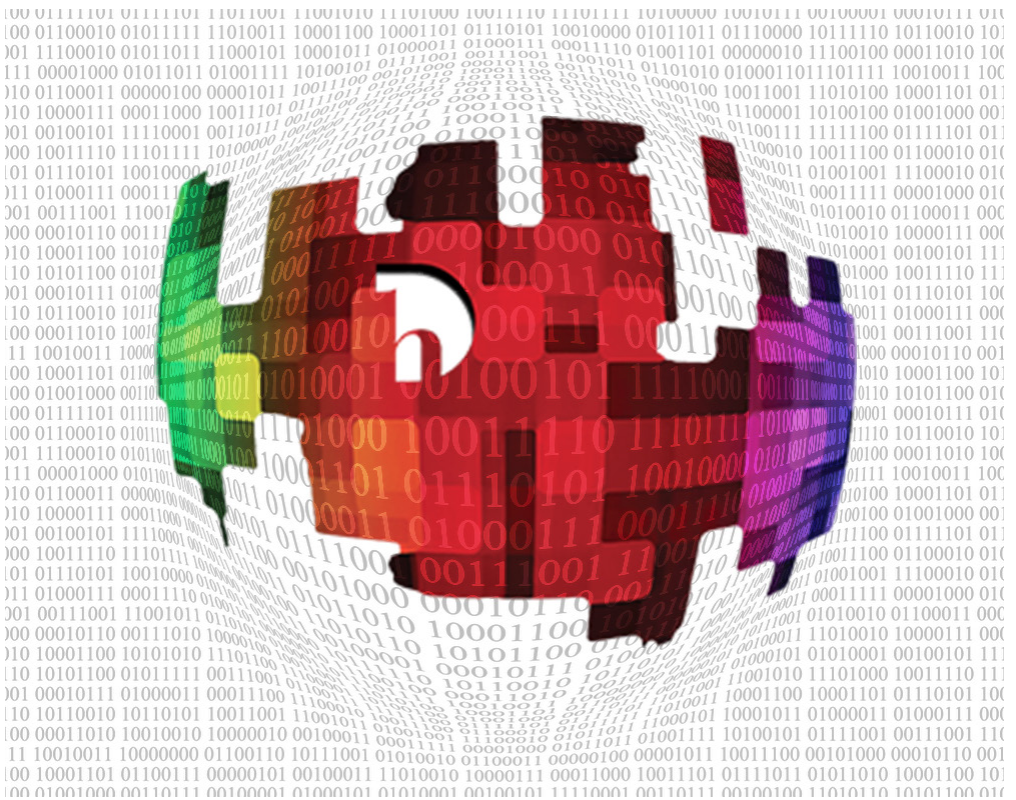




Atelier Digit_Hum 2017

Data deluge

Quelles compétences pour quelles données ?

JEUDI 5 OCTOBRE 2017

ECOLE NORMALE SUPERIEURE | SALLE DUSSANE





Atelier Digit_Hum 2017

Data deluge

Quelles compétences pour quelles données ?

Cet atelier s'inscrit dans le cadre de journées d'études explorant les projets et les solutions techniques développés en « *Digital Humanities* » qui mettent en œuvre de nouveaux outils pour l'analyse et la valorisation de corpus. Le CAPHÉS et le labex TransferS ont souhaité mettre en place une journée d'études par an sur ce sujet fortement interdisciplinaire où les chercheurs, enseignants, ingénieurs et techniciens travaillant à l'École normale supérieure de Paris partagent leur expérience avec des responsables de projets innovants en humanités numériques.

Le thème de l'Atelier 2017, *Data deluge* ou explosion de l'information, clin d'œil au *Big Data*, se réfère à la situation où le volume de nouvelles données disponibles submerge la capacité des institutions et des chercheurs à les gérer et en faire usage. Sans aller jusque-là, la question des compétences nécessaires dans le traitement de ces données reste à aborder dans toute sa complexité. Et ce n'est pas une question propre aux humanités, même si la place des humanités numériques au sein des « *digital studies* » doit être interrogée.

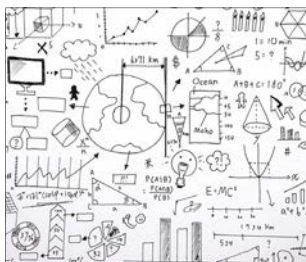
Accueil autour d'un café à partir de 9h30

Introduction par Emmanuelle SORDET (Directrice des bibliothèques de l'ENS)

10h00-10h30

Métier *data scientist*

par Serge ABITEBOUL (Inria – ENS)



La *data science* (la « science des données ») et ses liens avec le *Big Data* (parfois traduit par la « datamasse ») ont été beaucoup discutés. Le terme « *data scientist* » est pris ici au sens large pour ce qui est des compétences, comme scientifique et ingénieur, peut-être technicien, et son domaine d'application. Il recouvre des besoins : des entreprises (ingénieurs, techniciens) et des institutions de recherche (scientifiques). La France disposant d'excellents enseignements en mathématiques, statistiques et informatique, on devrait pouvoir rapidement former d'excellents *data scientists*. La question est : « comment s'y prendre ? ».

10h30-10h45

Présentation de l'IRIS Science des données, données de la science

par Florence WEBER (Département de Sciences Sociales de l'ENS – Centre Maurice Halbwachs – Présidente du Conseil académique de PSL)



L'IRIS « Science des données, données de la science » vise à favoriser le dialogue interdisciplinaire entre informatique théorique, sciences de la nature, sciences de la société et humanités numériques. Au-delà des solutions techniques, loin d'être également à la portée de toutes les disciplines, un tel dialogue renouvelle la réflexion sur les enjeux éthiques et épistémologiques des transformations en cours dans la production des données et dans leur diffusion. Révolution scientifique majeure, recomposition des occasions de profit, risques

pour la démocratie, émergence de nouveaux métiers, démocratisation de l'accès aux savoirs, nouvelles représentations de l'humanité individuelle et collective : ces enjeux seront éclairés par l'organisation – qui reste à construire – d'une réflexion commune au sein de PSL.

<https://www.univ-psl.fr/fr/iris-sciences-donnees>

10h45-11h15

Filigranes pour tous : développement d'un outil de reconnaissance d'identification partagées pour le papier médiéval et moderne

par Olivier PONCET (ENC)



Le projet «Filigranes pour Tous», financé par PSL (2017-2019) et développé par ses institutions-membres (École nationale des chartes, École normale supérieure et Institut de recherche et d'histoire des textes), vise à construire un nouveau corpus et développer une nouvelle méthodologie pour l'identification de filigranes médiévaux et de la première modernité. Le corpus initial sera constitué en collaboration avec des services d'archives et des bibliothèques, en se concentrant principalement sur des filigranes français.

La technologie de reconnaissance optique par les réseaux de neurones (*deep learning*) sera utilisée pour procéder aux identifications à partir des images collectées sur des sources originales diverses et non à partir de dessins réalisés par les chercheurs. Le projet utilisera un équipement composé de smartphones et des feuilles lumineuses pour les prises de vues. Des études plus approfondies

seront nécessaires pour vérifier si les corpus disponibles en ligne pourraient être inclus dans la base de données servant aux comparaisons/identifications.

Lorsque le logiciel correspondant aura atteint un stade de développement satisfaisant, nous planifions de le mettre en œuvre sous forme d'une application pour des smartphones, afin de permettre à des chercheurs, des experts, des conservateurs et des marchands de pouvoir identifier des filigranes instantanément, en ajoutant leurs propres images et des métadonnées de base au corpus (*crowdsourcing*). Les utilisateurs seront invités à suivre des directives communes pour l'acquisition d'image, en utilisant des outils inclus dans l'application afin de collecter de manière cohérente les nouveaux filigranes et d'enregistrer les échelles avec lesquelles les images seront prises.

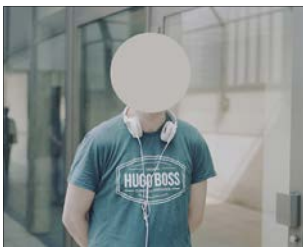
Illustration : Filigrane au motif de cadran dans un acte du 20 avril 1656 (Arch. nat., Min. centr., XIX, 456, non fol. / cl. O. Poncet)

11h15-11h30 : Pause

11h30-12h00

Données de sciences sociales : le projet ArchEthno du labex TransferS

par Florence WEBER (ENS – labex TransferS), avec Agnès TRICOCHÉ (CNRS – labex TransferS) et CARLO ZWÖLF (Paris Astronomical Data Centre – Observatoire de Paris)



Le projet ArchEthno est issu de l'archivage par les chercheurs d'une expérience originale d'ethnographie collective menée en 2003-2005 : pour vérifier à l'échelle de la population résidant en France des hypothèses issues d'enquêtes ethnographiques classiques au sein de familles confrontées à la dépendance cognitive d'un parent (troubles de type Alzheimer), l'équipe est passée par une étape d'ethnographie réduite (délégation de la passation des questionnaires à des enquêteurs rémunérés, chargés de remettre aux chercheurs leur journal d'enquête) avant de participer à l'élaboration des enquêtes Handicap Santé 2008 en population générale (INSEE-DREES). La

construction d'une base de données consultable sous conditions, fournie à l'appui des publications scientifiques *peer-review* des membres de l'équipe, va bientôt permettre de publier une partie de ces données en ligne. L'architecture originale d'ArchEthno, qui évolue avec les enquêtes archivées en

fonction des besoins des chercheurs (la dernière enquête archivée concerne la prise en charge en service fermé de patients de la psychiatrie et la stigmatisation des malades mentaux, moteur de l'aggravation de la maladie), a conduit à une collaboration avec l'Observatoire de Paris pour la mise en œuvre d'une solution informatique qui pose les bases d'une extension prudente en SHS, et pour des données qualitatives, de la méta-analyse de corpus de données à caractère personnel, courante en sciences biomédicales.

12h00-12h30

Villa de Diomède, Pompéi : des archives aux restitutions 3D

par Hélène DESSALES (ENS, AOROC UMR 8546)



Comment restituer l'évolution d'un monument de Pompéi au fil du temps, de sa création à l'époque romaine aux restaurations contemporaines ? La Villa de Diomède constitue un cas très emblématique, puisqu'il s'agit d'un des premiers bâtiments découverts sur le site (1771-1775) et un des plus représentés et décrits par les voyageurs au cours du XIX^e siècle. Engagé en 2012, un programme pluridisciplinaire a pour objectif de modéliser la fabrique de cette grande villa, des chantiers de construction successifs qui ont rythmé sa vie jusqu'à l'éruption du Vésuve en 79

ap. J.-C. aux fouilles et restaurations contemporaines qui en ont transformé l'image jusqu'à nos jours. Pour ce faire, différentes compétences se sont associées : archéologie de la construction, ingénierie des structures, gestion des archives historiques et archéologiques par bases de données, cartographie SIG, informatique avec photogrammétrie et modélisation 3D.

<http://villadiomede.huma-num.fr/3dproject/>

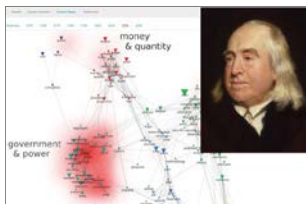
*Illustration : Villa de Diomède. Superposition d'une vue gravée de Philippe Benoist (1846) et du modèle photogrammétrique (2015).
© Villa Diomedes Project. Infographie 3D : Alban-Brice Pimpaud (archeo3D) | Yves Ubelmann & Philippe Barthelemy (Iconem)*

12h30-14h00 : Buffet

14h00-14h30

Mapping Bentham : une cartographie du fonds Bentham de UCL (partenariat entre le Lattice et le laboratoire Digital Humanities de University College London)

par Thierry POIBEAU (Directeur du Lattice), avec Pablo RUIZ et Frédérique MÉLANIE (Lattice)



University College London (UCL) possède une collection de 40 000 manuscrits inédits du philosophe anglais Jeremy Bentham (1748-1832). Le projet Transcribe Bentham (en cours depuis 2010) a permis de numériser, transcrire et encoder suivant le standard TEI 18 000 de ces manuscrits grâce à un large effort de *crowdsourcing* (myriadisation). Le projet doit se poursuivre jusqu'à la numérisation de l'intégralité du fonds Bentham. Mais qu'y a-t-il dans ces manuscrits ? De quoi parlent-ils ? Quels sont les documents sur un

même sujet ? Le projet Mapping Bentham, mené par le laboratoire Lattice en collaboration avec UCL, vise à répondre en partie à ces questions, en identifiant les principaux sujets traités dans les documents et en produisant des cartes susceptibles de révéler les proximités entre les documents et les thèmes de recherche abordés par Bentham au cours du temps.

<http://blogs.ucl.ac.uk/transcribe-bentham/>

14h30-15h00

Design / arts / sciences – expérimental

par Emmanuel MAHÉ (ENSAD, Directeur de la Recherche)



Fondé en 2007, EnsadLab est le laboratoire de recherche de l'École nationale supérieure des Arts Décoratifs — Paris. Les débats suscités par les différents classements mondiaux des grandes écoles et des universités montrent à quel point la Recherche est devenue une activité clé. Il faut y préparer la nouvelle génération de designers et d'artistes en leur proposant une formation spécifiquement dédiée. Dans ce contexte, les étudiants-chercheurs d'EnsadLab sont amenés à développer leurs recherches pratiques au sein de projets collaboratifs afin : d'acquérir des compétences complémentaires au cursus initial, d'explorer de nouvelles méthodologies, d'expérimenter des hypothèses inédites, d'apprendre à travailler avec des scientifiques, de proposer des solutions à divers problèmes contemporains, c'est-à-dire inventer de nouvelles formes (relationnelles, technologiques, esthétiques, sociales).

15h00-15h30

Les Big Data du passé : la machine à remonter le temps européenne

par Frédéric KAPLAN (EPFL, Directeur du Digital Humanities Lab)



Un consortium de plus de cent institutions propose aujourd'hui la création d'un programme de recherche et d'une infrastructure internationale permettant l'extraction massive et réalignement de données historiques selon une logique de mise en réseau et d'indexation multi-échelle directement inspirée des méthodes de traitement des Big Data. La construction de cette machine à remonter

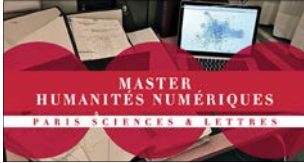
le temps européenne, planifiée entre 2020 et 2030, pourrait être financée par la Commission européenne et les pays membres à hauteur d'un milliard d'euros. Cet exposé présente certains enjeux scientifiques, technologiques, épistémologiques et politiques de ce projet. Il s'agit en effet d'anticiper les opportunités, mais aussi les risques de cette nouvelle cartographie du passé en termes de transformation des méthodologies, de représentations des incertitudes et des incohérences, d'opportunités de mémoire et de droit à l'oubli.

15h30-15h45 : Pause

15h45-16h00

Le master « Humanités numériques et recherche » de PSL

par Jean-Baptiste CAMPS (ENC – Responsable du master « Humanités numériques et recherche »)



L'université Paris Sciences & Lettres ouvre à la rentrée 2017 un nouveau master « Humanités numériques », fruit de la collaboration, notamment, de l'École nationale des chartes, de l'École normale supérieure et de l'École pratique des hautes études. Ce master a pour ambition de former à l'exploitation et l'analyse des données des sciences humaines des étudiants titulaires d'une licence qui approfondissent également la connaissance de leur discipline (histoire

et philologie, lettres, archéologie, etc.) afin de les doter de la double compétence indispensable pour mener, de manière autonome, des recherches s'inscrivant dans le champ des humanités numériques et computationnelles.

Illustration : Plaquette du Master Humanités numériques PSL

16h00-17h30 – TABLE RONDE

Traitement des données : métiers, compétences et formations à PSL

Animée par Richard WALTER (ITEM) et Marie-Laure MASSOT (CAPHÉS)

avec

Jean-Baptiste CAMPS

ENC – Responsable du master « Humanités numériques et recherche » de PSL

Hélène CHAUDOREILLE

Directrice du pôle ressources et savoirs de PSL

Frédéric KAPLAN

EPFL, Directeur du Digital Humanities Lab

Emmanuel MAHÉ

ENSAD, Directeur de la Recherche – Vice-doyen de la formation de PSL

Emmanuelle SORDET

Directrice des bibliothèques de l'ENS

Anne-Marie TURCAN-VERKERK

Vice doyenne Recherche SHS de PSL – Equipex Biblissima

Florence WEBER

Présidente du Conseil académique de PSL – labex TransfErS

Les acteurs en humanités numériques à l'ENS

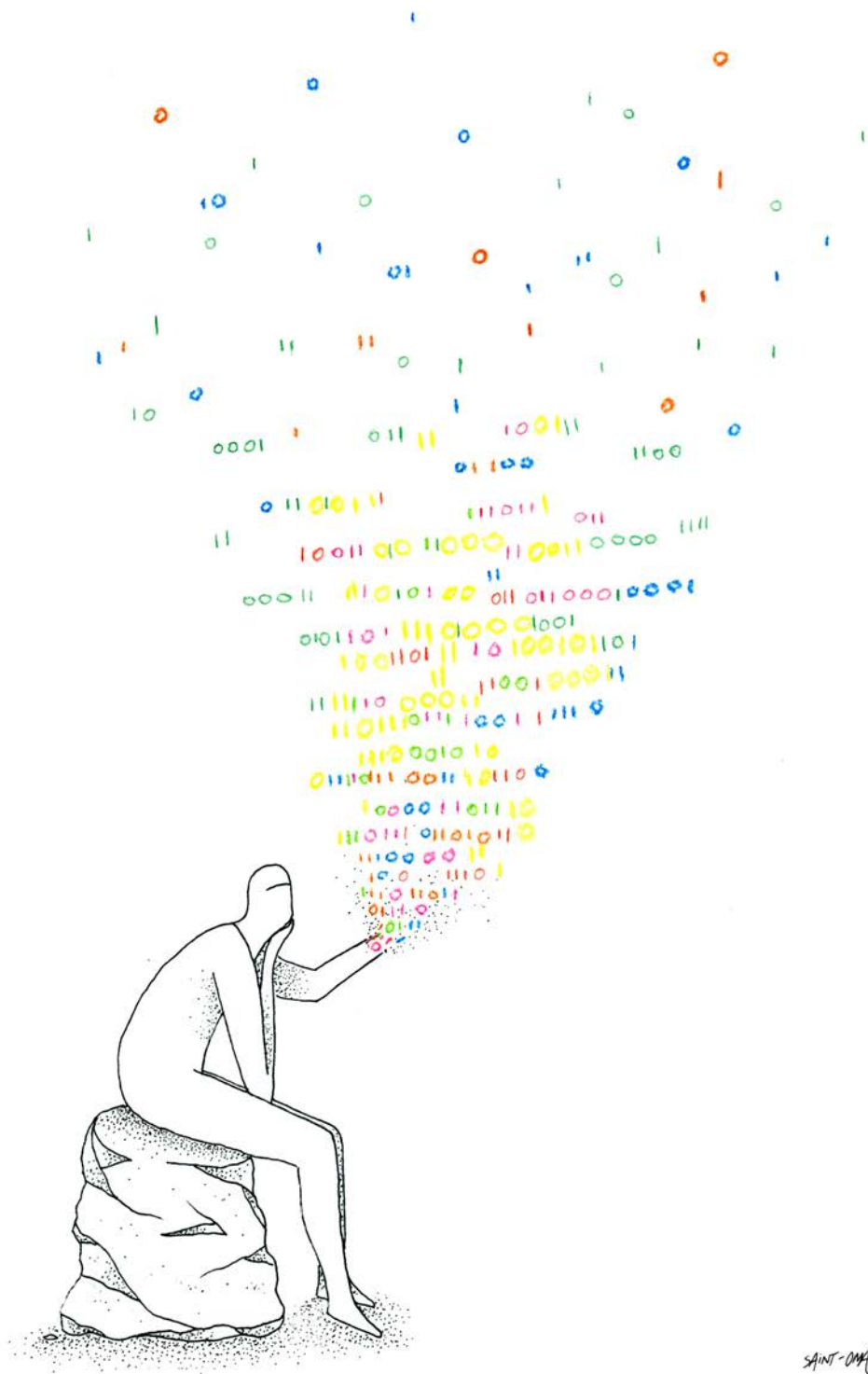
Entre savoir et savoir-faire

En extension aux Ateliers Digit_Hum, un projet de recensement des métiers et compétences en humanités numériques existant au sein de l'ENS est lancé à la rentrée 2017.

Une première étape consiste en une enquête réalisée en coopération avec la Mission de coordination des humanités numériques de l'ENS et visant à dresser un inventaire des projets, métiers, compétences et outils en humanités numériques.

Ce projet bénéficie d'un cofinancement du CAPHÉS, du labex TransferS et de la Bibliothèque LSHS de l'ENS.

Déclinaison graphique, en collaboration avec l'artiste Saint-Oma, et audiovisuelle, avec O.H.N.K. Production.





Atelier Digit_Hum

Journée organisée par :

Marie-Laure MASSOT

Ingénieure d'Études CNRS, UMS 3610 CAPHÉS, ENS Paris

Julien CAVERO

Ingénieur cartographe, labex TransferS, ENS Paris

digit_hum@ens.fr

École normale supérieure, 45 rue d'Ulm, Paris 5^e

