



HAL
open science

Analyser la structure de dispositifs numériques de mise à disposition de données géographiques sur le Web

Johan Desbonnet, Françoise Gourmelon, Christophe Claramunt

► To cite this version:

Johan Desbonnet, Françoise Gourmelon, Christophe Claramunt. Analyser la structure de dispositifs numériques de mise à disposition de données géographiques sur le Web. SAGEO 2018, 2018, Montpellier, France. halshs-02320468

HAL Id: halshs-02320468

<https://shs.hal.science/halshs-02320468v1>

Submitted on 18 Oct 2019

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.

Analyser la structure de dispositifs numériques de mise à disposition de données géographiques sur le Web

Johan Desbonnet¹, Françoise Gourmelon¹, Christophe Claramunt²

1. LETG-Brest (UMR 6554 CNRS), Université Bretagne Occidentale
IUEM, Rue Dumont d'Urville, 29280 Plouzané cedex
johan.desbonnet@univ-brest.fr ; francoise.gourmelon@univ-brest.fr
2. Institut de Recherche de l'Ecole Navale (IRENaV)
BCRM Brest, CC 600, Ecole Navale, 29240 Brest, cedex 9
christophe.claramunt@ecole-navale.fr

RESUME. Dans le cadre du suivi et de la gestion du trait de côte, le besoin d'informations et de connaissances accessibles sur le Web est actuellement décrit comme un enjeu par les pouvoirs publics français. Néanmoins sur cette thématique, la stratégie des promoteurs des Infrastructures de Données Géographiques (IDG) et des plateformes de données ouvertes de même que leurs contenus restent encore peu connus. Cet article décrit une approche méthodologique pour explorer les volets structurels de l'accès à l'information géographique numérique à partir de l'analyse des liens hypertextuels et des services de consultation de dispositifs numériques susceptibles d'être mobilisés dans le suivi et la gestion du trait de côte. L'analyse des composants structurels des dispositifs sur la toile tels que le volet hypertextuel et les services Web révèle un cheminement parfois complexe et peu diversifié pour accéder à une information géographique pertinente sur cette thématique.

ABSTRACT. In the context of shoreline management, inferring information and knowledge from the Web are considered by French public authorities as a crucial issue. However, Spatial Data Infrastructures promoters' strategies and content are still unknown. The preliminary research presented in this paper provides a methodological approach that explores the availability of geographical information as exhibited by the hypertext and Web services structure of geographical digital infrastructures. Analysis of hypertext and web services shows sometimes complex path to access the geographical information on digital platforms studied.

MOTS-CLES : Information géographique numérique, Infrastructures de Données Géographiques (IDG), plateformes de données ouvertes, structure hypertextuelle, services Web, mise en réseau, gestion et suivi du trait de côte

KEYWORDS: Digital geographical information, Spatial Data Infrastructures (SDI), Open Data platforms, hypertext structure, Web services, network, shoreline management

1. Introduction

Face aux enjeux présents dans les espaces bordant le trait de côte (risques, érosion de la biodiversité, aménagement du territoire), la mise en place et le renouvellement de projets en faveur de la production et de l'accès à une information géographique numérique (Litto 3D, indicateur national d'érosion, ortholittorale, observatoires, etc.) sur des dispositifs en ligne sont érigés en axes prioritaires par les pouvoirs publics depuis le début des années 2000 (Bersani et al., 2006 ; Georis-Creuseveau, 2014). Ces dispositifs, qui regroupent dans notre étude les Infrastructures de Données Géographiques (IDG)¹ et les plateformes de données ouvertes² ont permis la mise en ligne de l'information géographique sur le Web. Ils sont le résultat d'un investissement progressif des pouvoirs publics, des sociétés privées et des organisations non gouvernementales dans le domaine de l'accès à l'information à l'échelle mondiale, européenne (convention d'Aarhus, directive Inspire) et nationale (directives Cada et PSI, Loi Notre, Loi pour une république numérique dite aussi loi Lemaire).

Ces dispositifs, de plus en plus nombreux sur le Web 2.0, ont pris au cours du temps, de leurs évolutions techniques, et des objectifs fixés par leurs promoteurs (services de l'État, établissements publics) des dénominations variées : bases de données géographiques, API cartographiques³, Géoportails... Ils sont souvent décrits comme des technologies visant à faciliter l'accès et les usages de l'information dans un cadre d'action publique (suivi et gestion des zones côtières, urbanisme, risques, etc.) et peuvent être développés pour et par des acteurs spécifiques (Pornon, 2015 ; Georis-Creuseveau et al., 2016). Les travaux portant sur ces infrastructures de données géographiques numériques ont montré que de tels dispositifs pouvaient être influencés par différents facteurs techniques (normes, standards), politiques, organisationnels, économiques et sociétaux (Rajabifard, 2002 ; Noucher, 2013 ; Georis-Creuseveau, 2014). L'Association Française pour l'Information Géographique (AFIGEO) recensait 65 dispositifs (IDG) en 2014⁴ tandis que le programme de recherche Geobs⁵, qui s'intéresse aux dynamiques de circulation de

¹ Les Infrastructures de Données Géographiques, régies par des politiques et des normes, ont l'objectif de favoriser la mise en réseau des producteurs et des utilisateurs de l'information géographique (Rajabifard, 2002 ; Georis-Creuseveau, 2014 ; Noucher et al., 2017).

² Les plateformes de données ouvertes étudiées sont le résultat d'un investissement des gouvernements et agences gouvernementales dans l'ouverture de leurs données publiques. Ce sont des dispositifs qui s'appuient le plus souvent sur des algorithmes, des méthodes au service d'une politique et qui mettent à disposition ou renvoient vers des données de producteurs sous une licence libre pour permettre le développement d'autres applications (Colin et Verdier, 2015).

³ Une API est une interface de programmation pouvant être considérée comme une application technique support de fonctionnalités spécifiques (cartographiques, statistiques et graphiques) et qui prend le plus souvent la forme de service web (Mericksay, 2011).

⁴<http://www.afigeo.asso.fr/documentation/publications.html?download=718:catalogue-des-idg-03-juin-2014>

⁵ <http://geobs.cnrs.fr/>, <http://www.geobs.cnrs.fr/manifesto.pdf>

l'information géographique au sein de ces dispositifs, en recensait 45 aux échelons national et régional (Noucher et al., 2017). En parallèle et en lien avec la mouvance « open data », des plateformes de données ouvertes telles que data.gouv et geodatagouv⁶ ont été mises en place par le gouvernement français. Enfin, dans le cadre spécifique français de la stratégie nationale de gestion intégrée du trait de côte, un certain nombre d'infrastructures de données géographiques qualifiées « d'observatoires et de systèmes d'information sur le littoral (SILs) » (Suanez et al., 2012), ont été identifiées par les pouvoirs publics (en gras et soulignés dans le Tableau 1).

Ces différentes définitions et recensements sur les IDG et plateformes de données ouvertes françaises nous ont conduits à sélectionner un corpus de 43 dispositifs dont certains sont déployés spécifiquement dans le cadre de la thématique du suivi et de la gestion du trait de côte (Tableau 1).

TABLEAU 1. Corpus étudié (n=43)

Type	Echelle	Nom
IDG identifiées comme « observatoires et systèmes d'information sur le littoral » (n= 15)	Supranationale (n=3)	<u>Emodnet ; Sedatanet ; API « Coastline » (absence de site Web dédié)</u>
	Nationale (n=10)	<u>API « Rivages » ; Anemoc ; Candhis ; Bosco ; Geolittoral ; Onml ; Datashom ; Sextant ; Cartomer ; Dynalit</u>
	Régionale/ Interrégionale (n=2)	<u>Observatoire du Littoral Bas-Normand (Oliban) ; Réseau d'Observation du Littoral Normand et Picard (Rolnp)</u>
Autres IDG généralistes et thématiques françaises (n = 27)	Supranationale (n=2)	Geoportail Inspire Europe ; Portail de données de l'Agence Européenne de l'Environnement
	Nationale (n=17)	API Vigicrues ; Base de données Historiques sur les Inondations (BDHI) ; Georisques ; Infoterre ; EauFrance ; Inventaire National du Patrimoine Naturel (INPN) ; Géoportail national de l'urbanisme ^o ; Carmen Nature France ; ADéLie ; Cartélie, Geo-IDE ; Geocatalogue national ; Géoportail national ; Observatoire des territoires ; Portail de données publiques de Météofrance ; Applications Geoclip géo (hors ligne depuis début 2018) et Geoclip stat du Service de l'Observation et des Statistiques
	Régionale/ Interrégionale (n=7)	Geobretagne ; Bretagne Environnement ; Kartenn Bretagne ; Geonormandie ; Geopal ; Indigeo ; Osuna

⁶ <https://www.data.gouv.fr/fr/> <https://geo.data.gouv.fr/fr/>

	Communale/ Intercommunale (n=1)	Géocatalogue de Lorient Agglomération
Plateforme de données ouvertes (n =1)	Nationale (n=1)	Datagouv

S'intéresser aux questions de disponibilité de l'information géographique sur une thématique donnée et sur des dispositifs numériques tels que les IDG et les plateformes de données ouvertes permet de traiter des questions relatives à ses modes de communication dans « l'écosystème numérique » qu'est le Web (Cardon, 2011). Les aspects dits structurels de tels dispositifs numériques s'appuient sur les principes de fonctionnement de l'Internet et du Web 2.0 et constituent à ce titre des enjeux de recherche pour appréhender les processus qui orientent le parcours de l'information géographique numérique dans la gouvernance informationnelle.

Le Web peut être considéré comme un espace façonné par des intentionnalités et des actions d'acteurs qui affichent, diffusent et partagent de l'information dans un réseau interconnecté (Beaude, 2012 ; Graham et Dutton, 2014) et sur des thématiques fixées. À ce titre, Internet et la toile constituent un objet d'investigation (Gautreau et al., 2013 ; Desbois, 2015 ; Severo et Venturini, 2016 ; Noucher, 2017) permettant de caractériser certaines pratiques et stratégies autour de la mise à disposition, de la visibilité et de la mise en réseau de l'information numérique (Gautreau et al., 2013 ; Graham et Zook, 2013 ; Noucher et al., 2017).

Cette recherche vise donc à analyser la contribution d'un ensemble de dispositifs mettant à disposition de l'information géographique numérique sur le Web utilisable sur des enjeux de suivi et de gestion du trait de côte en France. Pour ce faire, nous proposons, à partir de méthodes de recherche dites numériques (Rogers, 2009, 2015 ; Plantin et Monnoyer-Smith, 2014) d'explorer des aspects renvoyant aux caractéristiques structurelles de ces dispositifs. Cette démarche est développée à partir d'une analyse de la structure hypertextuelle et des services Web disponibles (Rogers, 2015) que nous quantifions à partir d'indicateurs statistiques et de l'étude des liens entre ces dispositifs.

2. Le Web comme espace structurel d'investigation de dispositifs numériques en lien avec l'information géographique : éléments méthodologiques

2.1. L'analyse hypertextuelle : un indicateur des ressources offertes par les dispositifs numériques

La structure hypertextuelle des sites Web des IDG et des plateformes de données ouvertes est une première porte d'entrée technique sur leurs contenus. Elle permet à

l'utilisateur de parcourir et d'accéder à des ressources informationnelles sur le site Web de l'IDG ou de la plateforme de données ouvertes qu'il consulte, à partir de liens structurés renvoyant vers différents espaces de consultation sous la forme de pages (par exemple, document sur un autre site Web, espace de téléchargement, cartes, données en ligne etc.). Une telle armature hypertextuelle offre la possibilité d'observer la trame globale du dispositif et non de se concentrer sur un seul composant que serait le visualiseur cartographique, le catalogue ou bien encore l'information géographique diffusée (carte, graphique, etc.). Les hyperliens sont considérés par certains courants de recherche sur le numérique (Digital Methods Initiative⁷) comme des indicateurs de richesse et de diversité informationnelle, notamment les liens sortants (Rogers, 2009). Ils mettent en évidence la redirection du site vers d'autres sources potentielles d'information auxquelles l'utilisateur n'aurait pas accès dans le dispositif visité. Les hyperliens, en permettant la mise en réseau des différents espaces informationnels, peuvent favoriser l'accès à « de nouveaux espaces de connaissances » (Pinède, 2014). Pour les chercheurs en informatique (Kleinberg, 1999 ; Page, Brin et al., 1999 ; Chakrabarti, 2002), la nature documentaire du Web peut être caractérisée par sa structure topologique dont les liens entrants et sortants sont révélateurs de phénomènes de centralité et de hiérarchie. Les liens entrants sont considérés comme des indicateurs « d'autorité » sur le Web c'est-à-dire que la page Web d'un dispositif possède une autorité d'autant plus forte que la somme des degrés sortants des pages qui la mentionne est élevée. Les liens sortants mettent en évidence le phénomène de « hubs » définis comme des pages Web contenant un grand nombre de liens sortants vers d'autres pages Web qui elles-mêmes possèdent une forte autorité. C'est pourquoi les liens hypertextes sont aussi définis dans la littérature comme le résultat de « politiques » (Rogers, 2009) et de stratégies (Gautreau et al., 2013 ; Pinède, 2014) qui peuvent caractériser l'association ou une distance entre les acteurs porteurs des dispositifs.

Afin d'étudier la structure hypertextuelle de notre corpus, la méthode appliquée consiste à explorer, indexer et catégoriser les pages Web, les liens renvoyés et les sites Web découverts à partir d'un programme similaire à ceux utilisés par les moteurs de recherche : un « crawler » ou moissonneur (exemple : le « Googlebot » du moteur de recherche Google). Un « crawler » est un logiciel qui permet de collecter les pages et les liens du site Web visité et ainsi de reconstituer « un espace de cheminement pour l'utilisateur du Web » (Rogers, 2009). Le « crawler » indexe les pages jusqu'à une certaine distance dans le site (ou niveau de profondeur), c'est-à-dire un nombre de clics en partant de la page du site Web que l'on renseigne au « crawler » (ici $n = 2$)⁸. Le « crawler » utilisé est le logiciel Hyphe⁹. Une fois le

⁷ Le « Digital Methods initiative » ou « courant de l'Internet » est un groupe de recherche européen qui s'intéresse aux stratégies de diffusion de l'information par les hyperliens, <https://www.digitalmethods.net/Digitalmethods/WebHome>

⁸ Ce choix est motivé par le fait que la profondeur de 2 correspond à l'espace où l'on trouve les liens les plus importants (prioritaires) dans l'environnement des sites Web (Jacomy, Ghitalla, 2007). C'est généralement à ce niveau que les pages les plus importantes se situent, ce qui peut s'expliquer par le fait que plus les pages sont profondes moins les moissonneurs des moteurs de recherche leur accordent de l'importance.

crawl réalisé en partant de la page d'accueil de chaque dispositif (IDG ou plateforme de données ouvertes), on obtient des résumés sur le nombre de pages Web crawlées, le nombre total de pages Web et de sites Web découverts et le nombre de liens hypertextes présents sur les pages du dispositif visité. Par exemple en partant de la page d'accueil de l'IDG Geolittoral, le programme Hyphe a moissonné 434 pages Web internes à une distance de 2 clics de profondeur. Il a découvert et indexé 14230 liens hypertextes, 13593 adresses de pages Web et 47 sites Web extérieurs. Ces valeurs extensives permettent d'analyser statistiquement les ressources hypertextuelles offertes par chaque dispositif. Afin de caractériser cette ressource, on propose d'étudier la dispersion de la structure hypertextuelle des dispositifs numériques (calcul des valeurs centrales et des paramètres de dispersion, forme de la distribution, représentation graphique).

2.2. Analyse de graphe sur les liens hypertextes et sur les services de consultation des données (services Web, conventions) : un indicateur de la structure en réseau des dispositifs sur la toile

Les notions de réseau renvoient à des questions d'interopérabilité pour les dispositifs numériques sur le Web (Gasser, 2015), notamment ceux dédiés à l'information géographique (Rajabifard, 2002 ; Noucher, 2013 ; Georis-Creuseveau, 2014). En France, la mise en réseau par l'interopérabilité des IDG institutionnelles est promue par la directive Inspire et par la politique de la donnée de l'État. La structure en réseau de ces dispositifs peut dès lors être caractérisée par une analyse et une comparaison des graphes au niveau hypertextuel et au niveau des services de consultation de la donnée (services Web, autres). Le graphe est la figure la plus employée dans la représentation de la topologie du Web. C'est avant tout un objet mathématique défini « comme un ensemble fini de sommets N et un ensemble fini de liens L pouvant être vide, noté $G = (N ; L)$ » (Beaugitte, 2010).

Afin de représenter le réseau hypertextuel, un crawl a été lancé à partir du logiciel Hyphe sur les pages d'accueil des 43 dispositifs du corpus en août 2017. À partir des résultats des enquêtes du projet de recherche Geobs¹⁰ sur les services Web de moissonnage (standard CSW) des IDG institutionnelles françaises et du recensement des autres possibilités de consultation (services OGC standardisés de visualisation WMS, WFS et de téléchargement WPS, conventions), on a cartographié l'ensemble des liens entre les dispositifs. Les relations hypertextuelles et de service ont été cartographiées à l'aide du logiciel Gephi 0.9.1. La spatialisation des dispositifs, c'est-à-dire la représentation de l'ensemble de leurs relations au niveau hypertextuel ($n = 43$ dispositifs avec 114 liens) et au niveau des relations de visualisation, de catalogage et de téléchargement ($n=40$ ¹¹ avec 152 liens) vise à faciliter la lecture du positionnement des dispositifs sur le Web à un instant donné.

⁹ Hyphe est développé par le Sciences Po Médialab dans le cadre du projet Hypertext corpus initiative en 2010 (Girard et al., 2012), <http://hyphe.medialab.sciences-po.fr/>.

¹⁰ <http://geobs.cnrs.fr/>

¹¹ $N = 40$ dispositifs dans le réseau de services et non 43, car les IDG thématiques et généralistes européennes n'ont pas été intégrées à ce niveau d'analyse.

Cette visualisation s'appuie sur un même algorithme de spatialisation basé sur les forces (Futcherman, Reingold, 1991) qui permet la comparaison à partir d'une représentation des nœuds selon des principes d'attraction (les plus liés s'attirent) et de répulsion (les moins liés sont repoussés à la périphérie). Enfin, on a appliqué des algorithmes employés dans la caractérisation de la nature topologique du Web et dans l'indexation de son contenu pour mesurer la centralité de chaque IDG et de la plateforme de données ouvertes (« Autorité » et « Hub » de Kleinberg (1999) et algorithme de classement « PageRank » de Page et al. (1999)).

3. Analyse de la structure hypertextuelle et de la structure en réseau des dispositifs numériques : résultats

3.1. La distribution des pages Web, liens et entités Web : une première mesure de la ressource hypertextuelle

Concernant notre corpus, les résumés statistiques des distributions hypertextuelles (liens, pages Web et entités¹² Web) montrent qu'une personne souhaitant consulter les dispositifs jusqu'à une distance de 2 clics, a accès en moyenne à 22449 pages Web, 11791 liens hypertextuels et à 129 sites Web externes. La valeur modale dans les trois distributions est assez faible, c'est-à-dire que la valeur qui revient le plus fréquemment pour les dispositifs dans les trois séries statistiques est égale à 0 pour les liens découverts, égale à 2 pour le nombre de pages Web découvertes et égale à 0 pour le nombre d'entités Web découvertes. On trouve dans les trois cas des distributions dissymétriques à gauche (médiane très inférieure à la moyenne) c'est-à-dire qu'une majorité des dispositifs se situe dans les valeurs faibles. Ainsi comme on peut le voir sur les diagrammes en bâtons (Figure 1), certains dispositifs ne présentent aucune ou peu de pages (18 dispositifs renvoyant à un nombre de pages Web découvertes inférieur à 100), peu ou aucun lien (19 dispositifs avec un nombre de liens découverts inférieur à 50) et ne renvoient qu'à peu de sites Web différents (20 dispositifs renvoyant à un nombre de sites Web externes inférieur à 5). Seuls quelques dispositifs nationaux et régionaux détiennent des valeurs élevées en nombre de liens, de pages découvertes et d'entités Web (INPN, EauFrance, Datagouv, Geobretagne, Bretagne environnement).

Si l'on compare la répartition des dispositifs dans les trois distributions (pages Web, liens hypertextes, entités Web) selon un découpage en effectifs égaux (méthode des quantiles), on distingue un premier groupe de dispositifs caractérisé par un faible nombre de liens et de pages Web (dispositifs dont la valeur est inférieure à Q1 sur la figure 1). On retrouve dans ce premier groupe 5 dispositifs en lien avec le suivi et la gestion du trait de côte. Le second groupe (dispositifs avec une valeur supérieure à Q3) rassemble des dispositifs qui présentent un grand nombre de pages Web et de liens. Ce sont des IDG thématiques nationales (INPN, SIE, Vigicrues) et régionales (Bretagne environnement), une IDG généraliste régionale (Geobretagne) et la plateforme nationale de données ouvertes (Datagouv).

¹² Une entité web est un site Web.

Concernant les dispositifs identifiés dans la thématique du suivi et de la gestion du trait de côte, aucun dispositif ne se situe dans les 25 % des effectifs aux valeurs les plus élevées pour l'ensemble des indicateurs (nombre de liens, de pages Web et d'entités Web externes). Ce sont l'IDG européenne (Emodnet), les IDG nationales (Geolittoral et Sextant) et l'IDG interrégionale (ROLNP) qui présentent la structure hypertextuelle à la plus grande diversité en termes d'indicateurs hypertextuels (dans 2 indicateurs sur 3, les valeurs sont supérieures à Q1).

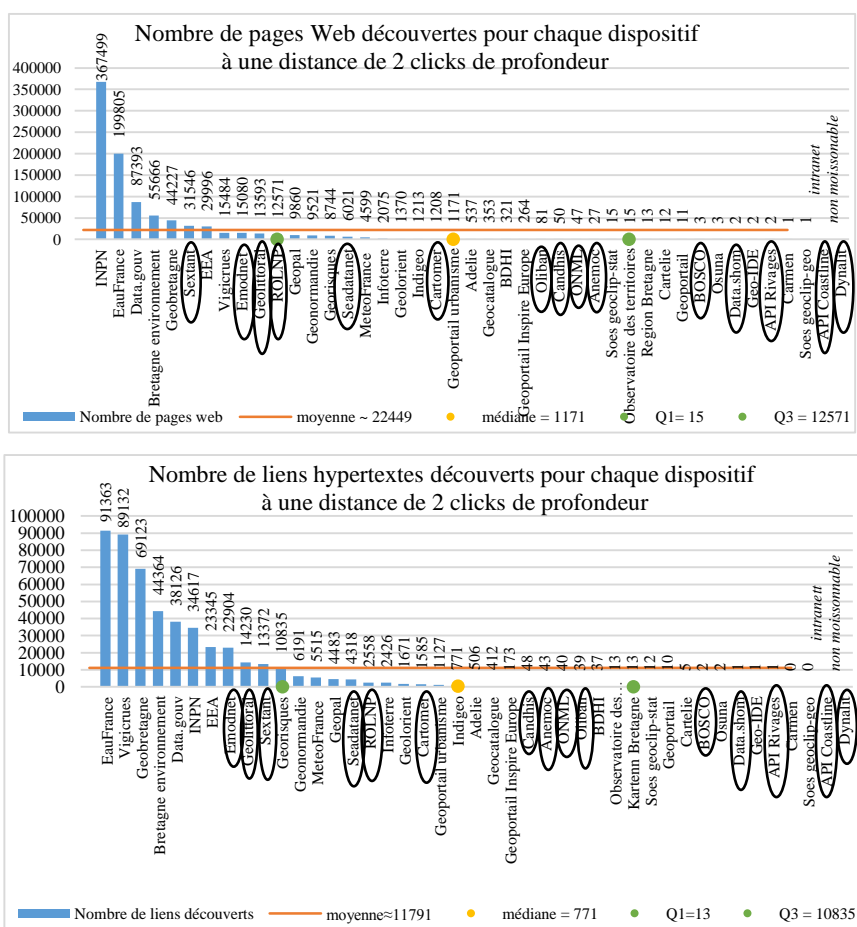


FIGURE 1. Deux des trois indicateurs de la structure hypertextuelle : nombre de pages Web (en haut) et nombre de liens (en bas) (les systèmes d'information sur le littoral – observatoires sont entourés en noir)

3.2. Analyse de la composante réseau des dispositifs numériques : un second indicateur pour caractériser l'aspect structurel

Le second indicateur structurel met en regard le niveau hypertextuel avec le niveau de services de consultation de la donnée (services Web et conventions) par l'analyse de graphe des liens entre dispositifs numériques (Figure 2). Dans les deux niveaux de graphe, la densité¹³ est relativement faible sur l'ensemble du réseau (< à 0,1), ce qui indique que les dispositifs du corpus n'exploitent pas le potentiel maximal de connexion. De plus, on trouve un nombre élevé de composantes¹⁴ connexes c'est-à-dire de sous-graphes (25 au niveau hypertextuel et 23 au niveau des services) pouvant être identifiés comme des communautés ; et des composantes faiblement connexes (7 au niveau hypertextuel et 4 au niveau des services). Les dispositifs les plus centraux dans le graphe (Figure 2) sont des IDG nationales thématiques (INPN, EauFrance) ainsi que des infrastructures généralistes nationale (Géocatalogue) et régionale (Geonormandie) et la plateforme de données ouvertes (Data.gouv). Dans les distributions de degrés entrants et sortants, on observe qu'une majorité de dispositifs situés dans le premier quartile de la distribution (25% des effectifs aux plus forts degrés entrants et sortants) sont des infrastructures nationales généralistes et thématiques.

Ce sont les dispositifs nationaux Geolittoral et l'ONML identifiés par la stratégie nationale de gestion intégrée du trait de côte et les systèmes d'information mer et littoral (Sextant, Cartomer et Datashom) qui sont les plus centraux dans les deux graphes (Figure 2). D'un point de vue hypertextuel, si la majorité des dispositifs (thématiques, généralistes, territoriaux) sont connectés aux IDG identifiées par la stratégie nationale de gestion intégrée du trait de côte, cela ne saurait masquer une faible intensité des liens (31 liens sur les 114 liens recensés). Au niveau du graphe des services, d'une part les possibilités de consultation des données d'un dispositif du trait de côte à un autre dispositif restent limitées (52 liens sur 152 liens recensés). D'autre part, la majorité des relations par l'intermédiaire des services de consultation de la donnée sont le plus souvent cantonnées à une possibilité de consultation unique par la présence d'un seul type de service (téléchargement ou moissonnage ou visualisation). Enfin au niveau de l'analyse de graphe, 6 dispositifs sont en situation d'isolat¹⁵ au niveau hypertextuel. Deux dispositifs identifiés dans la stratégie nationale de gestion intégrée du trait de côte (Bosco et Oliban) et 2 dispositifs en lien avec l'information géographique numérique mer et littoral (Candhis et Anemoc) sont isolés au niveau du réseau des services.

Si l'on compare les 10 scores les plus élevés des indicateurs de centralité (Autorités, « hubs », « PageRank »), ce sont les IDG nationales thématiques

¹³ « La densité (*density*) d'un graphe désigne le rapport entre le nombre de liens présents et le nombre de liens possibles elle varie entre 0 (graphe vide, sans aucun lien) et 1 (graphe complet). » (Beaugitte, 2013)

¹⁴ « Un composant correspond à un sous graphe non connecté aux autres sous-graphes. » (Beaugitte, 2010).

¹⁵ Un nœud ou sommet est dit en situation d'isolat lorsqu'il n'est connecté à aucun autre nœud.

(EauFrance, INPN) et généralistes (Géoportail, Géocatalogue) qui se partagent les plus forts scores de centralité au niveau topologique. Parmi les dispositifs dédiés à l'information géographique numérique sur le trait de côte, on retrouve ainsi quatre IDG nationales (Geolittoral, Onml, Datashom et Sextant) dans les plus forts scores de centralité. Au niveau des indicateurs d'autorité et de « hubs », seul un dispositif (Datashom) possède un score élevé (autorité) dans les deux réseaux (hypertexte et service). Les 4 dispositifs Geolittoral, Onml, Datashom et Sextant possèdent un score élevé de « PageRank » dans les deux réseaux.

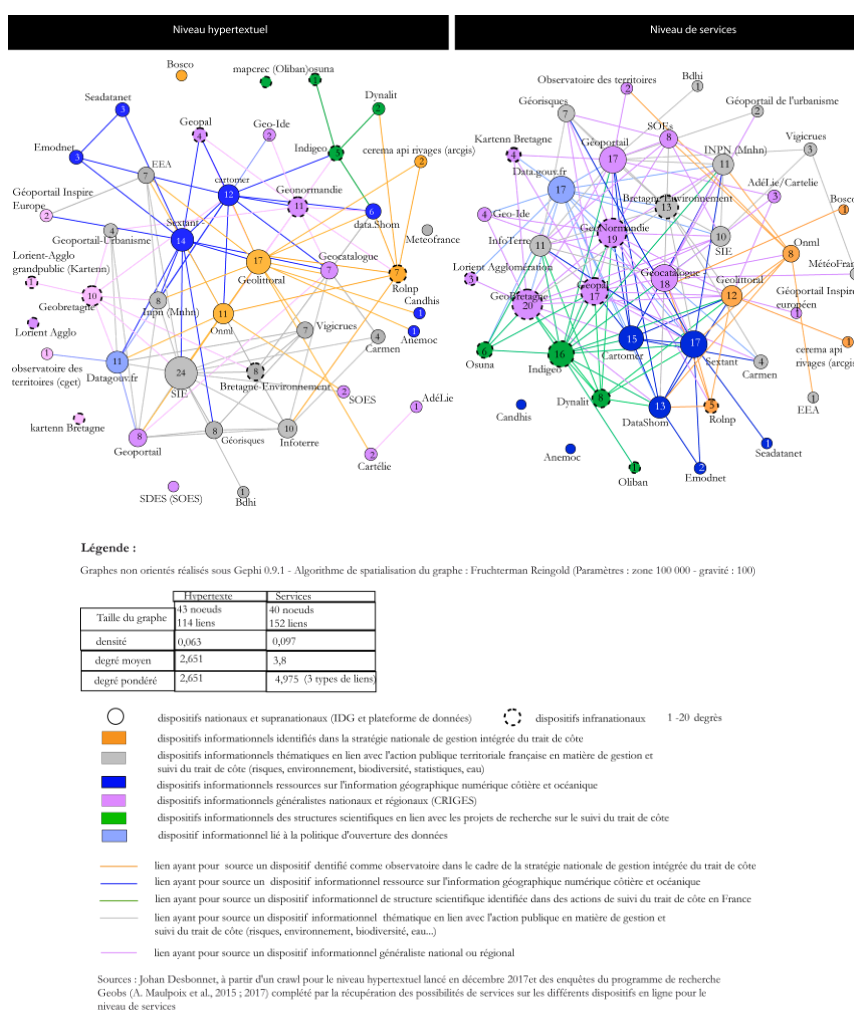


FIGURE 2. Mise en regard des graphes des liens entre dispositifs au niveau hypertextuel et au niveau des services de consultation de données

4. Discussion

Nous proposons un protocole méthodologique visant à fournir des indicateurs en lien avec les caractéristiques structurelles de dispositifs d'accès à l'information géographique numérique sur le Web (IDG-plateforme de données ouvertes). Ce protocole a été testé sur un corpus de 43 dispositifs institutionnels français, avec un focus sur ceux concernant la thématique du suivi et de la gestion du trait de côte en France. Il convient de rappeler que les méthodes numériques employées constituent une première phase d'analyse souvent exploratoire permettant de dégager des tendances globales sur la thématique étudiée et non de faire émerger des modèles généraux (Plantin et Monnoyer-Smith, 2013 ; Ash et al., 2016). Elles doivent être accompagnées de méthodes d'analyse complémentaires comme l'analyse de contenu et l'analyse thématique des pages renvoyées par certains dispositifs numériques.

D'un point de vue des limites de notre approche méthodologique, nous avons été confrontés à une dynamique de la donnée (en volume, mise à jour, et supports) et à des logiciels évolutifs (langages, librairies, systèmes d'exploitation), qui demandent de plus en plus de ressources logistiques et mettent en difficulté l'emploi d'outils traditionnels tels que les suites logicielles traditionnelles. Le volume de données hypertextuelles traitées a souvent entraîné des difficultés pour stabiliser la base de données (MongoDB). En effet, certaines infrastructures comme la plateforme Data.gouv ou l'IDG Vigicrues contiennent près d'un million de pages à une profondeur de 2 clicks, ce qui peut affecter les capacités de stockage et de traitement. D'autres dispositifs au contenu dynamique de type « Content Management Systems », à l'instar de Dynalit n'ont pas pu être moissonnés par le logiciel. Pour autant, si l'utilisation du caractère exploratoire d'un logiciel de « crawl » tel que Hype doit s'inscrire dans une problématique de recherche initiale, celui-ci peut aussi contribuer à la faire évoluer comme cela a été le cas dans la sélection de notre corpus qui a été redessiné au cours du projet de recherche.

5. Conclusion

Le recensement des liens hypertextes, pages Web et sites Web renvoyés par chaque dispositif étudié révèle à ce jour une structure très hétérogène. En effet, un premier groupe de dispositifs renvoie à un faible nombre de liens, de pages et d'entités Web parmi lesquels figurent 5 dispositifs en lien avec l'information géographique numérique sur le trait de côte ($n = 5$ dispositifs sur les 15 dispositifs en lien avec l'information géographique numérique sur le trait de côte présentent des valeurs inférieures à Q1). Ce sont le plus souvent des dispositifs spécialisés dans une forme de consultation unique de la donnée géographique (visualisation, génération de graphiques, catalogage de données). Ces infrastructures en ligne apparaissent dans les années 2005-2010 au moment de l'émergence des premiers dispositifs de cartographie grand public sur le Web (Mericksay, 2011 ; Noucher, 2017). D'autres au contraire sont très récents comme les applications de suivi du trait de côte par smartphone (information géographique volontaire). Il s'agit de dispositifs peu éditorialisés (faible quantité de liens et de pages Web au niveau hypertextuel) et peu

visibles dans le réseau des IDG institutionnelles (faible score d'autorité et peu interconnecté), au sein desquels le cheminement pour l'utilisateur est restreint à la fonction initiale du dispositif. Au contraire, certaines IDG nationales thématiques en lien avec les politiques publiques environnementales (biodiversité, eau, risque inondation) ainsi que la plateforme de données ouvertes nationale data.gouv sont plus visibles et offrent à l'utilisateur la possibilité d'élargir sa consultation à un grand nombre de ressources.

Enfin la composante réseau de ces dispositifs, deuxième indicateur structurel souvent mis en avant pour caractériser les IDG et autres plateformes de diffusion de données géographiques (Noucher et al., 2017), fait apparaître une connexité encore perfectible pour faciliter l'interopérabilité et l'accès aux différentes ressources offertes par l'ensemble des dispositifs. Ainsi, au sein des dispositifs identifiés par les pouvoirs publics dans la stratégie nationale de gestion intégrée du trait de côte, seules les IDG nationales Geolittoral et l'ONML se positionnent dans les scores les plus élevés en termes de centralités de réseaux (« Autorités », « Hubs », « PageRank »), tout en jouant des rôles différents dans les deux niveaux. Geolittoral, dispositif mis en avant dans la stratégie nationale de gestion intégrée du trait de côte, présente un profil de « hub » dans le niveau hypertextuel et d'autorité au niveau des services (synthèses, rapports) et à un public plus large présente un profil d'autorité au niveau hypertextuel et de « hub » au niveau des services. Ces infrastructures exhibent donc une structure sur le Web que l'on peut qualifier de niveau unique : « hub » et autorité uniquement à un niveau de réseau (hypertexte ou services web) et services de consultation de la donnée d'un seul type. L'autre constat qui émerge de cette recherche d'un point de vue structurel est la présence quasi systématique de dispositifs généralistes, thématiques nationaux et territoriaux aux indices de structure hypertextuelle et aux niveaux d'interopérabilité les plus élevés, mais qui sont pourtant encore peu pointés dans la stratégie nationale de gestion intégrée du trait de côte française.

Ces résultats, qui concernent les données hypertextuelles et les services de consultation des dispositifs sur le Web, seront prochainement complétés par une analyse des données géographiques cataloguées par ces dispositifs en ligne et par des entretiens menés avec des promoteurs et des utilisateurs afin de mieux comprendre l'aspect structurel et la disponibilité de l'information géographique sur ces dispositifs¹⁶.

¹⁶ Ce travail s'inscrit dans le cadre de la thèse AInfoGéo (<http://www-iuem.univ-brest.fr/pops/projects/ainfoge>) dont l'objectif est de proposer un cadre méthodologique pour étudier la contribution de l'information géographique numérique disponible sur des dispositifs en ligne comme les IDG et plateformes de données ouvertes dans les politiques publiques en lien avec le suivi et la gestion du trait de côte.

Bibliographie

- Ash J., Kitchin R. & Leszczynski A. (2016). Digital turn, digital geographies? *Progress in Human Geography*. vol. 42, p. 25-43, <https://doi.org/10.1177/0309132516664800>.
- Beaude B. (2012). *Internet, changer l'espace changer la société les logiques contemporaines de synchronization*, éditions Fyp, collection société de la connaissance.
- Beauguitte L. (2010). Graphes, réseaux, réseaux sociaux : vocabulaire et notation : Version 1 - Septembre 2010. Groupe f.m.r. (flux, matrices, réseaux) animé par César Ducruet et Laurent Beauguitte (CNRS, UMR , 2010 <halshs-00541898>)
- Beauguitte L. (2013). *D comme densité*, <https://groupefmr.hypotheses.org/2653>
- Bersani C., M.-L. Simoni, Y.-M. Allain, G. Ribière, J. Denègre, P. Planques, Y. Tugayé (2006). *Rapport relatif au schéma d'organisation des dispositifs de recueil de données et d'observation sur le littoral*, Ministère de l'intérieur et de l'aménagement du territoire, Ministère des transports, de l'équipement, du tourisme et de la mer, 119 p, http://www.shom.fr/fileadmin/data/DOPS/MIP/PEP/BATHY/Litto3D/Besoins/Rapport-Bersani_recueil-donnees-littoral.pdf
- Cardon D. (2011). L'ordre du Web. *Médium*, 2011/4 vol. n° 29, p. 191-202. DOI: 10.3917/mediu.029.0191. URL: <https://www.cairn.info/revue-medium-2011-4-page-191.htm>
- Chakrabarti S. (2002). *Mining the Web: Discovering knowledge from hypertext data*, Elsevier.
- Colin N. et Verdier H. (2015). *L'âge de la multitude-2e éd., Entreprendre et gouverner après la révolution numérique*. Armand Colin.
- Desbois H. (2015). *Les mesures du territoire : aspects techniques, politiques et culturels des mutations de la carte topographique*, Presses de l'ENSSIB, Nancy.
- Fruchterman, T. M. J., Reingold, E. M. (1991), Graph Drawing by Force Directed Placement, *Software: Practice and Experience*, vol. 21, no 11, p. 1129-1164.
- Gautreau, P., Severo, M., Giraud, T., Noucher, M. (2013), Formes et fonctions de la « donnée » dans trois webs environnementaux sud-américains (Argentine, Bolivie, Brésil), *Networks and Communication Studies (NETCOM)*, vol. 27 (2013), n° 1-2 pp. 22-59.
- Gasser U. (2015). *Interoperability in the Digital Ecosystem*, Berkman Klein Center for Internet and Society Research Publication No. 2015-13. Available at SSRN: <https://ssrn.com/abstract=2639210> or <http://dx.doi.org/10.2139/ssrn.2639210> <https://dash.harvard.edu/handle/1/28552584>
- Georis-Creuseveau. J. (2014). *Les Infrastructures de Données Géographiques (IDG) : développement d'une méthodologie pour l'étude des usages. Le cas des acteurs côtiers et de la GIZC en France*. Thèse de Géographie, Université de Bretagne Occidentale.
- Georis-Creuseveau J., Claramunt C., Gourmelon F. (2016). A modelling framework for the study of Spatial Data Infrastructures applied to coastal management and planning. *International Journal of Geographical Information Science*, vol. 31, no 1, p. 122-138. DOI: 10.1080/13658816.2016.1188929
- Girard P., Jacomy M., Baneyx A. (2012). The Hypertext Corpus Initiative: methods and tools for Social Sciences to build corpus from the Web, *Digital Humanities Luxembourg 2012*

Websites as Sources, Luxembourg, http://www.medialab.sciences-po.fr/publications/HypertextCorpusInitiative_webassources_abstract.pdf.

Graham M, and Zook M. (2013). Augmented Realities and Uneven Geographies: Exploring the Geo-Linguistic Contours of the Web. *Environment and Planning A* 45, n°1, p. 77–99.

Graham M., & Dutton W. H. (Eds.) (2014). Introduction, *Society and the internet: How networks of information and communication are changing our lives*, OUP Oxford, p. 1-19.

Jacomy M., Ghitalla F. (2007). *Méthodologie d'analyse de corpus en Sciences Humaines à l'aide du Navicrawler*. Rapport final, Fondation de la Maison des Sciences de l'Homme Programme TIC-Migrations , <http://webatlas.fr/wp/share/navicrawler/Guide%20m%E9thodo%20NC%202007.pdf>.

Kleinberg J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, vol. 46, no 5, p. 604-632.

Mericksay, B. (2011). Les Sig et la cartographie à l'ère du géoweb Vers une nouvelle génération de Sig participatifs. *L'Espace géographique*, 2011/2 Tome 40, p. 142-153.

Noucher M. (2013), Infrastructures de données géographiques et flux d'information environnementale, *Networks and Communication Studies (NETCOM)*, 27-1/2 | 2013, p. 120-147.

Noucher M., Gourmelon F., Gautreau P., Georis-Creuseveau J., Maulpoix, A., Pierson J., Rouan M. (2017). Spatial Data Sharing: A Pilot Study of French SDI, *ISPRS International Journal of Geo-Information*, vol. 6, no 4, p. 99.

Noucher M. (2017). *Les Petites Cartes du web. Approche critique des nouvelles fabriques cartographiques*, Editions Rue d'Ulm, collection Actes de la recherche à l'Ens n°18.

Page, L., Brin, S., Motwani, R., Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the web*, Stanford InfoLab, <http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>.

Pinède N. (2014). Le réseau, un « entre-deux » liens ? De quelques facettes et dynamiques du lien hypertexte. *Sciences de la société*, 91 | 2014, 94-105.

Plantin J-C, Monnoyer-Smith L. (2014). Ouvrir la boîte à outils de la recherche numérique. *tic&société* [En ligne], Vol. 7, N° 2 | 2ème semestre 2013, mis en ligne le 01 juin 2014, Consulté le 02 juin 2014. URL : <https://journals.openedition.org/ticetsociete/1527> ; DOI : 10.4000/ticetsociete.1527.

Pornon H. (2015). *SIG : La dimension géographique du système d'information (2ème édition)*, Management des systèmes d'information, Paris, Dunod.

Rajabifard, A., Feeney, M.-E., Williamson, I.P. (2002). Future directions for SDI development. *International Journal of Applied Earth Observation and Geoinformation*, vol. 4, n° 1, p. 11-22.

Rogers R. (2009). Situating Digital Methods in Internet Research, *The End of the Virtual: Digital Methods*, Amsterdam University Press.

Rogers, R. (2015). Digital Methods for Web Research, *Emerging Trends in the Social and Behavioral Sciences* (eds R. A. Scott and S. M. Kosslyn). doi:10.1002/9781118900772.etrds0076.

- Severo M., Venturini T. (2016). Enjeux topologiques et topographiques de la cartographie du Web. Le cas du patrimoine culturel immatériel français. *Réseaux*, 1/2016 n° 195, p. 85-105.
- Suarez S., Garcin M., Bulteau T., Rouan M., Lagadec L. & David L. (2012). Les observatoires du trait de côte en France métropolitaine et dans les DOM. *EchoGéo* [En ligne], 19 | 2012, mis en ligne le 10 février 2012, consulté le 12 décembre 2016. URL : <http://journals.openedition.org/echogeo/12942> ; DOI : 10.4000/echogeo.12942.