

Les aventures du Big data au CNRS. Des résistances à la numérisation des données en SHS

Yannick Magnien, Dominique Vinck

Ce chapitre propose de faire un retour sur les conditions de réalisation d'une infrastructure numérique dédiée à la recherche en sciences humaines et sociales : la plateforme ISIDORE (Intégration de Services, Interconnexion des Données de la Recherche et de l'Enseignement)¹ créée au sein du CNRS, entre 2007 et 2010, par le TGE ADONIS (Très Grand Équipement pour L'Accès unifié des DOcuments et données Numériques des Sciences humaines et sociales). Cette réalisation procède d'une grande ambition d'ingénierie technique, à savoir le fait de tirer profit du Web des données ou web sémantique pour assurer plus d'interopérabilité entre les communautés de recherche en SHS. Toutefois, elle résulte également d'actions et de décisions qui tiennent à une gouvernance et à un pilotage politico-administratif complexe associant une multiplicité d'acteurs institutionnels (Ministère de l'enseignement supérieur et de la recherche, CNRS, mise en place de la loi LRU sur l'autonomie et la responsabilité des Universités, etc.).

Les acteurs impliqués dans la conception et la mise en œuvre d'ISIDORE, dont l'un des auteurs de ce chapitre qui fournit une bonne part des matériaux de cette enquête (YM²), se sont parfois heurtés à des changements apparus dans l'environnement du projet ainsi qu'à des oppositions au projet lui-même, ou à diverses formes de résistance de la part de certains chercheurs en SHS ou de responsables d'institutions de recherche, notamment vis-à-vis de l'ambition de mutualiser des moyens informatiques d'une discipline et d'un laboratoire à l'autre.

En rendant compte du cheminement sinueux et expérimental de la mise en œuvre de cette infrastructure de recherche, nous traiterons de la capacité de collectifs nombreux et hétérogènes à innover au sein d'un ensemble complexe d'institutions en vue de réaliser une infrastructure numérique d'envergure au moins nationale. Ce travail conduira à s'interroger sur les évolutions en cours dans le domaine des humanités numériques (*digital humanities*) et de l'innovation technologique dans les SHS.

ADONIS I : le rêve d'un grand projet d'infrastructure de recherche pour les SHS

En juillet 2004, la direction générale du CNRS français crée un comité des « très grands équipements scientifiques » et « grandes infrastructures » (TGE/TGI) pour programmer son action en termes de « grands équipements » qu'ils relèvent de son autorité ou auxquels le CNRS participe comme dans le cas du réseau RENATER créé en 1993. Ce

¹ <http://www.rechercheisidore.fr/> (dernière consultation le 15 novembre 2014).

² Que l'un des co-auteurs de cet article ait été acteur de ce projet introduit évidemment un biais, sinon une approche partielle, mais faute d'évaluation officielle et indépendante sur cette période, pourtant demandée, il est difficile d'en être autrement.

A paraître dans Sainsaulieu I., Saint-Martin A. (éd.). *L'innovation en eaux troubles, Vulaines sur Seine, Ed. du Croquant, 297-314.*

comité travaille avec les directeurs des instituts nationaux et des départements scientifiques du CNRS pour assurer la veille et préparer la programmation. Les TGE/TGI sont définis comme des outils ayant un impact scientifique à long terme, avec un effet d'entraînement technologique et de retombées économiques et sociales, à la pointe du développement scientifique et technologique et porteurs d'innovation scientifique, ouverts à une communauté d'utilisateurs au niveau national, voire mondial selon les cas, requérant des engagements pluriannuels nationaux ou internationaux. L'histoire de la physique (accélérateurs de particules et radiotélescopes internationaux), qui sert ici de modèle, et les développements récents dans les sciences de la vie notamment (avec les plateformes de génomique et de protéomique) laissent penser que ces TGE/TGI sont porteurs d'une forme de développement scientifique. Dans sa « charte », le comité précise ainsi que les TGE/TGI doivent recouvrir un ensemble cohérent d'outils mutualisés et géographiquement distribués dans le cadre d'une stratégie nationale.

La direction du CNRS n'est pas en reste vis-à-vis des SHS qui ne disposent jusqu'alors d'aucun TGE/TGI de ce genre. Elle promeut alors un rêve de grandeur et de visibilité internationale pour ces disciplines en créant le TGE ADONIS³, afin de constituer un espace unifié de navigation et d'intertextualité entre les documents produits et utilisés dans les SHS : thèses, revues, catalogues (d'exposition notamment), actes de colloques, inventaires de collections, publication de fouilles, littérature grise, archives, bases de données, etc. Dans son document fondateur, daté du 30 août 2004, ce TGE est supposé remédier à la dispersion qui existe dans ces disciplines (multiplicité des documents numériques, des expériences de numérisation et fragmentation des corpus en ligne). Dans les SHS, en effet, les chercheurs et les enseignants-chercheurs, depuis parfois plusieurs dizaines d'années, s'efforcent de tirer profit des technologies de l'information (création de corpus numériques et numérisation d'archives, bibliothèques numériques, stratégies d'informatisation des données, intégration de données de géo-référencement, outils d'indexation et d'édition numérique, conception d'application multimédias, etc.). Cela se traduit par une pléthore de réalisations dont la dispersion est estimée, *par le CNRS notamment, mais plus largement par toutes les institutions ayant vocation interdisciplinaire (EHESS, EPHE, Collège de France, réseau des MSH, etc.)*, être un obstacle à l'émergence d'un corpus documentaire unifié en ligne. Le TGE ADONIS devrait remédier à ce problème et assurer l'accessibilité comme la circulation de l'information pour les chercheurs – comme c'est supposé être le cas dans les autres disciplines. Étant le seul TGE en SHS, il est doté *a priori* d'une très grande importance pour ces disciplines. Les responsables du TGE ADONIS, désignés au sein du CNRS, se donnent alors pour ambition de faire changer la situation en combinant logique documentaire et logique de production de connaissances nouvelles. Ils se soucient aussi de prendre en compte les nouvelles formes de lecture afin de définir des interfaces adaptées.

De nombreuses études et chantiers sont engagés pour construire différentes « briques de solution » à intégrer au sein d'une architecture technique englobante, toutefois sans qu'aucun débat ne soit impulsé au préalable avec la communauté scientifique. L'ambition de la direction du CNRS est alors de mettre en ligne la production scientifique française dans les SHS. Après avoir évalué et fait le ménage (en suscitant des regroupements éditoriaux) parmi les 196 revues qu'elle subventionne (2,7 millions d'euros par an), elle crée un Centre d'édition numérique scientifique (CENS) pour aider les revues scientifiques à

³ Le Président de son Comité de pilotage M. Spiro, président du Conseil du CERN, était physicien des particules.

A paraître dans Sainsaulieu I., Saint-Martin A. (éd.). *L'innovation en eaux troubles, Vulaines sur Seine*, Ed. du Croquant, 297-314.

passer de l'édition papier à l'édition électronique. Elle enrôle 49 ingénieurs dans cette aventure. Les « meilleures » revues *en terme de visibilité disciplinaire fondée sur des indicateurs bibliométriques* (77/152) reçoivent des subventions pour leur éditeur « papier » ; les autres ne sont désormais soutenues que pour leur édition électronique. A la même époque, d'autres initiatives animent les communautés de chercheurs en SHS, notamment la création de *Revue.org*, plateforme collaborative d'édition électronique de revues en SHS, qui n'est pas intégrée au réseau du TGE ADONIS pas plus que ne sont mobilisés ses partenaires officiels (INIST et Persée)⁴.

Deux ans plus tard, en 2006, le TGE ADONIS est néanmoins considéré comme un échec⁵. Seules 14 revues sont alors mises en ligne et seulement de façon très partielle. Beaucoup d'argent semble avoir été dépensé (entre 600 000 et plus de deux millions d'euros). Un audit est engagé par la nouvelle direction scientifique du CNRS en SHS (Marie-Françoise Courel et Isabelle de Lamberterie) et confié à un Comité d'audit composé de quatre personnalités scientifiques (Jean-Pierre Dalbera, MCEM, Paris ; Benoît Habert LIMSI, Orsay ; Alexandre Moatti, BNUE, Paris ; Yves Pétrouff)⁶ qui choisit les consultants Daniel Confland et Norbert Paquel pour préparer un nouveau rapport d'orientation, livré en 2007, qui conduit à mettre en évidence des problèmes de pilotage, l'absence de débat large dans la communauté scientifique concernée, des choix technologiques de solutions éditoriales discutables et non évalués de manière indépendante mais aussi l'absence de stratégie claire de la part du CNRS (par exemple, le fait de soutenir à la fois la plateforme concurrente *Revue.org* et la plateforme *Érudit* du CENS). Les revues de SHS ont, de fait, résisté à l'ambition de numérisation du TGE ADONIS alors que les chercheurs en SHS, au niveau de leurs laboratoires et universités, passent progressivement à l'édition électronique et au dépôt de leurs références et publications sur le site Internet d'archives ouvertes HAL de la recherche française.

ADONIS II : la fondation socio-scientifique d'une infrastructure de recherche

En 2006, les auteurs de l'audit avancent une proposition de réorientation du TGE ADONIS⁷. Elle impliquerait une structure fédérative devant permettre une concertation plus étendue et la possibilité d'associer des institutions culturelles (INA, BNF)⁸, d'autres ministères (en particulier le Ministère de la culture, des partenaires privés (notamment la plateforme d'édition CAIRN) et de redonner sa place à l'INIST en charge de l'information scientifique. Le projet n'est évidemment pas simple. Le parallèle avec les sciences de la nature et la grande production de données en physique et en biologie (*Big Data*) est, de ce

⁴ Institut de l'information scientifique et technique (INIST - <http://www.inist.fr/>) ; Programme de publication électronique de revues scientifiques en sciences humaines et sociales (PERSEE - <http://www.persee.fr/web/support/apropos>).

⁵ Miserey Y., CNRS : le scandale d'une numérisation ratée, *Le Figaro*, 10/11/2006 http://www.lefigaro.fr/sciences/2006/11/10/01008-20061110ARTFIG90064-cnrs_le_scandale_d_une_numerisation_ratee.php (dernière consultation le 15 novembre 2014).

⁶ <http://amoatti.pagesperso-orange.fr/RapportAuditAdonis-Juillet2006.pdf>

⁷ <http://amoatti.pagesperso-orange.fr/RapportAuditAdonis-Juillet2006.pdf> (dernière consultation le 5 juin 2015).

⁸ Institut National Audiovisuel (INA - <http://www.ina.fr/>) ; Bibliothèque Nationale de France (BNF - www.bnf.fr/).

A paraître dans Sainsaulieu I., Saint-Martin A. (éd.). *L'innovation en eaux troubles, Vulaines sur Seine*, Ed. du Croquant, 297-314.

point de vue, moins pertinent que la prise en compte de la complexité qualitative, due à l'extrême hétérogénéité des données en SHS.

Quoi de commun, en effet, entre les données de l'archéologie byzantine (notices des objets présents dans les collections et descriptions des collections ; catégories et mots-clefs ; photographies d'objets, de parures, de monnaies, de monuments, d'icônes et de vues panoramiques ; descriptifs et vidéos de fouilles), les collections d'images et de translittération de textes anciens en philologie et les données quantitatives et longitudinales portant sur des cohortes en science sociales ?

Il n'est pas certain non plus qu'il y ait vraiment continuité entre humanités et sciences sociales. Là où des milliers de chercheurs physiciens sont sur un seul site, celui du LCH au CERN, à la poursuite du Boson de Higgs, en discriminant statistiquement des milliards de données produites par les détecteurs afin d'éliminer les bruits de fonds et confirmer la théorie énoncée par un groupe de physiciens au cours des années 1960, les SHS développent une multitude d'outils spécifiques, en particulier des bases de données ayant leurs propres règles d'organisation et ontologies⁹ en fonction de leurs objets (textes anciens, photographies, descriptifs, données de géolocalisation, fichiers audio et vidéo, etc.) et approches scientifiques.

En reprenant l'expérience du TGE ADONIS et la création d'un outil de recherche (ISIDORE) qui permet d'interroger une diversité de bases de données en SHS, nous voudrions montrer que l'hétérogénéité interne des collections d'objets traités par les SHS interroge la notion d'infrastructure de recherche et pèse inévitablement sur la dynamique de l'innovation dans ces disciplines. Les difficultés manifestes rencontrées lors de la réalisation de cette infrastructure (contraintes institutionnelles, changements politiques, idiosyncrasies des communautés de chercheurs, pratiques budgétaires, etc.) renvoient aussi à des questions sociales, matérielles et épistémologiques qui font la spécificité des SHS.

Début 2007, le CNRS recrute deux nouveaux responsables (Yannick Maignien, directeur, corédacteur du présent chapitre, et Benoît Habert, directeur adjoint), pour le TGE ADONIS¹⁰. Ce recrutement était organisé par une annonce publiée dans *Le Monde*, signe de la volonté du CNRS d'ouvrir la recherche de compétences. Benoît Habert, chercheur linguiste spécialisé dans les bases de données et les formats numériques textuels, avait participé à l'audit de la précédente expérience du TGE ADONIS. Yannick Maignien, professeur certifié de philosophie, détaché au service des études du Ministère de la culture, avait été responsable de la politique scientifique de numérisation de la BNF (création de la bibliothèque numérique *Gallica*) entre 1992 et 1997, aux côtés d'Alain Giffard.

L'héritage de cet échec conduit à vouloir différencier le régime de production des données et l'édition des articles scientifiques. Le paysage complexe des acteurs publics et privés dans l'édition des revues en SHS incitait les nouveaux responsables du TGE ADONIS à

⁹ C'est-à-dire l'identification de l'ensemble des objets considérés comme constitutifs d'un domaine particulier ; les concepts structurant ainsi la représentation du domaine permettent d'organiser des systèmes de connaissances.

¹⁰ Le TGE ADONIS, Unité Propre de Service (UPS), liée seulement au CNRS, deviendra en mars 2011 Unité Mixte de Service (CNRS/Université) sous le nom d'Huma-Num (<http://www.huma-num.fr/>). Elle est dirigée par Marc Renneville, auteur du projet numérique criminocorpus (<https://criminocorpus.org/>). >> Olivier Baude (un linguiste) a repris la direction d'Huma-Num, depuis cet été je crois.

A paraître dans Sainsaulieu I., Saint-Martin A. (éd.). *L'innovation en eaux troubles, Vulaines sur Seine, Ed. du Croquant, 297-314.*

procéder plutôt à la mise en place de larges concertations. Des acteurs comme *Revue.org*, *Cairn*, mais aussi l'INIST, ou *Persée* et les collègues québécois d'*Erudit*, étaient en plein développement de leurs propres plateformes et cette dynamique, souvent concurrentielle, demandait surtout à être accompagnée, aidée et coordonnée plutôt que dirigée. Le TGE ADONIS travaille désormais en réseau avec divers acteurs du numérique au sein des SHS, notamment des chercheurs porteurs de projets qui ont besoin d'infrastructures, mais aussi en direction des infrastructures d'information d'autres sciences¹¹ et de réseaux de documentalistes et bibliothécaires¹² (Pinol, 2011). Le rapprochement vis-à-vis des chercheurs-utilisateurs et producteurs de données et de publications (notamment le Centre pour l'édition électronique ouverte - CLEO) permet d'engager une dynamique d'apprentissage nourrie par leurs retours d'expérience. Ce faisant, les responsables du TGE ADONIS ont voulu poser la question du devenir numérique des outils intellectuels dans les SHS, au niveau de la production, de la communication et de la conservation des données de la recherche. Cela revenait aussi à poser aux chercheurs la question de la mutualisation et de l'interopérabilité, non seulement d'un point de vue technique (standardisation des formats, systèmes informatiques, etc.), mais aussi en prenant en considération leurs souhaits en matière de coopération. Cela engageait aussi la discussion au niveau des possibles complémentarités entre les recherches, les approches, les disciplines et les institutions. Les « études des sciences » (*science studies*) nous enseignent que, dans les autres sciences (chimie et biologie notamment), le chemin fût particulièrement long avant que les approches, méthodes et langages intra et inter disciplinaires ne convergent et ne forment la base des grandes collaborations internationales et des grandes infrastructures de recherche que l'on connaît aujourd'hui (Vinck, 1994, 1999 et 2013 ; Bowker & Star, 1999 ; Mongili & Pellegrino, 2014).

En d'autres termes, l'objectif qui était assigné à ce grand équipement (« unifier » les accès aux données et documents des SHS) devait d'abord prendre la mesure de l'hétérogénéité des conditions de production des données en SHS. Pour mener à bien ce projet, l'équipe de direction du TGE ADONIS version 2 s'appuyait sur le rapport d'audit, demandé fin 2006, qui préconisait la réalisation d'un méta-portail pour assurer cette unification. Il s'agissait de soutenir l'implication des communautés de recherche. Cette conception conduisait à faire du TGE ADONIS une structure « légère », provisoire (une équipe-projet), plutôt que de créer une institution de plus. Le projet était d'accompagner l'interaction entre des institutions existantes, souvent cloisonnées. Un bon exemple est le chantier de conservation pérenne des données numériques qui a fait intervenir, dans une structure opérationnelle, des organismes comme le CINES¹³, l'IN2P3¹⁴, la Direction des Archives de France, au service des premiers corpus expérimentaux de données d'archives sonores des laboratoires d'Aix et de Paris. Dans d'autres domaines de ce vaste chantier, des collaborations étaient ouvertes avec l'INIST, l'ABES et la BNF notamment, donc de façon

¹¹ Le Centre Informatique National de l'Enseignement Supérieur (CINES - <http://www.cines.fr/>) et le centre de calcul de l'Institut de physique nucléaire et de physique des particules (IN2P3 - <http://cc.in2p3.fr/>).

¹² Le Système universitaire de documentation (SUDOC - <http://www.sudoc.abes.fr/>), le Centre pour la communication scientifique directe (CCSD - <http://www.ccsd.cnrs.fr/>) qui gère le site d'archives ouvertes HAL-SHS, et l'Agence bibliographique pour l'enseignement supérieur (ABES - <http://www.abes.fr/>).

¹³ Centre Informatique National de l'Enseignement Supérieur.

¹⁴ Institut national de physique nucléaire et de physique des particules.

A paraître dans Sainsaulieu I., Saint-Martin A. (éd.). *L'innovation en eaux troubles, Vulaines sur Seine, Ed. du Croquant, 297-314.*

transversale à plusieurs administrations spécifiques de la culture, de l'enseignement supérieur ou de la recherche.

Le mode de gouvernance du projet allait toutefois se modifier rapidement et profondément, à mesure que se développaient des débats politiques portant sur le rôle « d'agence de moyens » que serait amené à tenir le CNRS, dans le cadre de la nouvelle loi LRU accordant plus d' « autonomie » aux Universités. Cela conduisit alors à une contestation du TGE par les tutelles, la Direction générale de la recherche et de l'innovation (DGRI), celle du Ministère de l'Enseignement Supérieur et de la Recherche (MESR), la cellule des TGE du CNRS et l'Institut national des SHS du CNRS. Une contestation d'autant plus paradoxale que le TGE ADONIS était en voie de devenir une infrastructure de recherche largement au service des équipes universitaires, où se fait la majorité de la recherche en SHS. Un jeu complexe de pouvoirs s'engage alors entre le ministère, le CNRS et la CPU (Commission des Présidents d'Université) sur les questions d'infrastructure de recherche et d'allocation des moyens, enjeux de territoire¹⁵ et de positionnement stratégique. Les collaborations européennes, notamment le projet d'ERIC DARIAH¹⁶ relèvent alors du ministère, même si celui-ci peine à comprendre la logique et les prérogatives des partenaires européens (allemands, anglais ou hollandais par exemple, partenaires du TGE). Cette période troublée, marquée par trois changements de direction scientifique en SHS en trois ans et par des conflits, fut peu propice à la réflexion que l'équipe du TGE ADONIS était en droit d'attendre.

En 2010, lorsque la direction du CNRS confie la direction du TGE à Jean-Luc Pino¹⁷, trois objectifs lui sont fixés, par le Directeur de l'Institut national des SHS (CNRS), dans la continuité de ce que le TGE a réalisé : structurer le réseau des centres de ressources numériques spécialisés (sur texte, image, son, vidéo, 3D, etc.) chargés d'accompagner les laboratoires dans leurs projets numériques (constitution de corpus, conception d'outils et méthodes), de promouvoir la standardisation des modes de production et de diffusion des données numériques SHS (harmonisation et interopérabilité) ; assurer l'accès et la préservation des données numériques au niveau de la grille de services (services interconnectés) offerts par le TGE ADONIS (hébergement, diffusion, conservation à long terme, traitement des données) ; et mettre en exploitation la plateforme de recherche ISIDORE. Nous voyons ainsi se profiler l'infrastructure de recherche comme assemblage socio-scientifique.

ISIDORE : conception d'une plateforme de recherche respectant l'hétérogénéité des données en SHS

Malgré ces conditions peu favorables, le TGE ADONIS était dès 2007 rapidement passé à l'action. La méthode utilisée associait les laboratoires les plus avancés dans des réalisations numériques, quels que soient leurs champs disciplinaires et donc l'hétérogénéité

¹⁵ Au sens de Andrew Abbott (1988).

¹⁶ ERIC (European Research Infrastructure Consortium) DARIAH (Digital Research Infrastructure for the Arts and Humanities, <https://www.dariah.eu/>). L'ERIC DARIAH sera créée seulement de novembre 2014.

¹⁷ En 2009, B. Habert quitte le TGE ADONIS pour des responsabilités sur les aspects collaboratifs du document numérique au sein d'EDF. Fin 2010, le CNRS met fin aux fonctions de Y. Maignien. En même temps, décembre 2010, la version bêta de la plateforme ISIDORE est définitivement ouverte.

A paraître dans Sainsaulieu I., Saint-Martin A. (éd.). *L'innovation en eaux troubles, Vulaines sur Seine*, Ed. du Croquant, 297-314.

de leurs données (géo-data, données textuelles, iconiques, sonores, statistiques, 3D, etc.) afin de définir un cahier des charges pour une infrastructure pertinente. La sélection et le soutien apporté à des projets pilotes ou innovants contribuaient à la fois au développement d'un réseau d'acteurs engagés dans le développement de l'utilisation du numérique dans les SHS et d'un moyen d'apprendre d'eux comment rendre interopérables leurs données et préparer ainsi, *in vivo*, les réalisations futures. Cette recherche-action par l'implication des laboratoires s'est faite par deux appels à projets successifs (une centaine de laboratoires répondant à chaque fois) ; à l'issue de la sélection, environ vingt-cinq projets par appel ont reçu un soutien financier pour un montant total d'environ 1,6 million d'euros.

Afin de donner toute l'efficacité à ce travail de définition des choix et contraintes de réalisation d'un méta-portail¹⁸, la direction du TGE s'était adjointe une assistance à maîtrise d'ouvrage (remportée et menée par une équipe d'Atos consulting). Cette aide en matière de gestion de projet¹⁹ a permis de préparer un appel à projet par marché public²⁰ qui a conduit au développement de solutions originales dans le cadre du Web sémantique pour le Web de données et l'interopérabilité recherchée pour ce méta-portail. Le choix d'un modèle commun (les triplets du modèle Ressource Description Framework) a permis de réaliser un moteur de recherche adapté aux SHS, tout en restant ouvert sur la pluralité des productions et éditions en SHS sur le Web. Ce moteur devait offrir la possibilité d'une interconnexion entre plusieurs services auxquels pourraient avoir recours les équipes SHS²¹ ainsi qu'entre les services que des centres SHS peuvent rendre à d'autres communautés. La question de la réutilisation des données pour de nouvelles recherches était au cœur de ce projet à la fois contributif (faisant appel aux contributions de nombreux utilisateurs chercheurs ou amateurs, parfois de type *crowdsourcing*) et collaboratif (renvoyant à l'implication de plusieurs groupes de recherche)²².

La procédure d'appel, rapidement lancée grâce à l'aide efficace de l'administration du CNRS - Paris B, a abouti au choix d'un consortium conduit par les sociétés Antidot²³ et Mondeca²⁴ pour le développement informatique et a permis de réaliser, dans les temps requis, la plateforme ISIDORE (Intégration de Services, Interconnexion des DONnées de la

¹⁸ Portail IWEB unifié, selon la recommandation du Rapport Confland-Paquel de 2006.

¹⁹ Hervé Gibert et Gautier Poupeau avaient l'expérience des systèmes d'information dans le domaine culturel (bibliothéconomie, édition, musée, archives, etc.). G. Poupeau rejoindra ensuite la société Antidot et jouera un rôle majeur dans la réalisation ISIDORE (Voir son site <http://www.lespetitescases.net>). Il est maintenant DATA Architect au sein de l'INA.

²⁰ Gestion renforcée par le recrutement de Stéphane Pouyllau, ingénieur CNRS.

²¹ Des disciplines comme l'archéologie par exemple, utilisant massivement la géolocalisation ou le traitement des formes et des images, ou les disciplines liées à l'étude et au traitement textuels sont souvent motrices. L'aller-retour entre les projets numériques de domaines spécifiques soutenus par des appels à projets et par leur généralisation top-down est l'une des caractéristiques majeures du TGE ADONIS.

²² Cet axe ne semble pas avoir connu depuis de véritable avancée, malgré le développement de « l'ouverture des données publiques ». Voir aussi le développement par Huma-Num d'un nouveau service d'exposition des données, Nakala <http://www.huma-num.fr/service/nakala>.

²³ <http://www.antidot.net/nos-clients/services-publics/antidot-information-factory-et-antidot-finder-suite-sont-au-coeur-disidore/>

²⁴ Mondeca est une société centrée sur les questions d'ontologie et leur expression dans le Web sémantique.

A paraître dans Sainsaulieu I., Saint-Martin A. (éd.). *L'innovation en eaux troubles, Vulaines sur Seine, Ed. du Croquant, 297-314.*

Recherche et de l'Enseignement)²⁵. Cette plateforme de recherche devait rendre possible le moissonnage des métadonnées et des données numériques produites par les bibliothèques, centres de recherche, centres de ressources numériques spécialisés et plateformes de l'édition électronique.

Paradoxalement, le contexte politique peu favorable évoqué plus haut et certaines réticences venant des milieux de la recherche vis-à-vis de l'*open access* (transparence des données, altruisme contributif que cela suppose) ont stimulé la conduite du projet (tenue des délais et des objectifs). Ainsi, alors que le budget CNRS du TGE devenait « virtuel », l'utilisation « éclair » de budgets du « Plan de relance » impulsé par le gouvernement de N. Sarkozy a permis de contourner les obstacles et de lancer les marchés pour une réalisation d'ISIDORE.

L'innovation en matière d'infrastructure de recherche supposait d'instruire des discussions de fond

La réalisation d'ISIDORE a bénéficié d'un « vide de pouvoir » dans la direction scientifique. Par contre, elle s'est heurtée à des résistances d'une partie des milieux de la recherche.

Au cœur des interrogations et des réticences d'une partie de la communauté scientifique figure la question de l'*open access* (ou *open data*), c'est-à-dire de la mise en commun des données de la recherche, et pas seulement de ses résultats. Partager des données jugées sensibles pour différentes raisons, notamment dans un contexte de compétition scientifique ou éditoriale, va d'autant moins de soi que l'appel à le faire est parfois porté par les prosélytes de l'*open access* faisant de celui-ci une bannière de la contestation et/ou de l'affirmation de nouveaux pouvoirs institutionnels. L'ouverture « publique » des données de la recherche ne peut faire (facilement) abstraction des conditions même de l'organisation du travail et des institutions des chercheurs. Un changement aussi radical des règles du jeu se heurte aux réticences d'une partie de l'institution universitaire et de recherche. Ni en physique, ni en astronomie, ni en biologie, les chercheurs n'ont mis spontanément en commun, du jour au lendemain, leurs données (Millerand & Bowker, 2009) ; ce fut, au contraire, un long processus d'exploration et de construction de consensus qui s'est déployé sur plusieurs décennies. La mise en commun de données, d'autant plus qu'elles sont hétérogènes et très sensibles à leurs conditions de production, conduit inévitablement les chercheurs à s'interroger sur leurs manières de produire de la connaissance robuste et sur leurs pratiques de recherche, à faire preuve de prudence vis-à-vis des tentatives de renversement accéléré de leurs référentiels épistémologiques et méthodologiques. La simple et belle idée de mutualiser ses données afin qu'elles puissent être réutilisées ne va pas plus de soi que la mutualisation des instruments de recherche coûteux en nanosciences et nanotechnologies (Hubert, 2014). Elle conduit à redéfinir potentiellement et dans une mesure qu'il reste à appréhender (Vinck et Natale, 2015), la discipline et les institutions du savoir en rapport notamment avec le

²⁵ Ce parti-pris de sous-traitance externe n'était pas sans poser de nombreuses difficultés internes au CNRS dont sa composante INIST, dédiée à l'information scientifique et technique, avait vocation à contribuer à un tel projet mais avait été écarté après d'âpres évaluations. Inversement, une collaboration étroite sur la maîtrise d'ouvrage informatique était nouée avec le CCSD de Villeurbanne, plateforme support des archives ouvertes HAL, dirigée alors par Daniel Charnay.

A paraître dans Sainsaulieu I., Saint-Martin A. (éd.). *L'innovation en eaux troubles, Vulaines sur Seine, Ed. du Croquant, 297-314.*

développement du Web. Or, ces débats n'ont pas été instruits avec la promotion d'une TGE pour les SHS.

Le recours au web sémantique pour ancrer une infrastructure de recherche pour les SHS soulève toutefois des myriades de questions portant sur la catégorisation des savoirs et l'indexation des données (conception et re-conception des thesaurus, référentiels et ontologies), qui conduisent les chercheurs à repenser collectivement les normes de la production et de la validation scientifique, les règles de la mise en circulation des informations, jusqu'aux missions et contours des institutions scientifiques. De même, les traces des relations (réseaux sociaux et appartenance) des chercheurs et de leurs publications, exprimables selon l'ontologie/le vocabulaire FOAF²⁶ dans le Web de données, pouvaient ouvrir d'intéressantes perspectives d'analyse mais aussi rendre possible l'évaluation de ces relations, ce qui pose légitimement question. Le numérique et le web semblent « dé-livrer » les savoirs du livre des institutions qui le tenaient (Vinck & Clivaz, 2014) avec son lot d'interrogations portant sur les formes de clôtures du savoir (le livre, l'article, l'encyclopédie, l'université, la discipline) et sur les risques liés à cette libéralisation des savoirs, comme l'emprise croissante de très grandes entreprises (Maignien, 2013), la montée en puissance de mouvements sociaux, notamment communautaires, dont certains farouchement opposés au primat de la rationalité scientifique dans la conduite du monde.

Les méfiances institutionnelles se comprennent d'autant plus qu'un tel projet d'infrastructure ne se limite en rien à un projet de moyens ; c'est bel et bien un projet scientifique, voir un nouveau projet de société, prônant par exemple l'ouverture de l'accès aux données. L'innovation n'est dès lors plus seulement une affaire d'optimum ou de progrès technique et d'ouverture au changement ; elle soulève des questions épistémologiques et sociétales fondamentales (Vinck, 2013). Comment, dès lors, penser le pilotage de l'innovation ? Il n'est dès lors pas surprenant aussi que se soient manifestées des formes de suspicion de la légitimité des porteurs du projet, perçus comme des « ingénieurs » s'arrogeant en sous-main des pouvoirs de « chercheurs ».

On comprend dès lors la portée de la remarque de M. Spiro, physicien et Président du Comité de pilotage du TGE ADONIS, au moment de la définition du cahier des charges d'ISIDORE, lorsqu'il pronostiquait : « *Si vous réussissez [à réaliser cette plateforme d'interopérabilité], alors ce sont toutes les sciences qui devraient bénéficier de cette avancée* » (intervention orale lors d'un Comité de pilotage qu'il présidait). En réalité, la conception et la réalisation de la plateforme ont réussi mais elles n'ont pas encore dissipé les interrogations de fonds. Le projet n'a d'ailleurs pas fait d'émules dans les autres disciplines²⁷.

Le débat ouvert par les infrastructures numériques en SHS illustre le fait que le numérique « déborde » les projets voulant l'instrumentaliser pour faire mieux et plus de choses. Au contraire, l'usage du numérique conduit à interroger l'activité scientifique et ses

²⁶ FOAF (*Friend Of A Friend*) est une ontologie ~~langage informatique~~ > non, ce n'est pas un langage informatique qui définit un vocabulaire de propriétés liées aux personnes et les relations structurées qu'entretiennent entre elles ces propriétés afin de décrire sur le web les personnes, leurs relations et leurs activités, notamment dans le cadre des réseaux sociaux.

²⁷ Cf. La prise de conscience du Big data au CNRS Mastodons 2012. *Revue du CNRS*, 2013 <http://www.cnrs.fr/mi/spip.php?article53&lang=fr> cf. aussi : « La déferlante des données » <http://www.cnrs.fr/fr/pdf/jdc/269/#/20>.

A paraître dans Sainsaulieu I., Saint-Martin A. (éd.). *L'innovation en eaux troubles, Vulaines sur Seine, Ed. du Croquant, 297-314.*

fondements parce qu'il ouvre de nombreuses incertitudes que les chercheurs doivent encore explorer collectivement avant de s'accorder sur de nouveaux référentiels et à requalifier éventuellement cette activité. La conception et la réalisation d'une « simple infrastructure de recherche » supposée « accompagner » les chercheurs dans leurs projets s'apparentent, en fait, à l'ouverture d'une boîte de Pandore. On comprend alors pourquoi des acteurs surgissent et se confrontent, des rêveurs, des prosélytes parfois arrogants, mais aussi des gardiens du temple, des chercheurs prudents, voire méfiants. Les collectifs de recherche avancent à tâtons sur un chemin qui n'est pas encore bien éclairé. Certains craignent le fait de voir le numérique subvertir le savoir et de rendre difficilement discernables le savoir et son outil. Certains mouvements autour du *Big data*, auxquels désormais les SHS entendent aussi contribuer, renforcent les interrogations, surtout cette idée que les nombreux agrégats de données statistiques remplaceraient les hypothèses et produiraient d'eux-mêmes des connaissances (*data-driven science* ?) (cf. le chapitre consacré au *Human Brain project* dans cet ouvrage)²⁸.

Conclusion

Cette étude de cas montre qu'en matière d'innovation au niveau d'infrastructures de recherche, si la communauté scientifique concernée est très hétérogène (en termes d'objets, d'approches et d'institutions), la dynamique d'innovation technologique devient un défi en termes de construction politique, sociale et épistémologique ; la réussite de l'infrastructure dépend largement de la capacité à créer un « monde commun » (Callon, Lascoumes, Barthe, 2001) à partir d'acteurs engagés dans des dynamiques sociotechniques *a priori* très peu convergentes. En outre, le cas laisse entrevoir ce que l'on sait déjà grâce aux *infrastructure studies* (Edwards et al, 2009) à savoir qu'une infrastructure n'est pas seulement une base matérielle ou un cadre au sein duquel se déploie une activité. Non seulement, l'infrastructure tient à de nombreuses formes et activités sous-jacentes, telles que la constitution de catégories, de classements et d'ontologies que la matérialité de l'infrastructure ne fait parfois qu'incarner, mais, en outre, l'infrastructure est constituée par des activités. Comme l'énoncent Star et Ruhleder (2010), l'activité des uns constitue l'infrastructure des autres. Il y a un lien de récursivité entre les pratiques qui se déploient dans et avec les infrastructures et les performances du processus d'infrastructure (Mongili & Pellegrino, 2014).

La réalisation d'une infrastructure de recherche telle que la plateforme ISIDORE est au croisement d'enjeux épistémologiques et politiques (notamment la définition et la reconnaissance des disciplines). Elle soulève des questions de politique de la recherche qui vont bien au-delà de la mise en œuvre de bonnes méthodes de gestion de projet. Elle interroge les régimes de science en train de se redéfinir (régime disciplinaire versus régime utilitaire ou technico-instrumental) (Shinn & Marcovich, 2012) et les frontières des sciences et de leurs institutions en train d'être redéfinies. Ces conditions affectent la réalisation d'une infrastructure de recherche et questionnent le pilotage d'innovations qui changent les fondements scientifiques et institutionnels des manières de « faire science » en SHS. De ce point de vue, le pilotage politico-administratif (ministère, CNRS) apparaît avoir été défaillant eu égard à de tels enjeux. Ce « pilotage » vertical, centralisé et étatique, ne répond pas aux

²⁸ Benoît Habert, « Face aux big data, les SHS sauront-elles prendre le tournant numérique ? », 29 janvier 2014 <http://archinfo01.hypotheses.org/857>, (dernière consultation le 5 juin 2015).

A paraître dans Sainsaulieu I., Saint-Martin A. (éd.). *L'innovation en eaux troubles, Vulaines sur Seine*, Ed. du Croquant, 297-314.

régulations complexes qu'un tel projet d'infrastructure requiert. Les SHS, sur ce registre, nécessitent des interactions plus démocratiques, plus constantes sur le plan scientifique et plus attentives socialement pour que l'innovation technologique puisse se déployer²⁹. Le débat porté par les « humanités numériques » devrait probablement s'attacher à ces questionnements plutôt que de contribuer à la course aux prouesses technologiques et à la reconnaissance de ceux qui ont « les mains dans le code ». En tout état de cause, l'expérience relatée ici devrait inciter les chercheurs engagés dans les humanités numériques à s'interroger sur le sens même des processus d'innovation auxquels ils contribuent.

Bibliographie

ABBOTT A., 1988, *The System of the Professions. An Essay of the Division of Expert Labour*, Chicago, Chicago University Press.

BOWKER G., STAR S., 1999, *Sorting Things Out: Classification and Its Consequences*, Cambridge, MA: The MIT Press.

CALLON M., LASCOUMES P., BARTHE Y., 2001, *Agir dans un monde incertain. Essai sur la démocratie technique*, Paris, Le Seuil.

EDWARDS P., BOWKER G., JACKSON S., WILLIAMS R., 2009, Introduction: An Agenda for Infrastructure Studies, *Journal of the Association for Information Systems*, n° 10, p. 364-374.

HUBERT M., 2014, *Partager des expériences de laboratoire*, Paris, EAC.

MAIGNIEN Y., 2000, Quel travail intellectuel dans l'ère numérique ?, *Esprit*, vol. 262, N° 3-4.

MAIGNIEN Y., 2011, « ISIDORE, de l'interconnexion de données à l'intégration de services ». HAL https://halshs.archives-ouvertes.fr/sic_00593320

MAIGNIEN Y., 2013, Quelles redistributions de pouvoirs autour des « automates sémantiques » ?, M. CARMES et J.M. NOYER (dir.), *Les débats du numérique*, Paris, Edition de l'Ecole des Mines.

SHINN T., MARCOVICH A., 2012, Regimes of Science Production and Diffusion : Towards a Transverse Organization of Knowledge, *Scientiae Studia*, vol. 10, p. 33-64

MONGILI A., PELLEGRINO G., 2014, *Information Infrastructure(s): Boundaries, Ecologies, Multiplicity*, Newcastle u/T, Cambridge Scholars Publishing.

MILLERAND F., BOWKER G., 2009, Metadata Standards. Trajectories and Enactment in the Life of an Ontology., M. LAMPLAND and S.L. STAR, *Standards and Their Stories. How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*, Ithaca and London, Cornell University Press, p. 149-165.

PINOL J.-L., 2011, Une infrastructure pour les SHS : le TGE Adonis, *Revue d'histoire moderne et contemporaine*, vol. 5, n° 58-4bis, p. 90-100.

STAR S.L., RUHLER K., 2010, Vers une écologie de l'infrastructure. Conception et accès aux grands espaces d'information, *Revue d'Anthropologie des Connaissances*, vol. 4, n° 1, p. 114-161.

²⁹ On trouvera un écho de cette exigence dans le petit livre de Michel Viewiorka, « L'Impératif numérique », CNRS Editions. 2013.

A paraître dans Sainsaulieu I., Saint-Martin A. (éd.). *L'innovation en eaux troubles, Vulaines sur Seine*, Ed. du Croquant, 297-314.

VINCK D., 1994, La normalisation au cœur des réseaux de recherche, *Culture Technique*, n° 29, p. 167-181.

VINCK D., 1999, Les objets intermédiaires dans les réseaux de coopération scientifique. Contribution à la prise en compte des objets dans les dynamiques sociales, *Revue Française de Sociologie*, vol. XL, n° 2, p. 385-414.

VINCK D., 2013, Pour une réflexion sur les infrastructures de recherche en sciences sociales. A propos de Brian KLEINER, Isabelle RENSCHLER, Boris WERNLI, Peter FARAGO and Dominique JOYE (eds). *Understanding Research Infrastructures in the Social Sciences*. Zurich : Seismo, *Revue d'Anthropologie des Connaissances*, vol. 7, n° 4, p. 991-999.

VINCK D., CLIVAZ C., 2014, Les humanités délivrées. Savoir et culture réinventés hors du livre, *Revue d'Anthropologie des Connaissances*, vol. 8, n° 4, p. 681-703.

VINCK D., NATALE E., 2015, La transformation des sciences historiques. La part du numérique, In A. GORGA et J.Ph. LERESCHE (éd), *Disciplines académiques en transformation : entre innovation et résistance*, Paris, Edition des Archives Contemporaines, pp. 89-104.

Wieviorka M., 2013, *L'impératif numérique ou La nouvelle ère des sciences humaines et sociales ?* CNRS Editions.