



Au cœur des données numériques

GAGNANTS DU SECOND CYCLE (2011)

Cascades, Islands, or Streams? Time, Topic, and Scholarly Activities in Humanities and Social Science Research

(Chercheurs principaux : Cassidy R. Sugimoto, Ying Ding, Staša Milojević, Indiana University, Bloomington, NSF; Mike Thelwall, University of Wolverhampton, AHRC, ESRC et JISC; Vincent Larivière, Université de Montréal, CRSH.)

Ce projet vise à étudier le cycle de vie de sujets dans le cadre de corpus hétérogènes, et ce, non seulement dans la littérature scientifique, mais également pour des réseaux sociaux, des carnets Web et d'autres documents. Même si la croissance de vastes ensembles de données a permis l'étude d'ensembles de données scientifiques, il y a peu de travaux de recherche qui sont axés sur divers ensembles de données. L'équipe analysera l'importance qu'ont différentes activités de recherche pour la production, la gestion et la diffusion de nouvelles connaissances, la comparaison et la triangulation des résultats de méthodes d'analyse de sujets ainsi que la création d'outils transparents et accessibles. Ce travail devrait permettre de déterminer les activités de recherche qui concernent de nouveaux secteurs et les ensembles de données qui ne devraient plus être marginalisés, mais faire partie de la compréhension et de la mesure des activités de recherche.

ChartEx


(Chercheurs principaux : Robert C. Stacey, University of Washington, IMLS; Arno Knobbe, Leiden University, NWO; Sarah Rees Jones, University of York, AHRC, ESRC et JISC; Michael Gervers, University of Toronto, CRSH. Autres établissements participants : University of Brighton et Columbia University.)

Ce projet permettra de trouver de nouveaux moyens d'étudier tout le contenu de dossiers historiques numériques. On fera appel à des chartes médiévales, qui sont nombreuses du 12^e au 16^e siècles et qui représentent l'une des plus riches sources d'étude de la vie des sociétés d'autrefois.

Digging into Connected Repositories (DiggiCORE)

(Chercheurs principaux : Andreas Juffinger, The European Library Office, NWO; Zdenek Zdrahal, The Open University, AHRC, ESRC et JISC.)

Ce projet permettra d'analyser un vaste ensemble de publications de recherche à libre accès grâce au traitement des langues naturelles et à des méthodes d'analyse de réseaux sociaux afin de déterminer des tendances quant au



comportement de communautés de chercheurs et aux disciplines de recherche, d'acquérir de nouvelles perspectives au sujet des habitudes de citation des chercheurs et de découvrir des caractéristiques qui permettent de faire ressortir des ouvrages dont l'impact est important. Cela permettra de trouver de meilleures méthodes d'étude exploratoire et de consultation de collections numériques ou de nouveaux moyens d'évaluer des travaux de recherche ou l'impact de leur auteur.

Digging by Debating

(Chercheurs principaux : Colin Allen et Katy Börner, Indiana University, Bloomington, NEH; Andrew Ravenscroft, University of East London, Chris Reed, University of Dundee, et David Bourget, University of London, AHRC, ESRC et JISC.)

Ce projet vise à créer et à mettre en œuvre un atelier à multiples échelles appelé « InterDebates » afin d'étudier des données fournies par des centaines de milliers – et un jour des millions – de livres numérisés, de bases de données bibliographiques d'article de revue et de documents de référence complets préparés par des experts. Les hypothèses de l'équipe sont les suivantes : des arguments précis et particuliers peuvent orienter de nombreux aspects de la recherche en sciences humaines; on peut extraire des structures argumentatives de vastes ensembles de données à l'aide de diverses techniques automatisées et liées à l'informatique sociale; la disponibilité de telles analyses permettra des travaux de recherche interdisciplinaire novateurs et pourrait favoriser la tenue de débats critiques plus informés chez les étudiants et le grand public.

Digging into Human Rights Violations: Anaphora Resolution and Emergent Witnesses

(Chercheurs principaux : Ben Miller, Georgia State University, NSF; Lu Xiao, University of Western Ontario, CRSH. Autres établissements participants : University of North Florida.)

Ce projet permettra de créer un lecteur automatique pour de vastes archives de textes portant sur des abus liés aux droits de la personne – lequel reconstituera des histoires à l'aide de fragments issus d'une collection – et de concevoir une interface permettant de consulter ces histoires. En améliorant les techniques de résolution d'anaphores quant au traitement des langues naturelles pour l'établissement de liens entre des pronoms et des noms précis, ce système aidera les chercheurs et les tribunaux à découvrir des témoins et des tendances dans leur collection d'ouvrages.



Digging into Metadata: Enhancing Social Science and Humanities Research

(Chercheurs principaux : Mick Khoo, Drexel University, IMLS; Diana Massam, University of Manchester, AHRC, ESRC et JISC. Autres établissements participants : University of Glamorgan.)

Ce projet permettra d'automatiquement générer de nouvelles formes de fiches de métadonnées à l'aide de dossiers de métadonnées et de ressources connexes qui favoriseront les découvertes dans divers répertoires. Dans le cadre du projet, on utilisera quatre répertoires dont la taille, le domaine, la communication et le mode de création de métadonnées ainsi que la qualité varient. On utilisera PERTAINS, un outil créé par l'une des écoles partenaires, afin d'analyser les dossiers de métadonnées de chaque répertoire et de générer des fiches selon le système de classification décimale de Dewey. De plus, on fera appel à des algorithmes de regroupement afin de produire un index de similarité et de correspondance de ressources de divers répertoires. Après avoir effectué une recherche, l'utilisateur extraira une liste de ressources des différentes collections ayant été définies de façon similaire. On fera appel à des techniques de visualisation afin de présenter les résultats d'une manière qui améliore le processus de recherche.

Electronic Locator of Vertical Interval Successions (ELVIS): The First Large Data-Driven Research Project on Musical Style

(Chercheurs principaux : Michael Scott Cuthbert, Massachusetts Institute of Technology, NEH; Frauke Jürgensen, University of Aberdeen, AHRC, ESRC et JISC; Julie E. Cumming, Université McGill, CRSH. Autres établissements participants : Yale University.)

Ce projet vise à étudier des changements liés au style musical occidental des années 1300 à 1900 à l'aide de collections numérisées de vastes répertoires musicaux. L'équipe remarque que, pour comprendre le changement de style de la musique polyphonique occidentale, il faut être en mesure de décrire des cadences mélodiques et des sonorités (accords) verticales acceptables de chaque période ainsi que de montrer comment elles changent au fil du temps. Le projet vise à y parvenir pour la polyphonie européenne des années 1300 à 1900 en faisant appel à des techniques poussées de recherche de renseignements musicaux dans le but d'étudier des styles de musique très différents qui sont tout de même liés par des concepts communs de tonalité, de consonance et de dissonance ainsi que de conduite de voix.



An Epidemiology of Information: Data Mining the 1918 Influenza Pandemic

(Edward T. Ewing, Bernice L. Hausman, Bruce Pencek et Narendran Ramakrishnan, Virginia Polytechnic Institute & State University, NEH; Gunther Eysenbach, University of Toronto, CRSH.)

Ce projet vise à faire appel à des techniques d'exploration de données du point de vue de l'analyse interprétative des sciences humaines afin de comprendre comment les journaux ont orienté l'opinion publique et ont constitué un symbole de connaissances faisant autorité au cours de cette mortelle pandémie. Le projet fait appel à plus de 100 titres d'article de journal de 1918 tirés de *Chronicling America* de la Bibliothèque du Congrès des États-Unis et de la collection Peel's Prairie Provinces de l'University of Alberta Library. L'utilisation de techniques algorithmiques permet à un expert de ce domaine d'étudier un vaste répertoire de données de façon systématique ainsi que de définir des caractéristiques qualitatives de la pandémie à petite échelle et la généalogie de la communication de renseignements à grande échelle. Cette recherche peut fournir des moyens de comprendre la communication de renseignements et la progression de la maladie au sein d'autres sociétés devant faire face à la menace que représentent les pandémies.

Imagery Lenses for Visualizing Text Corpora

(Chercheurs principaux : Katharine Coles, University of Utah, NEH; Min Chen, University of Oxford, AHRC, ESRC et JISC.)

Ce projet vise à étudier de nouvelles techniques de visualisation utilisées pour de vastes corpus linguistiques et littéraires à l'aide des collections du British National Corpus et de diverses archives de poésie de plus modeste envergure. L'équipe tentera de déterminer si certaines techniques de visualisation de pointe peuvent permettre aux chercheurs du domaine des sciences humaines d'utiliser les connaissances de leur domaine de façon dynamique tout en tirant profit de la capacité de calcul d'ordinateurs. Plus particulièrement, la visualisation de données peut-elle aider les utilisateurs à faire de nouvelles découvertes et à proposer de nouvelles hypothèses? Le projet consiste à répondre à la question de recherche méthodologique ci-dessus et à créer un ensemble de nouveaux outils de visualisation pour de futurs travaux de recherche.



IMPACT Radiological Mummy Database

(Chercheurs principaux : Randall Thompson, Saint Luke's Mid America Heart Institute, NEH; Andrew Nelson, University of Western Ontario, CRSH. Autres établissements participants : Al Azhar Medical School, Caire, Quinnipiac University, Musée canadien des civilisations, University of Southern California, University of California, San Diego, Mount Sinai School of Medicine, South Coast Radiological Medical Group, Newport Diagnostic Center, University of California, Irvine, Wisconsin Heart Hospital.)

Ce projet vise à offrir aux chercheurs qui s'intéressent aux momies et à la médecine une vaste base de données comparative d'imagerie médicale de restes humains momifiés. Le fait de s'éloigner d'un modèle d'étude de cas concernant l'étude des momies entraînera un grand paradigme comparatif et épidémiologique dans ce domaine. L'équipe canadienne étudiera les aspects de l'éviscération et de l'excérébration de la tradition de momification égyptienne, et les équipes des États-Unis utiliseront la base de données pour une plus vaste étude de l'athérosclérose des anciennes momies égyptiennes dans le cadre de l'Ancient Health Research Group d'IMPACT (Internet-based Mummy Picture Archive Communication Technology) et pour la mise au point d'un nouveau système de diagnostic par consensus pour les restes momifiés.

Integrated Social History Environment for Research (ISHER)-Digging into Social Unrest


(Chercheurs principaux : Dan Roth, University of Illinois, Urbana-Champaign, NSF; Antal van den Bosch, Tilburg University, NWO; Sophia Ananiadou, The University of Manchester, AHRC, ESRC et JISC. Autres établissements participants : International Institute of Social History.)

Ce projet permettra de créer un environnement intégré à l'aide d'outils sophistiqués d'exploration de textes afin de faciliter la découverte de renseignements en matière de recherche sur l'histoire sociale. De plus, il offrira aux historiens et aux scientifiques des sciences sociales la possibilité de trouver et de lier des événements, des tendances, des personnes, des organisations et d'autres éléments qui intéressent particulièrement les historiens des sciences sociales.

Integrating Data Mining and Data Management Technologies for Scholarly Inquiry

(Chercheurs principaux : Ray R. Larson, University of California, Berkeley, et Richard Marciano, University of North Carolina at Chapel Hill, IMLS; Paul B. Watry, University of Liverpool, AHRC, ESRC et JISC. Autres établissements participants : Internet Archive et JSTOR.)

Ce projet intégrera de vastes collections, notamment JSTOR et les collections de livres de l'Internet Archive qui sont entreposées et administrées dans un milieu permettant de les préserver. De plus, il comprendra l'utilisation d'un logiciel d'exploration de textes et de traitement des langues naturelles permettant de générer des liens dynamiques menant à des ressources portant sur les mêmes personnes, endroits et événements. Dans le cadre de



ce projet de 17 mois, on ira au delà d'une analyse de base pour offrir un système prototype permettant de fournir du soutien aux chercheurs en matière de systèmes experts.

Mining Microdata: Economic Opportunity and Spatial Mobility in Britain, Canada and the United States, 1850-1911

(Chercheurs principaux : Evan Roberts, University of Minnesota, NSF; Kevin Schürer, University of Leicester, AHRC, ESRC et JISC; Kris E. Inwood, University of Guelph, CRSH. Autres établissements participants : University of Alberta, Université de Montréal et University of Essex.)

Dans le cadre de ce projet, on fera appel à une technologie d'exploration de données afin de tirer profit de l'une des plus grandes bases de données de population du monde. Il s'agit d'une vaste collection de microdonnées de recensement harmonisées des 19^e et 20^e siècles provenant de la Grande-Bretagne, du Canada et des États-Unis qui, à l'origine, avait été numérisée pour la recherche généalogique. L'objectif consiste à mieux comprendre l'impact qu'ont les possibilités économiques et la mobilité sur la structure sociale de l'Europe et de l'Amérique du Nord.

Trading Consequences

(Chercheurs principaux : Ewan Klein, University of Edinburgh, AHRC, ESRC et JISC; Colin M. Coates, Université York, CRSH. Autres établissements participants : University of St Andrews.)

Dans le cadre de ce projet, on étudiera les conséquences économiques et environnementales des opérations sur marchandises du 19^e siècle. L'équipe de projet fera appel à des techniques d'extraction de données afin d'étudier de vastes corpus de documents numérisés du 19^e siècle. Cette ressource numérique novatrice permettra aux historiens de découvrir de nouvelles tendances et d'examiner de nouvelles hypothèses grâce à une recherche structurée et à divers outils de visualisation.

* * * * *

Le **National Endowment for the Humanities (NEH)**, créé en 1965 en tant qu'organisme fédéral indépendant, appuie l'apprentissage dans les domaines de l'histoire, de la littérature, de la philosophie et d'autres secteurs des sciences humaines. Les subventions du NEH permettent d'enrichir l'apprentissage en salle de classe, de créer et de préserver des connaissances ainsi que de concrétiser des idées grâce à la télévision, à la radio, à de nouvelles technologies, à des expositions de musée ainsi qu'à des programmes offerts en bibliothèque et dans d'autres établissements communautaires. Des renseignements supplémentaires sur le NEH et sur ses programmes de subventions sont présentés dans son site Web (www.neh.gov).

Chaque année, l'**Arts & Humanities Research Council (AHRC)** offre environ 112 millions de livres Sterling provenant du gouvernement afin d'appuyer des travaux de recherche et des études supérieures dans le domaine des arts et des sciences humaines, notamment les langues, le droit, l'archéologie, la littérature anglaise, le design ainsi que les arts de la création et du spectacle. Chaque année, l'AHRC offre environ 700 bourses de recherche et approximativement 1 300 bourses d'études supérieures. Ce financement est attribué au terme d'un rigoureux processus d'évaluation par les pairs, ce qui permet de s'assurer que seules les meilleures demandes sont financées. Non seulement la qualité et la portée des travaux financés par cet investissement de fonds publics procurent des avantages sociaux et culturels, mais elles contribuent à la prospérité du Royaume-Uni.


L'**Economic and Social Research Council (ESRC)** est la plus grande organisation de financement de la recherche portant sur des enjeux économiques et sociaux du Royaume-Uni. Il appuie des travaux de recherche indépendants de grande qualité qui ont un impact sur le milieu des affaires ainsi que sur les secteurs public et tertiaire. Le budget de 2011-2012 de l'ESRC s'élève à 203 millions de livres Sterling. En tout temps, l'ESRC appuie plus de 4 000 chercheurs et étudiants de cycle supérieur d'établissements universitaires et d'instituts de recherche indépendants. Pour obtenir des renseignements supplémentaires sur l'ESRC, veuillez consulter son site Web (www.esrc.ac.uk).

L'**Institute of Museum and Library Services (IMLS)** constitue la principale source de financement fédéral des 123 000 bibliothèques et des 17 500 musées du pays. La mission de l'IMLS consiste à mettre sur pied d'efficaces bibliothèques et musées qui permettent de communiquer des renseignements et des idées aux gens. L'IMLS travaille au niveau fédéral en collaboration avec l'État et des organisations locales afin de préserver le patrimoine, la culture et les connaissances, d'augmenter l'apprentissage et l'innovation ainsi que d'appuyer le développement professionnel. Pour en savoir plus sur l'IMLS, veuillez consulter son site Web (www.imls.gov).

Le **Joint Information Systems Committee (JISC)** est un comité conjoint d'organismes de financement d'études postsecondaires et supérieures du Royaume-Uni qui est chargé de soutenir l'utilisation novatrice de technologies de l'information et des communications afin de promouvoir l'apprentissage, l'enseignement et la recherche. Il est surtout connu parce qu'il offre un réseau d'infrastructure nationale, diverses formes de soutien, de contenu et de services de consultation ainsi qu'un ensemble de ressources de grande qualité. Pour obtenir des renseignements supplémentaires sur le JISC, ses services et ses programmes, veuillez consulter son site Web (www.jisc.ac.uk).

La **National Science Foundation (NSF)** est un organisme fédéral indépendant qui appuie des travaux de recherche fondamentale et la formation dans tous les domaines des sciences et du génie. Au cours de l'exercice financier 2009, son budget était de 9,5 milliards de dollars, dont trois milliards de dollars affectés en vertu de l'American Recovery and Reinvestment Act. La NSF finance les 50 États grâce à des subventions attribuées à plus de 1 900 universités et établissements. Chaque année, la NSF reçoit environ 44 400 demandes de financement et procède à plus de 11 500 nouvelles attributions de financement. En outre, elle offre plus de 400 millions de dollars sous forme de contrats professionnels et de services chaque année. Pour obtenir des précisions sur la NSF, veuillez consulter son site Web (www.nsf.gov).





*La **Netherlands Organisation for Scientific Research (NWO)** finance des milliers de remarquables chercheurs d'universités et d'instituts. De plus, elle oriente les activités scientifiques du pays grâce à des subventions et à des programmes de recherche.*

*Le **Conseil de recherches en sciences humaines (CRSH)** est un organisme gouvernemental fédéral indépendant qui finance la recherche universitaire et la formation de cycle supérieur dans le cadre de concours nationaux d'évaluation par les pairs. En outre, le CRSH établit des partenariats avec des organisations des secteurs public et privé afin d'orienter la recherche et de contribuer à l'élaboration de politiques et de pratiques plus efficaces dans d'importants secteurs de la vie canadienne sur les plans social, culturel et économique. Pour obtenir des renseignements supplémentaires sur le CRSH, veuillez consulter son site Web (www.sshrc-crsh.gc.ca).*

