



HAL
open science

De l'open data à l'open science : retour réflexif sur les méthodes et pratiques d'une recherche sur les données géographiques

Nathalie Pinède, Matthieu Noucher, Françoise Gourmelon, Karel Soumagnac-Colin

► To cite this version:

Nathalie Pinède, Matthieu Noucher, Françoise Gourmelon, Karel Soumagnac-Colin. De l'open data à l'open science : retour réflexif sur les méthodes et pratiques d'une recherche sur les données géographiques. *Revue française des sciences de l'information et de la communication*, 2017, 27 (11), pp.11-36. 10.4000/rfsic.3200 . halshs-01616790

HAL Id: halshs-01616790

<https://shs.hal.science/halshs-01616790v1>

Submitted on 1 Jun 2022

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.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

De l'open data à l'open science : retour réflexif sur les méthodes et pratiques d'une recherche sur les données géographiques

*From open data to open science: reflections on the research methods and
practice of geographic data*

Nathalie Pinède, Matthieu Noucher, Françoise Gourmelon et Karel
Soumagnac-Colin



Édition électronique

URL : <https://journals.openedition.org/rfsic/3200>

DOI : 10.4000/rfsic.3200

ISSN : 2263-0856

Éditeur

Société Française de Sciences de l'Information et de la Communication

Ce document vous est offert par Université Le Havre Normandie



Référence électronique

Nathalie Pinède, Matthieu Noucher, Françoise Gourmelon et Karel Soumagnac-Colin, « De l'open data à l'open science : retour réflexif sur les méthodes et pratiques d'une recherche sur les données géographiques », *Revue française des sciences de l'information et de la communication* [En ligne], 11 | 2017, mis en ligne le 01 juillet 2017, consulté le 01 juin 2022. URL : <http://journals.openedition.org/rfsic/3200> ; DOI : <https://doi.org/10.4000/rfsic.3200>

Ce document a été généré automatiquement le 29 septembre 2020.



Les contenus de la *Revue française des sciences de l'information et de la communication* sont mis à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International.

De l'open data à l'open science : retour réflexif sur les méthodes et pratiques d'une recherche sur les données géographiques

From open data to open science: reflections on the research methods and practice of geographic data

Nathalie Pinède, Matthieu Noucher, Françoise Gourmelon et Karel Soumagnac-Colin

Introduction

- 1 Il y a 25 ans, J.-B. Harley proposait un texte fondateur des recherches en cartographie critique intitulé « Deconstructing the Map ». En prenant de la distance à l'égard des approches fonctionnalistes et en privilégiant leur portée cognitive, il a invité les sciences sociales à repenser les cartes comme des formes de savoir socialement construit, subjectif et idéologique (Lascoumes, 2007). Prônant une analyse du « pouvoir des cartes » (Wood, 1992), il a ouvert la voie à de nombreux travaux en géographie et plus généralement en sciences sociales démontrant l'intentionnalité sociale et le pouvoir d'iconisation du média cartographique.
- 2 Mais, depuis l'apparition du géoweb, il y a dix ans¹, le paysage de la cartographie a considérablement évolué (Plantin, 2014). Les mutations technologiques et les changements d'usages qu'elles ont induits, conduisent à une profusion de données géographiques aujourd'hui diffusées en libreaccès sur des infrastructures de données institutionnelles² (comme le géoportail de l'IGN) ou des plateformes collaboratives (comme OpenStreetMap). Dès lors, la carte échappe aux seules mains des cartographes. La massification (à défaut de parler de démocratisation) de la cartographie sur le web se traduit par une complexification du processus de fabrication cartographique (Noucher,

2017). La démultiplication des sources et la dispersion des acteurs potentiellement impliqués dans les différentes étapes de la mise en carte rendent aujourd'hui toute tentative de décryptage de l'intentionnalité cartographique particulièrement périlleuse. Comme le souligne Henri Desbois (2015) : « *La rhétorique propre aux globes virtuels et à d'autres objets néogéographiques ne se déchiffre pas aussi aisément que celle de la carte papier* ».

- 3 Ce phénomène de massification et de diversification des données géographiques disponibles sur Internet reconfigure dès lors les dynamiques de recherche selon trois axes de transformation clefs : les objets, les méthodes et les pratiques de recherche. Nous illustrerons ces trois axes en nous appuyant sur un projet de recherche en cours. Le projet de recherche « GÉOBS : Les Infrastructures de Données Géographiques (IDG) dans la gouvernance informationnelle de l'environnement » est financé par la Nouvelle Aquitaine (2015/2017) et coordonné par deux chercheurs du CNRS (Matthieu Noucher et Françoise Gourmelon). Il réunit trois laboratoires de recherche en géographie (Passages-Bordeaux, LETG-Brest, PRODIG-Paris), un laboratoire en sciences de l'information et de la communication (MICA) et un laboratoire en informatique (LaBRI-Bordeaux). Il vise à analyser les stratégies, contenus et usages de 45 IDG françaises (16 nationales et 29 régionales). Pour ce faire il mobilise une approche mixte qui associe entretiens auprès des responsables des IDG, enquêtes auprès des utilisateurs, analyses lexicales et sémiotiques des sites web et fouilles de données, statistiques uni et multi variées et analyses réseaux sur les 160 000 fiches de métadonnées associées aux jeux de données géographiques qui sont diffusées sur les sites web de ces IDG. L'objet de cet article n'est pas d'en présenter les résultats détaillés³ mais de mobiliser cette expérience pour revenir de manière réflexive sur l'évolution de nos pratiques de recherche. Tout d'abord, nous soulignerons comment les enjeux de pouvoir autour de la cartographie se sont déplacés avec l'avènement du web et de l'*open data* : du produit « carte » à sa mise en circulation. Ce faisant, nous soulignerons les impacts au plan de la méthodologie de recherche, et notamment le déplacement d'analyses initialement focalisées sur le discours des cartographes et la sémiologie des cartes à des analyses désormais centrées sur le contenu des sites web et les flux de données. Enfin, nous verrons comment, en s'inscrivant dans une démarche d'*open science*, ce projet de recherche qui diffuse ses résultats sous la forme d'articles dans des publications en *open access*, de visualisations interactives sur un site web, de jeux de données dans un portail *open data* et de codes informatiques dans un dépôt en ligne, a reconfiguré les relations entre les parties prenantes (promoteurs et usagers des IDG, objets de notre recherche) et les chercheurs. Nous montrerons donc en quoi les nouveaux terrains d'expérimentations sur le web, inscrits dans des problématiques sociétales, modifient les conditions de la pratique scientifique, des objets aux méthodes, de l'*open data* à l'*open science*.

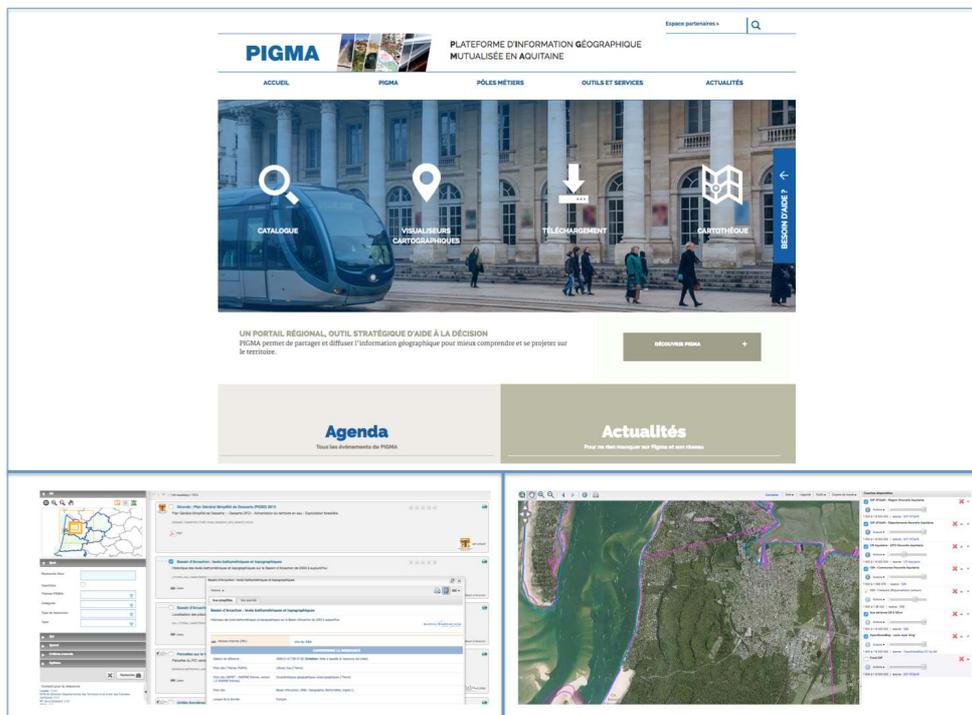
Reconfiguration des objets

Infrastructure de données : du moteur de recherche à l'objet de recherche

- 4 La volonté de faciliter l'accès sur un même territoire à des cartes issues de fournisseurs différents a conduit, depuis les années 1990, au développement croissant d'Infrastructures de Données Géographiques (IDG) nationales (Crompvoets *et al.*, 2004) puis à leur déclinaison à tous les échelons territoriaux (Masser, 2010). Ces plateformes

rassemblent aujourd'hui les bases de données, les réseaux informatiques, les normes et standards, les accords organisationnels et les ressources humaines nécessaires pour faciliter et coordonner le partage, l'accès et la gestion des données géographiques (Rajabifard *et al.*, 2003). Les IDG sont désormais au cœur de l'écosystème des données en libre accès dans la mesure où elles structurent les données géographiques institutionnelles qui constituent une part essentielle des données publiques libérées sous forme d'*open data* (Noucher et Gautreau, 2013). Souvent réduites à de simples sites web permettant de consulter les données cartographiques d'un territoire à travers un géoportail et de les télécharger via un géocatalogue, les infrastructures de données géographiques sont, en fait, de véritables dispositifs sociotechniques où se joue une partie des évolutions majeures affectant le régime actuel de production et de circulation de l'information géographique numérique. La figure 1 présente donne un exemple des sites web analysés à partir de trois composantes principales : la partie éditoriale permet d'accéder aux actions d'animation des communautés de pratique (en haut), le géocatalogue permet d'accéder aux métadonnées et de télécharger les données (en bas à gauche) et le géoportail permet de visualiser les données (en bas à droite).

Figure 1 – Trois composantes clés des sites web analysés



D'après l'IDG PIGMA – www.pigma.org

- 5 Majoritairement pilotées par les autorités publiques, mais s'adressant à l'ensemble de la société, ces infrastructures de données permettent d'observer la recomposition des rapports de pouvoir qui se joue autour du partage de données sur le web (Gautreau et Noucher, 2016).
- 6 Ainsi, alors que la cartographie critique (Harley, 1990 ; Wood et Fels, 1992) a, pendant longtemps, favorisé une analyse de l'objet « carte », de son pouvoir interne (résultant des choix des cartographes et de leurs commanditaires de mettre en (in)visibilité certaines ressources ou populations) et de son pouvoir externe (résultant de la capacité

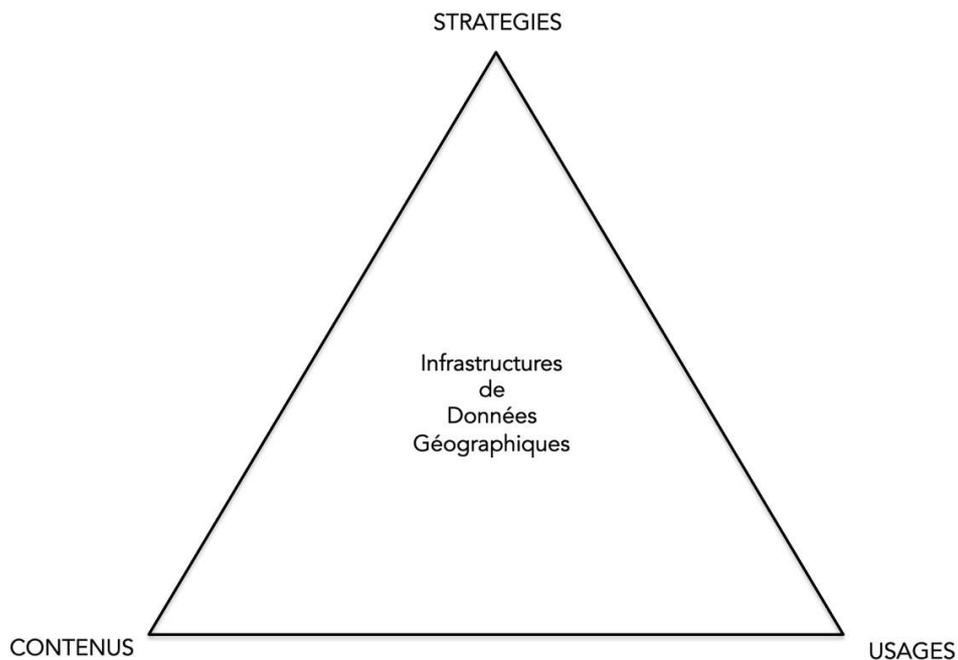
performative de la carte qui tend à s'imposer dès lors qu'elle est diffusée), l'émergence d'infrastructures de données géographiques interconnectées sur le web vient complexifier la donne. Deux enjeux de recherche s'imposent actuellement : d'une part être en mesure de circonscrire le patrimoine de données disponibles (tant le *data deluge* ne permet plus d'établir facilement un état des lieux des connaissances sur un territoire donné⁴) et, d'autre part, suivre la circulation des données cartographiques (tant la compréhension de la généalogie des sources est devenue complexe face à des données malléables qui tendent à s'hybrider). Ce faisant, il s'agit d'interroger les enjeux de la mise en réseaux de l'information géographique à travers l'étude de ces infrastructures, afin de comprendre les modalités selon lesquelles les représentations spatiales des territoires s'établissent et se diffusent. Le projet GÉOBS propose donc un basculement dans l'appréhension des infrastructures de données géographiques en les considérant comme un objet de recherche permettant de questionner les modalités contemporaines de construction, circulation, confrontation des savoirs (institutionnels) sur l'espace (Noucher, 2013).

Dépasser les incantations : confronter discours, données et usages

- 7 Les recompositions techniques (normalisation des données et développement de l'interopérabilité⁵), juridiques (renforcement du cadre réglementaire⁶) et organisationnelles (démultiplication de communautés de pratique⁷) autour des infrastructures de données pourraient laisser croire qu'un vaste chantier d'ouverture des données est engagé de façon consensuelle, sans être remis en cause. Ainsi, l'essor d'initiatives de type *Open Data* est souvent perçu comme emblématique des « révolutions numériques » en cours, matérialisant à la fois la réponse des pouvoirs publics à la transformation des exigences démocratiques et les possibilités offertes par la technique au déploiement et à l'actualisation de ces aspirations sociales. Mettant au premier plan la notion de « partage » des données, les Infrastructures de Données Géographiques sont ainsi censées représenter une forme de renouvellement du pacte entre Etat et citoyens, autour de l'échange d'informations.
- 8 Pourtant, les contours et les modalités de ce renouvellement sont loin d'être homogènes d'un territoire (et d'une échelle) à l'autre, et le sens même de cette ouverture est au cœur de controverses : « *moyen de pression pour la tendance libérale, elle témoigne d'une véritable méfiance envers toute institution publique pour les libéraux-libertaires tandis que d'autres acteurs y voient une opportunité de développer la démocratie participative* » (Giraud, 2014). Ainsi, ceux qui se présentent volontiers comme les militants du « mouvement open data » (Chignard, 2012) sont finalement porteurs de valeurs, d'objectifs et de stratégies qui diffèrent de ceux de la puissance publique. L'« open » d'Open Data désigne sans doute moins ce qui rassemble les différents acteurs de ce mouvement que ce qui les différencie⁸.
- 9 Face à cette diversité d'appropriation des infrastructures et des discours sur l'ouverture des données, il devient délicat d'envisager un décryptage d'intentionnalités potentiellement multiples. Il s'agit alors de dépasser le paradigme communicationnel pour ne pas limiter les infrastructures de données à de simples vecteurs d'informations, dont il suffit d'étudier en amont l'encodage et en aval le décodage. Du fait de ces différentes appropriations mouvantes et plastiques, le sens des données s'élabore dans la relation à son contexte d'usage (Georis-Creuseveau *et al.*, 2015). Autrement dit, au gré de ses réappropriations multiples, les données géographiques tendent à se métamorphoser,

dans leurs apparences comme dans leurs fonctions ou leurs différents sens. Une analyse des flux de données devient dès lors essentielle pour comprendre comment s'opère leur mise en circulation (Georis-Creuseveau *et al*, 2017). En privilégiant une entrée par les usages, peu développée dans les travaux de Harley et des tenants de la cartographie critique, il s'agit alors de prendre la mesure de la diversité des fonctions remplies par le brassage et la fragmentation des données géographiques. En proposant une lecture fine et nuancée de ce que les *usagers* font effectivement avec les infrastructures de données, l'objectif est de prolonger la cartographie critique en proposant une relecture des cadres habituels d'analyse de l'intentionnalité sociale du média cartographique. Pour ce faire, nous avons cherché, au-delà des incantations autour de l'*open data*, à confronter les stratégies des promoteurs (leurs discours et textes de références) de ces infrastructures, leurs contenus (les données effectivement diffusées) et les usages opérationnels qui y sont associés (figure 2).

Figure 2. Les trois axes d'analyse des IDG : stratégies, contenus, usages



Reconfiguration des méthodes

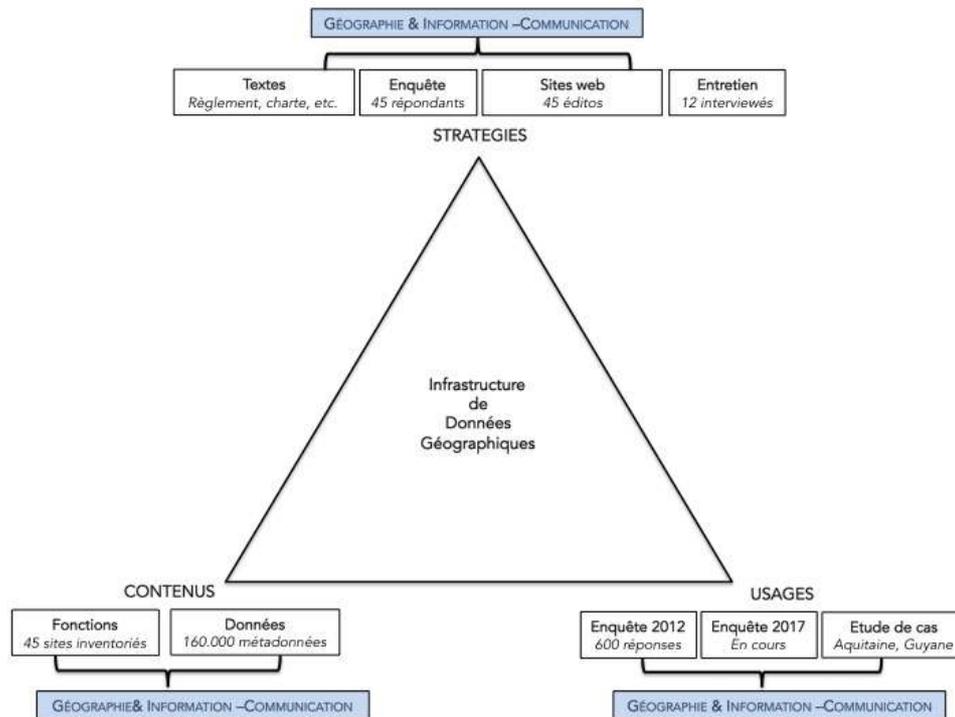
- 10 L'étude de notre objet de recherche dans ses trois axes d'analyse mobilise des corpus inédits et hybrides, nécessitant des investigations croisées. Dans cet environnement numérique ouvert de données, porteurs d'enjeux politiques et stratégiques territorialisés, plusieurs éléments émergent au plan de cette reconfiguration des méthodes : la nécessité d'une approche interdisciplinaire, des cadres d'action d'une « recherche numérique » ainsi que le recours à des méthodologies mixtes.

Interdisciplinarité

- 11 Si l'on reprend la terminologie proposée par Basarab Nicolescu (1996), « la pluridisciplinarité concerne l'étude d'un objet d'une seule et même discipline par plusieurs disciplines à la fois » (Nicolescu, p. 64). L'interdisciplinarité permet de franchir un cran supplémentaire dans l'entrelacement des disciplines, puisqu'elle « concerne le transfert des méthodes d'une discipline à une autre », avec des bouleversements induits variables, du simple transfert-applicatif à l'engendrement de nouvelles disciplines, en passant par des modifications de nature épistémologique (Nicolescu, p. 65). Enfin, stade ultime, la transdisciplinarité, qui transcende les disciplines du fait de sa finalité, à savoir la compréhension du monde, « dont un des impératifs est l'unité de la connaissance » (Nicolescu, p. 66).
- 12 En complémentarité de cette typologie relativement classique, A. Bénel propose trois dimensions constitutives de l'interdisciplinarité (Bénel, 2014), telles qu'elles peuvent émerger au cœur de projets de type humanités numériques et dans l'articulation entre STIC et SHS. *L'immersion* constitue la première dimension, renvoyant à une facette quasi ethnologique de l'interdisciplinarité. Elle suppose une temporalité longue de collaboration, permettant de transcender le strict cadre du travail pour construire une connaissance de l'autre participant de la compréhension des objets de recherche partagés. *Le transfert de méthodes* constitue la deuxième dimension proposée par A. Bénel et incarne la valeur instrumentale associée à ce transfert. Enfin, la *discussion épistémologique* représente la dernière dimension de l'interdisciplinarité et prend forme lors des débats et questionnements qui vont se poser à l'occasion de séminaires communs et des choix qui vont en résulter.
- 13 Piloté par deux géographes dont les recherches s'inscrivent dans le courant de la géomatique (interface entre géographie et informatique), le projet GÉOBS rassemble également des chercheurs en informatique et sciences de l'information et de la communication, avec des principes de collaboration proches des dimensions de l'interdisciplinarité telles que définies précédemment par A. Bénel. En ce qui concerne la dimension de *l'immersion*, celle-ci s'inscrit dans la durée du projet (trois années), tout en s'appuyant également sur des pratiques de collaboration pré-existantes au projet. Dans le fil du projet, l'immersion se traduit par les échanges lors de séminaires communs réguliers ou encore, à l'occasion de l'élaboration d'un modèle conceptuel partagé. En matière de *transfert de méthodes*, les dispositifs complexes que sont les IDG nécessitent pour être analysés selon les trois dimensions précitées (figure 1) de croiser des corpus diversifiés acquis et étudiés soit par les géographes, soit par les chercheurs en SIC. La stratégie des promoteurs de 45 plateformes françaises (16 nationales et 29 régionales) est appréhendée à partir d'enquête en ligne, d'entretiens semi directifs ciblés sur des acteurs nationaux et régionaux mais aussi par l'analyse des textes de références (chartes, guide de bonnes pratiques, cadres réglementaires, texte des sites web). L'analyse des contenus des infrastructures se fait à partir de l'inventaire des fonctionnalités disponibles et de l'exploration des métadonnées⁹ qui documentent les données géographiques diffusées. Enfin, l'étude des usages repose sur deux enquêtes en ligne réalisées en 2012 (Georis-Creuseveau, 2014) et début 2017. Les protocoles d'analyses sont mis en œuvre par les géographes (statistiques) et les informaticiens (analyses réseaux, modèles...). C'est la prise en compte de ces corpus variés qui permet d'appréhender l'objet et les jeux d'acteurs

dans leur ensemble et dans toute leur complexité (figure 3) et nous amène à confronter, partager et faire évoluer nos appareillages méthodologiques et instrumentaux, à l'occasion notamment d'ateliers techniques (autour du logiciel TULIP, du logiciel ManageChart pour la gestion de la base de données de métadonnées ou encore de la cartographie des emprises territoriales, par exemple).

Figure 3. Constitution des corpus et disciplines mobilisés pour l'étude des IDG



- 14 Enfin, la troisième dimension énoncée par A. Bénel, celle de la *discussion méthodologique*, se joue à l'occasion des débats et questionnements posés lors des séminaires transversaux, mais aussi des publications à venir en fin de projet.

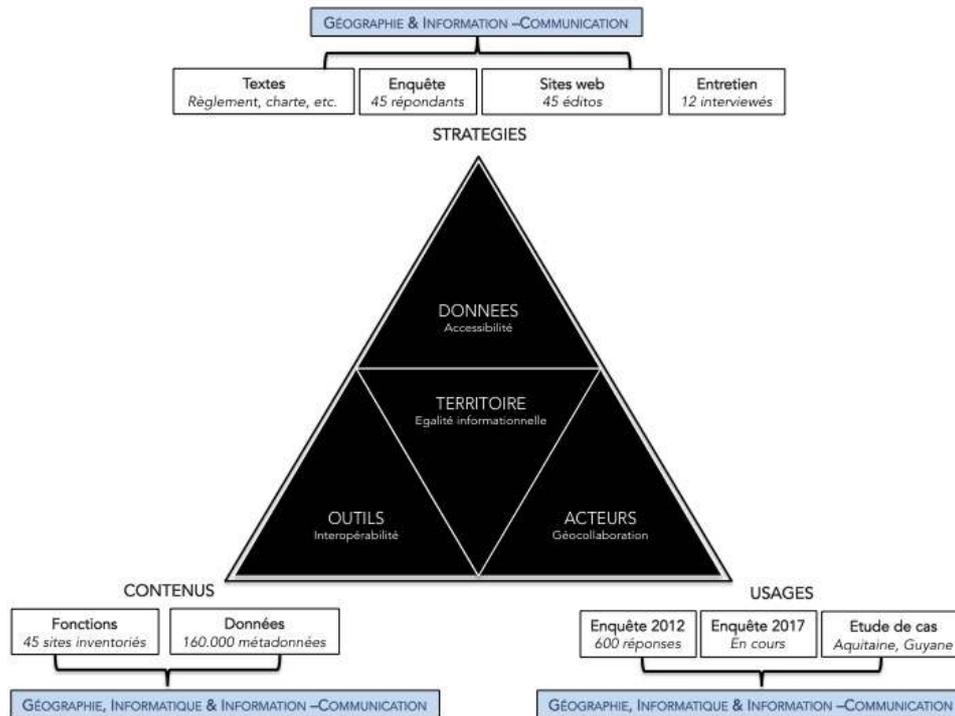
Recherche numérique

- 15 Jean-Christophe Plantin et Laurence Monnoyer-Smith distinguent trois types de redistribution de méthodes dans ce qu'ils appellent la « recherche numérique », soit « les recherches en sciences humaines et sociales qui portent sur des objets numériques ou numérisés, tout en ayant recours à des méthodes elles-mêmes numériques » (Plantin, Monnoyer-Smith, 2013, p. 42) : les *Digital Humanities* ; les *Cultural Analytics* ; et les *Digital Methods*. Toujours selon J-C Plantin et L. Monnoyer-Smith, les *Digital Humanities* s'appuient traditionnellement sur un croisement épistémologique et méthodologique entre éditorialisation de corpus en sciences humaines et sociales et méthodes informatisées d'analyses textuelles. Cette approche s'est élargie aujourd'hui jusqu'à prendre en compte les « modalités de création, de partage et de publication de données sur le web » (ibid., p. 47). Les *Cultural Analytics*, quant à elles, mettent l'accent sur l'analyse visuelle de productions culturelles en ligne, type de recherche dont Lev Manovich est l'un des fers de lance (Manovich, 2008, 2009). Enfin, les *Digital Methods* restreignent le domaine d'étude au web en interrogeant la façon dont les « objets nativement numériques » tels les liens

hypertextes, les sites web, les réseaux sociaux, etc. structurent, en tant que « medium », les pratiques de communication (Rogers, 2013). Bien entendu, cette présentation est trop concise pour ne pas être elliptique, et elle omet un certain nombre d'aspects, notamment en lien avec les définitions, débats et controverses qui animent le champ des *Digital Humanities*¹⁰. L'intérêt de la distinction opérée par J-C Plantin et L. Monnoyer-Smith est de resserrer le regard sur les méthodes associées aux recherches sur les environnements et données numériques.

- 16 À ce titre, et sur la base de la synthèse proposée, le projet GEOBS relève plutôt d'une approche orientée *Digital Methods*, car il aborde les IDG selon quatre dimensions : l'accessibilité des données géographiques, la mise en réseau des outils, la mise en réseau des acteurs et l'égalité informationnelle des territoires (figure 4). L'*accessibilité des données* a pour but d'évaluer la capacité effective des IDG à donner accès à des contenus réutilisables en identifiant les données téléchargeables et correctement documentées permettant leur mise en exploitation dans un Système d'Information Géographique (SIG). Le *décloisonnement des systèmes d'information* fait référence à la capacité des IDG à favoriser les échanges entre systèmes experts pour rompre avec les logiques en silos des années 90 (démultiplication de SIG isolés par métier et par territoire sans inter-connexion). La *mise en réseau des acteurs* est évaluée par l'existence de pratiques de géocollaboration au sein des IDG. Enfin, nous questionnons l'un des arguments récurrents de la justification de la mise en œuvre de ces plateformes, à savoir l'*égalité informationnelle des territoires*. Il sous-entend que, s'inscrivant dans une logique de service public, les plateformes institutionnelles couvrent de manière homogène leur territoire contrairement aux plateformes commerciales, ce qui est loin d'être le cas (Noucher *et al*, 2017). Ainsi, en abordant la déconstruction des IDG par le prisme des données, des systèmes d'information, des acteurs et des territoires, nous proposons une analyse critique d'un contexte national qui permette de confronter les discours théoriques aux contenus effectifs. Ce modèle d'analyse suppose le recours à des méthodologies mixtes.

Figure 4. Dimensions d'analyse, corpus et disciplines pour les 3 axes

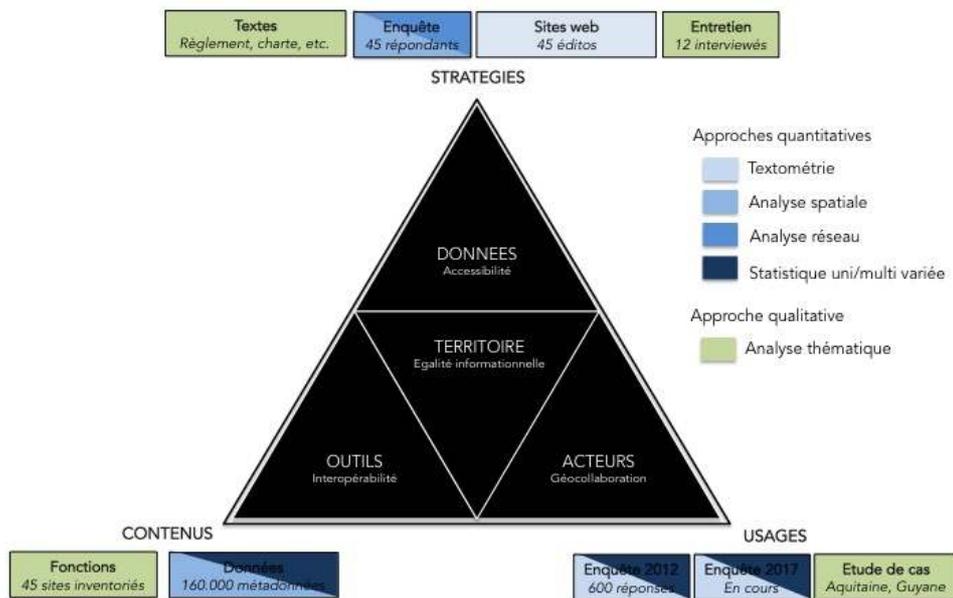


Méthodologies mixtes

- 17 Le décloisonnement et l'ouverture de jeux de données géographiques, inscrits dans des interfaces web, modifient les conditions de leur analyse. Mobilisant une logique interdisciplinaire, renouvelant les cadres de l'examen de l'objet, elle s'inscrit à ce titre dans une approche orientée « *mixed methods research* ». « *Mixed methods research [...] uses quantitative and qualitative research methods, either concurrently (i.e., independent of each other) or sequentially (e.g., findings from one approach inform the other), to understand a phenomenon of interest* » (Venkatesh et al., 2013, p. 23). Le recours conjoint, en synergie, à des méthodologies croisées permet de dépasser une opposition entre méthodes qualitatives et quantitatives, pour bénéficier des avantages des uns et des autres, constituant ainsi un troisième paradigme de recherche (Johnson et Onwuegbuzie, 2004). C'est également l'une des conditions permettant non seulement de comprendre et d'interpréter, mais aussi d'élaborer de nouveaux instruments méthodologiques, participant de facto à la construction de nouvelles modalités de sens.
- 18 À propos du phénomène de *l'open data* et de la massivité des données qu'il entraîne, D. Cardon parle de cet équilibre délicat et nécessaire entre « zoomer » et « dézoomer » (Cardon, 2014). C'est bien ce dont il s'agit dans le projet GEOBS, « zoomer » sur des éléments fins et individualisés d'analyse, permettant de dresser un portrait fouillé d'une IDG, « dézoomer » pour avoir une vision globale du « système IDG », permettant de faire émerger des traits saillants, par exemple en matière de flux, d'égalité territoriale en termes de métadonnées et de données ouvertes. À ce titre, la mobilisation croisée d'approches qualitatives (analyses de discours, lectures sémiotiques et stratégiques des dispositifs web) et quantitatives (analyses statistiques, spatiales et réseaux) s'impose pour

comprendre notre objet de recherche de façon dynamique et systémique. La confrontation entre ces différentes approches ne se fait pas de façon frontale : notre souci d'étudier ensemble les contenus, stratégies et usages en contextes numériques nécessite un entremêlement des méthodes où les analyses statistiques nourrissent les entretiens et inversement (figure 5).

Figure 5. Diversité des approches méthodologiques mobilisées pour traiter les corpus



Reconfiguration des pratiques du travail scientifique

Pratiquer une science ouverte...

- 19 La mobilisation de l'ensemble de ces corpus dont la plupart est issue de données directement accessibles en ligne, nous a conduit, en accord avec notre objet d'étude et du fait d'un consensus au sein de l'équipe de recherche, à adopter une démarche de type *open science*. Notre souhait n'était pas d'explorer seuls, enfermés en laboratoire, nos données mais, au contraire, tout au long du projet de les mettre à disposition pour susciter des réactions des promoteurs et usagers des infrastructures de données, d'autant plus qu'une partie de notre corpus (enquêtes et entretiens) leur est due. Notre approche part d'un double constat : d'une part, nous considérons que les acteurs des IDG sont les mieux placés pour évoquer leur pratique, d'autre part, de nombreux écarts peuvent exister entre leur discours et la réalité des pratiques souvent « invisibilisées » derrière des infrastructures qui s'apparentent à de véritables boîtes noires algorithmiques. Notre démarche repose donc sur la mise à disposition des matériaux de recherche collectés tout au long du projet. Ils mettent en visibilité les contenus, stratégies et usages pour amener les acteurs à adopter une posture réflexive au sujet de leur expérience pour en faire émerger les intentionnalités. Sachant que nous avons affaire à un public d'experts (les

professionnels de l'information géographique sont les principaux pourvoyeurs et utilisateurs des données des IDG), nous avons organisé la diffusion de nos résultats sous plusieurs formes relativement techniques :

- Production de livrables techniques : diffusion sur le dépôt institutionnel <https://halshs.archives-ouvertes.fr/> de rapports intermédiaires présentant les résultats des analyses statistiques ;
- Développement de (géo)visualisations interactives sur le site <http://geobs.cnrs.fr> permettant aux internautes d'explorer interactivement les données ;
- Mise en ligne des données sur le portail <http://www.data.gouv.fr/fr/>
- Dépôt de codes informatiques pour l'extraction (scrapping) et l'analyse des données sur la plateforme <https://github.com>

...pour susciter les interactions

- 20 Cette ouverture multiforme des données de la recherche vise à susciter les interactions avec les parties prenantes. En devenant des supports de discussion, les visualisations, livrables statistiques, données et autres rapports constituent des moyens privilégiés d'échange et d'expression qui facilitent les rencontres avec les acteurs impliqués dans les infrastructures de données étudiées. Ils permettent l'instauration d'une relation triangulaire (chercheur/parties prenantes de l'objet de recherche/données de la recherche) qui se substitue au face-à-face classique (chercheur/enquêté) et en amoindrit les éventuels effets inhibiteurs. Ces supports sont ainsi mobilisés lors des entretiens et permettent dès lors d'ancrer les propos sur des cas pratiques en fournissant un point de référence concret (Schwartz, 1989). La production d'un discours associé à l'analyse de document issu d'un travail scientifique permet alors à l'enquêté de collaborer étroitement à la recherche et l'incite même à s'y investir davantage encore, notamment lors de l'entretien mais parfois même aussi a posteriori. Ainsi, des contre-cartes ou contre-analyses statistiques ont pu être échangées avec différents enquêtés permettant de prolonger l'échange engagé lors de l'entretien. Ces supports multiples peuvent être assimilés à des « objets intermédiaires » au sens de Vinck (1999) qui relie les acteurs de la recherche entre eux.
- 21 Ces supports en ligne permettent aussi de reconnaître d'éventuels décalages entre le discours et les données, entre les usages affichés et les usages effectifs. L'identification de ces décalages peut alors permettre d'affiner l'analyse ou révéler des contradictions. Par exemple, dans le cadre du projet GEOBS, un institut de cartographie se présentant, lors de la première phase de l'entretien, comme le garant de la diffusion de référentiels homogènes sur son territoire a dû modifier son discours suite à la présentation de la couverture spatiale hétérogène de son patrimoine de données. Plus qu'une interprétation détaillée des supports proposés, c'est bien la mise en récit des usages qui est ici intéressante à analyser car elle permet de révéler les intentionnalités multiples des promoteurs et usagers des infrastructures de données.

Conclusion

- 22 Nous avons souhaité, dans cette contribution, « dézoomer » par rapport à une problématique purement centrée sur *l'open data* pour montrer comment l'information géographique voit ses cadres traditionnels bousculés par sa mise en publicisation sur le

web. Cette mise en visibilité et en circulation de données massives ne se fait pas sans repositionnements des différentes catégories d'acteurs du domaine, qu'ils soient étatiques, professionnels, civils ou scientifiques. Le regard des chercheurs sur les IDG, dispositifs sociotechniques hybrides, se doit dès lors d'agréger différents niveaux de lecture et de compréhension, de l'analyse de la réalité de l'*open data* sur différents terrains à la communication « transparente » des résultats, et ce, dans un dialogue permanent (et parfois délicat) avec les parties prenantes. L'ensemble suscite un véritable intérêt du côté des acteurs de terrain, qui ne va pas sans créer des effets de réflexivité et des demandes de justification, demandant d'autant plus de rigueur dans la production des résultats en général et sur l'origine des données en particulier, par exemple celles ayant permis la réalisation de visualisations ou l'analyse des stratégies. Si le décloisonnement des données géographiques s'inscrit dans un mouvement bien plus vaste de décloisonnement massif des données, illustré par la multiplication des portails open data (Labelle, Le Corf, 2013), il illustre aussi le décloisonnement du processus de recherche, tant en interne au niveau disciplinaire, qu'en externe avec les acteurs (institutionnels, professionnels, civils) en lien avec les dispositifs numériques étudiés. Il ne s'agit pas pour autant de céder aux sirènes de l'imaginaire vertueux, tendant vers l'impératif de transparence et d'ouverture, car le processus ne se fait pas sans « frictions » (Jaton, Vinck, 2016), sans remises en questions et difficultés, à différentes échelles. De la fabrication de la donnée géographique à sa mise en accès et « socialisation » numérique, nombre d'enjeux se jouent en permanence...

BIBLIOGRAPHIE

BÉNEL Aurélien, « Quelle interdisciplinarité pour les « humanités numériques » ? », *Les Cahiers du numérique*, vol. 10, n° 4, 2014, p. 103-132.

CARDON Dominique, « Zoomer ou dézoomer ? Les enjeux politiques des données ouvertes », dans Stiegler Bernard (sous la dir. de), *Digital Studies. Organologie des savoirs et technologies de la connaissance*, Limoges, FYP éditions, 2014, p. 79-94.

CHIGNARD Simon, *L'open data, comprendre l'ouverture des données publiques*, Limoges, Fyp Éditions, 2012.

COURBET Didier, « L'expérimentation en sciences de l'information et de la communication », dans Courbet Didier (sous la dir. de), *Objectiver l'humain ? Volume 2 : Communication et expérimentation*, Paris, Hermès-Lavoisier, 2011, p. 31-69.

CROMPVOETS Joep, BREGT A., RAJABIFARD Abbas, WILLIAMSON Ian, « Assessing the world wide developments of national spatial data clearinghouses », *International Journal of Geographical Information Science*, 18, 2004, p. 665-689.

DESBOIS Henri, « La carte et le territoire à l'ère numérique », *Socio*, [En ligne] 4, 2015. Disponibilité et accès <http://socio.revues.org/1262>

GAUTREAU Pierre et NOUCHER Matthieu, « Sharing Platform in Digital Geographic Information : Everything it promises ? », *Justice Spatiale/Spatial Justice*, n° 10, 2016. Disponibilité et accès <http://www.jssj.org/>

GAUTREAU Pierre et NOUCHER Matthieu, « Gouvernance informationnelle de l'environnement et partage en ligne des données publiques », *Netcom* [En ligne], 27-1/2, 2013 [consulté le 27 mars 2017]. Disponibilité et accès <http://netcom.revues.org/1234>

GEORIS-CREUSEVEAU Jade, *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*. Université de Bretagne Occidentale, 2014, 262 p. Disponibilité et accès <https://tel.archives-ouvertes.fr/tel-01092922v1>

GEORIS-CREUSEVEAU Jade, GOURMELON Françoise et CLARAMUNT Christophe, « Infrastructures de données géographiques : quelle contribution à la Gestion intégrée des zones côtières françaises ? », *Vertigo - la revue électronique en sciences de l'environnement* (En ligne), vol. 15, n° 1, mai 2015 [consulté le 31 janvier 2017]. Disponibilité et accès <http://vertigo.revues.org/15857>

GEORIS-CREUSEVEAU Jade, GOURMELON Françoise et CLARAMUNT Christophe, « A modelling framework for the study of Spatial Data Infrastructures applied to coastal management and planning », *Int. J. Geogr. Inf. Sci.*, 2017, 37, p. 122-138.

GIRAUD Pierre-Amiel, « Appartenir à des lieux d'une mouvance : le cas du Libre », dans Saule-Sorbé Hélène et Peylet Gérard (sous la dir. de), *L'appartenir en question*, Presses Universitaires de Bordeaux, coll. Eidolôn, 2014.

HARLEY John Brian, « Deconstructing the map », *Cartographica*, 26/2, 1990, p. 1-20.

JATON Florian et VINCK Dominique, « Processus frictionnels de mises en bases de données », *Revue d'anthropologie des connaissances*, vol. 10, n° 4, 2016, p. 489-504.

JOHNSON R. Burke et ONWUEGBUZIE J. Anthony, « Mixed methods research : A research paradigm whose time has come », *Educational researcher*, vol. 33, n° 7, 2004, p. 14-26.

LABELLE Sarah et LE CORF Jean-Baptiste, « Modalités de diffusion et processus documentaires, conditions du " détachement " des informations publiques. Analyse des discours législatifs et des portails open data territoriaux », *Les Enjeux de l'information et de la communication*, vol. 2, n° 13, 2012, p. 209-59.

LASCOUMES Pierre, « Gouverner par les cartes », *Genèses*, n° 68, 2007, p. 2-3.

MANOVICH Lev, *Cultural Analytics : Analysis and Visualisation of Large Cultural Data Sets*, Software Studies Initiative, 2008.

MANOVICH Lev, « Cultural Analytics : Visualizing Cultural Patterns in the Era of "More Media" », 2009. Disponibilité et accès http://manovich.net/content/04-projects/063-cultural-analytics-visualizing-cultural-patterns/60_article_2009.pdf

MASSER Ian, *Building European Spatial Data Infrastructures*, ESRI Press, Redlands, 2010, CA, USA.

MORILLON Laurent, « Recueil et analyse des données en sciences humaines et sociales : un panorama », dans Leleu-Merviel Sylvie (sous la dir. de), *Objectiver l'humain*, Paris, Hermes-Lavoisier, 2008, p. 31-65.

NICOLESCU Basarab, *La transdisciplinarité : manifeste*, Paris, Ed. du Rocher, 1996.

NOUCHER Matthieu, GOURMELON, Françoise, GAUTREAU Pierre, GEORIS-CREUSEVEAU Jade, MAULPOIX Adeline, PIERSON Julie, PINÈDE Nathalie, PISSOAT Olivier, ROUAN Matthias, « Spatial

Data Sharing : A Pilot Study of French SDIs », *ISPRS International Journal of Geo-Information*, 6 (4), 99, 2017.

NOUCHER Matthieu, *Les petites cartes du web. Approche critique des nouvelles fabriques cartographiques*, Paris, Éd. Rue d'Ulm, 2017.

NOUCHER Matthieu, « Infrastructures de données géographiques et flux d'information environnementale », *Netcom* [En ligne], 27-1/2, 2013 [consulté le 27 mars 2017]. Disponibilité et accès <http://netcom.revues.org/1404>

PLANTIN Jean-Christophe, *La cartographie numérique*, ISTME Editions, 2014.

PLANTIN Jean-Christophe et MONNOYER-SMITH Laurence, « Ouvrir la boîte à outils de la recherche numérique », *tic&société* [En ligne], vol. 7, n° 2, 2^{ème} semestre 2013, p. 38-66.

ROGERS Richard, *The end of the virtual : Digital methods* (Vol. 339). Amsterdam University Press, 2009.

ROGERS Richard, *Digital methods*, The MIT press, Cambridge, Massachussets, 2013, 274 p.

ROUAN Matthias et PIERSON Julie, « CSW Harvester : une chaîne d'extraction et d'analyse d'informations à partir des services web de catalogage (CSW) des infrastructures de données géographiques nationales et régionales en France », Workshop EXCES – Colloque SAGEO 2016, Déc. 2016, Nice, France.

SCHWARTZ Dona, « Visual ethnography : using photography in qualitative research », *Qualitative sociology*, vol. 12, n° 2, 1989, p. 119-154.

VENKATESH Viswanath, BROWN Susan A. et BALA Hilol, « Bridging the qualitative-quantitative divide : Guidelines for conducting mixed methods research in information systems », *MIS quarterly*, vol. 37, n° 1, 2013, p. 21-54.

VINCK Dominique, « Les objets intermédiaires dans les réseaux de coopération scientifique : contribution à la prise en compte des objets dans les dynamiques sociales », *Revue française de Sociologie*, vol. 40, n° 2, 1999, p. 385-414.

WOOD Denis et FELS John, *The power of maps*, New York, Guilford Press, 1992.

NOTES

1. Le géoweb peut se définir comme la généralisation du géoréférencement de l'information sur Internet. Il se traduit par une démultiplication de cartes (thématiques, statistiques, de localisation...) accélérée après le lancement de Google Maps en 2005 aux États-Unis, 2006 en France.
2. La Directive européenne 2007/2/CE impose à chaque État de l'UE la mise en œuvre d'infrastructures de données géographiques (IDG) pour faciliter l'accessibilité des données institutionnelles.
3. Ces résultats sont accessibles dans d'autres publications, notamment Noucher *et al.*, 2017.
4. À titre d'exemple, le catalogue de l'infrastructure institutionnelle de données géographiques de la région Aquitaine rassemblait 3 134 métadonnées en mars 2015 et 7 154 en mars 2017 (soit + 228 % en 2 ans).
5. L'adoption progressive de données dont les modèles sont standardisés et de systèmes d'information exploitant des architectures orientées services (SOA) rend les interconnexions, via le web, plus faciles.

6. On peut citer, au niveau européen, la Directive du 14 mars 2007 dite Directive INSPIRE qui oblige les administrations et collectivités de tous les pays de l'Union Européenne à mettre à disposition en consultation leurs métadonnées et les données utilisées ayant un caractère environnemental.

7. Un inventaire établi en 2015 sur 45 infrastructures de données géographiques françaises (d'échelon national et régional) met en évidence qu'au-delà des considérations techniques, ces dispositifs fédèrent des communautés de pratique : 132 groupes de travail ou pôles métier ont ainsi été recensés et sont particulièrement actifs en matière de géocollaboration (co-conception de nouveaux modèles de données, co-production ou co-contrôle de données, etc.).

8. C'est ce que note J. Chibois, dans un billet intitulé « Que signifie l'« open » de l'Open Data ? Du patch au hack, et de la publicisation à l'ouverture » paru sur : <http://laspic.hypotheses.org/1896>

9. Une chaîne d'extraction a été développée pour extraire, visualiser et analyser les 160 000 métadonnées (conforme à la norme de catalogage ISO 19115) récupérées sur les 37 géocatalogues opérationnels des 45 IDG étudiées (Rouan *et al.*, 2017).

10. Pour un panorama bien plus exhaustif, voir notamment le numéro spécial de la RFSIC consacré à ce sujet « Humanités numériques et sciences de l'information et de la communication », n° 16, 2016 (<https://rfsic.revues.org/1778>) ou encore le « Manifeste pour un positionnement des Sciences de l'Information Communication (SIC) vis-à-vis des *Digital Studies* (DS) et autres mutations du numérique » (<https://rfsic.revues.org/2543>)

RÉSUMÉS

Nous mobilisons ici l'expérience d'un projet de recherche en cours pour analyser la façon dont les nouveaux terrains d'expérimentations sur le web, modifient les conditions de la pratique scientifique, des objets aux méthodes, de l'*open data* à l'*open science*. La massification des données géographiques disponibles sur le web reconfigure les dynamiques de recherche selon trois axes de transformation : les objets, les méthodes et les pratiques de recherche. Tout d'abord, nous soulignerons comment les enjeux de pouvoir autour de la cartographie se sont déplacés avec l'avènement du web et de l'*open data*. Nous développerons ensuite les impacts en matière de méthodologie de recherche dans un contexte d'approche interdisciplinaire. Enfin, nous montrerons comment ce projet de recherche s'inscrit dans une démarche de type *open science*.

We are using here the experience of an ongoing research project to analyze the way in which the new experimental fields on the web modify the conditions of scientific practice, from objects to methods, from open data to open science. The massification of geographic data available on the web reconfigures research dynamics according to three axes of transformation : objects, methods and research practices. First, we will highlight how the power stakes around cartography have shifted with the advent of web and open data. We will then develop the impacts of research methodology in an interdisciplinary approach. Finally, we will show how this research project takes part of an open science approach.

INDEX

Keywords : geographic information, interdisciplinarity, data infrastructures, mixed methods, open data, open science, web

Mots-clés : information géographique, interdisciplinarité, infrastructures de données, méthodologies mixtes, open data, open science, web

AUTEURS

NATHALIE PINÈDE

Nathalie Pinède est maître de conférences en sciences de l'information et de la communication à l'université Bordeaux Montaigne. Elle effectue ses recherches au laboratoire MICA (Médiation, information, communication, art - EA 4426). Ses spécialités de recherche concernent l'analyse des sites web, notamment dans une perspective stratégique, les médiations hypertextuelles et les questions d'usage à partir des problématiques d'accessibilité numérique. Courriel : nathalie.pinede@u-bordeaux-montaigne.fr

MATTHIEU NOUCHER

Matthieu Noucher est chercheur au CNRS au Laboratoire Passages sur le campus de l'Université Bordeaux Montaigne. Géographe, spécialiste de la cartographie critique, ses recherches portent sur l'évolution des représentations de l'espace à travers l'analyse des nouvelles modalités de production, d'utilisation et de diffusion de l'information géographique numérique. Il s'intéresse en particulier aux liens entre la fabrique et la circulation des cartes (institutionnelles, commerciales, participatives, radicales, etc.), les différentes formes d'appropriation de l'espace et les rapports de pouvoir qui y sont associés. Courriel : matthieu.noucher@cnrs.fr

FRANÇOISE GOURMELON

Françoise Gourmelon est chercheuse en géographie au CNRS au sein de l'UMR LETG (Littoral, Environnement, Télédétection, Géomatique). Après avoir produit et utilisé de l'information géographique numérique pour répondre à diverses problématiques environnementales, elle s'intéresse maintenant à son utilisation en dehors de la sphère scientifique. Le rôle de l'information géographique et des technologies associées en tant que vecteur d'interdisciplinarité et de lien entre différents pans de la société est au cœur de son projet de recherche. Courriel : francoise.gourmelon@cnrs.fr

KAREL SOUMAGNAC-COLIN

Karel Soumagnac-Colin est maître de conférences en sciences de l'information et de la communication à l'université de Bordeaux. Elle mène ses recherches dans l'équipe RUDII (Représentations, Usages, Développements et Ingénieries de l'information) du laboratoire IMS - UMR CNRS 5218. Ses axes de recherche portent sur les usages et les pratiques informationnelles en contexte professionnel ainsi que sur les représentations des dispositifs numériques dans les organisations. Courriel : karel.soumagnac-colin@u-bordeaux.fr