



HAL
open science

L'IDHN : une structure innovante au service de la polysémie du numérique

Marianne Froye, Olivier Belin, Julien Longhi, Boris Borzic, Claudia Marinica, Michel Jordan, Basile Michel

► To cite this version:

Marianne Froye, Olivier Belin, Julien Longhi, Boris Borzic, Claudia Marinica, et al.. L'IDHN : une structure innovante au service de la polysémie du numérique. *Humanistica* 2020, May 2020, Bordeaux, France. hal-02875614

HAL Id: hal-02875614

<https://hal.science/hal-02875614v1>

Submitted on 19 Jun 2020

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

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

L'IDHN : une structure innovante au service de la polysémie du numérique

M. Froye, O. Belin, J. Longhi, B. Borzic, C. Marinica, M. Jordan, B. Michel

Humanistica 2020

Bordeaux, mai 2020

Plan

1. Présentation du contexte : origine, motivation et fonctionnement
2. Recherche hors les murs : émergence d'une méthodologie scientifique, interdisciplinarité structurante, variabilité scientifique
3. Success story à poursuivre : enseignement, développement d'outils et de partenariats

Créée dans le cadre des universités nouvelles en 1991, CY Cergy Paris Université est une université de l'Ouest parisien. Implantée principalement dans trois villes du Val-d'Oise : Cergy, Pontoise et Neuville-sur-Oise, elle comporte 5 pôles thématiques, 23 laboratoires et 2 structures fédératives dont l'Institut des Humanités numériques (IDHN). L'IDHN est une structure fédérative en Humanités numériques (HN) qui a été créée en 2017 en lien avec le Domaine d'intérêt Majeur (DIM) Sciences du texte et connaissances nouvelles (STCN) porté par Paris Sorbonne Université et dont CY Cergy Paris Université est membre fondateur et partenaire. Quatre laboratoires y intègrent d'emblée leurs projets et actions en HN : les trois équipes d'accueil « Lexiques, Textes, Discours, Dictionnaires » (LT2D), « Mobilités, Réseaux, Territoires, Environnement » (MRTE) et Agora représentant un potentiel d'une centaine de chercheurs en Lettres et SHS et « Équipes Traitement de l'Information et Systèmes » (ETIS), Unité Mixte de Recherche travaillant dans le domaine informatique, qui regroupe 60 permanents dont 8 ingénieurs et 40 doctorants. Ces quatre laboratoires couvrent la majorité des disciplines en SHS : géographie, Histoire et Civilisations, Littérature, sciences du langage, ainsi que des domaines de l'intelligence artificielle : la science de données, la représentation de connaissances, et la gestion de données. Après avoir rappelé les origines de la création de la structure fédérative, il convient de montrer les enjeux d'une recherche hors les murs des laboratoires pour finalement comprendre les défis qui attendent ses membres dans les prochaines années.

1. Présentation du contexte

a. Origine et motivation

Partant du constat que de nombreux projets fleurissaient en HN¹ à CY et que des chercheurs de différentes disciplines se posaient des questions similaires, il a semblé opportun de favoriser les confrontations scientifiques et méthodologiques dans une structure institutionnalisée. Les HN présentent des défis de plusieurs ordres aux chercheurs. Scientifiques, certes puisqu'elles apparaissent bien souvent comme une transdiscipline. Il convient alors aux chercheurs de franchir les frontières de leur section disciplinaire² pour s'ouvrir aux pratiques et aux coutumes d'autres disciplines plus ou moins proches. Épistémologiques puisqu'elles renouvellent et interrogent en profondeur les disciplines, leurs frontières et leurs domaines d'extension. Culturels puisqu'elles remettent en cause les pratiques et les coutumes des chercheurs. La naissance de l'IDHN est suscitée par la volonté de partager des outils et des approches. Elle est vécue par ses personnels comme la possibilité d'inventer de nouvelles formes de recherche et de collaborations de façon à créer un écosystème favorable à un dialogue interdisciplinaire³ et pluridisciplinaire. L'enjeu *in fine* est de développer l'innovation scientifique et culturelle.

CY Cergy Paris Université est engagée depuis de nombreuses années dans les innovations du numérique. Elle est membre fondateur du LabEx Patrima et de l'EquipEx Patrimex, réunis désormais en la Fondation des Sciences du Patrimoine. Participer à la création de la fondation dont les objectifs mêmes portent sur les procédés de conservation et de restauration amènent nécessairement à pratiquer et à réfléchir aux enjeux du numérique⁴. Il a nécessairement une place de choix dans les pratiques scientifiques de recherche en SHS, pour améliorer les techniques de diffusion des connaissances. L'usage du numérique est tout

¹ « Humanités numériques » entendues au sens où François Rastier les définit comme « sciences de la culture » in « Doxa et lexique en corpus pour une sémantique des idéologies », *Texto !*, www.revue-texto.net/Inedits/Rastier/Rastier_Doxa.html et P. Mounier, *Les humanités numérique. Une histoire critique*, Paris, Éditions de la maison des sciences de l'homme, 2018.

² Entendu comme section du Conseil national des universités.

³ Cf. Edgar Morin, « Sur l'interdisciplinarité », <http://ciret-transdisciplinarity.org/bulletin/b2c2.php> où il rappelle qu'« il arrive même qu'un regard naïf d'amateur, étranger à la discipline, voire même à toute discipline, résolve un problème dont la solution était invisible au sein de la discipline. »

⁴ Entendu au sens de Pierre Mounier, *Ibid.*, de « méthode ». Le numérique apporte une méthode et modifie les enjeux épistémologiques de la recherche en humanités.

désigné pour participer à cette progression. De fait, le numérique a accompagné bon nombre de projets très divers comme PatrimEph¹, #ideo2017², PARCOURS³ ou les Testaments de Poilus⁴. Dans tous les cas, la numérisation et le traitement informatique des données de sciences humaines suscitaient des interrogations similaires aux chercheurs. Très rapidement, le besoin de fédérer les différents projets en HN au sein de l'université devenait prégnant pour mettre en commun les savoir-faire et les pratiques de chaque chercheur dans des disciplines différentes. La structuration en HN Lettres-SHS qui s'est rapidement imposée permet de donner une plus grande visibilité à ce champ de recherche en interne et envers des partenaires extérieurs et d'apporter des solutions plus rapidement efficaces aux nouveaux projets. Une dynamique renforcée de ces projets par leur rapprochement dans le cadre de la programmation stratégique de l'IDHN, avec le souci de la valorisation, de la formation doctorale et de la diffusion scientifique conduit progressivement à la mise en place d'une véritable équipe d'ingénierie et de recherche (dont les doctorants) en humanités numériques.

b. Fonctionnement

D'un point de vue structurel, pour assurer son fonctionnement, l'institut s'est doté d'un conseil de direction et d'un conseil scientifique où les principaux projets en HN sont représentés. L'IDHN bénéficie du soutien de deux ingénieurs, l'un d'étude dans le cadre du DIM et l'autre de recherche mis à disposition par le CNRS à ETIS. L'IDHN dispose de fonds propres pour assurer l'achat de matériel comme un casque de réalité virtuelle ou des serveurs et de services comme les noms de domaines. Son existence est sous-tendue par deux objectifs principaux : technologiques et techniques pour garantir la possibilité de déployer des projets en HN ; scientifiques, épistémologiques et doctoraux.

¹ PatrimEph vise à valoriser et à étudier le patrimoine méconnu des éphémères. <https://www.fabula.org/atelier.php?PatrimEph>

² #ideo2017 a pour objectif de créer un outil d'analyse des tweets politiques lors de campagnes politiques. <http://ideo2017.ensea.fr/>

³ PARCOURS a pour objectif de développer un référentiel commun dans le domaine de la conservation et restauration des œuvres d'art et de développer un système d'interrogation des bases de données dans ce domaine. <http://projet-parcours.eu/>

⁴ Testaments de poilus s'attache à faire connaître un corpus inédit : les testaments des poilus morts pour la France. Le programme de recherche a abouti à la création d'une plateforme de transcription collaborative. <https://cahier.hypotheses.org/testaments-de-poilus>

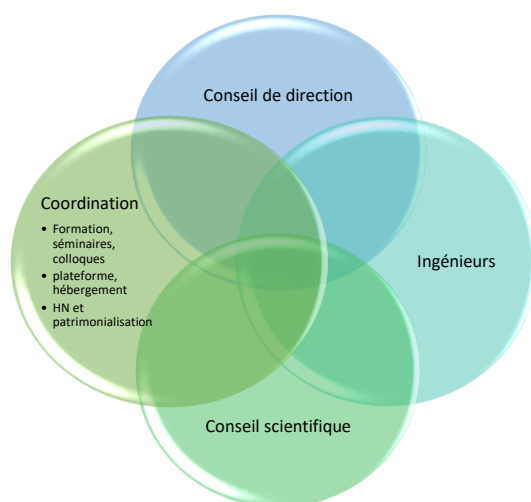


Figure 1 - Organigramme de l'IDHN

En externe, l'IDHN devient l'interlocuteur privilégié pour répondre à des appels d'offres. Ainsi, les chercheurs porteurs de projets peuvent bénéficier par l'entremise de l'IDHN des services et outils de la TGIR Huma-Num puisque certains projets en sont membres par le biais des consortiums. La mise en commun des outils facilite l'essor et les avancées des projets. Depuis le début de l'année 2018, l'IDHN a structuré la recherche en HN initialement au sein de la COMUE et du Grand établissement Paris Seine, puis de CY Cergy Paris Université. La visée scientifique est construite autour d'objectifs scientifiques, épistémologiques, et technologiques. L'IDHN a vocation à construire un hébergement commun des recherches à CY Cergy Paris Université en HN Lettres-Sciences humaines et sociales et leur donner une visibilité réelle ; à permettre des synergies pour démultiplier les moyens d'actions par des réponses communes aux appels à projets internes (Fondation partenariale, LabEx Patrima et sa Fondation, CYAS¹) et externes (DIM, investissement région, PhD région, SESAME, CoFund, Dariah, DAHLIA²). Conçu dans son démarrage comme un incubateur autant qu'un accélérateur de projets, l'IDHN a déjà travaillé à des rapprochements scientifiques internes à l'université et les premiers échanges ont d'emblée mis en valeur les synergies possibles. Il sert donc à développer les recherches en HN et à en diffuser et valoriser les résultats pour imaginer de nouveaux projets ; il permet d'accueillir et d'héberger le site et les recherches fédératrices des DIM STCN et PRADA pour lesquels les ingénieurs de recherche ont un rôle structurant et

¹ CY Advanced studies, institut des études avancées.

² <http://dahlia.egc.asso.fr/>

formateur. Ils font bénéficier de leurs compétences et pratiques l'ensemble de la communauté scientifique du site par capillarité. Inversement, ils bénéficient eux-mêmes des apports successifs d'autres personnels et chercheurs recrutés dans différents projets. L'existence de l'IDHN permet donc une mutualisation des moyens humains dans le domaine des HN : les différents projets de recherche sont gérés non plus seulement à l'échelle du seul projet, mais dans la perspective d'un écosystème global, ce qui permet de recruter des spécialistes de tel ou tel aspect des HN, et de les faire travailler sur plusieurs projets de recherche de manière transversale.

Une petite dizaine de projets de dimension, maturation et nature variées ont été identifiées en première approche, qui mobilisent des collaborations à géométrie variable entre les quatre unités. À ce titre, l'IDHN assure une expertise en formation dans le domaine des HN et en est un prestataire référent. Pour cela, des séminaires et écoles de printemps sont organisées : elles articulent l'intervention de chercheurs de CY spécialistes des différents domaines, et l'invitation de chercheurs, notamment grâce au système de professeur invité proposé par CYAS. L'IDHN favorise le développement des projets déjà engagés à l'interface entre les quatre laboratoires ou contribue à les faire aboutir dans de meilleurs délais. PatrimEph, « Testaments de Poilus » soutenus par le LabEx Patrima ou Lemel¹ qui bénéficie d'un financement européen ont connu leur aboutissement grâce à l'IDHN, « Frénaud numérique »² soutenu par le consortium Cahier³ ou le projet MEDET-LAT de cartographie de la circulation et de la réception internationale de la littérature d'Amérique Latine voient leur maturation arriver à terme grâce à l'IDHN.

¹ Lemel, « L'Europe dans les Médias en ligne » est une étude comparative, internationale et en réseau. Son objectif est de connaître l'Europe présentée, décrite, analysée, représentée, imaginée dans les médias en ligne dans plusieurs pays européens. <https://lemel.hypotheses.org/>

² « Frénaud numérique » est un projet de numérisation, de transcription et d'encodage de deux séries de manuscrits d'André Frénaud déposés à la Bibliothèque Littéraire Jacques Doucet à Paris. www.frenaudnumerique.com

³ Cahier : consortium de la TGIR Huma-Num « Corpus d'Auteurs pour les Humanités : Informatisation, Édition, Recherche ». <https://cahier.hypotheses.org/>

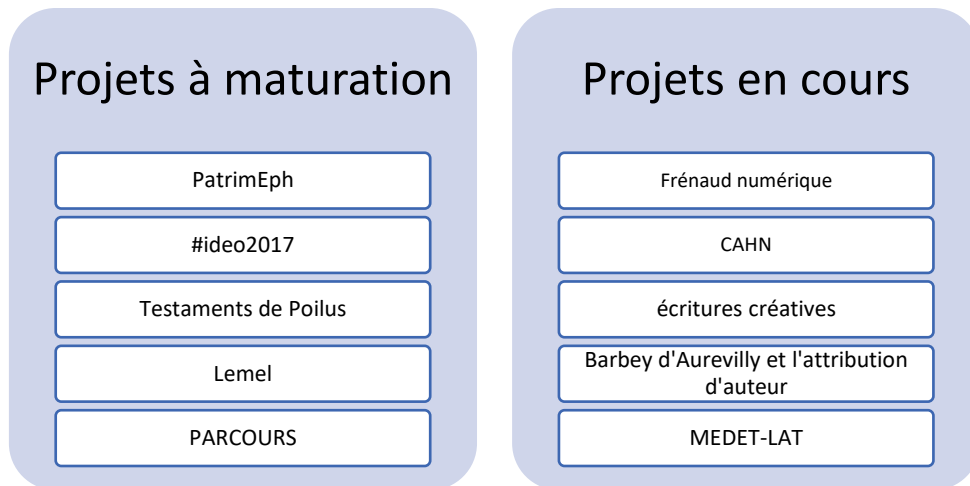


Figure 2 - État des lieux des projets

L'IDHN s'est doté de moyens informatiques propres, dont un serveur de calcul installé dans les locaux d'ETIS. Ce serveur, polyvalent et évolutif, héberge les sites web des projets de recherche de l'IDHN, mais aussi les bases de données ou de documents partagées entre les partenaires de ces projets (par exemple les données des projets « Frénaud numérique » et MEDET-LAT). Les capacités de calcul de ce serveur permettent également d'indexer et de fouiller ces bases de documents.

2. Recherche hors les murs

a. Émergence d'une méthodologie

Si le laboratoire est perçu comme une « entité sociocognitive pertinente »¹ par la majorité des acteurs de la recherche, fusionner ou transformer le paysage organisationnel de la recherche locale pouvait éventuellement apporter des réserves de la part de ses acteurs, craignant un niveau administratif additionnel. Or, loin d'être perçue comme une entité organisationnelle supplémentaire, la structure fédérative a au contraire créé des dialogues inter-laboratoires. L'enjeu principal était de susciter, certes des collaborations au sein de différents laboratoires en SHS, mais surtout de favoriser un travail effectif entre les chercheurs de SHS et ceux d'informatique. L'intention première était de dépasser l'idée de l'informatique comme outil pour innover ensemble. La structure met donc en commun des synergies qui

¹ Dominique Vinck et Gloria Zarama, « La fusion de laboratoires. Processus de gestion et constitution d'une entité pertinente de l'activité scientifique », *Revue d'anthropologie des connaissances*, volume 1, n°2, 2007, p. 268.

permettent les transferts méthodologiques en dehors des limites disciplinaires propres à chacun des chercheurs. Le lieu de rencontre physiquement déterminé par le local qui lui est dédié offre plusieurs effets positifs comme la mise en commun des méthodologies en faveur de l'émergence harmonieuse de bonnes pratiques. Dans le cadre de différents projets en histoire et en littérature, plusieurs chercheurs ont recours à l'encodage de leurs données en XML-TEI. L'ingénieur recruté grâce au DIM a permis de centraliser les protocoles d'encodage, de faire du lien entre des questionnements numériques similaires sur des données très différentes. Il s'agissait par exemple dans un cas d'encoder les testaments des poilus et de la poésie contemporaine dans l'autre. Mais le choix des balises se rejoignait : comment encoder les noms de personnes selon que c'est une entité culturelle, que sa dénomination se fait par le biais d'une reprise anaphorique ou que le lien familial permet de l'identifier ? Les types de dénomination sont les mêmes quels que soient les supports historiques ou littéraires, le choix des balises aussi.

Autre apport méthodologique : le recours à la visualisation cartographique des données. Si la représentation spatiale était usuelle pour les géographes, elle s'est avérée productive pour toutes les autres disciplines pour montrer des résultats d'analyse des données. La cartographie est une nouvelle pratique interdisciplinaire favorisée par l'IDHN. Elle permet par exemple de suivre le cheminement d'un poète dans Rome, d'analyser les trajectoires des visiteurs dans un musée ou de suivre les représentations de la Grande Guerre dans les médias à travers l'Europe. C'est pourquoi l'IDHN a développé le projet CAHN (Cartographie et Atlas en Humanités Numériques), qui a permis de renforcer la cohésion des travaux de l'équipe autour de cet aspect, et d'envisager la cartographie numérique non seulement comme outil de visualisation et de valorisation de données ou de corpus, mais aussi comme instrument heuristique pour stimuler de nouvelles hypothèses. Un des principaux centres d'intérêts des chercheurs d'ETIS est l'analyse de données et la gestion des connaissances. En collaboration avec les collègues des sciences humaines, ces derniers s'attaquent aux verrous scientifiques qui se dégagent de la rencontre entre les techniques d'analyse de données et les données disponibles dans le domaine des SHS.

b. Interdisciplinarité structurante et variabilité scientifique

Le va-et-vient interdisciplinaire structurant favorisé par le pôle ingénierie d'ETIS qui prend en charge une partie de l'ingénierie des projets de recherche de l'IDHN permet une efficacité accrue. La concentration en un même site intellectuel et physique permet une présence optimisée sur le web et facilite l'hébergement des données. Cette pratique de l'interdisciplinarité a rendu possible la conception d'outils spécifiques et transférables, par exemple pour l'analyse de données multimédia. La diversité des corpus : images, sons, vidéos, textes et modèles 3D posent des questions de visualisation, d'indexation et de fouille dans les bases de données. Elle nécessite d'organiser et de classer les connaissances. L'élaboration d'une plateforme générique à tout type et tout format de données est en cours. Elle favoriserait une visibilité commune de tous les projets et bénéficierait aux futurs projets qui n'auraient plus à anticiper la diffusion de leurs résultats. Les données du patrimoine historique et culturel sont un champ d'application majeur pour ces projets. Le numérique change le paradigme herméneutique, puisque, comme le rappelle Damon Mayaffre¹, il change la perception des textes et plus largement des corpus. L'ordinateur n'est plus seulement le moyen de trouver ce que l'on pressentait d'un corpus donné ou de confirmer une intuition *a priori*, mais bien de découvrir de nouveaux éléments saillants. Les principaux projets s'organisent désormais autour d'un concept essentiel à l'élaboration d'une réelle interdisciplinarité : la variabilité scientifique à travers l'archivage des données des chercheurs, leur accompagnement vers la maîtrise, la structuration, la normalisation et la documentation des dépôts, la valorisation et la diffusion des ressources déposées. Il s'agit donc de veiller à leur accessibilité, leur standardisation et surtout leur interopérabilité. La FAIRisation des données est par exemple un moyen pour partager et mutualiser les données produites avec l'ensemble de la communauté, indépendamment et au-delà des sites dédiés aux projets.

Plus généralement, ces ambitions seront articulées au développement des différents projets mentionnés, afin de les coordonner, et de mutualiser certaines tâches et compétences qui sont transversales à leur réalisation. Il s'agira aussi d'impliquer les chercheurs et les

¹ Cf. Damon Mayaffre, « Philologie et / ou herméneutique numérique : nouveaux concepts pour de nouvelles pratiques ? », XXVIIe colloque d'Albi Langages et Signification, juillet 2006, Albi, France, pp. 15-25.

doctorants, à travers l'organisation d'un séminaire, que l'institut propose aussi au catalogue des écoles doctorales. Comme prévu dans le règlement intérieur de la structure, les chercheurs qui publieront dans le cadre des projets de l'IDHN indiqueront l'appartenance à la structure, ce qui contribuera à sa visibilité et à son référencement.

3. Success story à poursuivre

a. Enseignement

L'offre d'enseignement de CY Cergy Paris Université intègre dorénavant les humanités numériques dans ses filières en SHS. Le Master des Sciences du langage prend désormais une orientation très nette en faveur du numérique puisque son appellation même a évolué en Master Sciences du langage : humanités numériques, linguistique et informatique (HNLII). La deuxième année de Master propose un accompagnement renforcé vers les métiers liés à l'ingénierie linguistique et les HN, notamment par le biais du projet tuteuré et du stage. L'interaction entre recherche et enseignement, entre chercheurs confirmés et chercheurs débutants est effective quand l'IDHN emploie comme stagiaire des étudiants de l'Université. Cela a été le cas par exemple pour le projet « Frénaud numérique » pour lequel un stagiaire issu de ce Master a assuré en partie la transcription des feuillets sous Transkribus.

L'offre d'enseignement se poursuit aussi de manière transversale par la tenue de Master Class comme celui organisé dans le cadre du projet Medet-Lat auprès des étudiants du Master « Ingénierie éditoriale et communication » sur les archives éditoriales et les archives numériques. L'interaction entre recherche et formation se prolonge également par des séminaires réguliers ouverts à l'ensemble des étudiants de Master en SHS. Ces séminaires pluridisciplinaires permettent d'une part de faire se rencontrer et dialoguer enseignants-chercheurs et étudiants qui sont amenés éventuellement à poursuivre leur cursus universitaire en doctorat, mais aussi les collègues de toutes les disciplines. Le chercheur présente les enjeux épistémologiques de sa recherche, en explique la méthodologie en ayant à l'esprit d'accentuer les éléments susceptibles d'être transférables à d'autres supports. Depuis 2019, ces séminaires sont organisés en alternance dans les différentes composantes et dans leurs départements, afin de pouvoir faire mieux connaître les HN aux collègues et étudiants qui ne se tournent pas spontanément vers l'IDHN.

b. Développement d'outils et de partenariats

Le développement d'outils ne doit pas faire oublier la place prépondérante de la recherche fondamentale en informatique pour le développement des humanités numériques¹. Il ne s'agirait pas de cantonner l'informatique à un aspect utilitaire. En l'occurrence, des chercheurs d'ETIS ont investi des objets de recherche fondamentale, notamment avec le groupe de travail DAHLIA qui « a pour but de réunir les acteurs (chercheurs ou institutions) qui s'intéressent, dans le cadre des humanités numériques, voire le patrimoine culturel, à la gestion des données mais aussi à leur analyse afin de produire des connaissances »². Des chercheurs issus du traitement des données et soutenus par l'Association Extraction de Gestion de Connaissances³ sont à l'origine de la création de ce groupe de travail (GT). Les ateliers annuels sont l'occasion d'un dialogue interdisciplinaire fructueux. Ce GT envisage l'élaboration d'une cartographie des projets en HN. Devant l'essor que ce champ de recherche a connu depuis quelques années, il devient nécessaire d'en rationaliser le paysage en ayant une vue globale des projets existants et des outils créés. L'avenir des HN est sans aucun doute dans l'effort de mutualisation qui doit dépasser les laboratoires d'un seul établissement. Pourquoi recommencer à concevoir un outil d'extraction de données, pourquoi ne pas plutôt reprendre les résultats d'un autre projet national ou international pour développer d'autres fonctionnalités ? Le dialogue est engagé en ce sens avec l'Université de Franche-Comté pour concevoir une plateforme informatique unique mais modulable selon les projets.

Par ailleurs, l'IDHN s'appuie sur la plateforme viky.ai pour ses projets intégrant des agents linguistiques et développe un partenariat de recherche avec la société Pertimm. viky.ai est une plateforme open source développée par la société Pertimm, pour créer et partager des agents linguistiques qui analysent des textes dans toutes les langues. La création d'agents d'analyse modifiables et réutilisables, qui peuvent être partagés en collaboration par la communauté, peut contribuer à simplifier considérablement la gestion de l'analyse textuelle sur des problèmes que l'on retrouve de façon similaire dans de nombreux domaines. viky.ai

¹ Cf. Julien Longhi, « Contours, perspectives et tensions des "humanités numériques" », *Sens dessus dessous*, n°24, 2019/2, pp. 123-135.

² <http://dahlia.egc.asso.fr>

³ Association Extraction et Gestion des Connaissances : <https://www.egc.asso.fr/>

offre une interface utilisateur unique et sans code. C'est une nouvelle façon de créer des modules sémantiques qui sont faciles à intégrer dans n'importe quel système par le biais d'une API. Les briques sémantiques, appelées « agents » de viky.ai, sont des assistants multilingues pour trouver des données pertinentes dans les textes. Ces briques sont constituées d'entités et d'interprétations et peuvent s'assembler à l'infini pour construire des modèles dans tous les domaines de l'analyse sémantique. Beaucoup d'experts du traitement du langage s'accordent à penser que les technologies de TAL alliant à la fois des règles et de l'apprentissage, seront la combinaison gagnante¹. C'est donc un choix fort opéré par l'IDHN qui s'inscrit dans une dynamique compatible avec celle de la communauté des HN.

Pour conclure, l'accueil favorable réservé aux projets en HN ne doit pas pour autant faire oublier les ambiguïtés du terme « numérique »² et ne doit pas réserver une place subsidiaire à la recherche informatique. Loin d'être un phénomène de mode, les HN invitent à de nouvelles modalités de recherche en SHS qui mettent en question l'identité de chercheur et qui mettent en jeu des enjeux épistémologiques, ontologiques et herméneutiques. L'expérience de la structure fédérative est fructueuse par la formation à des outils informatiques que les chercheurs de SHS ont effectuée et par la sensibilité que les chercheurs en informatique ont à l'égard du patrimoine.

¹ <https://www.ibm.com/blogs/research/2019/12/2020-ai-predictions/>

² Marin Dacos et Pierre Mounier, *Humanités numériques. État des lieux et positionnement de la recherche dans le contexte international*, Paris, Institut français, 2014.