

## Édito

de Patrice Bourdelais, Directeur de l'InSHS

Le changement d'année civile est toujours un moment propice aux bilans et aux perspectives pour l'année qui s'ouvre. 2016 a été, en particulier, l'année de la visite du

comité international d'évaluation du CNRS, en juillet [p2]

#### **NOUVELLES DE L'INSTITUT**

Bilan du test pilote VariSHS 2015 (Valorisation des Activités de Recherche des Ingénieur(e)s en SHS)

Mis en œuvre depuis 2010, l'outil RIBAC donne la possibilité aux chercheur.e.s en Sciences Humaines et Sociales (SHS) du CNRS de renseigner tous les ans l'ensemble de leurs activités qu'il s'agisse de recherche et d'enseignement, de communication ou de valorisation [p3]

### TROIS QUESTIONS A...

Cécilia Garcia Peñalosa, sur la fuite des cerveaux vers l'étranger

Directrice de recherche CNRS au sein du Groupement de Recherche en Économie Quantitative d'Aix-Marseille, Cécilia Garcia Peñalosa est membre de l'Ecole d'économie d'Aix-Marseille. Ses travaux de recherche portent notamment sur l'économie du développement, l'économie du travail et la macro-économie [p11]

#### **VALORISATION**

Le CVT Athéna, au service de la valorisation des sciences humaines et sociales. 3 ans d'actions déjà!

Le CVT Athéna (Consortium de Valorisation Thématique Athéna) a été créé par les pouvoirs publics dans le cadre des investissements d'avenir afin d'œuvrer dans le domaine des sciences humaines et sociales (SHS) [p21]

Denise Bernot ou la spécificité des études birmanes en France

Denise Bernot nous a quittés le 12 mai 2016 à Antony [p7]

#### **OUTILS DE LA RECHERCHE**

maladie. médecine prisme de la recherche en sciences humaines et sociales. Des propositions de financement innovantes

La recherche en SHS qui porte sur l'expérience de la maladie (...) concerne une science, la médecine, et des pratiques individuelles et collectives [p13]

#### **ENTRETIEN**

Les humanités numériques, fer de lance des sciences humaines et sociales. Rencontre avec l'équipe d'Huma-Num [p16]

#### ZOOM SUR...

Big data et information géographique

La nature de l'information scientifique pour les sciences sociales et les humanités a été profondément transformée par le développement des technologies numériques [p24]

#### UN CARNET À LA UNE

Le carnet de la MAFKF. Recherches archéologiques franco-koweïtiennes de l'île de Faïlaka (Koweït) [p38]

#### **À SIGNALER**

L'Institut de recherche stratégique de l'École militaire (IRSEM) recrute 5 chercheurs (Armement et économie de défense, Défense et société, Pensée stratégique) [...]

LIVRE



Mémoire vive. Chroniques d'un quartier, Bataclan 2015-2016, Sarah Gensburger, Anamosa. 2017

J'habite à mi-chemin entre la place de la République et la salle du Bataclan. Je suis sociologue au CNRS. Mes do-

maines de recherche sont la mémoire sociale et la santé et de la maladie : anthropologie crichez moi à 21 heures avec mon compagnon et mes enfants [...]

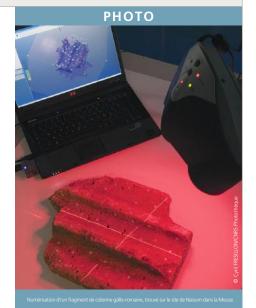
voir toutes les publications

#### REVUE



Anthropologie & Santé est une revue scientifique semestrielle, créée en 2010, qui se donne pour premier objectif de témoigner des recherches élaborées par les différentes tendances dans le champ de l'anthropologie de

sa localisation. Le 13 novembre, je suis rentrée tique des pratiques de soins, des systèmes de santé et des institutions ; anthropologie clinique ; anthropologie politique de la santé [...] voir toutes les revues





# Édito

de Patrice Bourdelais Directeur de l'InSHS

Le changement d'année civile est toujours un moment propice aux bilans et aux perspectives pour l'année qui s'ouvre. 2016 a été, en particulier, l'année de la visite du comité international d'évaluation du CNRS, en juillet. Elle avait été précédée d'une période de réflexion sur la politique de l'organisme et l'élaboration de nombreux documents de bilan et d'orientation. Le rapport de ce comité a été rendu au CNRS il y a quelques mois et il est à présent disponible sur le site de l'organisme. Une conférence de presse a permis d'en présenter les principales conclusions le 18 janvier 2017. Les principales remarques du comité portent sur le maintien d'une politique d'exigence de qualité internationale de la recherche co-pilotée par le CNRS, d'un meilleur financement des Très Grandes Infrastructures de Recherches (TGIR) et d'un meilleur accompagnement des jeunes chercheurs, ce qui a conduit à une recommandation globale aux pouvoirs publics d'augmentation substantielle du budget du CNRS.

Pour ce qui concerne les SHS, la politique de complémentarité par rapport aux Universités et aux grands établissements a été saluée, ainsi que les priorités définies notamment en direction de l'internationalisation et de l'interdisciplinarité. Le rôle des Maisons des Sciences de l'Homme (MSH) sur les sites a également été souligné tant pour ce qui concerne l'accès aux TGIR qu'à ce qui a trait aux transversalités et mutualisations encouragées à travers leurs programmes de recherche. Enfin, le comité s'est interrogé sur le sens d'un co-pilotage d'Unités Mixtes de Recherche (UMR) dans laquelle le CNRS compte moins de cinq agents. Il s'agit d'une question récurrente que nous connaissons bien et à laquelle la réponse ne peut être que circonstanciée. Les recommandations du Comité se sont traduites à l'InSHS par la décision d'allouer un crédit de 1 800 euros au laboratoire de tout(e) collègue recruté(e) en 2016, somme doublée par une contribution du laboratoire d'accueil. L'attention au suivi des MSH et aux possibilités de mutualisation des fonctions de support et de soutien, qui était déjà bien réelle depuis quelques années, sera renforcée. Enfin, le remplacement des personnels affectés aux unités d'adossement des TGIR constitue la première priorité de l'InSHS en ce qui concerne les ressources humaines.

L'année 2016 a aussi été marquée par la reconnaissance de l'importance de l'Information Scientifique et Technique dans nos domaines de recherche qui s'est traduite par l'attribution de l'une des quatre médailles de l'Innovation du CNRS 2016 à Marin Dacos, directeur du Centre pour l'édition électronique ouverte. 2017 sera, en outre, la première année d'application de la loi du 7 octobre 2016, « Pour une République numérique », favorisant le libre accès et ouvrant la voie à une véritable fouille des données.

L'année 2017 voit aussi le début des travaux du campus Condorcet qui va permettre de rassembler, après la deuxième phase, un potentiel de recherche considérable à l'échelle internationale en particulier sur trois grandes thématiques : les sciences de l'écrit ; les études aréales et les approches globales ; les sciences du territoire et l'étude des populations, spécialement des migrations. Le campus sera également un acteur majeur du basculement en cours vers les humanités numériques et la gestion des masses de données en SHS.

Notre investissement sur les grands sites à Paris et en région va aussi se poursuivre au gré des prochaines campagnes d'IDEX et d'ISITES et de l'approfondissement de notre pilotage sur les trois sites dont l'Idex a été confirmé (Université de Strasbourg, Université d'Aix-Marseille, Université de Bordeaux). Enfin, parce que la structuration de la recherche dans nos disciplines ne peut se satisfaire de la seule existence d'une douzaine de très grands ensembles, l'année 2017 sera aussi celle de la mise en place ou du renforcement de réseaux disciplinaires ou thématiques qui permettront à tous les collègues qui le souhaitent et n'appartiennent pas aux très grands centres, de participer à une dynamique scientifique forte, seule capable, dans une bonne articulation avec les grands sites, de développer des logiques d'ensemble vertueuses pour la qualité de la recherche française et sa participation aux grandes tendances mondiales.

2017 va donc nous permettre de développer plusieurs dossiers enthousiasmants et décisifs pour notre futur collectif. Dans une conjoncture difficile pour le pays mais où le CNRS a sauvegardé son budget global et sa capacité d'embaucher chercheurs et personnels de support et de soutien, je vous souhaite une excellente année faite d'initiatives nouvelles et de réussites insoupçonnées!

**Patrice Bourdelais,** Directeur de l'InSHS

## **NOUVELLES DE L'INSTITUT**

### Bilan du test pilote VariSHS 2015 (Valorisation des Activités de Recherche des Ingénieur(e)s en SHS)

Mis en œuvre depuis 2010, l'outil RIBAC donne la possibilité aux chercheur.e.s en Sciences Humaines et Sociales (SHS) du CNRS de renseigner tous les ans l'ensemble de leurs activités qu'il s'agisse de recherche et d'enseignement, de communication ou de valorisation. RIBAC permet une description la plus complète possible des productions et activités de la recherche en SHS. En 2012, pour la première fois, grâce aux données RIBAC, le rapport d'activités du CNRS a pu publier des données de productions scientifigues relatives aux unités de recherche de l'Institut des Sciences Humaines et Sociales (InSHS). Or, cet observatoire n'existe que pour les chercheur.e.s.

C'est pourquoi le Comité Scientifique et Technique RIBAC (CST RI-BAC) a préconisé la prise en compte des activités des ingénieur(e)s dans la construction des indicateurs de l'activité scientifique de l'InSHS et la mise en place d'un outil spécifique appelé VariSHS (Valorisation des Activités de Recherche des Ingénieur(e)s en SHS). En 2015, un test pilote a été lancé pour permettre d'évaluer si les rubriques actuelles du formulaire peuvent convenir pour répertorier les activités des ingénieur(e)s. Ce test pilote basé sur le volontariat a concerné, dans un premier temps, les 514 ingénieur(e)s (Ingénieur(e)s d'Etudes - IE et Ingénieur(e)s de Recherche - IR) des BAP D (Sciences Humaines et Sociales) et F (Information : Documentation, Culture, Communication, Edition, TICE) de l'InSHS. L'outil VariSHS a pour objectif la valorisation collective des activités des ingénieur(e)s de l'Institut dans toutes leurs richesse et variété.

Le lancement de ce test pilote a été accompagné de formations proposées non seulement par l'InSHS mais aussi au sein des laboratoires par les correspondants IST (Information Scientifique et Technique).

Dans l'ensemble, ce test pilote a été bien reçu. Le taux de participation, entièrement basé sur le volontariat, est bon puisque plus d'un ingénieur(e) sur deux s'est inscrit pour tester l'outil. Les critiques de fond ont été limitées et près d'un tiers des répondants a pris le temps d'expliquer leurs interrogations et propositions dans la rubrique Remarques / commentaires. La forte implication des ingénieur(e)s (IT) ressort aussi des échanges constructifs avec l'équipe VariSHS : plus de 250 courriels ont été échangés.

Les points évoqués dans la rubrique commentaires signalent l'impossibilité de recenser certaines activités non liées à la recherche ou encore des difficultés pour trouver où placer une activité dans le formulaire.

Lors de la présentation du projet VariSHS par Christine Kosmopoulos et Viviane Le Hay à la journée consacrée aux professionnels de l'IST en SHS du 7 décembre 2016, ces points ont été évoqués. En effet, VariSHS n'a pas vocation à recenser l'ensemble des tâches effectuées par les IT mais bien uniquement celles liées à la production scientifique. Ce recueil d'informations permettra à l'InSHS de disposer d'éléments objectifs pour valoriser les activités des IT et d'avoir une connaissance de la réalité de leurs activités. Il ne s'agit pas non plus de renseigner l'ensemble des

rubriques mais seulement celles qui concernent son activité. Les membres du CST VariSHS ont précisé qu'il y a plusieurs manières de saisir ses activités. Cela peut se faire de manière linéaire en suivant l'ordre des rubriques mais on peut également naviguer à l'intérieur des rubriques et constater qu'une part importante de l'activité des IT peut être répertoriée.

Certaines demandes des ingénieur(e)s ont été validées par le Comité Scientifique et Technique VariSHS pour faire évoluer l'outil pour la campagne 2016 :

- l'ajout d'informations manquantes dans la partie « identité » : ajout de la BAP (Branche d'activité professionnelle), de l'emploi type et de diplômes tels que Master, Licence, DUT, BTS;
- des modifications importantes dans la rubrique « productions multimédia » et des sous-rubriques, « images et sons », « sites web », « bases de données » et « cartographie » ;
- Des modifications dans la rubrique « activités éditoriales ».

#### Résultats du test pilote 2015

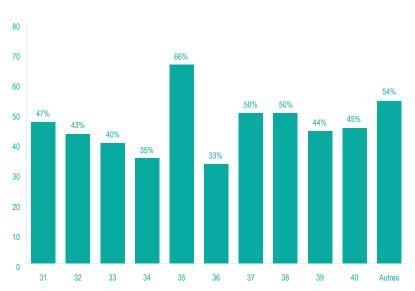
#### Participation au test

Le test pilote a été ouvert du 25 juin 2015 au 20 janvier 2016. Il a été adressé à 514 ingénieur(e)s d'étude et de recherche des BAP D (Sciences Humaines et Sociales) et F (Information : Documentation, Culture, Communication, Edition, TICE) de 193 unités SHS. 281 personnes de 133 unités se sont inscrites (54,7 %) et 226 personnes (44,0 % de répondants) ont renseigné leur dossier VariSHS.

Ce bon taux de participation permet de considérer les données obtenues comme pertinentes mais on observe des différences importantes selon les sections (Figure 1). Seulement un tiers des IT des laboratoires de linguistique (section 34) et de sociologie (section 36) ont répondu alors qu'ils sont deux tiers à avoir renseigné leur fiche VariSHS dans les laboratoires de la section 35 (sciences philosophiques et philologiques, sciences de l'art).

Il y a aussi des différences dans le taux de participation en fonction du grade (Figure 2).

Moins d'un tiers des Ingénieur(e)s d'étude hors classe (IEHC) ont répondu.



52% 49% 48% 48% 50 40% 40 30 20 10 0 IF1 IFHC IR2 IR1 IRHC IF2

60

Figure 1 : taux de participation des répondants par section du comité national

Figure 2 : taux de participation des répondants par grade (IE, Ingénieur(e) d'étude, IR, Ingénieur(e) de Recherche)

#### Domaines de compétences

Un des intérêts de la base est de permettre l'identification des compétences disciplinaires et des thématiques des IT. 170 ingénieur(e)s ont déclaré au moins une compétence. Ils pouvaient déclarer au maximum deux domaines de compétences.

Les trois compétences le plus représentées dans ce test sont l'archéologie (65 répondants), l'histoire (39 répondants) et l'art et l'histoire de l'art (27 répondants), (Figure 3).

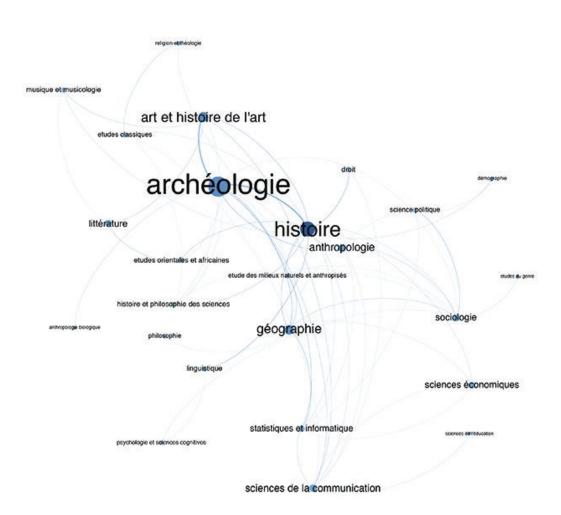


Figure 3 : Domaine de compétences des IT

	IT (nb publications)	% IT	% chercheurs
articles	66	33,7%	37,1%
chapitres	52	26,5%	29,6%
rapports	38	19,4%	7,3%
compte rendus	11	5,6%	6,5%
ouvrages	7	3,6%	3,8%
dir ouvrages	4	2,0%	3,7%
dir revues	8	4,1%	4,0%
notices	2	1,0%	3,6%
docs travail	1	0,5%	2,0%
traductions	3	1,5%	1,1%
HDR / Theses	2	1,0%	0,7%
éditions critiques	2	1,0%	0,4%
brevet	0	0,0%	0,0%

Tableau 1 : part des types de documents dans la production des IT et comparaison avec la production des chercheurs/chercheuses de l'InSHS (source VariSHS2015 et RIBAC2015).

		1 1	4.5	
$ \cup$	1.1	h	lication	-
			псанся	$\overline{}$

78 ingénieur(e)s (43 Ingénieur(e)s de recherche et 35 ingénieur(e) s d'étude) ont déclaré 196 documents publiés en 2015. La répartition de ces 196 publications par type de document est présentée dans le Tableau 1.

Elle est quasiment identique à celle des chercheur.e.s sauf pour le type de document « rapports » qui regroupe majoritairement les rapports de fouille des archéologues des sections 31 et 32. Si on compare la répartition par type de document à l'intérieur de la section 32, la part des publications sous forme de rapports représente environ 30 % pour les IT et est inférieure à 20 % pour les chercheur.e.s. Cette prédominance des rapports de fouille dans les publications des IT peut aussi résulter du fait que près de la moitié des publications (42,3 %) déclarées dans ce test ont été écrites par des ingénieur(e)s archéologues de la section 32 (Figure 4).

nbre auteurs par article	nbre d'articles CH	% articles CH nbre d'articles IT		% articles IT	
1	1177	50,6%	22	31,4%	
2	493	21,2%	20	28,6%	
3	265	11,4%	7	10,0%	
4	141	6,1%	4	5,7%	
5	81	3,5%	3	4,3%	
6	42	1,8%	3	4,3%	
7	33	1,4%	3	4,3%	
8	25	1,1%	2	2,9%	
9	9	0,4%	2	2,9%	
10	12	0,5%	1	1,4%	
>10	49	2,1%	3	4,3%	
	2327	100,0%	70	100,0%	

Tableau 2 : Répartition des articles de revue des chercheurs/chercheuses et IT en fonction du nombre d'auteurs par article.

Excepté pour les rapports de fouille, les usages et pratiques de publication entre chercheur.e.s et IT en SHS sont donc relativement proches.

Une différence dans les pratiques de publications entre les chercheur.e.s et les IT concerne le nombre d'auteurs par publication (Tableau 2).

Dans le domaine des SHS, plus de la moitié des articles de revue publiés par les chercheur.e.s est réalisée par un seul auteur (Tableau 2). Pour les IT, on retrouve cette prédominance de publication mono-auteur mais dans une moindre mesure puisque ces publications ne représentent qu'un peu plus de 30 % (31,4 %) pour les IT contre 50,6 % pour les chercheur.e.s.

Il n'y a pas de différences importantes entre les IE et les IR, les IR produisant 55 % des publications et les IE 45 %.

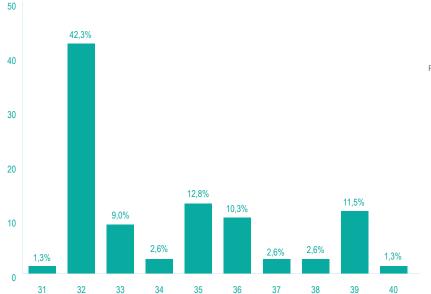


Figure 4 : Répartition des publications des IT en fonction de la section du comité national

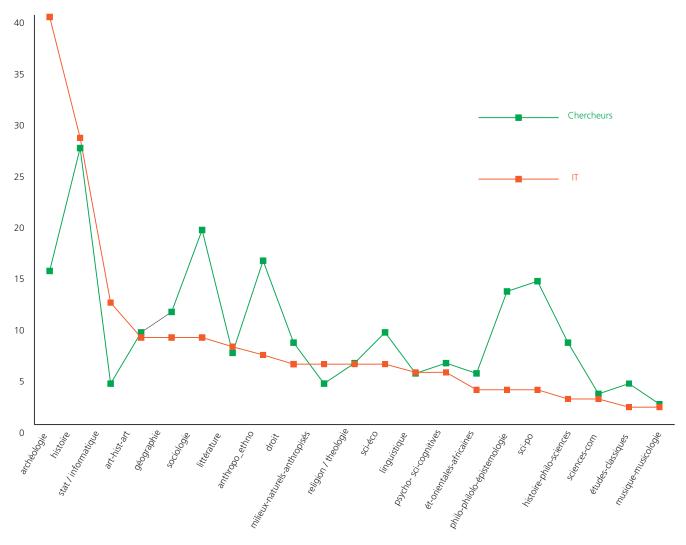


Figure 5 : Part des Publications (articles et chapitres) par discipline pour les chercheur.e.s (source RIBAC 2015) et les IT (source VariSHS 2015).

La répartition des publications (articles et chapitres) des Ingénieur(e)s (IT) et chercheur.e.s par discipline montre des spécificités disciplinaires des Ingénieur(e)s (IT). Les disciplines majoritaires de publication des IT sont l' « archéologie » (39,8 % des publications), « Stat/informatique » (28,0 %), « histoire » (11,9 %) et « arts et histoire de l'art » (8,5 %) (Figure 5). Il est cependant difficile d'aller plus loin dans l'interprétation de ces premiers résultats. En effet, il peut s'agir d'une vraie compétence spécifique des IT comme probablement dans le domaine « stat / informatique ». Pour l'archéologie, d'autres hypothèses peuvent être avancées comme une plus grande participation des IT aux publications (articles et chapitres) par rapport à d'autres disciplines SHS mais cela reste encore à confirmer.

Il faut attendre les résultats de la campagne VariSHS 2016 pour voir si ces tendances se confirment lors de l'étude des activités 2016 ou si les participants au test pilote variSHS 2015 ont un profil particulier.

#### Conclusion

Compte tenu de la forte participation au test pilote de 2015, il a été décidé de renouveler l'expérimentation VariSHS en 2016 et d'étendre cette phase pilote aux assistants ingénieur(e)s et à la BAP E (Informatique, statistiques et calcul scientifique) afin de

prendre en compte la diversité des contributions des IT aux activités de recherche des laboratoires SHS, d'observer la dynamique collective de production scientifique et d'être le plus exhaustif possible.

L'intégration des données des ingénieur(e)s dans les indicateurs de production scientifique peut se faire dès cette année pour les activités 2015 qui seront proposées pour le rapport d'activités du CNRS. Cette prise en compte des activités des ingénieur(e)s de même que celle des enseignants-chercheur.e.s représente un enjeu important non seulement pour valoriser leurs activités mais aussi pour produire des analyses et indicateurs au plus près de la réalité de la recherche dans les laboratoires SHS¹.

contact&info

► Michèle Dassa, InSHS
Michele.DASSA@cnrs-dir.fr

<sup>1.</sup> Je remercie Christine Kosmopoulos pour la relecture attentive de ce texte ainsi que Sylvie Lacaille pour la mise en conformité des données.

## **FOCUS**

### Denise Bernot ou la spécificité des études birmanes en France



Denise Bernot © Alexandra de Mersan

### Rappel biographique

Denise Bernot nous a guittés le 12 mai 2016 à Antony (France). Elle avait 94 ans. Diplômée de l'Ecole des chartes et de l'Institut national des langues et civilisations orientales (Inalco dit « Langues'O »), en Hindi, elle a effectué son premier terrain de recherche en linguistique dans la région des Chittagong Hill Tracts avec son époux Lucien Bernot<sup>1</sup> (ethnologue, titulaire de la chaire de sociographie de l'Asie du Sud-Est au Collège de France). C'était au temps où l'on parlait encore de Pakistan oriental, dans l'actuel Bangladesh... De ce premier voyage chez les Marmas, elle garda le goût du terrain et du travail en immersion.

Cette première expérience et les nombreux séjours en Birmanie qui suivirent lui permirent d'acquérir une connaissance hors pair du terrain, de son terrain et ce quels que soient les domaines : linguistique, littérature, histoire, botanique, ethnologie, philologie.

Un savoir éclectique qui lui permettait de traiter en profondeur et de façon transversale tous les sujets abordés. De sa formation d'archiviste, elle a en outre gardé une exceptionnelle rigueur de travail tout au long de sa carrière.

Entrée au CNRS, elle débuta sa carrière par le recensement des travaux ayant trait à la Birmanie dans les bibliothèques parisiennes, à commencer par la bibliothèque des Langues'O, pour laquelle elle mit au point son propre système de translittération du birman. Cette entreprise de longue haleine déboucha sur une impressionnante Bibliographie birmane<sup>2</sup> en plusieurs volumes.

Érudite, Denise Bernot possédait aussi une bibliothèque exceptionnelle que beaucoup d'entre nous ont fréquentée. De ses nombreuses missions en Birmanie, elle avait rapporté quantités de livres, pour elle mais aussi pour le fonds birman des Langues'O, auquel elle était particulièrement attachée, les sauvant ainsi d'une censure certaine et de la destruction des aléas climatiques en l'absence de politique de conservation dans la Birmanie d'alors.

Fondatrice de la chaire de birman aux Langues'O au début des

<sup>1.</sup> Pour une présentation de l'œuvre et de la vie de Lucien Bernot, voir : Toffin G. 1995, « Lucien Bernot (1919-1993) », in L'Homme, vol 35 :133,

p. 5-8. 2. Bernot D. 1968, *Bibliographie birmane - années 1950-1960*, CNRS Editions ; Bernot D. 1982-1984, *Bibliographie birmane, années 1960-1970* (partie méthodique : 2 volumes ; partie auteurs : 2 volumes), CNRS Editions.



Croquis d'une feuille de bauhinie (n.bot. césalpinacée). Extrait du cahier de terrain n° 'r9 (1976) © Archives Denise Bernot

années 1960, elle y assura l'enseignement de la langue et de la civilisation birmanes jusqu'en 1990, en créant elle-même les outils nécessaires. Elle se consacra alors sans compter aux études birmanes, à l'enseignement et aux étudiants.

À la fois linguiste, enseignant-chercheur, elle était aussi traductrice, ethnologue s'intéressant autant à la langue qu'aux savoirs et aux savoir-faire<sup>3</sup>.

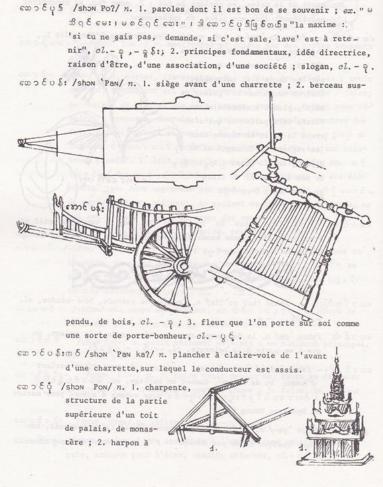
Forte de cette pluridisciplinarité, elle a su créer, avec son mari, une réelle communauté de chercheurs et de birmanologues en France. Elle était avant tout une dame de partage et de transmission<sup>4</sup>.

#### Son œuvre

Après un premier ouvrage sur les « Khyang » (Chin) de Chittagong, écrit en collaboration avec son mari et paru en 1958, Denise Bernot s'attela au recensement des travaux sur la Birmanie et à la constitution de sa *Bibliographie birmane* (1968-1984).

Puis, vint l'enseignement de la langue et la fabrication de ses outils de travail, un manuel d'apprentissage de la langue<sup>5</sup>, une grammaire<sup>6</sup> et bien sûr des dictionnaires. Car dans la communauté des birmanophones, Denise Bernot est surtout connue pour son extraordinaire dictionnaire bilingue (Birman-Français) en quinze volumes<sup>7</sup>. Ce dictionnaire a la particularité de suivre l'ordre alphabétique birman ce qui, bien qu'allant à l'encontre de la tradition occidentale, est cohérent avec la langue birmane. Parsemé de ses croquis exquis, très complet car couvrant des domaines divers (faune, flore, artisanat, gestes techniques, traditions populaires, etc.) hormis le domaine religieux, il est toujours un modèle à suivre — jusqu'en Birmanie — pour la réalisation des dictionnaires bilingues.

Les années qui suivent voient non seulement la publication de sa thèse *Le Prédicat en birman parlé*<sup>8</sup> en 1980, mais aussi la paru-



Croquis réalisé par D. Bernot de la partie avant d'une charrette Extrait du dictionnaire birman-français, vol. 5, p.135 © SELAF

tion de nombreux articles de linguistique, sur la phonologie du birman et sur des dialectes proches comme le marma, l'arakanais [arakanese], le tavoyen, le intha ou le maru<sup>9</sup>. Un travail comparatif de longue haleine, parallèle à son travail de traduction!

Même si Denise Bernot était incontestablement linguiste, comme l'attestent ses précieux travaux sur le lexique, la phonologie, ou encore la syntaxe de cette langue si différente de nos langues indo-européennes, elle était aussi une grande spécialiste de littérature. Elle écrivit sur la poésie birmane et publia plusieurs traductions de nouvelles contemporaines (2003)<sup>10</sup>.

#### Son héritage : une birmanologie de rigueur

L'héritage de Denise Bernot pourrait se résumer en deux points : la méthode et le terrain.

Sa méthode consiste à puiser dans tous les types de sources disponibles pour épuiser un sujet, à commencer par les sources en birman (érudites et populaires). Quant au terrain, c'est l'accès aux données dans leur contexte, la base de toute tentative de théorisation

<sup>3.</sup> Voir, à titre d'exemple, son savoureux article sur le bétel en Birmanie (1987) : Bernot D. 1987, « Epicurisme et nationalisme en Birmanie », De la voûte céleste au terroir, du jardin au foyer, Hommage à Lucien BERNOT, Editions de l'EHESS, pp.475-487.

<sup>4.</sup> Voir: Brac de la Perrière B. 2016, « In Memoriam, Denise Bernot (1922-2016) », in Moussons 28: 5-7

<sup>5.</sup> Bernot D., Cardinaud M-H, Yin Yin Myint M. 1990, Manuel de birman, langue de Myanmar, vol.1, L'Asiathèque.

<sup>6.</sup> Bernot D., Cardinaud M-H, Yin Yin Myint M. 2001, Manuel de Birman, Volume 2, Grammaire birmane, L'Asiathèque.

<sup>7.</sup> Bernot D. 1978-1992, *Dictionnaire birman-français*, Peeters Publishers.

<sup>8.</sup> Bernot D. 1980, Le prédicat en birman parlé, Peeters Publishers.

<sup>9.</sup> Une bibliographie complète de ses travaux est en cours. Pour des références complètes à ses travaux, notamment sur les dialectes birmans, voir la bibliographie publiée dans Pichard P., Robinne F. (dir.) 1998, « Etudes birmanes : en hommage à Denise Bernot », EFEO, pp.415-417.

<sup>10. 2003, «</sup> Le Rire de la terre Myé ga' yi çi ». Anthologie de nouvelles birmanes, Nouvelles présentées et traduites du birman par Denise Bernot, L'Asiathèque.



Illustration du peintre Bagyi Aung Soe. Extrait du magazine birman « Myawadi » © Yin Ker

Cette mise en avant du terrain est une qualité essentielle du travail de Denise Bernot et de son mari Lucien, que partagent d'autres collègues, chercheurs de renom, comme Georges Condominas (autre grand ethnologue de l'Asie du Sud-Est).

C'est ce que nous avons voulu mettre en avant par cet exercice d'admiration qu'a été, pour nous, l'élaboration du film « Denise Bernot, langues, savoirs et savoir-faire de Birmanie » (2015). C'est toujours ce qui nous tient à cœur dans le projet d'archivage du travail de terrain de Denise Bernot.

Ainsi, parallèlement au film réalisé en collaboration avec l'unité Cultures, Langues, Textes (UPS 2259, CNRS)11, nous avons commencé un travail d'archivage des données textuelles et sonores collectées par Denise Bernot en Birmanie entre 1952 et 2005. Recueillies avec le plus grand soin, ces données collectées sur la durée (50 ans) constituent la base empirique de la recherche de Denise Bernot. Elles concernent principalement le birman central, mais aussi d'autres langues et dialectes de Birmanie jusque-là non documentés, peu ou pas décrits. Ces documents linguistiques (enregistrements sonores, manuscrits, cahiers de terrain) constituent un fonds inestimable, essentiel pour la connaissance et le rayonnement des langues d'Asie du Sud-Est.

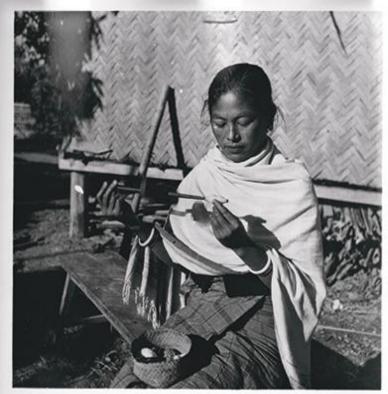
Dans le cadre du projet ArchiTeBir (ARCHIvage des données de TErrain sur la BIRmanie)<sup>12</sup>, les enregistrements sonores, dont les premiers datent de 1958, ont été numérisés, renseignés pour les documents concernant le birman standard. Certains d'entre eux sont déjà accessibles sur la plateforme d'archivage Cocoon, rassemblant des collections de corpus oraux numériques. Les cahiers de terrain et illustrations de Denise Bernot, eux aussi numérisés, seront progressivement intégrés à la plateforme d'archivage des données ethnologiques Asie-Pacifique ODSAS. À terme, un appariement des données textuelles et sonores mises en correspondance est prévu.

Ce projet s'inscrit dans une démarche plus globale de valorisation de la recherche de terrain en sciences humaines et sociales, en montrant d'une part à quoi ressemblaient les premiers terrains (fieldwork) linguistiques en Asie du Sud-Est et en donnant d'autre part l'accès à des enregistrements uniques de variétés de langue birmane et de langues de Birmanie, pays fermé pendant de nombreuses années.

<sup>11.</sup> Film : « Denise Bernot, Langues, Savoirs, Savoir-faire de Birmanie », sur une idée originale de Alice Vittrant et Alexandra de Mersan, réalisation Céline Ferlita & Maryline Leducq.

<sup>12.</sup> Le projet ArchiTerBir a bénéficié d'un financement du LabEx EFL (Empirical Foundations of Linguistics (ANR-10-LABX-0083) - Paris en 2014 pour la partie numérisation des données sonores, de financements du consortium Archives des ethnologues de la TGIR Huma-Num (2013-2016) pour la numérisation des données textuelles, puis d'un financement du LabEx ASLAN (ANR-10-LABX-0081) de l'Université de Lyon en 2016 pour le traitement des données (métadonnées, mise en ligne, etc).





Métier à tisser et travail du fil de coton dans les Chittagong Hill Tracts (1951-52) @ Archives Denise Bernot

#### **Conclusion**

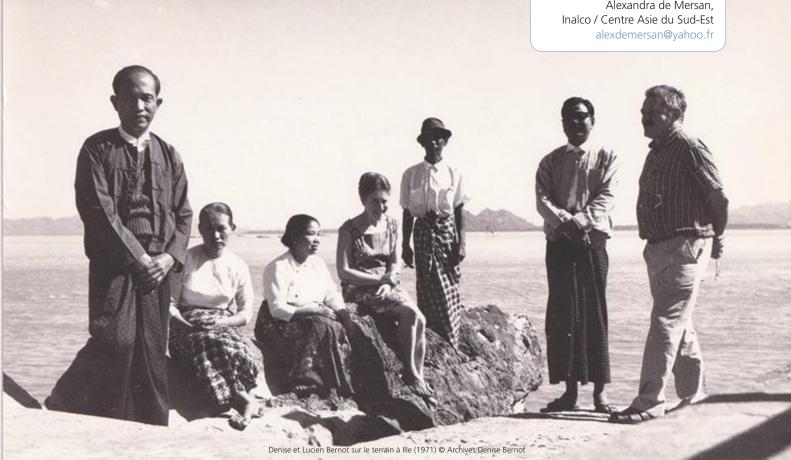
Les études birmanes en France se sont construites autour de deux figures exceptionnelles, Denise et son mari Lucien Bernot. Imbibés de connaissances disciplinaires solides et variées, nourris de langue et de culture, du terrain qu'ils ont arpenté toute leur vie, les Bernot ont produit, l'un et l'autre, des ouvrages exemplaires. Ces travaux, dépassant souvent les frontières des domaines disciplinaires, — à savoir la linguistique, la littérature, l'ethnologie, l'anthropologie — restent des modèles à suivre, surtout à l'heure du retour en grâce de l'interdisciplinarité. Nous ne pouvons que regretter que ceux-ci n'aient jamais été traduits en anglais, ce qui leur aurait donné le rayonnement qu'ils méritent, a fortiori à l'heure où la Birmanie connaît un regain d'intérêt depuis la

transition politique, et qu'un de leur terrain privilégié (la région de l'Arakan et de sa frontière) se trouve sous les feux de l'actualité.

Au cœur des études birmanes, il y a toujours eu Denise Bernot, sa maison d'Antony et sa bibliothèque ouverte à tous, sa rigueur dans le travail et sa grande disponibilité, sa volonté de transmettre son amour du terrain. Cette spécificité des études birmanes a sans conteste contribué au rayonnement actuel de la recherche française en sciences sociales sur la Birmanie.

### contact&info

▶ Alice Vittrant, AMU / Dynamique du langage alice.vittrant@cnrs.fr Alexandra de Mersan, Inalco / Centre Asie du Sud-Est alexdemersan@yahoo.fr



# TROIS QUESTIONS À...

### Cécilia Garcia Peñalosa, sur la fuite des cerveaux vers l'étranger

Directrice de recherche CNRS au sein du Groupement de Recherche en Économie Quantitative d'Aix-Marseille (GREQAM, UMR7316, CNRS / EHESS / AMU / Centrale Marseille), Cécilia Garcia Peñalosa est membre de l'Ecole d'économie d'Aix-Marseille. Ses travaux de recherche portent notamment sur l'économie du développement, l'économie du travail et la macro-économie. Pour le Conseil d'analyse économique (CAE), elle a produit en mai dernier, avec l'économiste Etienne Wasmer, une note sur la mobilité internationale des talents. Elle revient, pour l'InSHS, sur les principaux enjeux de cette mobilité.

Plus de trois millions de personnes nées en France vivent actuellement à l'étranger. L'expatriation de jeunes actifs aurait même tendance à augmenter ces dernières années. Quelle est la situation aujourd'hui en France et comment notre pays se positionne t-il parmi ses voisins européens ?

Effectivement, plus de trois millions de personnes nées en France vivent à l'étranger, selon les estimations de l'Insee. Parallèlement, les données de recensement indiquent que début 2013, 5,8 millions de résidents en France sont nés à l'étranger de nationalité étrangère, et que 39 % de ces personnes entrées en France en 2012 sont diplômés du supérieur. Le débat sur l'expatriation des Français a été en grand partie motivé par les évolutions récentes car notre pays connaît une forte augmentation des départs de jeunes actifs depuis trente ans, avec un taux qui a doublé, passant de 1 % à 2 % de la population. Mais ce chiffre reste très bas par rapport à ce qu'on voit chez nos voisins européens : en Allemagne, il est de 4 % environ et au Royaume-Uni, c'est 7 % de la population âgée de plus de 25 ans qui choisit de s'installer dans un autre pays depuis les années 90. En comparaison, la France exporte peu de travailleurs qualifiés. Pourtant, nous prévoyons des départs de plus en plus importants, car les expériences d'autres pays indiquent que les tendances que nous observons aujourd'hui vont probablement continuer dans les décennies à venir.

## Quel est le profil de ces actifs qui s'expatrient ? Fautil se réjouir ou bien s'inquiéter de cette mobilité ?

Ce sont ceux qu'on appelle les « talents ultra mobiles », majoritairement des scientifiques et artistes, ou encore des entrepreneurs. Les étudiants ayant suivi des formations techniques et les ingénieurs issus des grandes écoles sont surreprésentés dans les statistiques parce que la France en forme beaucoup, de très bon niveau, et qu'ils sont recherchés à l'étranger. Les écoles de commerce françaises placent aussi beaucoup d'étudiants à l'étranger. Dans les autres disciplines, on manque d'informations. Le système universitaire n'assure pas de suivi de ses étudiants et il n'y a aucune obligation pour un Français à l'étranger de se faire recenser.

Afin de mesurer la perte de « talents », on peut comparer la visibilité des scientifiques qui restent dans leur pays et de ceux qui partent, celle-ci étant mesurée à travers l'impact des publications. Les gagnants nets semblent être principalement le Royaume-Uni, les Etats-Unis et l'Espagne, les perdants nets des pays comme l'Italie. La France n'est, selon ces chiffres, ni gagnante ni perdante : la « visibilité » des scientifiques nés à l'étranger et travaillant en France est équivalente à celle des scientifiques nés en France et travaillant à l'étranger. Cependant, il y a des différences entre disciplines avec un déficit, par exemple, pour les physiciens. D'une façon plus générale, à partir des pages Wikipédia des personnes « notables », il apparaît que les personnes ayant quitté la France ont une « visibilité » plus grande que celles entrées en France. Ainsi, il semble que la France perd plus de « talents » qu'elle n'en reçoit, même si l'impact sur la productivité est difficile à évaluer.

En principe, la mobilité est positive à la fois pour l'individu et le pays. L'individu acquiert de nouvelles expériences, il rencontre des opportunités d'apprentissages, crée des contacts qui vont enrichir ses capacités professionnelles. Pour le pays, les études montrent que l'émigration des talents développe les liens commerciaux et favorise les exportations et la croissance. Une augmentation de 10 % du nombre de migrants entre deux pays stimule le commerce bilatéral de 1 %. C'est beaucoup!

En revanche, on constate déjà une perte nette de talents susceptible, à terme, d'avoir une incidence sur la croissance du pays. Même si ces départs pèsent encore peu sur le système de protection sociale français, il faut préparer l'avenir. À l'âge où, normalement, ils devront contribuer par leur travail à équilibrer les comptes sociaux et à rembourser le coût de leur formation initiale, ces individus qualifiés le feront dans d'autres pays développés, notamment anglo-saxons. Il n'est donc pas question de freiner les départs, mais de faire en sorte qu'ils soient remplacés. Il faut réfléchir à une adaptation de notre système de protection sociale à un monde mobile.

Quelles sont vos recommandations pour retenir ou faire revenir ces jeunes actifs ? Comment rendre notre pays suffisamment attractif pour attirer les jeunes étrangers qualifiés ? Quel rôle l'enseignement supérieur français peut-il jouer pour améliorer cette attractivité ?

D'abord, le système de retraites français représente un frein à la mobilité. La génération des 25-35 ans qualifiés veut bouger, changer d'emploi et de pays, acquérir des expériences variées. Or, le système français continue de fonctionner sur l'idée d'un emploi à vie dans le même domaine d'activités. Les salaires y dépendent beaucoup de l'âge, les systèmes de retraite ne sont pas les mêmes selon les secteurs, la mobilité entraîne une pénalité. Si vous travaillez une partie de votre carrière à l'étranger, votre retraite risque de ne pas être complète, vous aurez des pénalités importantes de retour en France, même si vous y avez cotisé pendant 20 ans. Il faut par conséquent renforcer les accords avec les pays qui, comme l'Allemagne, ont un système de retraite public. Mais cela ne suffit pas. Dans la note du CAE, nous proposons que la retraite puisse être calculée en proportion du nombre d'années travaillées en France. Faciliter les mobilités de courte durée est aussi important et un nouveau dispositif introduit en décembre 2015 va d'ailleurs dans cette direction. Un individu qui a travaillé moins de deux ans en France peut désormais demander à percevoir ses cotisations d'assurance vieillesse au moment de quitter le pays. Il n'abandonne pas derrière lui le montant de ses cotisations. C'est le type de politique qui favorise les retours/arrivées des individus qualifiés car ils peuvent passer une période en France sans que cette expérience ait un coût financier important pour eux.

En revanche, l'idée de quitter la France pour payer moins d'impôts est à mon avis dépassée. De nombreux facteurs contribuent à la

mobilité et la réflexion ne peut pas se limiter à un débat sur la baisse de l'impôt. D'ailleurs, un pays peut être attractif avec une imposition élevée — regardez la Suède! Ce qui fait la différence avec la France, c'est la visibilité. Le système fiscal en Suède est transparent et stable, on sait qu'on y paye des impôts en échange de services et d'une qualité de vie. Au contraire, le système fiscal français est complexe, peu lisible et changeant, ce qui crée de l'incertitude pour les salariés et ajoute une part de risque supplémentaire pour les entrepreneurs. La France communique peu sur ce qu'elle propose en échange des cotisations. Or, les jeunes professionnels ambitieux peuvent aussi fonder une famille, apprécier les modes de garde et chercher les écoles publiques de qualité qu'on trouve en France. Il devient fondamental de stabiliser le système fiscal et de faire comprendre que notre fiscalité élevée a une contrepartie qui peut être très attractive pour ces professionnels.

L'enseignement supérieur joue aussi un rôle important. La France est le troisième pays d'accueil en Europe pour les étudiants étrangers, mais ils sont peu nombreux à s'y installer ensuite pour travailler. Sur 100 jeunes étrangers qui font leurs études en France, nous n'en conservons que 4 comme travailleurs qualifiés, contre 23 aux Pays-Bas et 47 au Royaume-Uni. Deux voies s'ouvrent pour améliorer ce « taux de rétention ». D'abord, il faut renforcer la politique d'accueil de ces jeunes qualifiés. L'obtention d'un permis de séjour en France nécessite plusieurs passages en préfecture, des files d'attente. Des progrès ont été réalisés avec la création notamment du « passeport talent » qui vise à attirer les étrangers qualifiés en simplifiant les démarches mais il reste peu connu et il faut aller plus loin dans la simplification. Dans nombre de pays, tout va plus vite, un contrat de travail permet d'effectuer les démarches administratives en ligne avant d'arriver ou l'entreprise qui recrute peut s'en charger elle-même. De ce point de vue, la France n'est pas concurrentielle.

Ensuite, la France attire beaucoup d'étudiants étrangers du fait de la gratuité de son enseignement supérieur et non de la qualité de ses formations, ce qui n'est pas idéal. Former des étrangers et retenir ces jeunes qualifiés nécessite ainsi une politique d'éducation supérieure ambitieuse et sélective. La sélection est un aspect fondamental d'un système éducatif et plusieurs études ont montré — par exemple, dans des lycées en Israël — qu'elle améliore les résultats non seulement de ceux qui sont dans les institutions avec un niveau plus élevé mais aussi de ceux qui sont dans un lycée moins ambitieux. En effet, face à une population d'élèves plus homogène, les enseignants sont plus performants. Aujourd'hui, nous avons un système qui permet la sélection dans certaines formations du supérieur, par exemple les IUT, et l'empêche à l'Université. Ça n'a pas de sens !

L'amélioration de nos formations passe aussi par une augmentation des moyens. Dans le contexte des finances publiques actuel, il est difficile de voir comment les dotations publiques pourront augmenter. La solution anglaise a été de rendre les études supérieures payantes : l'individu paye sa formation et s'il décide de partir ensuite, cela n'a pas de conséquences sur le système. C'est aussi le modèle américain où les étudiants ont massivement recours à l'emprunt. Cela pose des problèmes d'inégalités d'accès à l'éducation et peut diminuer le niveau d'études général dans un pays. Des solutions sont envisageables dans un système majoritairement public. Il est possible de mettre en place une politique européenne de compensations entre les différents systèmes d'enseignement public. Les pays scandinaves ont déjà créé un tel système : le pays d'origine reverse un financement au pays qui

accueille. C'est un système vertueux car les universités ont intérêt à attirer des étudiants étrangers et donc à proposer des formations de qualité.

Une deuxième proposition est d'introduire, dans le système d'enseignement supérieur public, le principe de co-paiement différencié pour les étudiants hors Union européenne. Pourquoi le contribuable français devrait-il financer les études de personnes qui, en majorité, ne vont pas rester en France ? En payant leurs formations, ces étudiants étrangers contribueront à augmenter les ressources des universités, leur qualité et leur attractivité. Cela n'empêchera pas le Ministère des Affaires étrangères et du Développement international d'accorder des bourses aux étudiants de pays qu'il souhaite aider à se développer.

### contact&info

Cécilia Garcia Peñalosa, GREQAM

cecilia.garcia-penalosa@univ-amu.fr

Pour en savoir plus

http://www.cae-eco.fr/Preparer-la-France-a-la-mobilite-internationalecroissante-des-talents.html

## **OUTILS DE LA RECHERCHE**

### Santé, maladie, médecine au prisme de la recherche en sciences humaines et sociales. Des propositions de financement innovantes

La recherche en sciences humaines et sociales qui porte sur l'expérience de la maladie, les formes de prise en charge et les différentes compétences socio-professionnelles impliquées par celles-ci, les politiques de santé, concerne une science, la médecine, et des pratiques individuelles et collectives (de patients, proches, professionnels) qui ont une incidence directe sur la vie des personnes et leur état de santé.

Sans prendre nécessairement la forme de « recherche-action », cette recherche a d'entrée de jeu, potentiellement, une portée pratique significative. Depuis quelques années, elle a retenu l'attention de certains organismes qui ne sont pas, au premier chef, des opérateurs de la recherche, mais des organismes parapublics, des agences créées par les gouvernements, des associations dont l'objet est la santé humaine et la médecine

Trois exemples illustrent ici cet intérêt, tandis que l'entretien réalisé avec Martine Bungener (directrice de recherche émérite CNRS) permet d'éclairer l'évolution qui a conduit à ce type de financement, son sens et les enjeux qu'il recèle.

#### Quel est votre parcours en tant que chercheuse SHS dans les fondations/associations qui financent la recherche sur des thématiques de santé?

Mon parcours de chercheuse en SHS a été marqué par le fait qu'il s'est déroulé dans une unité mixte de recherche (UMR) entre le CNRS et l'Inserm. Première UMR de sciences humaines et sociales créée et pilotée entre ces deux institutions, le Centre de recherche Médecine, sciences, santé et société (Cermes) a été initié par la sociologue Claudine Herzlich en 1986. C'est une des rares unités en SHS labellisées par l'Inserm. Claudine Herzlich y a d'ailleurs été membre du conseil scientifique de l'Inserm avant de devenir conseillère pour les SHS sur les guestions de santé auprès de Philippe Lazar, alors directeur général de l'Inserm. Lui ayant succédé en tant que directrice du Cermes, j'ai été conduite à occuper des fonctions similaires tant au conseil scientifique qu'auprès de Claude Griscelli, puis de Christian Bréchot, qui ont successivement été à la tête de l'Inserm.

Le travail initial a été de persuader de l'intérêt et de la pertinence d'une recherche en sciences humaines. Celui-ci ayant été fécond, la demande interne et externe de recherche sur ces thèmes de santé et de médecine s'est renforcée et a très vite débordée les institutions parapubliques intéressées à les financer, comme le Comité d'organisation des recherches appliquées sur le développement économique et social (Cordes), le commissariat général au Plan, puis la Mission interministérielle entre les Ministères des affaires sociales et du travail (MiRe). Des appels d'offre ouverts aux SHS ont d'abord émané de mutuelles (MGEN) ou de la caisse nationale d'assurance-maladie (CNAMTS), puis d'initiatives gouvernementales ciblées sur des pathologies telles que le sida (Agence nationale de recherches sur le sida et les hépatites virales), le cancer et les maladies rares, les maladies chroniques, psychiques et mentales, ou encore le vieillissement. Ce sont enfin les grandes associations de maladies qui souhaitent soutenir la recherche en SHS. Pour faire fonctionner ces appels d'offre, écrire les appels à projets, évaluer les propositions des chercheurs, ma position d'interface m'a permis de participer à divers conseils scientifiques et, par mimétisme, a contribué à ouvrir cette possibilité à d'autres instances, par exemple à la Fondation Plan Alzheimer.

#### Plusieurs fondations/associations ont, ces derniers temps, développé des fonds de recherche consacrés aux SHS. Depuis quand et pourquoi?

Je crois qu'on assiste à un double mouvement. D'une part, on reconnaît davantage que les phénomènes de la maladie et de la santé, de leur conservation comme de leur prise charge, ont d'importantes dimensions sociales qu'il ne faut ni ignorer, ni négliger si on veut comprendre et intervenir dans ces domaines. La nature même des questions de santé, des formes dominantes de la maladie et des moyens de les prendre en charge s'est considérablement modifiée ces dernières décennies, avec la montée du vieillissement et des maladies chroniques, la transformation des techniques de soins, l'étendue des possibilités financières de recourir aux soins, même si persistent des exclus de certains soins. D'autre part, les travaux en SHS financés et reçus par les associations pionnières en ont intéressé d'autres concernées par des problématiques proches ; le vivier des chercheurs disponibles pour effectuer ces travaux s'est accru, ainsi que leurs compétences. On constate donc un phénomène de « boule de neige ».

Toutefois, il faut garder à l'esprit que ces chercheurs constituent encore un tout petit milieu. Les premières années, il est fréquent que certains de ces appels d'offre restent infructueux car les différentes pathologies concernées se font en quelque sorte « concurrence » entre elles. De plus, il existe pour les chercheurs en SHS un coût d'entrée important dans les spécificités de chacune des pathologies nouvelles sur lesquelles ils sont conduits à travailler au gré de ces nouveaux appels à projet. Comme je suis alors souvent amenée à le dire aux fondations face à la déception d'un premier appel d'offre ayant attiré peu de candidats, un délai de latence de deux, voire trois années, est donc tout à fait compréhensible. Il est nécessaire de persister au-delà pour voir arriver de bons projets! Le sida a été, pour les travaux en SHS, avant l'arrivée de soins efficaces, une maladie paradigmatique, au même titre que la maladie d'Alzheimer aujourd'hui.

#### Quels sont les grands enjeux vers lesquels devrait tendre, ces prochaines années, l'articulation entre recherche en SHS et expertise de ces fondations/ associations?

Le travail de recherche à concevoir avec ces associations et fondations doit se soucier du fait que leurs financements proviennent des dons de proches de malades et de citoyens concernées par la pathologie en cause. C'est un enjeu important dont il faut tenir compte. À ce titre, la parole de ces patients, de leurs proches doit être particulièrement entendue par les chercheurs, non seulement dans la conception et la formulation des questionnements de recherche prioritaires et des hypothèses à tester, dans l'élaboration des problématiques, dans l'établissement des méthodologies qui impliquent souvent de recueillir leurs expériences, leurs avis et opinions, mais aussi dans la conception même des résultats attendus et donc, au final, dans l'orientation même des projets proposés. Cela ne veut pas dire de renoncer à la liberté de sa recherche mais de se préoccuper du fait que les malades sont détenteurs d'un savoir propre sur leur maladie et d'une connaissance expérientielle qu'il est scientifiquement pertinent de considérer. De plus, dans un souci éthique et déontologique, il faut considérer que ces recherches demandent du temps et de l'implication de la part des personnes malades ou des proches dont le temps disponible est d'abord consacré aux soins. Ces recherches doivent donc avoir pour objectif d'être utiles pour eux ou pour le suivi ultérieur de leur maladie, et ne pas leur faire courir le risque de gâcher inutilement leur temps du fait de l'obtention de résultats déjà connus, inutiles ou inutilisables! On retrouve ici les principes essentiels de la promotion d'une recherche qui se veut co-produite entre

chercheurs et bénéficiaires potentiels, ceux qui sont ici les plus directement impactés par les résultats. C'est ce que demandent d'ailleurs vivement les associations de patients les plus concernées par la recherche, au point de vouloir la financer, mais aussi les procédures de la recherche participative ou collaborative qui peuvent servir de modèle paradigmatique pour ces travaux.

contact&info

Martine Bungener,
Cermes3
bungener@vjf.cnrs.fr



#### **Association France Parkinson**

L'association France Parkinson constate les impacts de la maladie sur les aspects psychologiques, sociaux, économiques, financiers de la vie du malade de Parkinson. Les recherches sont nombreuses, très actives et intenses pour comprendre et soigner la maladie de Parkinson d'un point de vue médical. Néanmoins, compte tenu d'une absence trop marquée des recherches sur les aspects sociaux, pourtant primordiaux à la vie de chacun, l'association a décidé de créer une commission dédiée aux sciences humaines et sociales.

#### Fonctionnement de la commission

La commission est composée de « spécialistes » sur certains sujets (neurologie, neuropsychologie par exemple), tous investis dans la prise en charge de la maladie de Parkinson, des malades de Parkinson eux-mêmes et des salariés investis sur cette cause.

Ils se réunissent pour déterminer les dossiers sélectionnés.

Tous les types de recherche sont acceptés : bourse de doctorat, subventions à projet, recherches ou dossiers d'entités privées ou de particuliers. À ce jour, les soutiens possibles sont compris entre 10 000 et 50 000 euros.

#### Les projets soutenus

Les dossiers de demande de bourses et subventions doivent concernés les thématiques suivantes :

- Malades (droits des personnes, statut et place, qualité de vie...);
- ► Aidants familiaux et professionnels ;
- ▶ Interventions psychosociales (évaluation d'interventions psychosociales, environnement et nouvelles technologies, outils...);
- ▶ Représentations sociales (représentations sociales de la maladie...);
- Contexte de la maladie (dispositifs, aspects socio-économiques...);
- Évaluation des actions non médicamenteuses ;
- ► Impact de la rééducation fonctionnelle ;
- ► Impact de l'activité physique.

De ces thématiques, 2 axes au choix des candidats :

- ▶ Un axe conceptuel étude selon plusieurs aspects (enjeux psychologiques, enjeux éthiques et juridiques, enjeux économiques et enjeux anthropologiques et sociologiques).
- Un axe innovations (utilité des approches psycho sociales, étude d'impact).

#### contact&info

► Florence Delamoye, f.delamoye@franceparkinson.fr

Pour en savoir plus

http://www.franceparkinson.fr/larecherche/commission-scienceshumaines-et-sociales/



#### **Fondation Médéric Alzheimer**

Depuis sa création en 1999, la Fondation Médéric Alzheimer, reconnue d'utilité publique, œuvre pour que la maladie d'Alzheimer soit abordée de manière globale et décloisonnée. C'est pourquoi elle soutient depuis plus de quinze ans la recherche en sciences humaines et sociales, qui permet d'analyser les conséquences de la maladie (qu'elles soient psychologiques, sociales, juridiques ou économiques), de concevoir de meilleures réponses d'accompagnement et d'envisager la maladie et les incapacités qu'elle occasionne dans toute leur complexité.

Afin d'encourager de jeunes chercheurs en SHS à consacrer leur thèse de doctorat aux handicaps cognitifs liés au vieillissement, la Fondation Médéric Alzheimer a attribué, depuis 2002, 46 bourses doctorales et 18 prix de thèse. Au cours de la même période, elle a également soutenu et accompagné 56 projets de recherche, ce qui lui a permis de constater que les chercheurs en SHS qui s'intéressent aujourd'hui à la maladie d'Alzheimer et au vieillissement cognitif sont nombreux, motivés à travailler de manière décloisonnée et qu'ils s'inscrivent dans une très grande variété de disciplines¹.

Forte de ce constat, la Fondation Médéric Alzheimer a souhaité constituer un réseau de recherche, réunissant des chercheurs de toutes les disciplines des SHS, ainsi que des professionnels et des opérateurs de l'accompagnement engagés dans des recherches appliquées). Et afin de créer une coalition de chercheurs, de professionnels, d'usagers et d'opérateurs, elle organise les 7 et 8 mars 2017, avec le soutien de la Fondation de France, les Assises de la recherche et de l'innovation sociale : pour relever le défi du vieillissement cognitif.

1. Gzil F. 2014, « Le point sur la recherche en sciences humaines et sociales sur la maladie d'Alzheimer », dans *Retraite et Société*, n°169: 159-173.

### contact&info

► Fabrice Gzil, gzil@med-alz.org

Pour en savoir plus

http://www.fondation-medericalzheimer.org/Nos-Actions/Soutien-a-la-recherche



### Agence de la biomédecine

L'Agence de la biomédecine est une agence publique d'Etat qui exerce ses missions dans les domaines du prélèvement et de la greffe d'organes, cellules et tissus, ainsi que dans les domaines de la procréation, de l'embryologie et de la génétique humaines.

Elle lance, chaque année, des appels d'offres « recherche et greffe » et « assistance médicale à la procréation, diagnostic prénatal et diagnostic génétique » ouverts non seulement aux projets de médecine et de biologie, mais également aux projets pluridisciplinaires et/ou relevant des sciences humaines et sociales

L'Agence accorde une attention particulière à ces derniers car ils permettent d'éclairer l'impact éthique et social des thérapeutiques et des innovations dans les domaines dont elle a la responsabilité. Elle constate cependant que les projets qu'elle finance dans ce domaine restent encore peu nombreux : « les sciences humaines et sociales apportent un regard nécessaire sur les innovations médicales relevant de l'Agence de la biomédecine », rappelle pourtant Anne Courrèges, directrice générale de l'Agence.

Outre le critère d'excellence, l'objectif de ces appels d'offres est de permettre à des projets d'accéder à des financements qui n'existent pas nécessairement par ailleurs. Il peut s'agir de sujets très spécifiques, comme l'amélioration de la conservation des greffons, mais aussi de projets portant, par exemple, sur les facteurs intervenant dans la décision du don ou sur l'exploration du vécu et des représentations sociales du don de spermatozoïdes.

La fenêtre de dépôt des dossiers s'étend chaque année d'octobre à novembre. Les dossiers sont évalués par deux experts extérieurs et sélectionnés par le comité médical et scientifique de l'Agence, puis soumis à la décision de la directrice générale de l'Agence. Les projets sont financés pour une durée allant de 1 à 3 ans et pour un montant maximal de 40 000 euros.

### contact&info

Nicolas Chatauret,
 nicolas.chatauret@biomedecine.fr
 ▶ Pour en savoir plus
 https://www.agence-biomedecine.fr/GT-recherche

## **ENTRETIEN**



### Les humanités numériques, fer de lance des sciences humaines et sociales. Rencontre avec l'équipe d'Huma-Num

La très grande infrastructure de recherche Huma-Num a organisé ses premières rencontres en octobre dernier sur le site de Valpré (Lyon). Ces rencontres ont réuni une centaine de participants et partenaires issus de communautés de recherche, d'institutions et d'infrastructures françaises, européennes et internationales afin de présenter les nouveaux services d'Huma-Num et d'échanger entre pairs. Olivier Baude et Stéphane Pouyllau, respectivement directeur et directeur technique de la TGIR Huma-Num, dressent un bilan de ces journées et confient leurs souhaits pour l'avenir.



#### Quels étaient les objectifs de ces journées et le public attendu ?

Les rencontres de la TGIR Huma-Num qui ont eu lieu à Valpré au mois d'octobre 2016 avaient pour principal objectif de réunir et de faire dialoguer entre eux l'ensemble des acteurs s'intéressant au numérique au sein de la communauté des sciences humaines et sociales: les porteurs de projet de recherche utilisant les services de la TGIR, les Maisons des Sciences de l'Homme (MSH),

les infrastructures européennes, les consortiums qui travaillent avec nous et nous aident à créer de nouveaux services et à mieux disséminer les bonnes pratiques et les enjeux du numérique.

Nous avons invité nos homologues européens à venir expliquer ce qu'ils font, à la fois dans le cadre des projets H2020 auxquels nous participons, mais aussi à travers un retour d'expériences des *European Research Infrastructure Consortium* (ERIC) Clarin et Dariah pour lesquels Huma-Num assure la coordination en France.

Nous avons réuni toutes ces communautés afin qu'elles échangent autour de questions méthodologiques, de la conception des outils, de la gestion des données numériques dans les processus de recherche.

L'ensemble des métiers de la recherche était présent lors de ces rencontres. Le public attendu était un public d'enseignants chercheurs, de chercheurs et de collègues ingénieurs spécialistes du traitement des données, documentalistes, bibliothécaires, archivistes. Comme il nous semblait important d'apporter un éclairage sur ce qu'il est possible de faire en matière de partage des données de la recherche, nous avions également convié des collègues juristes.

Nous avons réuni ces personnes autour de projets et de réalisations concrètes, il ne s'agissait pas de faire de grandes présentations institutionnelles, nous souhaitions aller directement au cœur des sujets et discuter de projets au centre de nos activités quotidiennes. C'est pourquoi nous avons voulu centrer les dis-

cussions sur des échanges de pratiques dans les SHS numériques. Nous avons notamment organisé un forum de posters présentant les réalisations des consortiums de la TGIR Huma-Num et des projets européens en cours de réalisation.

Pouvez-vous nous rappeler comment est structurée la TGIR Huma-Num, quel est son rôle et quelles sont ses spécificités ? Quels sont les services qu'elle offre et inversement ceux qu'elle ne propose pas ?

La TGIR Huma-Num est une très grande infrastructure de recherche, portée par une unité mixte de services associant le CNRS, Aix-Marseille Université et le Campus Condorcet. Elle est coordonnée par un comité de pilotage et accompagnée par un conseil scientifique qui joue un rôle essentiel. C'est en effet au sein de ce conseil — composé de chercheurs, de professionnels des bibliothèques, d'archivistes, d'universitaires — que se réfléchissent les grandes orientations de notre infrastructure. Sa mission première est de garantir une bonne adéquation entre les services que nous allons déployer et les besoins des communautés. Il évalue également chaque année et de manière quadriennale le travail de nos consortiums. C'est une étape indispensable pour nous permettre de travailler efficacement avec les communautés et de nourrir ainsi les services que nous offrons à tous. Dans cette perspective, le conseil scientifique fait des recommandations au comité de pilotage qui, quant à lui, joue son rôle exécutif.

Cette coordination entre le conseil scientifique, le comité de pilotage et la direction est un enjeu essentiel pour le bon fonctionnement de la TGIR Huma-Num. En effet, la volonté première de la TGIR n'est pas d'être une infrastructure *top-down* mais de proposer un dispositif qui s'alimente des attentes des chercheurs et répond aux besoins exprimés par les communautés de chercheurs.

Il ne faut pas oublier que nous travaillons pour l'ensemble des disciplines des SHS, nous sommes là pour accompagner l'ensemble des communautés. À ce titre, la TGIR est « agnostique », elle est neutre et ne privilégie pas telles ou telles disciplines ou tels ou tels cadres théoriques. Elle est à l'écoute des besoins et se saisit d'outils émanant d'une discipline particulièrement avancée pour les mettre à disposition d'autres communautés. Elle favorise ainsi une interaction basée sur l'entraide entre les différentes disciplines. Il s'agit de décloisonner les pratiques des disciplines tout en respectant les choix d'outils, parfois anciens, que certaines d'entre elles ont fait il y a plusieurs années.

Pour alimenter la réflexion sur les services dont les SHS ont besoin, nous animons des consortiums organisés de façon disciplinaire ou interdisciplinaire et ancrés sur tout le territoire à travers les Maisons des Sciences de l'Homme. Ces consortiums développent des programmes d'activité sur des périodes de quatre ans, dans lesquels ils mettent en commun des pratiques numériques, posent des questions sur les enjeux du numérique et alimentent la TGIR en réflexions afin que nos services collent aux besoins de la communauté de recherche.

Stocker la donnée, la documenter, la disséminer et la signaler : voilà concrètement comment nos services sont organisés. Au sein de la TGIR, nous proposons des services numériques qui vont du stockage des données numériques jusqu'à leur signalement dans les grands moteurs de recherche. Nous avons une palette de services numériques dédiée aux programmes de recherche en SHS. Ces services ne sont pas la juxtaposition de briques logicielles mises les unes à côté des autres. Nous avons en effet fait le choix de les organiser afin que cela corresponde aux besoins des différentes disciplines des SHS. Par exemple, sur les outils de stockage que nous offrons aux communautés, nous allons brancher des dispositifs de reconnaissance de caractères (OCR), avec des dictionnaires en grec ancien ou en latin qui correspondent à des besoins spécifiques de chercheurs en SHS. Nous pouvons également faire de la conversion audio/vidéo à la demande, directement dans les espaces de stockage. Ainsi, nous avons des outils pour créer des ontologies, des thésaurus, des bases de données, des corpus numériques en ligne, etc. Ces outils qui viennent d'ailleurs directement, pour la plupart, des communautés sont gratuits, comme l'ensemble des services de la TGIR.

Huma-Num n'est pas uniquement une infrastructure composée de services numériques destinés aux personnels IT. C'est avant tout une infrastructure dans laquelle les chercheurs peuvent s'impliquer et définir en concertation avec nous les services dont ils ont besoin, dans leur discipline ou dans des disciplines proches de la leur. Dans ce contexte, la TGIR Huma-Num n'est pas une agence de moyens. Nous ne finançons pas des programmes de recherche mais nous accueillons ceux qui ont besoin d'un environnement numérique pour se développer. Nous ne sommes pas non plus une agence d'intérim d'ingénieurs mis à disposition des chercheurs. Les équipes de recherche qui font appel à nos services et cherchent des informations sur les enjeux numériques viennent la plupart du temps avec leur équipe technique ou encore des prestataires qui travaillent de plus en plus au cœur des programmes de recherche. Notre rôle est d'accompagner le développement numérique de leur programme.

Quels sont les consortiums labellisés par la TGIR? Quelles sont leurs missions? Comment interagissent-ils entre eux et comment s'emparent-ils des nouvelles technologies pour proposer des concepts innovants?

Les consortiums de la TGIR sont actuellement au nombre de dix. Leur mission principale est de réfléchir et de mesurer les impacts que le numérique peut avoir sur les pratiques de recherche dans les différentes disciplines. Trois familles de consortiums cohabitent : ceux qui manipulent les données de la recherche au travers de plateformes de données, ceux qui s'intéressent aux transformations des objets eux-mêmes et ceux qui tissent un réseau de projets pour faire remonter des pratiques.

Parlons d'abord des consortiums interdisciplinaires organisés autour de plateformes de données numériques, en liaison avec les Maisons des Sciences de l'Homme. Ils réfléchissent à la fois à la façon de proposer des données de la recherche et à leur réutilisation possible dans les publications scientifiques et les projets de recherche.

Nous trouvons de ces consortiums en archéologie avec le consortium MASA ou en ethnologie avec le consortium Archives des ethnologues. Ce consortium propose par exemple une plateforme de transcriptions collaboratives des carnets de terrain en ethnologie et anthropologie. Il s'agit de la plateforme Transcrire. Il s'agit d'une expérience très intéressante qui s'appuie sur des outils standards — et c'est remarquable — venant du domaine des humanités numériques et que nous accueillons. L'intérêt de cette plateforme (et d'autres d'ailleurs), c'est d'être disponible pour d'autres programmes et d'autres disciplines des SHS.

D'autres types de consortiums réfléchissent aux transformations des objets modifiés par l'arrivée du numérique. Le consortium 3D s'intéresse, par exemple, à l'impact de la 3D dans toutes les branches des sciences humaines et sociales. En effet, l'usage de la 3D n'est pas propre qu'à l'archéologie ; on la retrouve également en histoire, en géographie, en architecture, etc. L'utilisation des drones pour réaliser des relevés tridimensionnels soulève une multitude de questions quant au volume de données que cela génère, à la pérennisation de ces données et à leur réutilisation dans un domaine dans lequel il y a une forte innovation et où il est complexe pour les chercheurs de s'y repérer. Le consortium 3D couvre un très grand nombre de laboratoires et d'équipes d'accueil qui manipulent des objets en 3D dans une optique de recherche scientifique et de valorisation patrimoniale, en lien avec des musées ou des organismes comme l'INRAP, par exemple. Ce type de consortium permet aux chercheurs de rester maitre dans la compréhension des enjeux et outils numériques, c'est-à-dire de construire une connaissance partagée en SHS sur ces questions complexes.

La dernière famille concerne les consortiums organisés sous forme de réseaux. C'est le cas du consortium Cahier qui a pour mission de fédérer un très grand nombre de projets autour des corpus d'auteurs et de les relier entre eux. Il s'agit en général de petits projets portés par une ou deux personnes que l'on met en relation pour qu'elles échangent sur les bonnes pratiques en matière de structuration de l'information, d'édition électronique, de corpus, de collections patrimoniales, etc.

Les consortiums ont pour mission de repérer quelles sont les transformations apportées par le numérique sur les données utilisées pour de la recherche. Ils doivent également réfléchir avec les communautés scientifiques aux enjeux de mutualisation et de pérennisation des plateformes de données. Enfin, ils ont un rôle d'animation et sont chargés de fédérer des équipes — qui ne sont pas forcément des grosses unités de recherche — et de les relier entre elles, de leur proposer des formations, etc.

Début 2016, les consortiums ont organisé une journée d'échanges inter-consortiums sur les archives numériques de la recherche. Cela nous semblait important de voir les consortiums collaborer entre eux et se poser des questions communes. Un focus particulier a été fait sur la question des problèmes éthiques et juridiques liés à la mise à disposition et à la réutilisation des données. Qu'est-ce que cela veut dire de réutiliser des données en archéologie, ethnologie, anthropologie, etc. ? Cette journée a aussi engendré des interrogations sur la mutualisation des plateformes de données. Elle a alimenté les réflexions autour des services proposés par Huma-Num, comme NAKALA qui est une

plateforme de stockage de données mise à disposition de toutes les communautés SHS.

Les rencontres de la TGIR Huma-Num ont permis de continuer le partage d'expériences. On a voulu confronter les travaux des consortiums à ceux de nos collègues des autres infrastructures européennes, leur donner la possibilité d'échanger sur des questions thématiques comme les enjeux du numérique en SHS, leur proposer des collaborations extérieures à la France, en lien avec les ERIC Clarin et Dariah, notamment.

La TGIR Huma-Num a une forte présence dans le paysage européen de la recherche. Elle est notamment impliquée dans l'infrastructure européenne Dariah dont les missions sont de développer et soutenir la recherche dans toutes les disciplines des sciences humaines et sociales. Pouvez-vous nous en dire plus ?

En effet, Huma-Num porte la participation de la France dans deux infrastructures européennes : Clarin, centrée sur les données de la langue et les données linguistiques et Dariah, centrée sur les humanités et les arts. Dans les deux cas, le rôle de la TGIR est d'impulser et de coordonner les contributions de la recherche française en SHS au sein de ces grands projets à moyens et longs termes.

Cette année, avec l'aide du conseil scientifique et du comité de pilotage, Huma-Num a proposé la réactualisation de la feuille de route de la participation de la France dans Dariah pour la période 2016-2019. Cette feuille de route, accessible sur notre site web en français et en anglais, se structure autour de trois ambitions. Il s'agit d'abord de constituer un vivier de projets en humanités numériques. Une enquête menée au printemps dernier auprès de toutes les universités françaises a permis de cartographier tous les projets en humanité numérique. Cette cartographie sera une contribution de la TGIR Huma-Num à Dariah.

Par ailleurs, nos consortiums vont proposer chacun une action phare en tant que contribution à Dariah.

Enfin, nous souhaitons proposer une coordination de l'ensemble des grands opérateurs français de l'IST dans la dimension spécifique des besoins en SHS (OpenEdition, CCSD, Persée, Huma-Num, etc.).

Aujourd'hui, la France au travers d'Huma-Num est un des fournisseurs les plus importants de l'infrastructure Dariah, à la fois en terme de services, mais aussi en termes d'interaction avec les communautés. L'idée de la feuille de route Huma-Num a été reprise par d'autres pays européens qui s'intéressent à la façon dont nous nous sommes structurés. N'oublions pas aussi que Dariah est une structure créée pour une vingtaine d'années. Les temporalités de mise en place ne sont pas les mêmes à l'échelon européen qu'à l'échelon national. Un des défis majeurs de Dariah est de gérer la complexité des modèles dans les différents pays européens. C'est dans ce contexte que le conseil scientifique de la TGIR Huma-Num a souhaité inviter récemment plusieurs pays membres de Dariah, lors de l'une de ses séances, pour partager de l'information et mieux connaître la façon dont certains d'entre eux implémentent Dariah chez eux.

Enfin, la TGIR Huma-Num a proposé qu'une personne recrutée par Dariah soit placée au sein de la TGIR afin de faire le lien entre l'ERIC et ce qui se fait dans les communautés scientifiques européennes.

Par ailleurs, Huma-Num participe à deux programmes financés par H2020. Le premier, *Parthenos*, est un projet européen qui réunit 17 pays de l'union autour de la question de l'interopérabilité, de la pérennité et de l'accès aux données numériques patrimoniales. Il s'interroge sur la manière de faire le lien entre des collections de musées et des collections de laboratoires et cherche à définir comment les chercheurs pourraient avoir accès de façon simultanée à ces objets là et ce dans toute l'Europe, avec des vocabulaires communs d'indexation, etc.

Le deuxième projet H2020, *Humanities at Scale* (HaS), a pour principale mission de doter Dariah de divers services venant de différents pays de l'union européenne impliqués dans l'ERIC et cela de la façon la plus distribuée possible. C'est le cas des outils développés par Huma-Num, mais aussi de ceux créés par OpenEdition (Hypotheses.org, Calenda) ou par le Centre pour la Communication Scientifique Directe – CCSD (Hal – Archives ouvertes).

## Existent-ils des modèles équivalents au niveau européen et international dont la TGIR pourrait s'inspirer ?

Il n'y a pas véritablement d'équivalent de la TGIR Huma-Num, tout simplement parce que les autres pays ne sont pas du tout structurés comme le notre.

Cela ne veut pas dire qu'il n'y a pas d'autres modèles. Il existe des infrastructures dédiées au numérique dans les sciences humaines et sociales, par exemple le DANS au Pays-Bas ou le LINDH de l'*Universidad Nacional de Educación a Distancia* (UNED) à Madrid. Mais les communautés de recherche sont organisées de manière très différente, il n'y a pas tout à fait le même mode de fonctionnement et d'organisation.

En Europe, nous sommes les seuls à proposer des services numériques conçus directement par les communautés SHS et directement adaptés avec elles. Même si les universités en Europe proposent un certain nombre de services, il ne s'agit pas de services intégrés à l'infrastructure, s'appuyant, pour exister, sur les communautés de recherche.

Au niveau international, le problème est un peu différent. Il existe des centres d'humanités numériques un peu partout dans le monde, comme celui de l'Université de Stanford. Ce sont des centres où l'on fait de la recherche sur les humanités numériques. Ces centres ne proposent pas forcément de services mais aident les chercheurs à réfléchir à la façon dont ils doivent eux-mêmes concevoir des services dans le cadre universitaire. Dans ces centres, on peut trouver plusieurs chercheurs qui aident les collègues à forger leurs outils. À l'Université de Montréal, la chaire Ecriture numérique, animée par le professeur de philosophie et de numérique Marcello Vitali-Rosati, est une petite équipe de recherche en humanités numériques qui propose de nouvelles façons d'éditer des textes scientifiques, pour dépasser le fonctionnement classique du simple article dans une revue.

Nous sommes actuellement en train de monter un partenariat au Canada avec la Cyberinfrastructure ouverte pour les sciences humaines et sociales (COSHS) qui est une infrastructure de recherche dédiée aux humanités numériques pour l'ensemble du Canada. Si Huma-Num n'est pas spécifiquement un modèle, elle intervient dans ce cas un peu comme un « conseiller spécial » afin de partager l'expérience que nous avons depuis 10 ans. C'est dans la perspective d'une collaboration privilégiée avec le Canada qu'a eu lieu le colloque *Infrastructures du numérique*, organisé à Toulouse en novembre 2015. Puis, lors des rencontres d'octobre dernier, nous avons fait venir Vincent Larivière, direc-

teur de la Cyberinfrastructure. Enfin, en novembre dernier, nous nous sommes rendus au Québec pour une semaine de travail technique avec les équipes de la Cyberinfrastructure, afin de développer des services en commun. Nous sommes ainsi en train de voir comment décliner ISIDORE et NAKALA avec leurs propres dispositifs. Nous avons un programme de travail sur les deux ans qui viennent où nous allons échanger, définir des services en commun entre la Cyberinfrastructure et Huma-Num et cela en relation avec les besoins que les chercheurs de nos deux pays peuvent avoir.

Cette Cyberinfrastructure est liée au consortium de la plateforme Erudit qui propose un accès numérique à la recherche et à la culture francophones d'Amérique du Nord en sciences humaines et sociales. Lors de notre séminaire de novembre dernier, nous avons créé la première chaîne d'enrichissement des textes d'Erudit, en s'appuyant sur le projet *Isidore à la demande* développé depuis un an par Huma-Num. Ce projet permet à des communautés de recherche d'utiliser les référentiels ISIDORE (ou les leurs) pour annoter eux-mêmes du texte intégral. Nous avons fait une chaîne de production complète en une semaine avec les équipes d'Erudit, à Montréal. Depuis, en France, un projet équivalent a été monté avec le CLEO pour les collections de livre d'OpenEdition.

Rappelons ici que, depuis 2015, ISIDORE est devenu multilingue. Il est capable d'indexer les documents en trois langues (français, anglais, espagnol), de proposer un enrichissement sémantique en trois langues, d'offrir des liens entre les documents en trois langues et de présenter une interface en trois langues. Ainsi, sur un document en français, une personne non francophone va pouvoir accéder à des concepts en anglais ou en espagnol et avoir ainsi des informations précises sur ce dont parle le docu-

ISIDORE est en train de devenir le moteur de recherche des sciences humaines et sociales internationales. De plus en plus de communautés de tous les pays du monde y mettent leurs données. C'est le cas par exemple de la plateforme brésilienne Scielo, ainsi que du réseau mexicain Redalyc. De plus, nos collèques Adeline Joffre et Nicolas Larrousse sont partis récemment présentés ISIDORE dans le cadre d'un colloque organisé par la bibliothèque nationale argentine. Peut-être une nouvelle collaboration en perspective.

Quel bilan pouvez-vous tirer de ces rencontres ? Quelles nouvelles pistes de réflexion ont-elles fait émerger ? De même, quelles suites envisagez-vous à ces journées ? Quels sont les prochains chantiers de la TGIR?

Les rencontres de la TGIR Huma-Num ont montré que les sciences humaines et sociales ont vraiment pris aujourd'hui le tournant du numérique. Les enjeux de stockage, de traitement, de signalement et de réutilisation de la donnée sont devenus des enjeux majeurs pour les SHS. Les programmes de recherche incluent de plus en plus souvent et de plus en plus tôt la gestion des données numériques et la dimension technologique dans leur phase préparatoire. Ces rencontres nous ont permis de réfléchir à la manière d'accompagner le mieux possible ces programmes de recherche dans leurs besoins numériques, en facilitant l'accès à nos services. Les données ne doivent pas être enchâssées dans des technologies fermées. Il y a tout un travail en amont à faire pour se préparer à cela.

En outre, ces rencontres nous ont permis de préparer une nouvelle phase de la TGIR, en anticipant sur les besoins des chercheurs en matière d'outils numériques, sur l'utilisation de ces outils et sur les enjeux éthiques et juridiques que cela engendre.

Les rencontres ont permis de mettre en lumière tous ces aspects là et de mieux relayer les actions de la TGIR en région en s'appuyant sur le réseau national des MSH (RnMSH). Des projets communs ont été initiés avec lui comme des formations, des ateliers d'échanges. Chaque année, nous organisons une journée dans 4 à 5 MSH pour discuter, présenter nos services et leurs évolutions. Nous sommes actuellement dans une dynamique importante qui doit accueillir à la fois de nouveaux programmes de recherche et des utilisateurs déjà anciens qui ont de plus en plus de besoins en matière d'outils et de services.

Comme il faut aussi penser aux chercheurs qui gravitent hors des réseaux MSH, consortiums, etc., Huma-Num va prochainement recruter une personne chargée spécifiquement de cette mission. Il s'agit de faciliter et d'accroitre l'appropriation des outils numériques que nous mettons à disposition. Cela sera pour nous l'enjeu de l'année 2017.

Nous entrons maintenant dans l'ère de la consolidation des dispositifs existants, consortiums et services. Aujourd'hui, Huma-Num couvre à peu près tous les besoins des communautés SHS. Nous continuons bien sûr à améliorer nos services et lancerons en 2017 la version 2 de NAKALA et la nouvelle interface d'ISI-DORE. De même, nous allons continuer à déployer la *Huma-Num* box dans les MSH afin de pouvoir proposer l'accès à un réseau de stockage de données distribuées. Huma-Num va également poursuivre le développement de ces formations et de son accompagnement aux chercheurs, améliorer son service de gestion de demandes, créer des guides de bonnes pratiques et des manuels pour mieux s'approprier les outils, proposer plus de logiciels permettant de faire des traitements sur l'audio, la vidéo, la 2D, la 3D. etc.

Nous souhaitons aussi continuer à faire de l'innovation numérique adaptée aux sciences humaines et sociales. Au sein d'Huma-Num, nous menons ainsi un petit programme de recherche et développement, principalement autour des enjeux du travail collaboratif sur les données numériques. Nous réfléchissons à la manière de réintroduire la notion de travail à plusieurs et en simultané autour d'objets numérisés, en 2D ou en 3D. Dans ce cadre, nous développons un projet qui s'appelle ISITable qui consiste à mettre à disposition des laboratoires et des bibliothèques de recherche une table tactile embarquant les services d'Huma-Num (ISIDORE dans un premier temps, puis sans doute NAKALA). Cette table tactile proposera des interfaces dédiées au travail en groupe afin de pouvoir analyser les données, poser des questions scientifiques communes autour d'un corpus, etc. Nous présenterons ces travaux lors de la prochaine édition d'Innovatives SHS, les 17 et 18 mai prochain à Marseille.

À l'heure d'une généralisation de la diffusion en accès ouvert, quelles seraient les recommandations de la TGIR sur la gestion des données de la recherche?

La principale recommandation que fait Huma-Num à toutes les communautés scientifiques tient en un mot : anticiper ! Plus les communautés de recherche viennent nous voir tôt, plus elles sollicitent nos services et plus nous pouvons anticiper la diffusion de bonnes pratiques et faciliter pour l'équipe de recherche la préparation des outils numériques nécessaires pour le traitement de ses informations.

Il est indispensable que les communautés SHS anticipent un maximum les enjeux de numérisation, de l'accès ouvert, de réutilisation, du stockage et d'archivage des données.

Une des grandes traditions des SHS est de travailler avec des corpus constitués de façon cumulative et de les analyser. Le numérique entraîne de nouvelles préoccupations dans la gestion de ces données, en particulier sur les enjeux de la conservation sur le long terme mais aussi sur les questions épistémologiques, éthiques et juridiques. Cela devrait donner lieu à la publication d'un guide de bonnes pratiques auquel plusieurs consortiums sont associés.

Anticiper les enjeux du numérique dans les programmes de recherche, mener une réflexion sur les objectifs poursuivis par le projet numérique, se poser les bonnes questions au bon

moment : voilà les préconisations que peut faire la TGIR Huma-Num qui est toujours à la disposition des chercheurs pour leur dresser un état de l'art des outils et services disponibles pour leurs programmes de recherche.

#### Propos recueillis par Armelle Leclerc

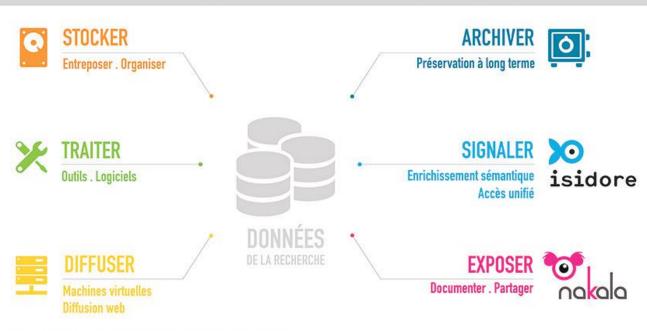
#### contact&info

➤ Olivier Baude olivier.baude@huma-num.fr Stéphane Pouyllau stephane.pouyllau@huma-num.fr

► Pour en savoir plus http://www.huma-num.fr

## SERVICES POUR LES DONNÉES NUMÉRIQUES

H



Partenariat avec le CC-IN2P3 le CINES, et le CCSC

## VALORISATION

## Le CVT Athéna, au service de la valorisation des sciences humaines et sociales 3 ans d'actions déjà!













Campus de Meudon-Bellevue, siège du CVT Athéna, 2016 © Camille Liégeois

#### Un acteur et son écosystème

Le CVT Athéna (Consortium de Valorisation Thématique Athéna) a été créé par les pouvoirs publics dans le cadre des investissements d'avenir afin d'œuvrer dans le domaine des sciences humaines et sociales (SHS) pour le compte du CNRS, de la Conférence des Grandes Ecoles (CGE), de la Conférence des Présidents d'Universités (CPU) et de l'Institut national des études démographiques (INED).

Les objectifs du CVT Athéna sont les suivants :

- ▶ Faire croître les liens entre la recherche en SHS et les acteurs économiques et sociaux ;
- ► Favoriser le dynamisme des échanges avec la sphère économique et sociale ;
- ➤ Communiquer en direction du monde socio-économique afin de faire connaître le potentiel appliqué des recherches effectuées dans les laboratoires en SHS ;
- ➤ Communiquer symétriquement en direction des laboratoires pour leur faire connaître les demandes en provenance des entreprises ;

▶ Identifier des projets innovants qui génèrent de l'activité économique, des emplois et de la croissance.

#### Missions du CVT Athéna

La fonction du CVT Athéna est d'agir concrètement sur le terrain afin de réaliser tout type d'actions de nature à favoriser la collaboration entre les laboratoires de recherche en SHS et le monde socio-économique et à jouer un rôle de catalyseur pour faire croître le nombre de contrats signés. Sa mission est donc clairement opérationnelle et passe notamment par :

## L'incitation et l'aide à la réalisation d'événements favorisant la valorisation

Différents moyens peuvent être utilisés pour communiquer sur l'offre des laboratoires et la mise en relation directe est quelquefois la plus efficace. Cependant, l'organisation d'évènements locaux permettant des rencontres entre laboratoires et entreprises est également un moyen d'accélérer les contacts. Il entre dans les missions du CVT Athéna de favoriser l'émergence de tels évé-



Couverture des Cahiers des Rencontres du CVT Athéna

nements et de faciliter leur organisation aussi bien sur le plan opérationnel que par le biais d'une aide financière limitée. En interne, le CVT Athéna a mis en place deux types d'événements : les « Rencontres du CVT Athéna » et les « Petits déjeuners du CVT Athéna ». Des chercheurs de diverses disciplines ont ainsi pu présenter des projets de recherche réalisés en partenariat avec des entreprises et exposer des projets en quête de partenaires capables de les financer.

## Le recensement et la mise en forme de l'offre des laboratoires

Les recherches accomplies dans les laboratoires sont la plupart du temps présentées d'un point de vue purement académique. Or, cette présentation n'est souvent pas compréhensible par les responsables du monde socio-économique et, surtout, les possibi-

lités d'application de ces recherches dans la résolution des problèmes que se posent les entreprises ne sont pas mises en avant, voire n'apparaissent pas du tout. Un travail de compréhension de la recherche effectuée, d'analyse de l'applicabilité de ces recherches et de traduction en offres pouvant intéresser les entreprises est donc indispensable en vue de les présenter aux entreprises. C'est ce que le CVT Athéna accomplit à travers la base de compétences qui a pour ambition de représenter de manière exhaustive l'ensemble des compétences valorisables issues des laboratoires de recherche en sciences humaines et sociales au niveau national. Actuellement, le moteur de recherche permet de filtrer les requêtes par des mots-clés, par zone géographique, par discipline ou encore par secteur d'application. Une commission interne a été mise en place afin de réfléchir à de nouvelles fonctionnalités pour rendre la base à la fois plus performante et plus ergonomique. À ce sujet, le CVT Athéna collabore avec des acteurs tels que les SATT et les services de valorisation des universités.

## La communication aussi bien en direction du monde socio-économique que des laboratoires

Les travaux menés dans nos laboratoires SHS sont très souvent méconnus des entreprises qui ne se rendent pas compte de la valeur ajoutée que pourrait leur apporter une collaboration avec les chercheurs pour résoudre beaucoup des problématiques qui se posent. Symétriquement, bien que la situation évolue favorablement ces dernières années, la culture des chercheurs SHS est encore quelquefois réticente à la valorisation et il est important de faire évoluer les états d'esprit pour les inciter à avancer dans cette direction. La vidéo « Parcours : des Sciences Humaines et Sociales à l'entreprise » a ainsi été réalisée afin de présenter des profils SHS ayant intégré aussi bien des grands groupes que des startups. Cette vidéo est à destination des entreprises comme des chercheurs ou même des doctorants SHS.



« Parcours : des Sciences Humaines et Sociales à l'entreprise », une vidéo qui présente des profils SHS intégrés à l'entreprise produite par le CVT Athéna, 2016 © Camille Liégeois

#### Orientations prévisionnelles

Le bilan global de l'action du CVT Athéna à ce jour est extrêmement positif. Depuis sa création, il a tout d'abord contribué à faire « bouger » les choses dans le domaine de la valorisation des SHS. A travers plus d'une quarantaine d'événements et par les contacts sur le terrain qu'il a organisés, il a d'une certaine manière dynamisé les structures locales existantes, dont certaines d'entre elles ne prenaient sans doute pas la mesure complète du challenge qui est posé pour inciter les laboratoires SHS à plus collaborer avec les entreprises. À travers le processus de constitution de la base de compétences, il a aussi mobilisé des acteurs et des chercheurs qui ont plus clairement pris conscience des enjeux.

Le CVT Athéna est aujourd'hui reconnu comme un acteur incontournable de la valorisation dans le domaine SHS. L'expérience de ces 3 ans d'activité, a permis de prendre conscience que c'est à travers le type d'actions qui a été mené que le CVT Athéna pourra jouer son rôle en étant un catalyseur de la valorisation des recherches SHS en France et que les résultats se concrétiseront de manière exponentielle en capitalisant sur plusieurs années.

La feuille de route du CVT Athéna pour ces prochaines années est claire :

- ▶ Organiser des événements locaux thématiques de rencontres permettant de susciter des liens entre laboratoires et entreprises débouchant sur des collaborations contractuelles ;
- Développer les contacts directs entreprises / laboratoires et jouer un rôle dans la médiation entre les deux mondes ;
- Développer un réseau de plus en plus dense d'entreprises par-
- Accroitre la communication en direction des deux mondes pour leur faire comprendre tout l'intérêt qu'ils ont de part et d'autre à collaborer;
- Continuer à développer la base de compétences pour la rendre la plus exhaustive possible et la faire connaitre comme un outil de présentation de l'offre des laboratoires auprès des entreprises ;
- Ètre au service de la communauté des laboratoires pour les aider dans leurs démarches et faire évoluer favorablement les relations avec le monde socio-économique.
- ► Abonnez-vous à la newsletter du CVT Athéna
- Retrouvez le CVT Athéna sur les réseaux sociaux :







#### Les actions du CVT Athéna en chiffres

- > 5 appels à projets ont été lancés à ce jour :

- ▶ 2 « Rencontres du CVT Athéna » se sont déjà déroulées au Ministère de l'Education nationale, de l'Enseignement supérieur et de la Recherche rassemblant plus de **200 participants** dans l'objectif de nouer des relations entre chercheurs en SHS et monde socio-économique. Ces rencontres ont donné lieu à la publication de **2 cahiers** qui en retranscrivent les échanges.
- ▶ 4 « Petits déjeuners du CVT Athéna » ont été organisés,
- Le CVT Athéna a participé à **15 événements** pour représenter la recherche en sciences humaines et sociales auprès du monde
- ▶ 5 formations « Valorisation des SHS dans les projets d'inno-
- Plus de 2000 fiches de compétences ont été rédigées, appor-
- ▶ 1 vidéo produite par le CVT Athéna « Parcours : des Sciences

### contact&info

► Hervé Zwirn. CVT Athéna

Herve.zwirn@cnrs.fr cvt-athena@cnrs.fr

Pour en savoir plus http://www.cvt-athena.fr

## ZOOM SUR...

### Big data et information géographique

La nature de l'information scientifique pour les sciences sociales et les humanités a été profondément transformée par le développement des technologies numériques, par la place qu'elles occupent désormais dans tous les compartiments de la vie quotidienne et par les gigantesques flux de données qu'elles génèrent. S'agissant de l'information géographique et de la manière de la mobiliser pour comprendre les espaces contemporains et les processus qui les structurent, la transformation est encore plus massive avec le développement du rôle joué par les systèmes d'information spatialisés et la diversité des formes de représentation numérique qui leur est associée.

La géographie et toutes les sciences sociales qui ont part à l'analyse des interactions espaces/sociétés sont en effet confrontées, comme de nombreuses autres sciences, à l'augmentation vertigineuse de l'information disponible. Le temps semble déjà lointain où il s'agissait seulement de rechercher et de construire l'information à partir d'un nombre limité de bases de données connues. Au data mining a fait place le data deluge. La géolocalisation et la généralisation de l'usage d'appareils connectés a transformé chaque utilisateur en un collecteur de données muni de plusieurs capteurs potentiels. Les réseaux sociaux génèrent des flux d'information dont l'usage apparaît pertinent pour de nombreux secteurs de la géographie et de l'analyse des territoires : modes de transports, géographie de la santé, géographie économique, flux illégaux de marchandises, inégalités... Face à ces flux de données pour la maîtrise desquels il faut inventer de nouveaux outils, l'enjeu pour l'analyse géographique est d'abord de savoir faire face à la quantité mais surtout de le faire au bénéfice d'une meilleure compréhension du monde qui nous entoure, et d'une meilleure maîtrise par les personnes concernées de la circulation et de l'usage des données.

Face à cet enjeu, les recherches menées dans le champ de l'information géographique et de l'analyse des espaces et des sociétés et dont les contributions réunies dans ce dossier donnent de passionnants exemples, se font en développant des interfaces fécondes autour de la modélisation et de l'analyse et de la représentation des systèmes complexes. Ces recherches interdisciplinaires sont menées souvent en étroite collaboration avec l'informatique et différentes ingénieries, notamment dans le domaine des risques, des mobilités et des dynamiques urbaines. Dans ces dialogues, les spécialistes de l'analyse géographique apportent des savoir-faire et une grande inventivité dans le domaine de la géomatique et de la représentation des données géographiques. La carte et l'atlas restent des outils plus que jamais pertinents à l'ère du big data mais ils se déploient dans une toute autre dimension. Des portails ouverts à tous publics permettent par exemple de représenter des phénomènes à une résolution spatiale et à une fréquence temporelle de plus en plus élevée. Au-delà des usagers experts, tels que les enseignants, les gestionnaires de l'espace, les personnes en charge de l'action publique, les agriculteur.trices, les militaires et les forces de l'ordre, l'ensemble des habitants connectés peut enrichir la connaissance de son milieu de vie en se connectant à des portails mis en place par les agences publiques. L'agence pour la protection de l'environnement aux États-Unis a ainsi mis en place un portail dédié aux inégalités environnementales qui permet de connaître les propriétés de son espace de vie sous le rapport des nuisances environnementales, afin d'exiger, au besoin, que la législation soit bien respectée. Plus généralement, les *big data* ouvrent l'espoir d'une amélioration de tous les services aux habitants des territoires notamment dans le cadre des villes intelligentes.

Parmi les horizons de recherche ouverts par les donnes massives, un des premiers est la recherche de dispositifs d'analyse et de modélisation parcimonieux. Modéliser, simuler et visualiser des interactions socio-spatiales fait croître considérablement la quantité des données à analyser. Il en résulte un besoin croissant en moyens technologiques d'analyse et de traitement de ces données qui stimule la créativité des modélisateurs. Il faudra en outre être capable de décrire mais aussi d'expliquer : les big data devront être maîtrisées pour mieux comprendre les processus sociospatiaux. Un autre défi est celui de la collaboration du public dans des études à caractère scientifique. Même si la plupart des smartphones contiennent des capteurs utiles pour mesurer de nombreux comportements spatiaux, leur collecte et leur utilisation par les scientifiques dans des actions de crowdsourcing posent des questions qui relèvent des relations entre science et société : comment concevoir des protocoles d'acquisition de données qui tiennent compte de la réticence de participants à faire connaître leur localisation malgré un encadrement juridique qui garantit l'anonymat ? Comment engager dans la durée des participants bénévoles au sein d'études longitudinales, notamment avec le risque de l'exploitation d'un travail gratuit ? À quel point faut-il intégrer des non spécialistes dans l'interprétation et la discussion des résultats?

Les big data et leur usage pour la compréhension des interactions espaces-sociétés posent aux disciplines des SHS et à leurs partenaires des défis techniques, méthodologiques, épistémologiques et éthiques passionnants. Les nouveaux horizons ouverts à la recherche obligent à renforcer encore la réflexivité dans la pratique de la recherche, en particulier au sujet du potentiel offert par l'intensification des interactions entre chercheur.e.s et habitants-usagers de l'espace et des territoires pour la production de connaissances, et de nouvelles formes de participation de tou.te.s aux décisions.

Pascal Marty, InSHS

### Données massives et information géographique

Timothée Giraud est géomaticien au sein du Réseau interdisciplinaire pour l'aménagement du territoire européen (Riate, UMS2414, CNRS / Université Paris Diderot / Commissariat Général à l'Égalité des Territoires). Directeur du Collège international des sciences du territoires (CIST), Claude Grasland est membre de l'unité Géographie-cités (UMR8504, CNRS / Université Paris 1 Panthéon-Sorbonne / Université Paris Diderot). Ses recherches portent entre autres sur l'aménagement du territoire européen et les dynamiques économiques et démographiques de l'espace mondial. Également membre de l'unité Géographie-cités, Marianne Guérois étudie les dynamiques d'étalement urbain et les interactions entre morphologies urbaines et pratiques de mobilité individuelles, analysées dans une perspective comparative. Au sein du Pôle de recherche pour l'organisation et la diffusion de l'information géographique (Prodig, UMR8586, CNRS / Université Paris1 Panthéon-Sorbonne / Université Paris Diderot / IRD), Malika Madelin mène des recherches en environnement (climatologie), sur l'information à l'échelle locale. Marta Severo, enfin, s'intéresse aux méthodes numériques pour les sciences sociales et les représentations de l'espace sur Internet, au sein du laboratoire Dicen-IDF (Dispositifs d'Information et de Communication à l'Ère Numérique – Ile-de-France, CNAM / UPEM / Université Paris Ouest Nanterre).

#### Introduction

Au-delà des effets de modes qui conduisent aujourd'hui à l'emploi obligé de *buzzwords* tels que *Digital Humanities* ou *Big Data* dans tout appel à projet qui se respecte... de nombreux chercheurs en sciences sociales, en informatique et en linguistique ont engagé depuis plusieurs années des recherches approfondies et cumulatives sur l'analyse des données massives issues du Web, des réseaux sociaux et plus généralement de l'ensemble des gisements de données ouvertes.

À la croisée de plusieurs disciplines, les spécialistes des sciences de l'espace¹ et des sciences territoriales² proposent un point de vue original sur l'exploitation de ces nouveaux gisements de données en privilégiant l'étude de leur contenu en termes d'information géographique. Plus précisément, ils développent des outils et des concepts permettant d'analyser à la fois des coordonnées spatiales de localisation de type latitude-longitude et des attributs territoriaux de nature sémantique désignant des lieux aux contours plus imprécis tels que des villes ou des pays.

Sans prétendre épuiser la richesse des travaux menés actuellement dans de très nombreuses équipes de recherche, le présent article souhaite illustrer la diversité des approches à l'aide d'une série d'exemples tirés de travaux du Collège international des sciences du territoire (CIST) et des équipes de recherche qui en sont membres.

## Traces numériques des plateformes de l'économie « collaborative » : l'exemple des données *AirBnb*

Les données AirBnb issues des plateformes Internet de location touristique entre particuliers offrent un exemple riche de traces numériques du Web 2.0 détournées en sources d'information géographique pour l'analyse des dynamiques métropolitaines. Les enjeux scientifiques, voire politiques, liés à ces informations sont en effet nombreux et ont déjà suscité plusieurs études portant sur le développement de cette offre d'hébergement, qu'elle appartienne à la sphère de l'économie touristique collaborative ou s'inscrive dans de nouvelles niches de spéculation immobilière<sup>3</sup>.



#### Qu'est-ce que le CIST ?

La mission du Collège international des sciences du territoire est de promouvoir, par ses activités, la constitution des sciences territoriales, associant étroitement la théorisation et la pratique, ainsi que l'exploration de l'information territoriale et de son impact sur la vie en société. Il s'appuie pour cela sur une communauté pluridisciplinaire de chercheurs et ingénieurs des 24 équipes de recherche qui le constituent, engagés dans ses 9 axes scientifiques, dont les axes Information territoriale locale et Médias et territoires, plus particulièrement impliqués dans des recherches mobilisant les *Big Data* (projets Géomédia, Grandes métropoles...). Créé en tant que GIS en 2010, le CIST devrait se transformer en Fédération de Recherche en 2017.

En savoir plus

D'un point de vue méthodologique, l'intérêt suscité par ces données tient tout d'abord à leur exhaustivité (a priori tous les biens de ce marché sont renseignés et localisés), à leur haute résolution temporelle (exploitable pour peu qu'un archivage permette d'en garder la mémoire) et spatiale<sup>4</sup>. De plus, dans un contexte où l'accès aux données immobilières localisées à l'adresse s'avère en général très coûteux, les données AirBnb sont facilement accessibles et peuvent être obtenues soit directement via l'extraction des données de la plateforme, soit par l'intermédiaire du site indépendant Inside AirBnb qui facilite le téléchargement d'un grand nombre d'informations pour les centres-villes d'une quarantaine de métropoles dans le monde, avec parfois plusieurs enregistrements temporels. Enfin, ces données se caractérisent par une grande richesse sémantique, aussi bien en termes de description qualitative et quantitative des caractéristiques du bien (type de location, capacité d'accueil, prix de la nuitée, propriétaires multiples ou non...) que de commentaires associés aux locations.

<sup>1.</sup> Goodchild M. F., Janelle D. G. (dir.). 2004, Spatially integrated social science, Oxford University Press.

<sup>2.</sup> Beckouche P., Grasland C., Guérin-Pace F. & Moisseron J. Y. (dir.). 2012, Fonder les sciences du territoire. Karthala.

<sup>3.</sup> Voir à ce sujet : Gutierrez J., Garcia-Palomares J. C., Romanillos G., Salas Olmedo M. H. 2016, « Airbnb in tourist cities: comparing spatial patterns of hotels and peer-to-peer accommodation », arXiv:1606.07138; Quattrone G., Proserpio D., Quercia D., Capra L., Musolesi M., 2016, « Who benefits from the « sharing » economy of AirBnb ? », *Proceedings of the 25th International Conference on World Wide Web*; travaux du groupe de recherche Net(h)no-graphies.

<sup>4.</sup> Le fait que l'information diffusée soit ponctuelle ne doit cependant pas faire illusion sur la précision absolue de la localisation qui est « floutée » à 150 mètres près pour des raisons de confidentialité. Les motivations de ce floutage font l'objet de plusieurs interprétations : protection des biens ou partage restreint d'une information soumise à des contrôles publics ?

Dans le cadre du projet Grandes métropoles du CIST, plusieurs pistes d'analyse territoriale du phénomène AirBnb ont été expérimentées à partir de jeux de données disponibles sur le site Inside AirBnb : les informations sur les caractéristiques des biens loués ont ainsi permis d'esquisser un tableau de la diffusion de la présence de ce service dans plusieurs grandes métropoles (Figure 1). Celle-ci, reconstituée à partir de la date du premier commentaire<sup>5</sup> associé à chaque hébergement, met en valeur la croissance exponentielle de l'offre et l'importance des cycles saisonniers. Dans Paris, l'offre AirBnb présente des spécificités très nettes en comparaison d'autres centres de grandes métropoles : en particulier, on y dénombre une part très importante de logements loués en entier (86 %), mais une très faible proportion d'hôtes « multipropriétaires » (7 %). Cette offre se concentre principalement autour de Montmartre et dans le 3e arrondissement (Figure 2), avec une répartition spatiale des prix des locations similaire à celle d'autres biens immobiliers (centre et ouest de Paris).

	Barcelone	Berlin	Chicago	Londres	New York	Paris	San Francisco
Population (millions d'hab.) Superficie (en km²)	1,6 100	3,5 890	2,7 590	8,4 1 570	8,6 790	2,2 105	0,9 120
Offre	17.040	15.070	5.1.47	40.040	40.007	50 705	0.410
Nombre de logements % avec commentaires		15 373 79,5 %	5 147 80,5 %	49 348 70,3 %	40 227 77,6 %	52 725 72,5 %	8 619 74,2 %
% avec disponibilité > 1/2 an	61,3 %	63,1 %	67,8 %	50,3 %	40,1 %	50,6 %	41,4 %
Logement entier	50,4%	60,7%	56,9%	51,2%	49,5%	85,7%	57,6%
Chambre privée	48,4%	38,0%	38,3%	47,3%	47,0%	13,3%	36,9%
Chambre partagée	1,2%	1,3%	4,8%	1,4%	3,4%	1,0%	5,5%
Hôtes Nombre	10 112	12 405	3 848	34 678	33 582	44 874	6 705
% avec plusieurs locations	27,0 %	8,3 %	15,6 %	16,0 %	12,0 %	7,4 %	13,7 %
Prix Moyen (en \$/pers) Médian		23,0 20,0	50,2 41,7	33,5 27,5	58,0 47,5	32,6 28,0	85,4 62,5
Date des données (sur <u>InsideAirBnb.com</u> )	12/16	10/15	10/15	10/16	12/16	07/16	07/16

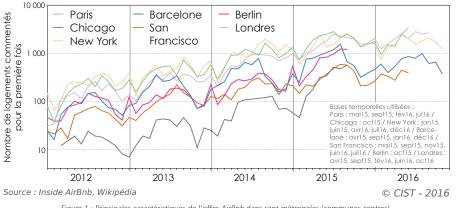
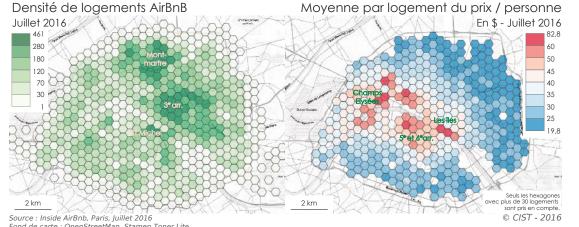


Figure 1 : Principales caractéristiques de l'offre AirBnb dans sept métropoles (communes centres) et évolution du nombre de locations commentées pour la première fois, entre 2012 et 2016



#### Le projet Grandes métropoles

Le <u>projet Grandes métropoles</u> du CIST vise à constituer une plateforme d'échanges autour des défis théoriques et méthodologiques soulevés par le croisement de données locales de plus en plus foisonnantes (ouverture des données publiques, multiplication de données localisées *via* les plateformes Web et les médias sociaux, diffusion de capteurs individuels — de pollution, de température...). Le projet a d'abord effectué un travail préliminaire d'harmonisation des périmètres, des maillages territoriaux et des sources statistiques utilisables pour comparer les métropoles retenues (initialement Paris, Chicago et Mexico). Puis, des ateliers ont été organisés dans le cadre de ce projet, qui permettent d'avancer sur différents fronts, à la fois théoriques et méthodologiques, pour poser *in fine* la question des connaissances originales apportées par cette profusion d'information et de leur interopérabilité.

#### Des fils RSS aux événements géomédiatiques

Le projet ANR Géomédia, qui vient juste de s'achever, se proposait l'objectif ambitieux de discuter l'existence d'un espace public mondial à travers la collecte et l'analyse d'un grand corpus de nouvelles de presse internationales. L'intérêt pour ce type d'objet était motivé, entre autres, par la disponibilité massive d'un nouveau type de données médiatiques, le fil RSS. Le fil RSS est un fichier XML, mis à disposition librement par un site Web pour diffuser les dernières nouvelles publiées en ligne. Cet outil de diffusion des actualités s'est largement répandu sur les sites Web des journaux de presse écrite qui l'ont adopté pour communiquer en temps réel. Par conséquent, ces données numériques offrent au/à la chercheur/ se en sciences sociales une source d'information alternative aux bases de données médiatiques payantes comme Factiva ou Europresse: non seulement les fils RSS sont librement disponibles sur Internet, mais ils peuvent être collectés, archivés et analysés grâce à leur structure standardisée (titre, date, résumé).

Pour étudier les relations internationales au prisme des médias, l'équipe Géomédia a stocké le contenu des fils RSS associés aux articles publiés par une centaine de quotidiens dans différents pays. L'analyse a privilégié les actualités internationales (publiées dans le fil RSS international du

journal), c'est-à-dire les informations ayant trait à des événements se produisant en dehors des frontières du pays. L'analyse des actualités collectées dans cette base a permis de déduire deux types d'information portant respectivement sur les flux entre États et sur les événements internationaux.

Figure 2 : Densité de l'offre et prix des locations AirBnb à Paris, en juillet 2016

<sup>5.</sup> Cet indicateur est discutable, la pratique du commentaire n'étant pas aussi forte d'un pays à l'autre.





Figure 3 : Geomedia Event Xplorer. Le référendum sur le mariage entre personnes de même sexe en Irlande (22 mai 2015) apparaît comme un événement géomédiatique dans le fil RSS du New York Times. Source : geomediaex.ums-riate.fr

Si les fils RSS présentent l'avantage d'être facilement accessibles, ils posent plusieurs défis quand l'objectif est d'étudier l'information géographique qu'ils véhiculent. En effet, à la différence des données AirBnb décrites dans la partie précédente, les données RSS ne contiennent ni coordonnées géographiques ni localisation clairement identifiable. Tout l'intérêt du travail mené dans le cadre du projet Géomédia réside alors dans la création d'une chaîne d'archivage, de traitement, d'enrichissement, de modélisation et de visualisation de ces données, afin d'identifier des localisations à partir d'une donnée purement textuelle. Plusieurs pistes ont été explorées : l'étude des flux internationaux en s'appuyant sur les cooccurrences entre pays cités dans les actualités ; les flux entre le pays du journal émetteur de l'actualité et le(s) pays cité(s) dans l'actualité... Cela a soulevé de nombreux problèmes à la fois empiriques (ambiguïtés dans la reconnaissance des entités nommées) et épistémologiques. Un journal (ou plusieurs journaux) peut-il être considéré comme représentatif d'un pays ? Comment une information géographique peut-elle être détectée dans le contenu d'une actualité (nom de l'État, du chef d'État, de la résidence du chef d'État...) ? Peut-on caractériser des types d'attitude des médias face à l'international?

Le projet Géomédia a ainsi permis de développer une réflexion interdisciplinaire sur l'emploi de données textuelles brutes qui ne contiennent pas une information clairement localisée pour étudier la dimension territoriale d'un objet comme l'événement médiatique. L'ensemble des résultats scientifiques de l'ANR Géomédia ainsi que la base de données collectées sont accessibles sur différents supports<sup>6</sup>.

Dans le cadre du projet, a été développée notamment une application Web, Geomedia Event Xplorer, qui propose une visualisation cartographique et sémantique des événements qui focalisent l'attention internationale des médias sur un pays plutôt qu'un autre au fil du temps (Figure 3). On détermine pour chaque quotidien un niveau moyen de médiatisation des pays du Monde au cours des 4, 12 ou 24 semaines précédentes. Puis, on détermine statistiquement les pays qui présentent un pic brutal de médiatisation par rapport au niveau antérieur. L'application permet ensuite de visualiser la localisation de ces pics d'attention pour chaque média et de les comparer. On associe à chaque pic des nuages de mots qui permettent de les interpréter tout en évitant de reproduire le contenu exact de la nouvelle pour des raisons de droits d'auteur des journaux.

## Médias sociaux et territoires : l'exemple des données *Twitter*

Ces dernières années, les données des médias sociaux ont également suscité l'intérêt des chercheurs travaillant sur la localisation spatiale des acteurs et des contenus. Sans doute les données *Twitter* se sont-elles avérées comme les plus adaptées pour deux raisons. En premier lieu, étant de nature généralement publique, leur collecte et leur traitement soulèvent a priori moins de questions éthiques et juridiques comparés à d'autres plateformes de réseautage social comme *Facebook* et *Instagram*. Ensuite, certains *tweets* sont associés à une localisation spatiale précise avec un couple de coordonnées longitude-latitude.

On peut tout d'abord prendre en compte la composante spatiale des *tweets* dotés de coordonnées spatiales, parmi d'autres caractéristiques (identifiant de l'utilisateur, nombre de « *followers* », langue, contenu du message...). Si la richesse de la résolution spatiale et temporelle de ces données individuelles est souvent mise en avant, le passage de ces données localisées à une information territoriale exploitable ne va toutefois pas sans difficulté.

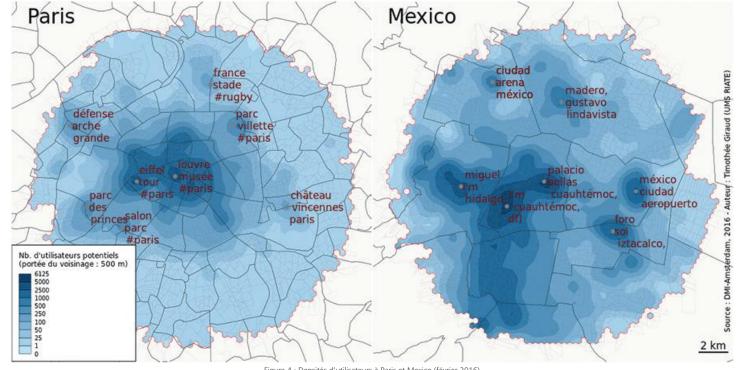


Figure 4 : Densités d'utilisateurs à Paris et Mexico (février 2016)
Sur cette figure, sont mises en valeur dans un périmètre comparable pour Paris et Mexico (cœur central des aires urbaines cumulant 5 millions d'habitants) les zones où l'on trouve le plus grand nombre d'utilisateurs actifs dans un rayon de 500 m. Les pics d'utilisateurs sont associés à des équipements touristiques, culturels ou sportifs majeurs des deux métropoles, comme le révèlent les principaux mots-clés extraits du contenu des messages. Cette méthode d'analyse croisée des localisations et du texte permet aussi, en sens inverse, de détecter les lieux où certains sujets sont abordés.

Une limite importante concerne la représentativité de ces données : les utilisateurs de *Twitter* sont de plus en plus nombreux, mais les *tweets* géoréférencés (qui possèdent des coordonnées de latitude et de longitude) ne représentent que 1 à 2 % du total des *tweets* publiés<sup>7</sup>. On peut de plus faire l'hypothèse que la souspopulation qui laisse accessibles ses données de géolocalisation a un profil spécifique par rapport au reste de la population (minorité d'utilisateurs jeunes). Enfin, la majorité des utilisateurs localisables publie très peu, tandis qu'une minorité (comptant la présence de « robots » programmés qui publient des messages automatiques — spams ou publicités pour la plupart — en grande quantité) s'avère très active. C'est pourquoi il est souvent préférable de raisonner à partir d'un indicateur de densité d'utilisateurs (Figure 4) plutôt que de densité de *tweets*.

Même si ces résultats sont encourageants, la faible représentativité des tweets géolocalisés spatialement peut conduire à analyser les informations territoriales éventuellement présentes dans le message même. La tâche n'est pas du tout anodine pour plusieurs raisons. D'abord, Twitter n'est pas une plateforme strictement liée au territoire. Au contraire, les échanges sous forme de tweets peuvent concerner n'importe quel sujet et, dans la plupart des cas, ils jouent surtout une fonction phatique, c'est-à-dire qu'ils servent pour construire et alimenter une relation. Par ailleurs, ces dernières années, les utilisateurs les plus actifs, qu'ils soient des personnes ou des robots, sont engagés dans une guerre de visibilité et de réputation. Enfin, l'analyse du contenu pose d'importants défis linguistiques parce que les twittos peuvent utiliser des hashtags ou des abréviations qui ne sont pas faciles à détecter avec des techniques d'analyse textuelle classiques. Tout cela produit un effet de bruit significatif qui rend difficiles l'identification univoque des localisations dans le texte et le repérage des informations qui concernent les territoires.

Malgré ces obstacles, les travaux réalisés dans le cadre du CIST ont permis de développer un certain nombre de techniques qui facilitent l'extraction d'informations utiles pour l'analyse territoriale à partir du contenu des *tweets*. Il est par exemple possible de

conduire l'analyse du contenu qualitatif et/ou quantitatif sur des sous-échantillons de *tweets* localisés dans des endroits-clés pour comprendre la raison d'une densité plus importante d'activité. Une deuxième approche qui a donné des résultats encourageants consiste à sélectionner de manière qualitative un groupe de *twittos* qui peuvent être considérés comme des « influenceurs » du territoire qu'on entend étudier. Enfin, nous avons expérimenté l'intérêt de reproduire la méthodologie développée dans le projet Géomédia pour le fil RSS sur les *tweets*, c'est-à-dire nous avons reconstruit les flux entre pays à partir des pays mentionnés dans les *tweets* géolocalisés dans un certain État.

#### **Conclusion**

À l'heure où l'information circule plus rapidement que jamais, il n'est sans doute pas inutile de plaider résolument en faveur d'une science lente qui prend le temps de construire et analyser précisément les données massives à la lumière de concepts et d'hypothèses explicites. Cette construction sera d'autant plus performante à long terme qu'elle aura pris le temps du dialogue entre plusieurs disciplines apportant des éclairages complémentaires : géographie, informatique, sciences de la communication, linguistique...

contact&info

Marion Gentilhomme,
CIST
marion.gentilhomme@gis-cist.fr

<sup>7.</sup> Gerlitz C. & Rieder B. 2013, Mining One Percent of Twitter: Collections, Baselines, Sampling, in M/C Journal, 16(2); Severo M., Giraud T. & Pecout H. 2015, Twitter data for urban policy making: an analysis on four European cities, in Handbook of Twitter for Research, EMLYON Press.

### Modélisation à base d'agents et données massives : un inévitable prix à payer

Directeur de recherche CNRS, Arnaud Banos est directeur du laboratoire Géographie-cités (UMR8504, CNRS / Université Paris 1 Panthéon-Sorbonne / Université Paris Diderot) et responsable scientifique et technique du Labex DynamiTe. Ses travaux de recherche portent notamment sur la modélisation et la simulation spatiales en géographie. Sébastien Rey-Coyrehourcq est ingénieur de recherche à l'Université de Rouen et membre du laboratoire Identité et Différentiation de l'Espace, de l'Environnement et des Sociétés (IDEES, UMR6266, CNRS / Université Caen Normandie / Université de Rouen Normandie / Université Le Havre Normandie)

La modélisation à base d'agents<sup>1</sup> a ouvert des perspectives considérables en géographie, en particulier dans le domaine de la modélisation des dynamiques socio-environnementales<sup>2</sup>. Cette approche computationnelle<sup>3</sup> présente le grand avantage d'être à la fois très proche des approches formelles dans les situations où celles-ci s'appliquent<sup>4</sup> mais également de pouvoir plus facilement être étendue à d'autres situations plus difficiles à formaliser. Elle augmente significativement de ce point de vue nos capacités à explorer des dynamiques socio-environnementales complexes, sur une base plus « immédiatement intelligibles pour les praticiens des sciences sociales »<sup>5</sup> que les approches mathématiques.

Cette extension du domaine d'exploration a toutefois un prix : l'évaluation de ces modèles, très souvent sur-paramétrés dans un premier temps de la démarche afin de rester proches des phénomènes modélisés, implique l'analyse d'une quantité considérable de données de simulation.

Prenons un exemple basique (Figure 1) : un modèle avec 5 paramètres présentant chacun 10 valeurs possibles possède déjà un ensemble de combinaisons de l'ordre de 10<sup>5</sup> = 100 000 combinaisons. Si on ajoute la nécessité de répliquer chaque simulation un nombre minimal de fois compte tenu de la nature stochastique<sup>6</sup> du modèle, par exemple 30 réplications par simulations, on obtient déjà 3 000 000 de combinaisons de paramètres en sortie pour ce simple modèle, chaque combinaison devant être caractérisée par au moins un indicateur macroscopique (dans le cas de la Figure 1, une loi rang-taille caractérisant la hiérarchisation du système urbain obtenu).

Le modélisateur se retrouve par conséquent dans la situation suivante : un modèle à base d'agents, même « simple », génère par essence une quantité de données supérieure à ce qui peut être facilement analysé de manière exhaustive : la taille de l'espace des paramètres à explorer croît en effet de manière diabolique avec le nombre de paramètres et il s'avère rapidement illusoire de chercher à le parcourir dans son intégralité (Figure 2), sauf à disposer de technologies adaptées.

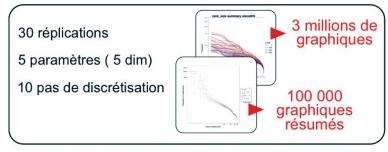


Figure 1 : Dimensionnement du volume de données en sortie de simulation à partir du nombre de paramètres en entrée (Réalisation : Sébastien Rey-Coyrehourcg).

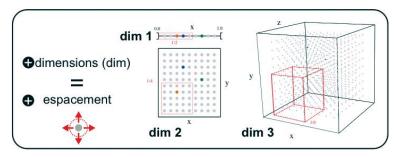


Figure 2 : L'impossible caractérisation exhaustive dans l'espace de ses paramètres d'un modèle à base d'agents (Réalisation : Sébastien Rey-Coyrehourcq).

Il s'agit là d'une vraie difficulté, qui ne doit en aucun cas être sous-estimée : une critique souvent formulée à l'encontre des modèles informatiques, à base d'agents notamment, concerne en effet la difficulté (parfois insurmontable) à laquelle le modélisateur est confronté lorsqu'il s'agit d'en caractériser le comportement. De ce point de vue, la plateforme OpenMole<sup>7</sup> est une contribution majeure : elle permet en effet de distribuer aisément et de manière transparente tout modèle de simulation sur une grille de calcul<sup>8</sup> et de repousser largement les limites imposées par la « malédiction de la dimension » (« curse of dimensionality »), phénomène identifié par le mathématicien Richard Bellman dès les années 1960 (Figure 3).

<sup>1.</sup> Egalement dénommée modélisation multi-agents, cette approche est issue des recherches menées en informatique. De manière simplifiée, un système multi-agents (SMA) peut être vu comme un système artificiel composé d'agents situés dans un environnement, au sein duquel ils interagissent en fonction de règles de comportement prédéfinies. Voir à ce sujet : Ferber J. 1995, Les systèmes multi-agents. Vers une intelligence collective, InterEditions.

<sup>2.</sup> Banos A, 2016, Modéliser, c'est apprendre : Itinéraire d'un géographe, Editions Matériologiques.

<sup>3.</sup> Le philosophe Franck Varenne parle à ce propos de « simulation algorithmique », simulations qui sont « fondées sur des systèmes de règles (un « modelé de simulation ») simulant chacune localement un comportement visible et qui opèrent en cumulant leurs interactions et leurs itérations » : Varenne F. 2014, Epistémologie des modèles et des simulations : tour d'horizon et tendances", dans Levy J.M. (dir.), Les Modèles, possibilités et limites. Jusqu'où va le réel ?, Matériologiques, pp. 13-46.

<sup>4.</sup> Epstein J. 2006, Generative Social Science – Studies in Agent-Based Computational Modeling, Princeton University Press.

<sup>5.</sup> Varenne F. 2010, Les simulations computationnelles dans les sciences sociales, dans Nouvelles perspectives en sciences sociales, Volume 5,

<sup>6.</sup> Même si un tel modèle peut être de nature strictement déterministe dans ses règles, son implémentation distribuée implique l'introduction d'un aléa dans l'ordonnancement de celles-ci, en raison de la nature séquentielle du processus de simulation informatique.

<sup>7.</sup> Plateforme libre développée par l'Institut des Systèmes Complexes de Paris Île-de-France (ISC-PIF, UPS3611, CNRS), en collaboration avec le laboratoire Géographie-cités (UMR8504, CNRS / Université Paris 1 Panthéon-Sorbonne / Université Paris Diderot) et BioMedIA.

<sup>8.</sup> Reuillon R., Leclaire M., Rey-Coyrehourcq S. 2013, OpenMOLE, a workflow engine specifically tailored for the distributed exploration of simulation models, in Future Generation Computer Systems, Volume 29, n°8:1981-1990.

Cette extension du domaine des possibles par une technologie adaptée ne peut toutefois pas s'affranchir complètement de cette « malédiction de la dimension ». Il faut nécessairement penser le problème différemment pour aller plus loin. De ce point de vue, le projet ERC GeoDiverCity coordonné par Denise Pumain a permis de réaliser des avancées tout à fait considérables<sup>9</sup>. Un principe général est de relâcher la contrainte d'exhaustivité dans l'exploration de l'espace des paramètres en introduisant une stratégie de recherche guidée par des objectifs. Parmi les nombreuses pistes explorées, deux d'entre elles illustrent de manière particulièrement frappante le changement de paradigme imposé par la complexité du problème posé.

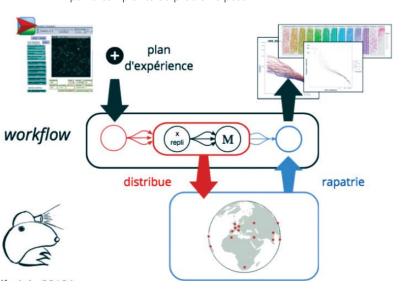
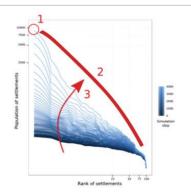


Figure 3 : Distribuer les simulations pour mieux les étudier, une contribution majeure de la plateforme OpenMole (Réalisation : Sébastien Rey-Coyrehourcq).

Une première approche consiste ainsi à définir *a priori* ce que l'on cherche à reproduire avec le modèle. On parle ici de « fonction objectif ». Dans le cas du modèle SimpopLocal<sup>10</sup>, trois objectifs simultanés on été identifiés (Figure 4) : la population maximale atteinte par le plus gros site de peuplement, la distribution statistique des tailles des sites de peuplement et la durée de la période simulée. Une fois cette « fonction objectif » définie, il devient possible de sélectionner les simulations réalisées à partir de leur distance à l'objectif. Compte tenu de la taille de l'espace des paramètres à parcourir, une approche par algorithmes génétiques<sup>11</sup> a été retenue et a ainsi permis, au moyen de plus de 500 millions de simulations sur OpenMole, de calibrer de manière fine le modèle SimpopLocal, c'est-à-dire d'identifier l'ensemble des combinaisons des paramètres permettant, pour une condition initiale fixée, d'atteindre l'objectif fixé<sup>12</sup>.



3 objectifs simultanés

**obj 1 -** population max = 10000 hab.

**obj 2 -** distribution villes = lognormalité

**obj 3 -** temps = 4000 ans

Figure 4 : Définition d'objectifs de simulation à atteindre par le modèle (Réalisation : Sébastien Rey-Coyrehourcq).

Cette approche, particulièrement efficace lorsque l'on peut définir un objectif à atteindre, devient toutefois caduque lorsque cette opération n'est pas possible. C'est dans cette perspective qu'une approche complémentaire a été proposée<sup>13</sup>. Reposant également sur des algorithmes génétiques, la méthode PSE (Pattern Space Exploration) permet d'explorer de manière quidée un espace des paramètres immense à la recherche non pas, cette fois, d'une structure (ou pattern) fixée, mais du plus grand nombre de structures différentes. L'enjeu est ici de dépasser la seule confrontation d'un modèle à des données observées ou des objectifs fixés, afin de permettre une caractérisation beaucoup plus large de la diversité des structures qu'il permet de produire par simulation. Là encore, ce sont des millions de simulations qui doivent être réalisées : ce qui pourrait toutefois passer pour une surenchère computationnelle n'est pourtant rien d'autre qu'une adaptation au contraire plutôt parcimonieuse, à cette « malédiction de la dimension » évoquée.

Les possibilités d'exploration, par la modélisation et la simulation à base d'agents, des systèmes complexes n'ont jamais été aussi grandes¹⁴. L'assouplissement des conditions de modélisation autorisé par les approches informatiques (notamment à base d'agents) par rapport aux approches mathématiques a toutefois un prix : l'augmentation considérable, d'une part, des simulations à réaliser et, d'autre part, des données à analyser en sortie des simulations. Ces étapes ne peuvent pas être dissociées, les modèles à base d'agents étant par construction des modèles de simulation. Se donner, individuellement et collectivement, les moyens d'avancer dans cette voie exigeante est, sans aucun doute, l'un des grands enjeux actuels de la recherche en sciences humaines et sociales.

contact&info

▶ Arnaud Banos,
Géographies-cités
arnaud.banos@parisgeo.cnrs.fr

11. Inspirés des mécanismes évolutionnistes identifiés en biologie, les algorithmes génétiques reposent sur un mécanisme simplifié de « sélection naturelle pour approcher — avec des temps de calcul souvent raisonnables — des solutions possibles dans le cadre de problèmes d'optimisation difficiles à résoudre au moyen d'approches plus conventionnelles.

12. Schmitt C, Rey-Coyrehourcq S, Reuillon R, Pumain D. 2015, Half a billion simulations: evolutionary algorithms and distributed computing for calibrating the SimpopLocal geographical model, in *Environment and Planning B: Planning and Design*, n°42(2): 300-315.

13. Chérel G., Cottineau C., Reuillon R. 2015, Beyond Corroboration: Strengthening Model Validation by Looking for Unexpected Patterns, in *PLoS ONE* 10(9): e0138212. doi:10.1371/journal.pone.0138212.

14. Banos A., Lang C., Marilleau N. 2015, Agent-Based Spatial Simulation with NetLogo, Volume 1, Introduction and Bases, ISTE. Banos A., Lang C., Marilleau N. 2016, Agent-Based Spatial Simulation with NetLogo, Volume 2, Advanced concepts, ISTE.

<sup>9.</sup> Pumain D, Reuillon R. 2017, *Urban Dynamics and Simulation Models*, Springer International Publishing.

<sup>10.</sup> Modélisation d'un système de peuplement sous contrainte environnementale à la période du néolithique, développé et évalué dans le cadre des thèses de Clara Schmitt et Sébastien Rey-Coyrehourcq : Schmitt C. 2014, Modélisation de la dynamique des systèmes de peuplement : de SimpopLocal à SimpopNet, soutenue le 28 février 2014, Université Paris 1 Panthéon-Sorbonne ; Rey-Coyrehourcq S. 2015, Une plateforme intégrée pour la construction et l'évaluation de modèles des insulation multi-agents, soutenue le 13 octobre 2015, Université Paris 1 Panthéon-Sorbonne.

## Entre big data et dispositif d'enquête ad hoc : un entre-deux qui reste à inventer...

Maître de conférences en aménagement et urbanisme, Thomas Buhler est membre du laboratoire Théoriser et modéliser pour aménager (ThéMA, UMR 6049, CNRS / Université de Bourgogne - Franche-Comté). Ses recherches portent principalement sur le rôle des habitudes dans les pratiques de déplacements urbains et dans l'usage domestique de l'énergie. Il s'intéresse également à la mise en forme et au "sens" des politiques publiques ayant pour objectif des changements de comportements dans ces deux domaines.

Au sein du laboratoire ThéMA, nous¹ travaillons depuis 2014 à la mise en place d'un dispositif permettant un suivi longitudinal, automatique et quotidien des pratiques de mobilité d'un panel de personnes volontaires. Ce dispositif prend appui sur les données des différents capteurs présents dans la plupart des smartphones courants (accéléromètre, GPS, Wi-Fi, GSM).

Suivre les pratiques quotidiennes de mobilité : un enjeu social, un défi technique

Deux constats plutôt simples sont à l'origine du projet « TELEM » (« Temps Long Energie Mobilité »). Tout d'abord, notons qu'en dépit de la mise en place de nombreuses politiques publiques ciblant une évolution des pratiques quotidiennes de mobilité<sup>2</sup>, il n'existe à l'heure actuelle aucun dispositif permettant un suivi longitudinal quotidien à l'échelle individuelle pour ces pratiques. Les données actuellement les plus utilisées par les chercheurs et les opérationnels de la mobilité quotidienne sont issues d'enquêtes « transversales » réalisées en général à l'initiative de la collectivité<sup>3</sup>. Dans ces enquêtes « classiques » il s'agit d'interroger en face-à-face une population statistiquement représentative d'une agglomération. L'opération est reconduite à une fréquence d'environ dix ans. Par rapport à la thématique des changements de comportements quotidiens, les limites de ces données sont claires. Les échantillons étant statistiquement « représentatifs », il s'agit de personnes différentes interrogées à chaque millésime. On ne sait donc pas concrètement, par exemple, qui a pu changer de mode de déplacement dans l'intervalle des dix ans. Autre limite flagrante : ces enquêtes se basent sur une « journée-type » (un mardi ou un jeudi) et ne prennent pas en compte la variabilité au sein de la semaine, par exemple.

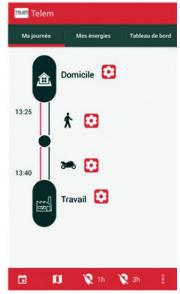
Le second constat concerne la généralisation de l'équipement en smartphone. En effet, une majorité de français possède désormais un smartphone<sup>4</sup>, avec *a minima* dans la poche : un accéléromètre, et des puces GPS, Wi-Fi et GSM. Tout ceci – potentiellement – permet d'identifier les modes utilisés par la personne, les trajets effectués (chemins et horaires) et les lieux fréquentés.

## Une interprétation des données de capteurs couplée à une validation par l'utilisateur

La confrontation de ces deux constats nous a amené à nous lancer dans la constitution d'un dispositif *ad hoc*. Celui-ci se veut longitudinal — de type panel — et repose sur le développement d'une application smartphone native (pour *Android* et *iOS*, qui représentaient respectivement 69,9 % et 20,5 % des ventes de smartphones en 2015) et d'une plateforme de données spatiotemporelles<sup>5</sup>.

Bien sûr, les « panels » existent depuis longtemps en sciences humaines et sociales. Cela dit, ils sont rarement utilisés pour analyser les pratiques quotidiennes. En effet, un retour d'information trop régulier de la part du panéliste peut représenter un risque grave d'attrition c'est-à-dire d'abandon du dispositif par la personne. Dans le cadre du projet TELEM, l'idée que nous défendons est de créer un suivi « semi-automatique », c'est-à-dire une phase d'interprétation des données des capteurs suivi d'une phase de modification et de validation très rapide de la part de l'utilisateur.

La partie d'interprétation et de synthèse des informations adressée à l'utilisateur n'est évidemment pas une mince affaire<sup>6</sup>. Elle a nécessité un travail d'environ deux ans. À l'aide d'un algorithme développé spécialement et s'appuyant sur les données des différents capteurs du smartphone (accéléromètre, GPS, Wi-Fi, GSM) l'application présente une *timeline* synthétique à l'utilisateur qui se veut être le plus proche possible des déplacements réellement effectués. Cette *timeline* permet alors de recueillir ses modifications en termes de modes utilisés, de lieux fréquentés et de motifs de déplacement.



La timeline permet au panéliste de modifier et/ou de valider l'interprétation automatique de ses déplacements.

- 1. L'équipe du projet est composée de Thomas Buhler, Philippe Signoret et de Florian Litot.
- 2. Pensons ici aux Plans de Déplacements Urbains (PDU), aux Plans de Déplacements d'Entreprise (PDE) et autres opérations d'aménagement de type « écoquartier ». Ces plans, politiques et projets visent tous le maintien ou la baisse de la part modale de l'automobile en ville.
- 3. Par exemple, les « Enquêtes Ménages-Déplacements » réalisées par toutes les grandes et moyennes agglomérations françaises.
- 4. En 2015, ils étaient 58 % à posséder un smartphone, d'après le CGEIET. Ils étaient 17 % quatre ans plus tôt. On peut donc parler d'un mouvement de généralisation de l'équipement.
- 5. Le consortium Mobi-Lise a assuré une partie du développement informatique.
- 6. Pour une présentation plus en détail du fonctionnement technique du dispositif, se référer à la publication à paraître prochainement : Buhler T., Signoret P. 2017, « Suivre les consommations énergétiques et les déplacements quotidiens sur le temps long. Revisiter la méthode du panel à l'ère de la généralisation du smartphone », dans Revue Internationale de Géomatique.

Deux logiques ont présidé à ce choix porté sur une timeline interactive. Tout d'abord, le travail d'analyse — même très poussé — des données brutes des quatre principaux capteurs du téléphone ne suffit pas pour arriver à une information détaillée totalement fiable en termes de modes, d'horaires, d'itinéraires, de lieux et de motifs. Les erreurs matérielles, les situations d'incertitude montrent les limites des big data lorsque celles-ci ne sont pas « qualifiées » et validées par l'utilisateur. Prenons l'exemple d'un cycliste en descente : ne pédalant pas, il n'effectue aucun mouvement latéral détectable par l'accéléromètre. Sa vitesse étant proche des 30 ou parfois 40 km/h, une analyse « froide » le considérera comme utilisateur d'un mode motorisé. De tels exemples sont légions et doivent inciter à la prudence quant aux données construites sur la base des big data brutes.

L'idée d'une timeline interactive repose également sur l'idée d'une correction des données par l'utilisateur qui ne doit pas être trop chronophage. Ce dernier doit pouvoir visualiser l'ensemble d'une semaine écoulée et pouvoir y ajouter des modifications en moins d'une minute. Au-delà, le risque d'épuisement et donc d'attrition est important.

## Recrutement et suivi du panel : les limites de la « science participative » ?

Une fois l'application devenue techniquement satisfaisante, à l'issue d'une série importante de tests, nous avons débuté en mars 2016 le recrutement du panel sur les agglomérations de Besançon et de Belfort-Montbéliard. Des moyens humains, médiatiques et matériels conséquents ont été mis en place pour y parvenir. Outre l'emploi de plusieurs personnes pour recruter en face-à-face, plusieurs leviers ont été activés : la distribution de petits objets (nettoyeurs d'écran de téléphone) pour récompenser symboliquement l'engagement à installer l'application, de nombreux passages médiatiques, l'utilisation des réseaux sociaux. La période « intensive » de trois mois de recrutement a débouché sur un total de 276 personnes utilisant l'application. Ce résultat est à la fois quantitativement un peu décevant (nous avions tablé sur le double) et qualitativement plutôt satisfaisant.



Un module de cartographie permet de visualiser l'ensemble des points recueillis lors des déplacements du panéliste sur la journée (points GPS en rouge, Wi-Fi en bleu).

Contrairement à ce que nous aurions pu penser initialement, le panel recruté se révèle être plutôt diversifié, en termes de catégories socioprofessionnelles. Les cadres et professions intellectuelles supérieures et les personnes de moins de 40 ans sont certes surreprésentés. Cela dit, l'ensemble des catégories sociales et des âges est représentée dans l'actuel panel TELEM.

La constitution et le suivi du panel se heurte à deux principaux problèmes. Premièrement, les retours des recrutements de terrain font apparaître qu'une bonne moitié des personnes rencontrées est catégoriquement opposée à l'idée d'être suivi en continu par une application, même si les données servent à la recherche publique, font l'objet d'une déclaration CNIL et garantissent l'anonymat de la personne. Le « vivier » potentiel restant est d'une « petite moitié », à laquelle il faut retrancher les personnes qui n'ont pas de smartphone ou qui ne sont tout simplement pas intéressées. Par ailleurs, comme pour tout panel, TELEM perd réqulièrement une partie de ses utilisateurs. Nous touchons probablement là aux limites de la « science participative ». Les personnes volontaires le sont pour la plupart pour aider la recherche. Passés plusieurs mois, elles peuvent se poser la question de l'utilité de leur engagement. Or, une application peut être désinstallée en moins de deux secondes sur un téléphone...

#### Vers une solution hybride?

En l'état, le panel TELEM répond à des enjeux importants en termes d'organisation et de qualification des données de mobilité quotidienne. Les données obtenues permettent déjà de travailler, par exemple, sur la constitution d'un indice de récurrence (mode, chemin, horaires) de trajets domicile-travail, sur les transitions modales saisonnières ou encore sur les rapports entre contexte météorologique et pratiques modales. Cela dit, la généralisation d'un tel dispositif pose une question fondamentale : la forme que peut prendre la participation d'un panéliste.

Cette participation est pour l'heure bénévole, ce qui explique certaines limites évoquées plus tôt. On pourrait imaginer qu'elle soit rétribuée, mais les volumes de journées et de personnes représentent un coût très important que ni les collectivités territoriales, ni les laboratoires de recherche ne sont en mesure de financer à grande échelle<sup>7</sup>.

Nous l'avons vu plus tôt, les données *big data* (issues des capteurs que sont l'accéléromètre, GPS, Wi-Fi, GSM) qu'il est possible d'acheter via des développeurs d'applications grand public ne suffisent pas pour identifier finement les modes, lieux, motifs et itinéraires sur une journée. La validation par l'utilisateur reste fondamentale. Il reste à trouver un dispositif sociotechnique adéquat pour générer des *big and smart data*...

7. Certaines pistes pourraient être creusées comme la rétribution aléatoire par tirage au sort.

contact &info

► Thomas Buhler,
ThéMA
thomas.buhler@univ-fcomte.fr

### Des big data pour analyser les mobilités, les dynamiques et les événements urbains

Le laboratoire Études des structures, des processus d'adaptation et des changements de l'espace (ESPACE, UMR7300, CNRS / Aix Marseille Université / Université d'Avignon et des Pays de Vaucluse / Université Nice Sophia Antipolis) est dirigé par Christine Voiron. Il développe des recherches sur les dynamiques urbaines à l'aide de méthodes d'analyse spatiale et de modélisations - systèmes multiagents, réseaux bayésiens, réseaux de neurones, etc., avec, pour objectif, de comprendre l'organisation de la mobilité quotidienne et d'explorer les conditions d'une mise en place d'une accessibilté durable.

Depuis le début des années 2010, de nouvelles sources de données dites « massives », ou big data, sont de plus en plus accessibles pour la communauté scientifique. Souvent géolocalisées ou géolocalisables à des résolutions variables, ces données sont produites, le plus souvent en continu, par des opérateurs de réseaux (à des fins d'exploitation commerciale), par les acteurs publics, mais aussi, consciemment ou pas, par les individus, notamment ceux équipés d'un smartphone<sup>1</sup>. Ces données viennent s'ajouter aux données issues de longues séries d'observations et de mesures systématiques, comme celles de l'observation de la terre ou de bases de données privées (banques, assurances, etc.) dont le volume rend obsolètes les systèmes traditionnels de gestion de bases de données.

Si elles sont théoriquement mobilisables par la recherche, elles n'ont toutefois pas été produites spécifiquement dans cette perspective et sont ainsi souvent incomplètes, hétérogènes, avec des définitions sémantiques floues. Certaines de ces données sont aujourd'hui aisément disponibles (la géolocalisation des Tweets), d'autres, notamment celles liées directement à un intérêt commercial (localisation des téléphones mobiles), le sont beaucoup moins. À cet égard, certaines données jusqu'à aujourd'hui exploitables librement (les Tweets par exemple), vont faire l'objet de restrictions d'utilisation de la part de la CNIL, notamment en ce qui concerne la géolocalisation. Par ailleurs, les filtres opérés par Twitter (part des tweets échangés et, parmi ceux-ci, part des tweets ayant des coordonnées, et utilisations différenciées selon les pays et les niveaux de liberté des populations) grèvent la représentativité socio-spatiale de ces informations<sup>2</sup> (Figure 1). Données massives il y a donc, mais pas toujours données disponibles ou utilisables.

Bien qu'hétérogènes, incomplètes et partiellement disponibles, le volume, la vitesse et la variété de ces données offrent la promesse d'une forme d'exhaustivité et de temps réel auxquels les sciences sociales n'ont jusqu'alors pas eu accès³, et ouvrent ainsi des perspectives inédites de mesure de certains phénomènes sociaux.

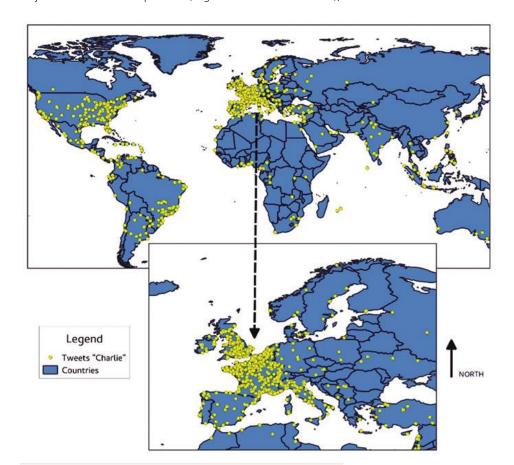


Figure 1. Analyse du Tweet « je suis Charlie » : des données socio-spatiales non représentatives

<sup>1.</sup> Goodchild M.F. 2007, Citizens as sensors: the world of volunteered geography, in GeoJournal, 69(4): 211-221.

<sup>2.</sup> Morchid M., Portilla Y., Josselin D., Dufour R., Altman E. et al. 2015, An Author-Topic based Approach to Cluster Tweets and Mine their Location, in Spatial Statistics 2015

<sup>3.</sup> Kitchin, R. 2013, Big data and human geography: Opportunities, challenges and risks, in Dialogues in Human Geography, 3(3): 262–267.

## Quel apport pour la modélisation des dynamiques urbaines ?

L'utilisation des *big data* dans le champ des études urbaines peut s'envisager selon deux grands objectifs : mieux connaitre et comprendre comment les individus vivent la ville et se servir de cette connaissance à des fins d'aménagement ou de développement économique.

Tout d'abord, les big data, et particulièrement celles qui comprennent des informations sur la géolocalisation, peuvent être utilisées pour documenter des phénomènes jusqu'alors difficilement mesurables par les méthodes usuelles. C'est notamment le cas pour l'étude des flux de mobilité quotidienne, pour savoir qui est où, à quelle heure. Les données de géolocalisation issues de la téléphonie mobile représentent une source d'informations utile, en particulier dans les pays qui ne disposent pas de grandes enquêtes. Dans le projet SPOT par exemple<sup>4</sup>, des algorithmes ont été développés pour permettre de reconstituer les emplois du temps des individus de Dakar (information n'existant pas par ailleurs) à partir de la géolocalisation relative des téléphones mobiles par rapport aux antennes GSM. Ces emplois du temps ont ensuite été mobilisés dans un modèle de simulation transporturbanisme en fonction des densités de population par quartier (Figure 2).

Au-delà du rôle de palliatif à des enquêtes coûteuses ou impossibles à réaliser, utiliser les données massives peut aussi permettre d'analyser le mouvement dans la ville, directement à partir des

localisations et/ou des itinéraires des individus, pour voir comment se font les polarisations, quels sont les déplacements réels, sans passer par des données déclaratives. De même, savoir à une échelle fine qui est où et à quelle heure, peut être très utile dans l'analyse des phénomènes de diffusion spatiale (de maladies, d'idées...).

Du point de vue des applications, mieux connaître les comportements grâce au *big data* est d'un grand intérêt aussi bien pour l'aménagement que pour le développement économique. La connaissance en temps réel des distributions de population est, par exemple, précieuse pour le développement de transports intelligents<sup>5</sup>. Il s'agit ici de savoir comment optimiser en temps réel l'offre de transport pour qu'elle réponde le mieux possible à la demande, rompant ainsi avec l'idée d'une offre de transport fixe correspondant à une vision fordiste du fonctionnement de l'espace aujourd'hui désuète.

La connaissance des mouvements de population en temps réel peut également s'avérer très utile pour répondre à des problématiques d'allocation de ressources. En effet, la localisation de services publics et d'équipements et leur affectation est le plus souvent basée sur la distribution de la population résidentielle — c'est-à-dire celle disponible dans les recensements ou les enquêtes, qui est loin de correspondre à la population présentielle (Figure 3). Ces données prenant en compte la géolocalisation précise des individus pendant la journée peuvent alors permettre de mieux affecter les ressources pour optimiser le service rendu.

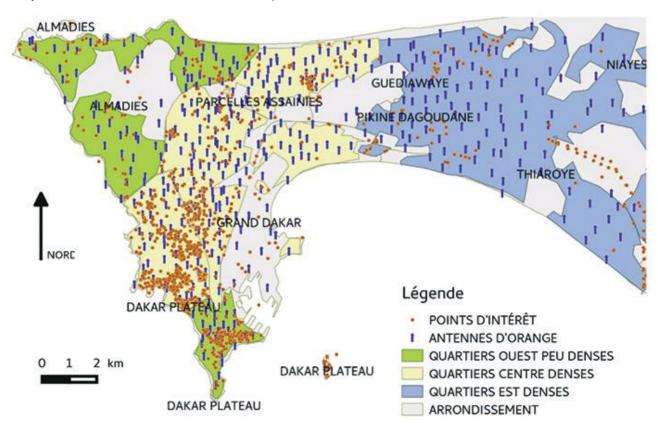


Figure 2. Éléments géographiques pris en compte pour la simulation d'emplois du temps et de mobilités à partir des données des téléphones mobiles d'Orange à Dakar (Sénégal, projet SPOT, lauréat du challenge D4D 2014 d'Orange).

<sup>4.</sup> Gueye S., Ndiaye B.M., Josselin D., Poss M., Faye R.M., Michelon P., Genre-Grandpierre C., Ciari F. 2015. *Using mobile phone data for Spatial Planning simulation and Optimization Technologies (SPOT)*, Prix DATA Crossing du challenge D4D.

<sup>5.</sup> Genre-Grandpierre C., Josselin D. 2005, « Des transports à la demande pour répondre aux nouvelles formes de mobilité. Le concept de Modulobus », in Montulet B. et al., *Mobilités et temporalités*, Facultés Universitaires Saint-Louis, pp. 151-164.

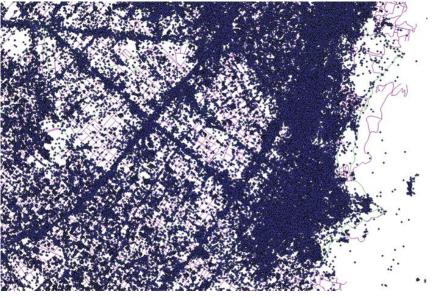


Figure 3. Répartition des tweets dans le centre-ville de Figure 3. Repartition des tweets dans le centre-ville Bogota en 2015 : la concentration de tweets dans la partie Est du centre-ville est représentative d'une localisation présentielle sous-estimée par les estimations de la population résidentielle (TOMSA, projet de coopération scientifique ECOS-Nord entre la France et la Colombie)

Parallèlement à ces problématiques relevant de l'aménagement, le big data est aussi mobilisable dans une perspective de personnalisation de services, qu'elle émane de la puissance publique ou, plus probablement, du secteur privé. L'objectif est ici de savoir comment adapter ou définir au mieux un service en fonction des besoins de l'individu. Par exemple, transmettre à un individu au temps t, en fonction de sa localisation, l'information utile ou souhaitable pour l'opérateur afin de l'avertir des opportunités à proximité (qu'elles soient fixes ou elles-mêmes mouvantes) peut avoir d'innombrables applications pour l'animation commerciale, l'optimisation de l'usage d'équipements publics, etc.

Pour autant, si l'émergence des données massives a coïncidé avec celle du concept de ville intelligente, c'est-à-dire une smart city<sup>6</sup> dont le développement (durable) s'appuie sur la mise en place de services numériques censés améliorer la qualité de vie, les utilisations potentielles du big data ne s'y résument pas.

### Quels impacts sur les pratiques de recherche?

Outre les problèmes de disponibilité réelle des données massives, l'usage du big data pour l'analyse des dynamiques urbaines ne va pas sans poser des questions méthodologiques et épistémologiques majeures.

Au niveau de la démarche scientifique, ces données massives conduisent à privilégier les méthodes inductives, notamment celles issues de l'intelligence artificielle (réseaux de neurones, réseaux bayésiens, etc.). Ces dernières sont en effet mieux à même de mettre en évidence les tendances, notamment les signaux faibles, qui constituent une information pertinente que l'on peut extraire des big data. On pourrait donc penser qu'il devient possible grâce au big data de se passer de théorie, pour se contenter de probabilités. Inutile, pour prendre ainsi les décisions adéquates, de savoir comment et pourquoi on contracte le cancer si, grâce à la fouille d'immenses jeux de données, on est capable de prédire la probabilité d'avoir telle ou telle maladie. Or, décrire et prédire n'est pas expliquer, qui reste l'objectif de la science. Il convient de ne pas perdre de vue cet objectif central qu'est l'explication (qui souvent passe par la modélisation) et de ne pas lui substituer seulement la description. Cette observation n'exclut cependant pas l'utilité des big data qui, dans une démarche explicative de modélisation, peuvent être utilisées aussi bien pour bâtir les modèles que dans une perspective confirmatoire.

En dehors de la finalité de l'utilisation des données massives, leur gestion, analyse et traitement posent des questions méthodologiques majeures. Les big data supposent d'adapter les outils traditionnels d'analyse (statistiques notamment), par exemple pour que les algorithmes intègrent d'emblée les problèmes liés à la grande taille des bases de données en trouvant des procédés permettant de limiter au strict minimum les étapes nécessitant la lecture de l'ensemble de la base. De même, les big data et les problématiques qui y sont liées, comme celle de la personnalisation des services pour un géomarketing très fin, nécessitent de revisiter de nombreux concepts, comme la gravitation, beaucoup plus basés sur l'établissement que

sur le mouvement. Pour ne prendre qu'un exemple, afin de mobiliser de façon pertinente les données en temps réel pour faire du géomarketing, tout reste à inventer ou presque en matière de méthodes. Les algorithmes nécessaires au traitement des données massives peuvent avoir une influence sur les résultats des analyses encore plus grande que les données elles-mêmes, sans que la méthode soit maîtrisée par l'analyste. Cela bouleverserait la connaissance géographique et les prises de décision qui se basent sur cette connaissance<sup>7</sup>. Plus largement, un des défis actuels est d'évaluer en quoi l'arrivée de big data peut permettre de revisiter les grandes questions de modélisation en géographie quantitative, telles que le problème écologique d'agrégation des données spatiales8.

Enfin, concernant le développement méthodologique, se pose la guestion de la place du géographe. Doit-il lui-même devenir informaticien pour développer des algorithmes toujours plus complexes ou est-il condamné à s'inscrire dans une logique de dépendance ? Plus que jamais des équipes pluridisciplinaires paraissent nécessaires.

#### Conclusion

Les big data représentent indéniablement une opportunité nouvelle pour l'analyse, la modélisation et l'aménagement des villes. Elles constituent une offre de données inédites qui, en complément des sources existantes, permettent de documenter des zones d'ombres des études urbaines et d'identifier des tendances émergentes.

Elles présentent en contrepartie certains risques9. La question de la dépendance de la recherche à des acteurs privés qui lui fournissent des données est ainsi posée, celle de l'adéquation de ces données et des algorithmes à la démarche scientifique l'est également. Enfin, les questions éthiques liées en particulier à la protection des données individuelles ne peuvent être ignorées.

> contact&info Christine Voiron. Espace Christine.VOIRON@unice.fr

<sup>6.</sup> Batty M. 2012, Smart cities, big data, in Environment and planning B: Planning & design, 39(2): 191-193.

<sup>7.</sup> Kwan M-P. 2016, Algorithmic Geographies: Big Data, Algorithmic Uncertainty, and the Production of Geographic Knowledge, in Annals of the American Association of Geographers, 106:2, 274-282

<sup>8.</sup> Louvet R., Josselin D., Genre-Grandpierre C., Aryal J. 2015, Impact du changement d'échelle sur l'étude des causes des feux de forêts du sud-est de la France, SAGEO'2015, Hammamet, Tunisie (2015).

<sup>9.</sup> Boyd D., Crawford K. 2012, Critical questions for big data, in Information, Communication and Society, 15(5): 662-679.

## Un programme international d'enseignement de la recherche en modélisation géographique

Marion Le Texier est spécialisée dans la collecte, le traitement et la modélisation de traces physiques et numériques des circulations internationales. Elle est membre du laboratoire Identité et Différentiation de l'Espace, de l'Environnement et des Sociétés (IDEES, UMR6266, CNRS / Université Caen Normandie / Université de Rouen Normandie / Université Le Havre Normandie). Geoffrey Caruso est spécialiste de la modélisation des formes et des dynamiques urbaines, d'économie géographique et de simulation spatialisée à base d'agents. Il est membre du laboratoire IPSE (Identités. Politiques, Sociétés, Espaces).



Du 6 au 17 février 2017, l'unité Identité et Différentiation de l'Espace, de l'Environnement et des Sociétés et l'unité Géographie-cités, en partenariat avec l'unité Théoriser et modéliser pour aménager, accueillent à Rouen deux semaines de formation dans le cadre du projet *Workshop and Moocs in Geographical Modelling* (MGM+e), qui réunit différentes universités en Europe (Université de Manchester, Université de Louvain-la-Neuve, Université du Luxembourg, Université de Lausanne, Université de Besançon, Université de Dijon, Université Paris 1 Panthéon Sorbonne, Université Paris Diderot, Université de Rouen). Il a pour objectif d'améliorer la qualité et la pertinence de l'apprentissage de la modélisation en géographie pour les étudiants en master et la création d'une plateforme de *e-learning*.

Ce partenariat entre trois laboratoires de géographie en lien avec d'autres laboratoires européens permet une mutualisation des formations de haut niveau pour les futurs doctorants en modélisation. Piloté par l'Université de Lausanne entre 2011 et 2013, ce programme KA2 – *Master Geographical Modelling Campus* + e, coordonné par l'Université du Luxembourg, a été renouvelé depuis 2015 dans le cadre du programme Erasmus +.

Chaque année, une trentaine d'étudiants issus de ces horizons divers sont réunis afin de bénéficier d'une formation intensive à la modélisation en géographie d'une durée de deux semaines. Le format pédagogique de la formation est conçu, chaque jour, comme une alternance de cours théoriques et de mise en pratique autour d'exemples issus des grands champs d'application, en géographie et en aménagement : gestion des risques, transports, écologie urbaine... Outre ces deux semaines de formation in situ, le projet vise à construire des contenus pédagogiques en ligne (e-tutorials) suivant une dynamique biannuelle : séance de cours et enregistrement la première année, test avec les étudiants et amélioration de l'e-tutorial la deuxième année.

### Principaux enseignements pour 2017

#### Modèles multi-agents

La modélisation à base d'agents est devenue l'un des outils phares pour étudier la complexité des systèmes spatiaux. Cette approche de modélisation permet de prendre en compte de façon explicite l'hétérogénéité et la spatialité des entités composant un système spatial. Ce tutoriel introduit les principaux concepts de la modélisation multi-agents spatialisée et propose une introduction pratique à la construction de modèles à partir d'un cas d'étude issu de l'épidémiologie.

Tutoriel animé par Arnaud Banos, Géographie-cités, et Patrick Taillandier, INRA Toulouse.

#### Automates cellulaires

Les modèles d'automates cellulaires ont un grand potentiel d'application aux études urbaines, d'utilisation du sol ou de transports car ils permettent de capter les interactions complexes au cœur de ces phénomènes. Ce tutoriel discute du concept et de la définition des automates cellulaires en général, puis de leur utilisation et de leur implémentation dans différents contextes géographiques et en recherche urbaine. Dans un premier temps, les étudiants apprendront, à partir d'une série de modèles simples, à comprendre et utiliser des automates cellulaires simulant l'émergence de structures complexes. Dans un second temps, ils passeront en revue les principaux facteurs de croissance urbaine et analyseront un modèle multi-scalaire à base de vecteurs développé pour l'aménagement urbain.

Tutoriel animé par Nuno Pinto, Université de Manchester, et Geoffrey Caruso, Université de Luxembourg.

#### Simulation appliquée à l'économie urbaine

Ce tutoriel vise à relier les modèles de simulation dynamique spatialisée dans lesquels des structures urbaines émergent d'interactions entre agents aux modèles standards d'économie urbaine.

<sup>1.</sup> IDEES, UMR6266, CNRS / Université Caen Normandie / Université de Rouen Normandie / Université Le Havre Normandie.

<sup>2.</sup> UMR8504, CNRS / Université Paris 1 Panthéon-Sorbonne / Université Paris Diderot.

<sup>3.</sup> Théma, UMR6049, CNRS / Université de Bourgogne / Université de Franche-Comté.

Un premier temps est dédié à la revue des concepts fondamentaux de l'économie urbaine (préférences, utilité, équilibre et optimum, etc.) et au modèle canonique d'Alonso, lequel est à la fois analysé mathématiquement et numériquement. Dans un second temps, des interactions spatiales locales sont introduites en 2D et de façon dynamique. Les étudiants utiliseront une version augmentée du modèle d'Alonso implémentée dans NetLogo afin d'analyser comment la préférence pour des espaces verts influence l'émergence de formes urbaines particulières. Ils analyseront enfin une version du modèle de Fugita-Ogawa implémentée dans NetLogo dans laquelle des interactions locales entre firmes existent et où plusieurs pôles infra-urbains émergent de ces interactions.

Tutoriel animé par Rémi Lemoy, Université de Luxembourg, et Justin Delloye, Université catholique de Louvain.

#### Flux et réseaux

Ce tutoriel vise à l'apprentissage de l'analyse, de la visualisation et de la simulation des réseaux géographiques. Il se décompose en cinq temps :

- initiation à l'approche multi-échelle des réseaux urbains,
- initiation à l'approche multidimensionnelle des réseaux urbains.
- mesures et visualisations de réseaux de villes,
- simulations des dynamiques de réseaux dans les systèmes urbains.
- exercices de visualisation.

Tutoriel animé par Céline Rozenblat, Université de Lausanne. Etutoriel construit avec Antoine Bellwald, Université de Lausanne.

#### Analyse fractale

Ce tutoriel vise à enseigner aux étudiants comment la modélisation fractale peut être utilisée afin de délimiter les agglomérations urbaines. Les objectifs pédagogiques sont triples :

- définir des critères (principalement morphologiques) mobilisables pour identifier des frontières urbaines,
- ▶ découvrir les fondamentaux de la géométrie fractale et de son application à la géographie urbaine,
- être capable d'utiliser deux logiciels d'application (Morpholim et FracGIS) dédiés à l'analyse fractale de zones bâties.

Tutoriel animé par Cécile Tannier, CNRS-Université de Bourgogne Franche-Comté, Chrono-Environnement, et Isabelle Thomas, Université de Louvain-la-Neuve, CORE.

#### Modèles de transport et d'utilisation du sol

L'objectif de ce tutoriel est d'améliorer la compréhension des comportements de mobilité quotidienne des individus à partir d'enquête Ménages et de données simulées. La première partie du tutoriel apporte une contextualisation théorique du lien réalisé entre la time-geography et les systèmes complexes. La seconde partie prend la forme d'un jeu sérieux dans lequel différents profils de villes durables sont conceptualisés à l'aide du prototype Smart access.

Tutoriel animé par Thomas Thévenin, Université de Dijon, Théma, Nicolas Marilleau, UMMISCO.

#### Interaction spatiale

Au cours de ce tutoriel, les étudiants recevront un apprentissage théorique général des concepts d'interaction spatiale (effets continus de la distance, du temps, de coûts, etc. sur les échanges et les flux) et d'interaction territoriale (effets discrets d'appartenance territoriale, culturelle, de voisinage, etc. sur les échanges et les flux). Un second temps d'apprentissage sera consacré à la (re-)découverte d'une famille de modèles statistiques (régressions

de Poisson sans contraintes, avec simple ou double contrainte) permettant de mesurer l'impact des facteurs et freins supposés à l'interaction spatiale et/ou territoriale. Ces enseignements théoriques seront complétés par des applications pratiques mises en œuvre par les étudiants à partir de bases de données variées (commerce international, corpus médiatique, etc.).

Tutoriel animé par Claude Grasland, Université Paris Diderot, Géographie-Cités, Marion Le Texier et Sophie de Ruffray, Université de Rouen, IDEES.

#### Modélisation spatialisée en écologie du paysage

Ce tutoriel présentera une introduction à l'écologie du paysage et à ces principaux concepts, ainsi qu'aux notions de graphes paysagers. Il comportera à la fois une entrée théorique et un apprentissage pratique avec l'usage des outils les plus récents d'analyse de graphes tels que Graphab.

Tutoriel animé par Xavier Girardet, Université de Besançon, Théma, et Céline Clauzel, Université Paris Diderot, Ladyss.

#### Systèmes de villes

L'objectif de ce cours est d'étudier les systèmes de villes des points de vue empirique, conceptuel et de modélisation. Les notions de complexité, d'auto-organisation, de bifurcation et d'émergence seront explicitées et leur sens sera discuté dans un contexte d'analyse et de modélisation de la dynamique d'un système de villes. Des sessions théoriques alterneront avec des applications en groupes : sous forme de brainstorming et de mise en œuvre de modèles de simulation.

Tutoriel animé par Lena Sanders, CNRS Géographie-Cités, Anne Bretagnolle, Université de Paris 1 Panthéon-Sorbonne, Géographie-Cités, Pierre Franckhauser, Université de Besançon, Théma, et Sébastien Rey-Coyrebourcq, Université de Rouen, IDEES.

#### Diffusion et épidémies

L'objectif de ce cours est un exposé des processus de diffusion définis selon différents contextes : sociologique, géographique et épidémiologique. Ils permettront aux étudiants de réfléchir à la construction d'une définition formelle de la diffusion et à son intégration dans des modèles agrégés, sous la forme d'équations différentielles et de modèles distribués, de type multi-agents. Un modèle épidémiologique de la dengue sera présenté, décomposé en trois sous modèles (hôte, vecteur et environnement). La construction pas-à-pas d'un modèle épidémiologique permettra de comprendre comment se fabrique un modèle de simulation multi-agents.

Tutoriel animé par Eric Daudé, Renaud Misslin, Thomas Huraux, CNRS, IDEES.

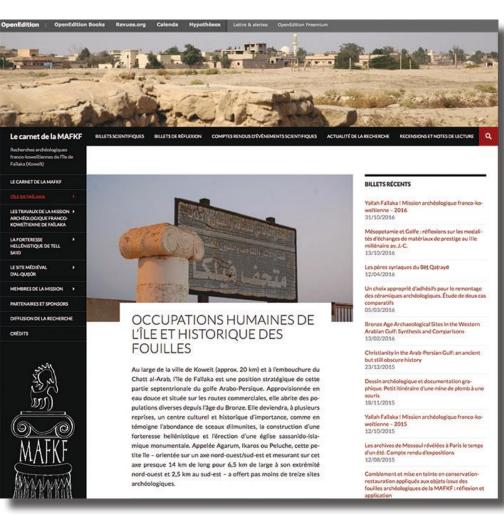
- ► Geoffrey Caruso, coordinateur du programme MGM+e (2015-2018)
- Marion Le Texier, UMR IDEES, coordinatrice du programme 2017

### contact&info

➤ Sophie de Ruffray, IDEES sophie.deruffray@univ-rouen.fr Marion Le Texier, IDEES marion.le-texier@univ-rouen.fr Geoffrey Caruso, Université du Luxembourg geoffrey.caruso@uni.lu ➤ Pour en savoir plus http://bit.ly/2jMx0hw

## UN CARNET À LA UNE

**Le carnet de la MAFKF.** Recherches archéologiques franco-koweïtiennes de l'île de Faïlaka (Koweït)



Créé en 2014, le carnet de recherche de la mission archéologique franco-koweïtienne de Faïlaka (MAFKF) accompagne la mission archéologique franco-koweïtienne de Faïlaka, au Koweït. Régulièrement occupée depuis l'âge du bronze, cette petite île de 43 km² a été un carrefour culturel et commercial d'importance qui concentre aujourd'hui un nombre exceptionnel de sites archéologiques. Les fouilles de la MAFKF portent sur deux de ses sites majeurs: la forteresse hellénistique de Tell Saʿīd et l'établissement chrétien médiéval d'al-Qusūr.

À travers ce carnet, les membres de la mission partagent l'avancée de leurs travaux avec les chercheurs, ainsi qu'avec le grand public. Au-delà, leur objectif est de mieux faire connaître l'histoire et l'actualité archéologique du Proche-Orient et de la région du Golfe.

Ainsi, le blog de la MAFKF accueille non seulement des publications produites par les membres de la mission, mais il offre également un espace d'information et de réflexion à tous les spécialistes de ce champ d'étude. Toutes les publications présentes sur le carnet de la MAFKF sont placées sous licence libre *Creative Commons*, facilitant ainsi leur diffusion vers un public le plus large possible.

Qu'ils soient experts ou non des questions abordées sur le carnet, les lecteurs du blog ont ensuite la possibilité de commenter les billets publiés et d'engager des discussions avec leurs auteurs sur des sujets variés, comme par exemple l'utilisation du dessin dans l'analyse archéologique.

Du fait de son fonctionnement largement ouvert et collaboratif, le carnet de recherche de la MAFKF participe activement au renouvellement des formes d'écritures scientifiques et de diffusion des connaissances

auprès d'un public qui dépasse les frontières académiques.

Céline Guilleux, Marion Wesely et François Pacaud

### contact&info

▶ Julie Bonnéric, Ifpo j.bonneric@ifporient.org

Pour en savoir plus

https://mafkf.hypotheses.org/ http://www.openedition.org/15630

contact&info

► François Pacaud CLEO / OpenEdition

francois.pacaud@openedition.org

Pour en savoir plus

http://www.openedition.org http://cleo.openedition.org

## <u>UN CARNET A LA UNE</u>

# la **ettre** de l'InSHS

- ▶ Directeur de la publication Patrice Bourdelais
- ▶ Directrice de la rédaction Marie Gaille
- ▶ Responsable éditoriale Armelle Leclerc armelle.leclerc@cnrs-dir.fr
- ▶ Conception graphique Sandrine Clérisse & Bruno Roulet, Secteur de l'imprimé PMA
- ▶ Graphisme Bandeau Valérie Pierre, direction de la Communication CNRS
- ► Crédits images Bandeau © Photothèque du CNRS / Hervé Théry, Émilie Maj, Caroline Rose, Kaksonen
- ➤ Pour consulter la lettre en ligne www.cnrs.fr/inshs/Lettres-information-INSHS/lettres-informationINSHS.htm
- ► S'abonner / se désabonner
- ► Pour accéder aux autres actualités de l'INSHS www.cnrs.fr/inshs

#### Institut des sciences humaines et sociales CNRS

• 3 rue Michel-Ange 75794 Paris cedex 16 •

ISSN: 2272-0243