.../...

dans lequel tous les chercheurs en sciences humaines peuvent intégrer leurs bases de données, qui deviennent des sous-graphes du graphe de connaissances global. La maintenance, la formation et le soutien à long terme associés à un tel écosystème sont mutualisés et donc grandement facilités. Nous décrivons la solution que nous avons mise en œuvre, ses avantages et ses inconvénients, ainsi que les difficultés que nous avons rencontrées.

16h10 Quentin Bernet, **Le médiéval mis en données : objets, vocabulaires et traditions disciplinaires**

À partir des travaux issus de notre thèse de doctorat en cours, nous nous intéresserons à la base de données comme matériau de recherche et objet épistémologique, et particulièrement au cas des bases sémantiques standardisées, décrivant de grands corpus d'objets rattachés à la production médiévale, dont l'Index of Medieval Art de Princeton, la base Images du groupe Ahloma (EHESS) et la base Palissy du ministère de la Culture. Nous nous demanderons de quelles façons ce matériau peut mettre à nu certaines des structures mentales qui le sous-tendent, voire soutenir une approche quantitative des réalités historiques et matérielles qu'il décrit. Bien qu'étant favorisée par le développement des méthodes quantitatives et des outils computationnels, il apparaît que la mise en données suit des schémas conceptuels et disciplinaires ancrés dans la longue durée, telles que les catégories nébuleuses d'art gothique, de Vierge à l'Enfant, et plus globalement une approche taxinomique et positiviste de la production matérielle – notamment par la standardisation et l'appauvrissement du discours que le format requiert. Mais si les bases de données présentent des similarités évidentes dans leur historique et leur formation, il sera également intéressant de voir que les vocabulaires et les catégories analytiques varient en réalité d'un acteur à l'autre, par leurs positions diachronique et synchronique au sein du champ de la médiévistique ou de l'histoire de l'art. Se pose alors l'enjeu de la cumulativité de ces différents systèmes : des points de jonction peuvent-ils être identifiés ? Si oui, sont-ils dus à la similarité des objets décrits par les bases, ou plutôt à celle des acteurs qui les conçoivent ?

FIN

17 h 30

MATINÉE (9h30-12h30)

CONSTRUCTION ET PÉRENNITÉ DES BASES

« Tous les historiens sont autodidactes » : la constitution de base de données entre injonctions et interrogations méthodologiques
Vincent Demont, Université Paris-Nanterre, Institutions et dynamiques historiques de l'économie et de la société (IDHE.S)

Cartes, enquêtes et minutes. À la recherche du potentiel meunier francilien à l'époque moderne
Raphaël Morera, CNRS, Centre de recherches historiques (CRH)

Évaluations et perspectives d'utilisations des « sources historiques numériques ». Le cas des *Annuaires historiques parisiens 1798-1914*
Pascal Cristofoli, EHESS, Centre de recherches historiques (CRH)

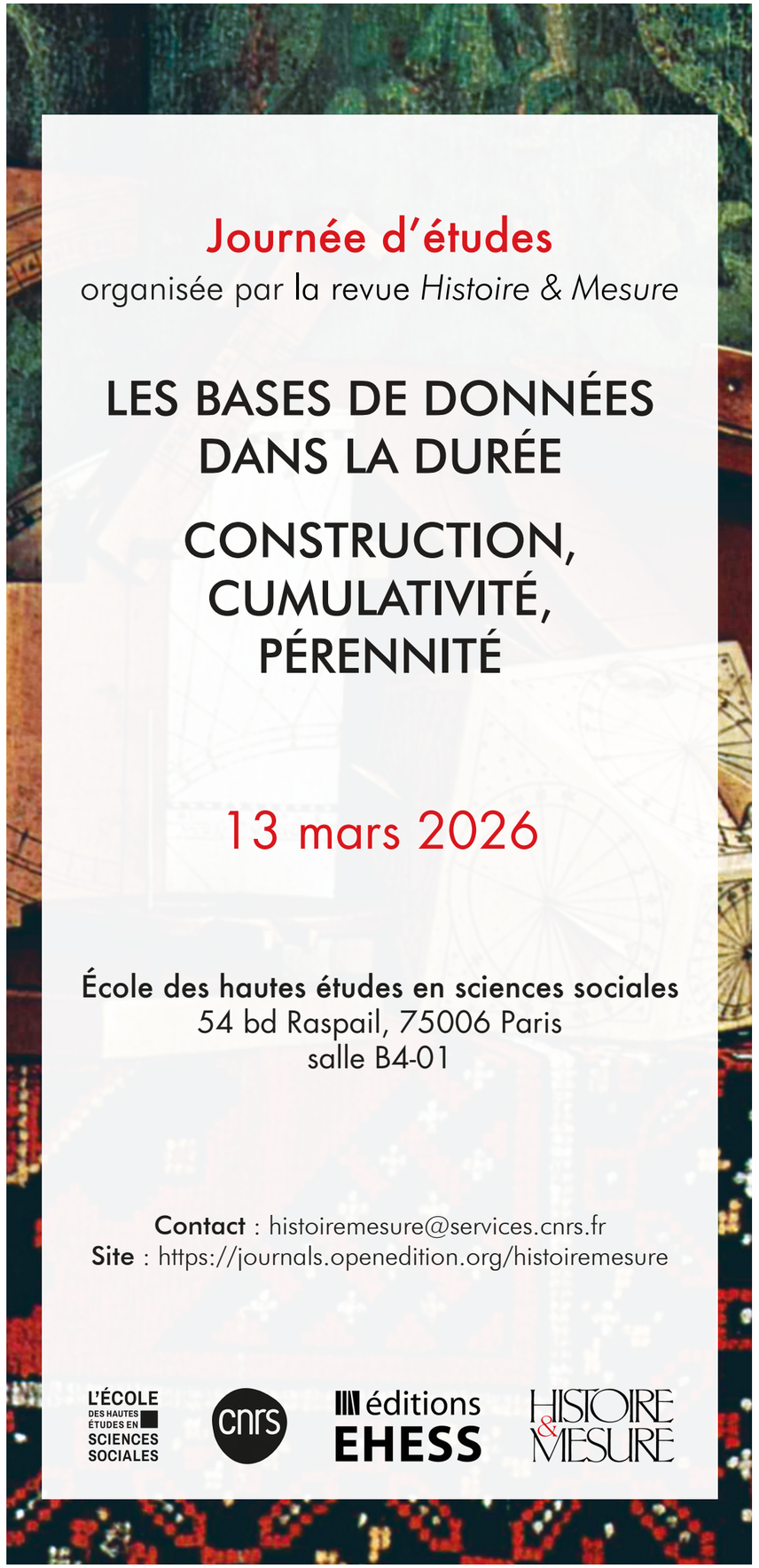
APRÈS-MIDI (14h30-17h30)

CUMULATIVITÉ DES BASES

Mettre le social en bases : une usine à gaz ?
Claude Chevaleyre, CNRS, Institut d'Asie orientale (IAO)

Décloisonner les données de recherche en sciences humaines : SPARQuLb, un écosystème RDF mutualisé
Sébastien de Valeriola, Université libre de Bruxelles

Le médiéval mis en données : objets, vocabulaires et traditions disciplinaires
Quentin Bernet, Université Paris-Saclay, Institut des sciences sociales du politique (ISP)



Journée d'études

organisée par la revue *Histoire & Mesure*

LES BASES DE DONNÉES DANS LA DURÉE

CONSTRUCTION, CUMULATIVITÉ, PÉRENNITÉ

13 mars 2026

École des hautes études en sciences sociales
54 bd Raspail, 75006 Paris
salle B4-01

Contact : histoiremesure@services.cnrs.fr
Site : https://journals.openedition.org/histoiremesure

L'ÉCOLE DES HAUTES ÉTUDES EN SCIENCES SOCIALES — cnrs — éditions EHESS — HISTOIRE & MESURE

MATINÉE **CONSTRUCTION ET PÉRENNITÉ DES BASES**

9 h 40

Vincent Demont, « Tous les historiens sont autodidactes » : la constitution de base de données entre injonctions et interrogations méthodologiques

L'interrogation des pratiques entourant les bases de données historiques fait apparaître un fort hiatus entre des injonctions incitant explicitement à leur constitution ou normant celle-ci, et les usages et devenirs de ces outils, qui restent largement incertains. Ce hiatus problématise le travail d'organisation et de mise à disposition des données de la recherche, en posant des questions que la présente communication, fondée sur la participation à la constitution collective de la base *Privilèges*, cherche à lister et décrire. L'objectif de cette description est de suggérer que le travail par base de données, en plus de soulever des questions scientifiques, implique un positionnement professionnel.

10 h 30

Raphaël Morera, Cartes, enquêtes et minutes. À la recherche du potentiel meunier francilien à l'époque moderne

L'approvisionnement en farine de Paris au cours de l'époque moderne nécessitait la mobilisation d'une importante force hydraulique : le blé était en effet moulu en majeure partie grâce à des moulins à eau. La croissance démographique observée entre les XVIe et XVIIIe siècles a ainsi mobilisé un potentiel meunier de plus en plus important. L'histoire des techniques meunières a fait l'objet de nombreux travaux et se distingue par une attention portée aux améliorations techniques, surtout sensibles à la fin du XVIIIe siècle. La trajectoire de chaque moulin ne peut quant à elle être reconstituée que de manière aléatoire en fonction de la conservation des archives. Comprendre le fonctionnement de l'économie de la farine en Île-de-France suppose donc un changement d'échelle afin de pouvoir analyser les situations particulières à l'aune d'une logique d'ensemble. Cela suppose la constitution et comparaison de diverses bases de données mettant en relation des informations complémentaires : situations, acteurs, valeurs. Ces bases peuvent être constituées à partir de plusieurs types de sources : des cartes, des enquêtes et des minutes notariées. Cette communication aura pour objectif de discuter les conditions d'élaboration de ces bases de données et leurs apports à la compréhension des dynamiques économiques et environnementales de la période moderne à travers l'estimation du potentiel meunier francilien.

11 h 20

Pascal Cristofoli, Évaluations et perspectives d'utilisations des « sources historiques numériques ». Le cas des *Annuaires historiques parisiens 1798-1914*

On assiste aujourd'hui à la multiplication des publications de données numériques ouvertes produites par des processus entièrement automatisés d'extraction d'informations au sein de sources documentaires anciennes. Si cette industrialisation de la production de données historiques numériques est riche en perspectives, l'ampleur et les propriétés de ces corpus posent des défis et problèmes inédits aux historiens et autres potentiels utilisateurs. Comment et à quelles conditions est-il possible de construire et d'utiliser efficacement de telles « sources historiques numériques » ? Nous nous efforcerons de répondre à ces questions en nous appuyant sur l'expérience de la création du corpus des *Annuaires historiques parisiens 1798-1914* [*] qui compte 23 millions d'entrées localisant annuellement, à l'échelle de l'adresse, les activités ou états des personnes et entités collectives enregistrées par les éditeurs de ces ouvrages.

[*] URL : https://nakala.fr/10.34847/nkl.98eem49t Corpus de données élaboré dans le cadre du programme de recherche « Dynamiques sociales en contexte urbain (SoDUCo, 2019-2023) », ANR-18-CE38-0013.

PAUSE

12 h 30

APRÈS-MIDI **CUMULATIVITÉ DES BASES**

14 h 30

Claude Chevaleyre, Mettre le social en bases : une usine à gaz ?

Comment transformer les informations contenues dans les sources historiques en données réutilisables par tous sans sacrifier la texture et la perspective des sources ? Quelles « ontologies », quels modèles de données et quel type de base utiliser ? Comment conserver la main sur la partie technique de la procédure ? Comment assurer la pérennité des données produites ? Depuis l'apparition et la démocratisation de l'informatique jusqu'à l'avènement des grands modèles de langage, ces questions n'ont cessé d'occuper les historiens et de susciter des attentes, parfois démesurées. Nous reviendrons dans cette intervention sur les choix opérés dans le cadre du programme « World of Related Coercions in Work » en prenant pour exemple la base de données du projet *China Human Trafficking and Slaving Database*. Nous reviendrons notamment sur les possibilités offertes par les bases graphes et les approches low-tech, ainsi que sur le problème des ontologies existantes et sur les limites de cette entreprise.

15 h 20

Sébastien de Valeriola, Décloisonner les données de recherche en sciences humaines : SPARQuLb, un écosystème RDF mutualisé

Pour de nombreux chercheurs en sciences humaines, la gestion des données structurées collectées ou produites dans le cadre de projets de recherche représente un défi technique. Par le passé, on faisait largement appel à des prestataires externes à cette fin, ce qui entraînait non seulement des coûts élevés, mais aussi la création d'un grand nombre de silos de données isolés dispersés sur tous les serveurs de l'université, qui n'étaient généralement pas entretenus de manière professionnelle et devenaient rapidement obsolètes. Dans cet exposé, nous présenterons une solution originale à ce problème, déployée à l'Université libre de Bruxelles par Præsto, la plateforme de soutien aux humanités numériques qui travaille en étroite collaboration avec les bibliothèques. Elle consiste à déployer un écosystème RDF partagé

.../...