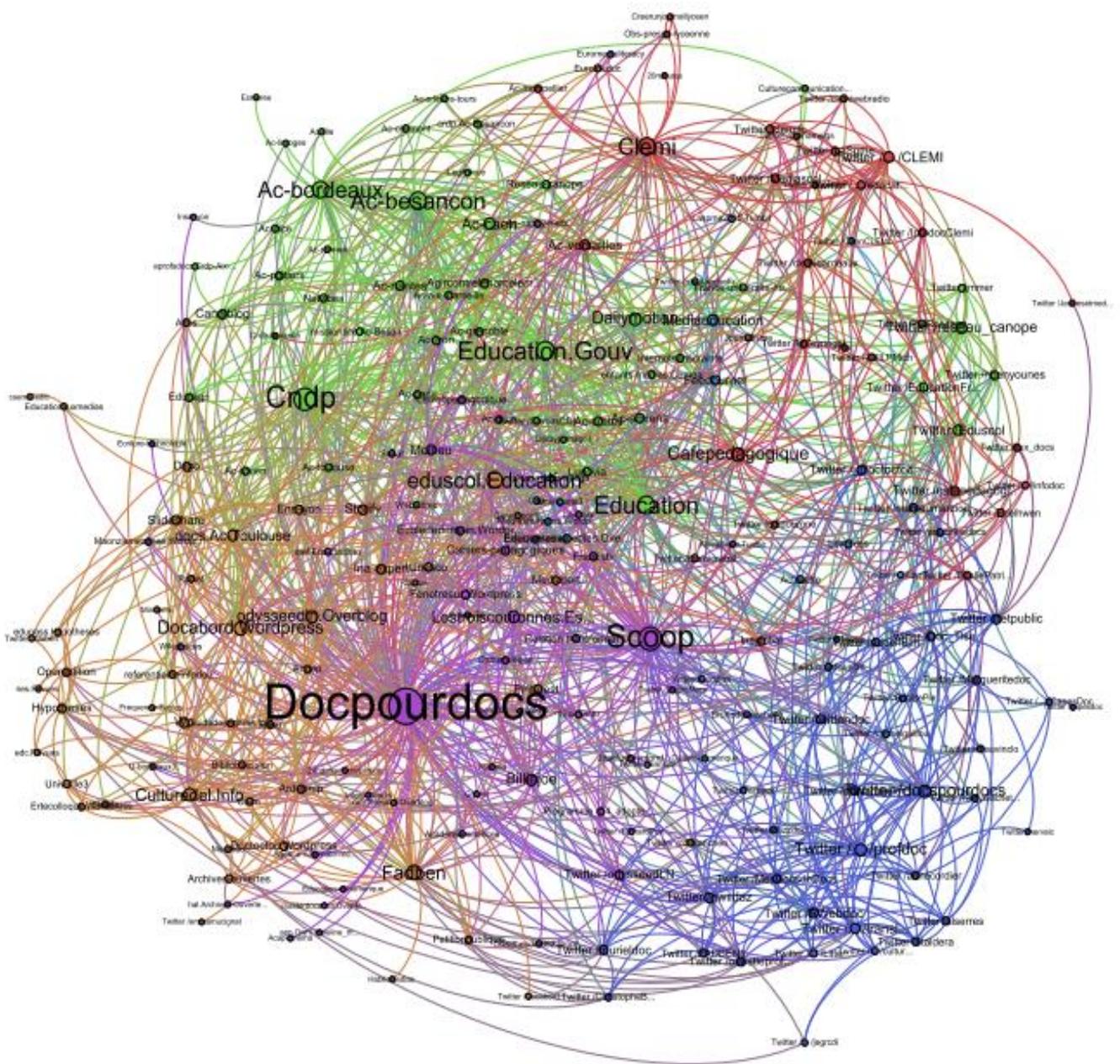
---

<sup>17</sup> Mathieu Jacomy, Franck Ghitalla, *Méthodologies d'analyse de corpus en Sciences Humaines à l'aide du Navicrawler*, 2007. p. 6. Disp. sur :

<http://webatlas.fr/wp/share/navicrawler/Guide%20m%E9thodo%20NC%202007.pdf>

<sup>18</sup> Voir sur le Carnet du GRCDI le fichier PDF, dans lequel il faut zoomer pour agrandir :

[http://grcdi.hypotheses.org/files/2015/10/Carto\\_EMI\\_globale\\_degre\\_2015-09-11.pdf](http://grcdi.hypotheses.org/files/2015/10/Carto_EMI_globale_degre_2015-09-11.pdf)



Carte n° 1, Domaine de l'EMI

### Que signifient les couleurs attribuées aux nœuds ?

Elles expriment une autre notion importante de Gephi, qu'il faut comprendre pour pouvoir « lire » ces cartographies : la notion de partitionnement. Le partitionnement est une répartition *qualitative* des nœuds, selon des attributs de catégorie (par ex. H/F, ou bien ici à partir des champs de description des ressources : Types d'acteurs, etc.). Dans notre métaphore

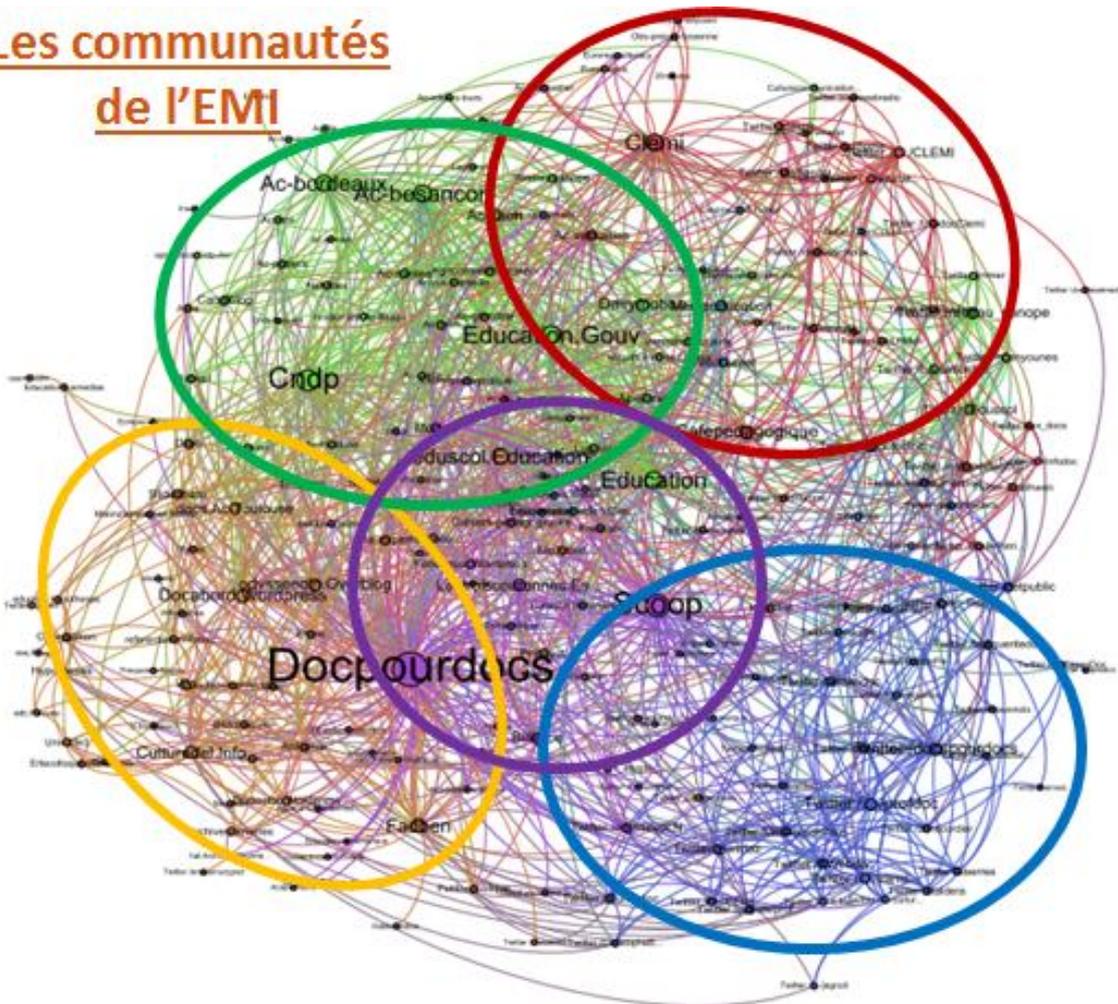
architecturale, si la spatialisation correspondait au plan global d'une maison, le partitionnement serait la répartition des pièces à l'intérieur de la maison, par exemple selon leur fonction principale (chambres, salle de bain, cuisine, etc.), ou selon d'autres critères décidés au préalable (pièces avec ou sans papier peint, etc.). On peut donc demander à Gephi de créer autant de cartographies que de paramètres de partitionnement, avec une attribution de couleurs aux nœuds, selon leur groupe de rattachement. Le premier critère de partitionnement, proposé par Gephi, est le calcul de modularité (*Modularity Class*), qui est basé sur un algorithme de détection de communautés, c.-à.-d. de sites proches. Gephi détecte ainsi, dans l'ensemble du corpus de nœuds, différentes « communautés » (calculées automatiquement) et va attribuer des couleurs à ces communautés.

Que nous montre ici ce partitionnement par communautés (*Modularity class*) ? Il révèle des éléments très intéressants sur la structuration du domaine de l'EMI, la proximité ou l'éloignement de ses composants. On peut y distinguer en effet, plus ou moins nettement, au moins cinq communautés de sites, réunis par la proximité de leurs liens, que nous avons interprétées et qualifiées ainsi :

- une communauté institutionnelle de l'Education Nationale (en vert) : les sites académiques, ministériels ;
- une seconde communauté institutionnelle, dominée par l'Education aux médias (en rouge), avec les sites du CLEMI ou proches du CLEMI, mais aussi avec les sites professionnels éducatifs (Café pédagogique) ;
- une communauté que l'on pourrait qualifier de scientifique (en orange), qui regroupe plusieurs sites et ressources scientifiques ;
- au milieu, en violet, semble apparaître une communauté professionnelle, avec plusieurs sites professionnels (dont DocspourDocs, les Trois couronnes, Billiejoe) ;
- et à droite en bleu, une communauté regroupant surtout des comptes Twitter, que nous avons appelée la « communauté Twitter ».

Notons la présence, en haut à droite de la carte, d'une communauté assez indéterminée, composée de plusieurs couleurs mélangées.

## Les communautés de l'EMI



Carte n° 2, Les communautés de l'E.M.I.

### **Que nous apprend le positionnement des nœuds sur la carte ?**

Tout d'abord, il faut se garder de « sur-interpréter » ces positionnements précis des nœuds et leur distance (éloignement/proximité), car ceux-ci peuvent varier d'une cartographie à l'autre. En revanche, ce qui reste assez constant dans les différentes cartographies, c'est l'environnement immédiat d'un nœud. Par exemple, les nœuds de la FADBN et du GRCDI peuvent être placés à différents endroits du graphe, ils resteront assez proches l'un de l'autre.

On peut faire quelques observations sur ce premier partitionnement :

- le positionnement du site de la FADBN dans la communauté scientifique ;
- le relatif éloignement entre la communauté CLEMI et la communauté scientifique ;
- la spécificité et la richesse de la communauté Twitter ;
- la proximité entre les communautés scientifique et professionnelle, avec des acteurs de poids (dont DocspourDocs) ;

- le recoupement entre les communautés professionnelle et institutionnelle (avec Eduscol et des acteurs de médiation).

### **Pourquoi certains nœuds sont-ils plus gros que d'autres ?**

Une autre notion essentielle de Gephi doit être introduite ici : celle de classement des nœuds. Si le partitionnement est une répartition qualitative des nœuds, en fonction des attributs qui leur sont donnés, le classement est une répartition *quantitative*, selon différents critères quantitatifs ; il se traduit par l'attribution d'une taille aux nœuds. Pour continuer notre métaphore de la maison, le classement consisterait à classer les pièces selon leur dimension, ou selon leur hauteur, etc.

Ce qu'il faut noter, c'est que Gephi propose de nombreux critères de classement. Ici, sur cette cartographie globale, le classement des nœuds choisi est le classement par *Degrés*, c.-à.-d. par le total des liens entrants et sortants.

Ainsi, un zoom sur le site de la FADBEN<sup>19</sup> montre bien les liens sortants (qui pointent vers d'autres sites) et les liens entrants (qui pointent vers la FADBEN, indiqués par une flèche). Gephi a recensé ici un total de 42 liens, exprimé par le critère *Degrés* : 20 liens entrants et 22 liens sortants. Précisons que ce comptage ne porte pas sur l'ensemble du site de la FADBEN, mais seulement sur les articles sélectionnés pour notre corpus (i.e. les articles traitant de l'EMI).

### **Quels sont les sites les plus cités ?**

On connaît l'importance de la citation, que ce soit dans le monde des publications scientifiques, depuis les travaux d'Eugen Garfield en 1958 sur la bibliométrie, ou dans le monde du web, avec le célèbre Page Rank de Google (d'ailleurs inspiré de Garfield). « Dis-moi qui tu cites, je te dirai qui tu es et dans quel réseau tu t'inscris »... Mais la cartographie des citations, ou plutôt ici des liens hypertextes, est aussi intéressante pour mesurer, sinon l'autorité, du moins la notoriété d'un site.

Nous avons donc fait cette cartographie<sup>20</sup>, en gardant toujours le partitionnement par *Modularity Class* (par communautés), et un classement des nœuds par *Degrés entrants* (i.e. par liens entrants), mais avec un filtrage pour faciliter la lecture, en ne gardant que les sites recevant au moins 10 liens entrants.

Cette cartographie montre donc un peu plus d'une quarantaine de ressources (d'entités web), sur un total de 233 nœuds, qui sont les plus citées dans le petit monde de l'EMI. Dans les dix premiers nœuds, on trouve, en tête, les grands sites institutionnels, Education (Eduscol) et Education.gouv, qui totalisent à eux deux 112 liens entrants ! Suivent en troisième position plusieurs sites Scoop.it, regroupés dans la même entité web, dont le topic du CLEMI, Infodoc-Press (ces Scoop.it font un total de 39 liens). Viennent ensuite, dans ces 10 premiers, le CNDP, le Café pédagogique, Daily Motion (qui regroupe plusieurs vidéos sur

<sup>19</sup> Voir la diapo 21 sur le [diaporama](#) de la communication.

<sup>20</sup> Voir sur le Carnet du GRCDI le fichier PDF : [http://grcdi.hypotheses.org/files/2015/10/Carto\\_EMI\\_liens-entrants\\_2015-09-02.pdf](http://grcdi.hypotheses.org/files/2015/10/Carto_EMI_liens-entrants_2015-09-02.pdf)

l'EMI), le CLEMI, DocspourDocs, les sites académiques de Versailles et Besançon. La FADBEN arrive en onzième position (avec 20 liens), et le GRCDI en quatorzième (17 liens).

### **Quels sont les types d'acteurs les plus cités ?**

Une autre cartographie<sup>21</sup> est intéressante pour montrer les types d'acteurs les plus cités. Nous avons donc fait un autre partitionnement des nœuds, sur l'attribut « Type d'acteurs », avec le même classement par « Degrés entrants » et le même filtrage (sur 10 liens entrants).

Que peut-on y voir ? Bien sûr, le même corpus que sur la cartographie précédente, puisque les sites sont les mêmes, mais leur couleur donne des indications sur les types d'acteurs auxquels ils appartiennent.

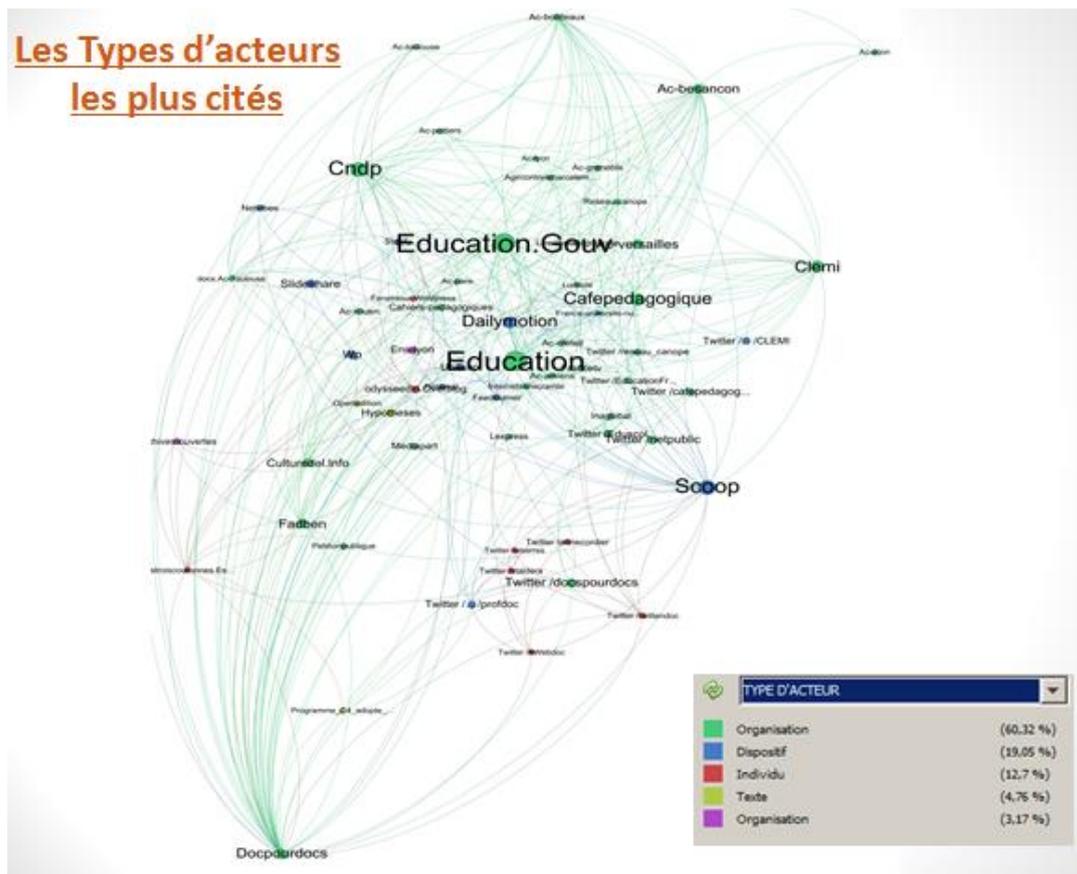
Ainsi la légende de Gephi indique :

- une nette prédominance des acteurs « organisationnels », regroupant aussi bien les sites institutionnels, associatifs, les collectifs professionnels ou scientifiques : ils constituent plus de 60 % des nœuds affichés ;
- l'importance des acteurs socio-techniques, c.-à.-d. les « Dispositifs » : sites de veille, comptes Twitter collectifs, sites de partage (SlideShare, DailyMotion), qui représentent près de 20 % des Acteurs affichés ;
- la part réduite des « Acteurs humains », c.-à.-d. des Individus : un peu plus de 12 % ;
- et la faiblesse des « Textes » : moins de 5 %.

---

<sup>21</sup> Voir sur le Carnet du GRCDI le fichier PDF :

[http://grcdi.hypotheses.org/files/2015/10/Carto\\_EMI\\_TypesActeurs\\_Liens-entrants\\_2015-09-02.pdf](http://grcdi.hypotheses.org/files/2015/10/Carto_EMI_TypesActeurs_Liens-entrants_2015-09-02.pdf)



Carte n° 3, Les types d'acteurs les plus cités

### Quels sont les sites les plus « populaires », selon un calcul de type PageRank ?

Dans cette cartographie<sup>22</sup>, nous avons utilisé un autre critère de classement, le PageRank. Selon l'explication fournie par Gephi, le PageRank permet de classer les sites en fonction de la probabilité d'arriver dessus en suivant les liens. Il s'agit donc d'une sorte d'indicateur de popularité, qui vient compléter le classement par citations.

La cartographie produite<sup>23</sup> affiche seulement 34 nœuds, et leur classement est sensiblement différent de celui du classement par Liens entrants. Si l'on retrouve plus ou moins les mêmes sites, le classement des dix premiers est différent : Education.gouv, DailyMotion, Eduscol, FeedBurner, CLEMI, Scoop.it, CNDP, France-Université numérique, Internetsanscrainte et le compte Twitter du Ministère de l'Education

<sup>22</sup> Paramètres techniques de la cartographie des sites populaires (PageRank) : partition selon Modularity class, classement par "PageRank", filtrage par Attributs, Plage, PageRank, de 0.0063.... à 0.07..... (34 sites affichés)

<sup>23</sup> Voir sur le Carnet du GRCDI le fichier PDF : [http://grcdi.hypotheses.org/files/2015/10/CartoEMI\\_influents-PageRank\\_2015-09-02.pdf](http://grcdi.hypotheses.org/files/2015/10/CartoEMI_influents-PageRank_2015-09-02.pdf)

On voit, dans ce classement, l'importance des dispositifs socio-techniques avec la place de DailyMotion, de Scoop.it, l'apparition de FeedBurner, de Netvibes ; un renforcement de la position du CLEMI et des sites institutionnels. On peut également voir une différence pour DocspourDocs entre le nombre de liens entrants et le PageRank (la popularité).

### **Quels sont les sites « médiateurs », au coeur des flux d'informations ?**

Cette cartographie<sup>24</sup> repose sur un autre critère de classement, fourni par Gephi, la « *Betweenness Centrality* », définie ainsi : « *Le calcul de l'intermédiarité permet d'afficher les nœuds qui sont au cœur des flux d'informations. Ils sont indispensables pour diffuser l'information au sein du réseau.* ». Il s'agit donc ici des sites-relais, qui jouent un rôle important de médiation, de rediffusion de l'information au sein du domaine.

Que montre cette cartographie<sup>25</sup> ? Un nouveau classement, avec en tête DocspourDocs, qui confirme son importance de « hub » et de site-relais, suivi de Scoop.it, du CNDP, du CLEMI, de Education.gouv. La FADBEN est en huitième position.

### **Quels sont les (sous-)domaines de l'EMI les plus cités ?**

Dans cette cartographie<sup>26</sup>, nous avons cherché à visualiser les différents sous-domaines, qui composent le corpus de l'EMI, en fonction de nos critères de qualification.

Que nous dit cette cartographie et sa légende<sup>27</sup> ?

- la partie « EAI », i.e. l'info-doc est très nettement majoritaire, et représente plus de 35% du corpus affiché ;
- les sites d'éducation, au sens large (non spécifiques à l'EMI, ou à l'EAI) arrivent en deuxième position, avec près de 22 % ;
- les sites d'EAM, explicitement orientés vers l'éducation aux médias, sont troisièmes (avec 17,6 %) ;
- et les sites spécifiquement EMI n'arrivent qu'en 4<sup>ème</sup> position, avec plus de 11 %.

A noter enfin la part très faible des sites spécifiquement dédiés à la culture numérique dans ce corpus. Ce qui tendrait à établir une assez nette coupure entre les deux domaines, de l'EMI et de la formation au numérique, contrairement aux souhaits de la Conférence EMIconf.

### **Quels types de sources les plus citées ?**

Cette nouvelle cartographie<sup>28</sup> porte sur les types de sources, établies à partir de nos qualifications : sources institutionnelles, scientifiques, professionnelles, etc. Précisons qu'il

---

<sup>24</sup> Paramètres techniques de la cartographie des sites médiateurs : partition selon Modularity class, classement par «Betweenness Centrality», filtrage par plages de Degrés, de 30 à 146

<sup>25</sup> Voir sur le Carnet du GRCDI le fichier PDF :

[http://grcdi.hypotheses.org/files/2015/10/Carto\\_EMI\\_intermediaires\\_2015-09-02.pdf](http://grcdi.hypotheses.org/files/2015/10/Carto_EMI_intermediaires_2015-09-02.pdf)

<sup>26</sup> Paramètres techniques de la cartographie des domaines les plus cités : partition selon les Domaines, classement par «Degré entrant», filtrage par Topologie, Plage «Degrés entrants» de 8 à 57

<sup>27</sup> Voir sur le Carnet du GRCDI le fichier PDF

[http://grcdi.hypotheses.org/files/2015/10/Carto\\_EMI\\_Domaines\\_Liens-entrants\\_2015-09-02.pdf](http://grcdi.hypotheses.org/files/2015/10/Carto_EMI_Domaines_Liens-entrants_2015-09-02.pdf)

s'agit bien ici des sources, à l'origine de l'information, et non du type de ressource. Cette cartographie<sup>29</sup> souligne une fois encore l'importance des institutions, mais aussi, et c'est un indicateur intéressant, la place importante des sources personnelles, qui arrivent en deuxième position. Les sources associatives et professionnelles arrivent bien après, et les sources scientifiques sont assez faibles (avec 6,5 %).

Ceci induit une conclusion intermédiaire, que nous avons observée dans les précédentes cartographies : l'EMI est d'abord produite par l'institution.

### **Quels types de ressources les plus citées ?**

Enfin, lorsque l'on observe les types de ressources les plus importantes, la dernière cartographie<sup>30</sup> vient à la fois confirmer et compléter les analyses précédentes<sup>31</sup>.

On perçoit d'abord nettement la force des ressources éducatives (près de 32%) et professionnelles (27,5%) : elles représentent près de 60 % du corpus des sites les plus cités. Ce qui signifierait que la préoccupation première est à la fois dans les dispositifs, le discours éducatif, l'appropriation et la mise en œuvre, dans une forme d'urgence, par la profession. Il faut souligner ensuite la relative importance des ressources scientifiques (en 3<sup>ème</sup> position avec 17,6 %), ce qui peut paraître contradictoire avec la cartographie précédente sur les sources. Enfin la faiblesse des ressources politiques, sociales, journalistiques : est-ce que cela signifierait que la question des enjeux de l'EMI n'est pas encore vraiment posée ?

### **Quelles conclusions principales peut-on tirer de l'interprétation de ces cartographies ?**

Sur la base de toutes les observations faites précédemment, à partir de ces cartographies qui sont autant de facettes du domaine de l'EMI sur le web, nous pouvons tirer une dizaine de conclusions synthétiques :

- l'EMI est un domaine très interconnecté, i.e. un agrégat ;
- on peut y repérer au moins cinq « communautés » de sites, en fonction de liens de proximité ;
- on peut dresser une sorte de typologie, certes provisoire, des acteurs selon leur positionnement : les acteurs « faisant autorité » (les plus cités), les acteurs « populaires » (ou fréquentés), et les médiateurs d'information ;
- nous avons repéré à maintes reprises la domination très nette des acteurs organisationnels et la force des dispositifs socio-techniques : l'EMI est pour le moment affaire d'organisations, d'institutions, de collectifs, plutôt que d'individus, et elle repose aussi sur les dispositifs de type web 2, ce qui accélère sa diffusion ;

---

<sup>28</sup> Paramètres techniques de la cartographie des Types de sources les plus cités : partition : Types de Sources, classement par "Degré entrant", filtrage par Topologie, Plage "Degrés entrants" de 8 à 57

<sup>29</sup> Voir fichier PDF : [http://grcdi.hypotheses.org/files/2015/10/Carto\\_EMI\\_TypesSources\\_Liens-entrants\\_2015-09-02.pdf](http://grcdi.hypotheses.org/files/2015/10/Carto_EMI_TypesSources_Liens-entrants_2015-09-02.pdf)

<sup>30</sup> Paramètres techniques de la cartographie des Types de Ressources les plus cités : partition : Types de ressources, classement par "Degré entrant", filtrage par Topologie, Plage "Degrés entrants" de 8 à 57

<sup>31</sup> Voir fichier PDF : [http://grcdi.hypotheses.org/files/2015/10/Carto\\_EMI\\_TypesRessources\\_Liens-entrants\\_2015-09-02.pdf](http://grcdi.hypotheses.org/files/2015/10/Carto_EMI_TypesRessources_Liens-entrants_2015-09-02.pdf)

- pour approfondir le point précédent, dans les organisations, on soulignera la force des sources institutionnelles : l'EMI est portée par un discours institutionnel ;
- on aura noté à la fois le paradoxe entre la faible visibilité des sources scientifiques et la relative importance des ressources scientifiques : peu de sources, mais qui produisent beaucoup ;
- autre phénomène notable : l'importance des sites de curation et de veille (Scoop.it, Twitter, Netvibes...) ;
- la très faible présence de la presse et des ressources politiques : selon notre corpus, l'EMI reste encore confinée aux cercles relativement étroits de l'école et des professeurs documentalistes, sans résonance réelle dans la société civile ;
- la présence anecdotique de la culture numérique : la séparation semble ici assez nette entre les domaines de la culture numérique, des TICE, etc. (qui doivent représenter certainement un très grand corpus) et ceux de l'EMI, EAI et EAM.

Une nouvelle hypothèse de recherche émerge à partir de ces premières observations : la nécessaire vérification de ces intuitions et de ces résultats en termes de places, de positions et de stratégies d'influence des acteurs, que la cartographie ne peut pas montrer.

## En conclusion

Pour conclure, nous souhaitons insister sur plusieurs points.

Tout d'abord, sur le choix de cartographier un domaine du web : c'est une méthode de recherche qui procède d'une démarche itérative et revêt une dimension heuristique.

Jean-Christophe Plantin, conduisant une comparaison entre la cartographie géographique et la cartographie du web, en décrit les intérêts avec trois verbes d'action : représenter, naviguer et analyser.

Représenter tout d'abord : il s'agit de représenter un territoire inconnu au sein du web, une mise en visibilité pour se repérer au sein dudit territoire : « *La carte constitue toujours une simplification nécessaire du phénomène étudié : y faire figurer l'ensemble d'un phénomène concourt rapidement à rendre la carte illisible. À l'inverse, simplifier la carte en sélectionnant les éléments et les variables à afficher permet de faire ressortir plus facilement les insights et suscite l'analyse. Si la carte se construit ainsi sur d'inévitables « petits mensonges » (Monmonnier 1993), tout le travail du cartographe consiste à encadrer ceux-ci par une méthode permettant d'éviter de mentir par omission ou approximation.* »<sup>32</sup>.

Naviguer ensuite, il s'agit alors de circuler dans cet espace en se constituant un réseau de voies de communication pour circuler et se repérer dans ce territoire ainsi circonscrit. On peut varier la focale d'analyse en débutant par une vue globale, pour l'identification de tendances générales ; puis faire des zoom, des focus sur une partie seulement selon les zones que l'on

---

<sup>32</sup> Jean-Christophe Plantin, « D'une carte à l'autre : le potentiel heuristique de la comparaison entre graphe du web et carte géographique », in Christine Barats (dir.), *Analyser le web en Sciences humaines et sociales*, Armand Colin, 2012. Chapitre 11.

veut explorer dans l'objet d'étude. Par exemple, les liens entre institutions et acteurs professionnels. Pour un autre niveau d'analyse, il faudrait entrer plus dans le détail.

Ce qui permet aussi de se situer dans l'espace ainsi circonscrit. Et Plantin de citer Ghitalla, « où suis-je ? » signifie par sous-entendu « et moi dans cet espace ? ». Par exemple, quelle est la place du GRCDI dans l'ensemble du champ de l'EMI ? Ce travail nous a permis de nous décentrer, de mettre à distance nos travaux.

Analyser enfin.

Jean-Christophe Plantin insiste sur un point dans sa comparaison entre carte géographique et carte du web : « *l'important n'est pas la position absolue d'une URL dans l'espace (en haut à gauche ou en bas à droite) mais sa position relative aux autres URL* »<sup>33</sup>. Peu importe où se trouve la FADBEN : ce qui compte, ce sont ses voisins sur la carte...

### **Quelles limites ?**

La principale limite est la part d'arbitraire, comme dans toute entreprise cartographique, dans toute tentative de représentation. Même si nous avons veillé à une neutralité maximale, méthodologique, nous avons dû procéder à des choix, entre le premier et le second corpus, pour l'algorithme de spatialisation, les critères de classement, etc.

### **Et maintenant ?**

Quels pourraient être les prolongements à envisager pour ce travail de recherche, à peine commencé ? Deux voies semblent s'ouvrir :

- d'une part, le renouvellement du corpus dans quelques mois : on pourrait étendre le corpus et poursuivre la fouille du web. D'autres textes ont été publiés depuis la rentrée, notamment à l'occasion de la publication de fiches guides institutionnelles sur le Parcours citoyen dans lequel l'EMI est inscrit. Celui de Françoise Chapron<sup>34</sup>, en prologue à ce Congrès : elle évoque l'EMI et sa place dans la pédagogie de l'information-documentation. D'autres vont suivre : un positionnement sur l'EMI dans InterCDI par Anne Cordier, des articles dans les actes du colloque international de « Former au monde de demain » par J. Kerneis et Stalder/Delamotte/Cordier, etc. Ainsi, les positions au sein du réseau des textes institutionnels et scientifiques seront-elles probablement modifiées. Redisons que cette cartographie de l'EMI est une photo à un instant T.

- d'autre part, et parallèlement, il faudrait conduire une analyse qualitative de ces textes. C'était l'une de nos intentions au départ, mais qui, faute de temps, n'a pu être réalisée, sauf de manière très empirique et provisoire. Nous sommes dans une phase de construction de l'EMI et donc dans des stratégies de positionnement des acteurs, qui développent des arguments, pour imposer ou développer l'EMI (les institutions), ou pour renforcer leur légitimité (les professionnels).

Certains des acteurs s'expriment pour se positionner comme référent, pour justifier l'existence de l'EMI (les sites institutionnels, notamment, pour l'imposer, mais aussi les sites professionnels pour ancrer leurs actions auprès des élèves dans une forme de prescription).

---

<sup>33</sup> Franck Ghitalla, « La « Toile Européenne » Parcours autour d'une cartographie thématique de documents web consacrés au thème de l'Europe et à ses acteurs sur le web francophone ». In *Communication & langages*, 2008, p. 66.

<sup>34</sup> Disp. sur <http://congres2015.FADBEN.asso.fr/De-la-pedagogie-du-document-au.html>

Une grille d'analyse a été constituée à partir de travaux de Patrick Charaudeau pour construire une *théorie de l'influence*. Il propose de s'interroger sur la stratégie de l'énonciateur pour persuader son interlocuteur de son discours : « *Comment entrer en contact avec l'autre, quelle position d'autorité adopter vis-à-vis de l'autre, comment toucher l'autre, et, conséquemment, comment ordonnancer son dire ?* »<sup>35</sup>

D'après les premiers résultats de cette étude qualitative, on voit se dessiner des éléments de langage chez certains acteurs, notamment les professionnels qui développent deux types d'arguments :

- une légitimation de l'intervention pédagogique des professeurs documentalistes, puisque l'EMI mobilise des concepts ancrés dans l'information-documentation ;

- une légitimation du fait d'une implication depuis le début de l'existence de la profession dans l'éducation aux médias et l'investissement dans des dispositifs du type « éducation à ».

Systématiquement dans les textes analysés, on voit un cadre de référence « scientifique » : de nombreuses références scientifiques sont mentionnées, pas seulement françaises, mais aussi belges. Ce recours est un élément qui appuie l'autorité des propos de l'acteur qui s'insère dans un réseau qui fait lui-même autorité.

### **Quelques derniers questionnements sur la cartographie :**

- quelle représentation de l'EMI apparaît à la lecture de ces cartographies ? Quel rôle des acteurs dans la constitution du domaine, avec la requalification de ressources anciennes, la réorganisation de contenus « Infodoc », ré-indexés « EMI » ?

- quelle signification donner à la place des dispositifs socio-techniques, des plateformes 2.0, dans l'EMI ?

- les absents : pourquoi certains sites professionnels, inscrits en Infodoc, apparaissent peu dans les cartographies ?

### **En guise de conclusion conclusive et finale**

A la question « A qui appartient l'EMI ? », nous pouvons donc apporter une première réponse en revenant sur le titre de cette communication : « L'EMI en partage ». La variété des acteurs montre en effet que c'est l'affaire de beaucoup d'acteurs, aux motivations différentes. La place dans le réseau cartographié, ainsi que le maillage des liens, montrent :

- que c'est d'abord une volonté politique forte portée par l'institution, qui s'appuie sur des relais médiateurs également institutionnels (Canopé, CLEMI, sites académiques) ;

- que c'est un territoire en construction et en bornage dans la mesure où chaque acteur sélectionne et met en avant volontairement ou non des dimensions de l'EMI. La visibilité ainsi opérée contribue à le construire comme un acteur. Pour situer son engagement sur le terrain de l'EMI, il faut dire ce que l'on prend de l'EMI et ce que l'on en refuse, là est le bornage. Les mobiles de ces processus ne sont pas toujours explicites pour les acteurs eux-mêmes. Ainsi, se situer comme acteur de l'EMI en se contentant de ré-indexer ses contenus, ses titres de page web, est un moyen d'être actif, visible rapidement sur le web mais manifeste une vision sans doute à courte vue de ce qu'est une réflexion collective. Les enjeux éducatifs

---

<sup>35</sup> Patrick Charaudeau, « L'argumentation dans une problématique d'influence », *Argumentation et Analyse du Discours* [En ligne], 1, 2008, mis en ligne le 02 octobre 2008, : <http://aad.revues.org/193>

sont pourtant à cette échelle. Par ailleurs, la poursuite d'une légitimation professionnelle, notamment par une blogosphère agissante qui participe à la diffusion de l'EMI, pose de manière accrue l'absence de prescription claire pour les professeurs documentalistes.

## Références bibliographiques

Bruillard, Eric, Durpaire, Jean-Louis (sous la coord.). *Cultures numériques, éducation aux médias et à l'information. Ecole Normale Supérieure de Lyon, 21-22 mai 2013*. Futuroscope : Scéren-CNDP, 2013. 143 p.

Charaudeau, Patrick. « L'argumentation dans une problématique d'influence », In *Argumentation et Analyse du Discours* [En ligne], 1, 2008, mis en ligne le 02 octobre 2008, : <http://aad.revues.org/193>

Ghitalla, Franck. « La « Toile Européenne » Parcours autour d'une cartographie thématique de documents web consacrés au thème de l'Europe et à ses acteurs sur le web francophone ». In *Communication & langages*, 2008, pp 61-75. Disp. sur : <http://www.necplus.eu/action/displayAbstract?fromPage=online&aid=2423488>

Jacomy, Mathieu, Ghitalla, Franck. *Méthodologies d'analyse de corpus en Sciences Humaines à l'aide du Navicrawler. Rapport final* (Rapport de recherche). Paris: Fondation de la Maison des Sciences de l'Homme, 2007. Disp. sur : <http://webatlas.fr/wp/share/navicrawler/Guide%20m%e9thodo%20NC%202007.pdf>

Jacomy, Mathieu. « *La cartographie du web entre démocratisation et maturité* », s.d. Présentation PDF. Disp. sur : <http://webatlas.fr/tempshare/pratic.pdf>

Jeanneret, Yves. « Informatique literacy : manifestations, captations et déceptions dans le texte informatisé », *SPIRALE - Revue de Recherches en Éducation*, 2001, n° 28

Ollivier, Guillaume. « Jalons pour une étude (con)textuelle du web », In *Socio-informatique et argumentation*, 16 janvier 2012. Disp. sur : [http://socioargu.hypotheses.org/3427/comment-page-1?lang=fr\\_FR](http://socioargu.hypotheses.org/3427/comment-page-1?lang=fr_FR)

Plantin, Jean-Christophe. « D'une carte à l'autre : le potentiel heuristique de la comparaison entre graphe du web et carte géographique ». In Barats, Christine (dir.). *Analyser le web en Sciences humaines et sociales*. Paris : Armand Colin, 2012. Chapitre 11. Version preprint disp. sur : <http://cartonomics.org/wp-content/uploads/2011/09/PLANTIN-2012-dune-carte-%C3%A0-lautre-1.pdf>

Severo, Marta. « La cartographie du Web : le lien social sur le Net » . mars 2012. Disp. sur : <http://halshs.archives-ouvertes.fr/halshs-00678768/document>