

LES SCIENCES SOCIALES ET LEURS DONNÉES



Du partage des données à la production de connaissances.

Sonia Guérin-Hamdi.

« Humanités Numériques et données ouvertes »
24-25 mai 2016, Lyon



Les sciences sociales et leurs de données

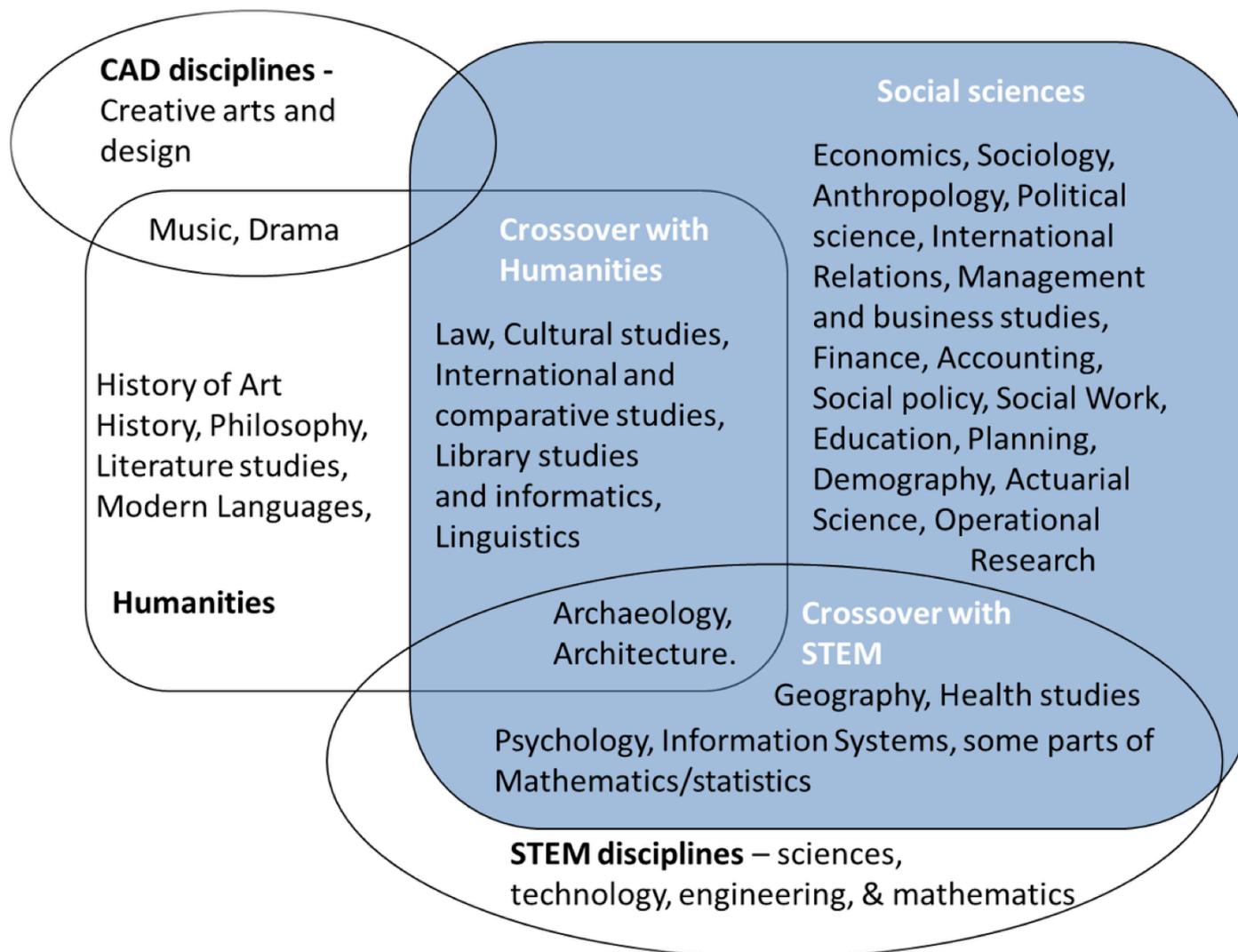
Du partage des données à la production de connaissances

1. La question des données
2. Partage de données dans processus de recherche
3. Cadre méthodologique et technologique
4. Institut des Sciences de l'Homme (ISH) - PANELS

Sciences sociales

- Sciences de l'observation et moins de l'expérimentation
- Le contexte d'une recherche en sciences sociales est **collectif, interdisciplinaire.**
-
- La démarche scientifique,
 - Objet de la recherche: la **Société** et les relations individuelles ou collectives qui s'y nouent.
 - Des disciplines qui se rapprochent par leur objet d'étude.

Sciences Sociales



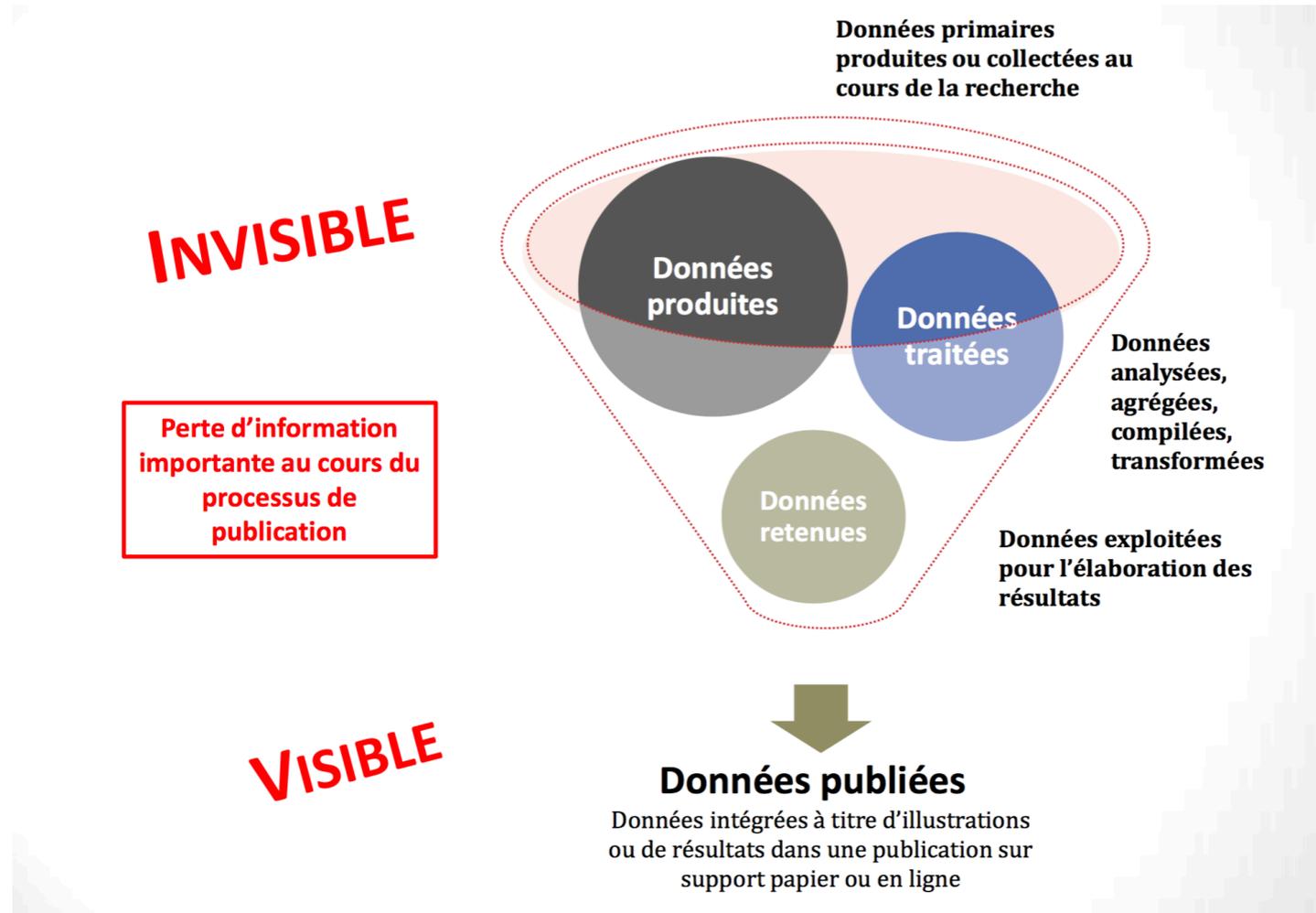
Partage de données

Production

- Tout processus de recherche implique **production** et **consommation** de données, qui peuvent être très diverses selon les disciplines.
- Données produites
 - Diversité de **nature** : issues d'observation de terrain, d'enquêtes, d'enregistrements, de collecte
 - Diversité de **format** : Images fixes, animées, son, texte, tableaux, chiffres, données complexes
 - Diversité de **support** : physiques (archives, document, objets) , numériques ...
 - Diversité **d'état** (brutes ou données finales?) : Il s'agit d'un continuum - cycle de vie de la donnée
- Données « **visibles** » - résultats + publications associées
- Données « **invisibles** » pour le reste de la communauté scientifique - Travaux préparatoires

Partage de données

Production



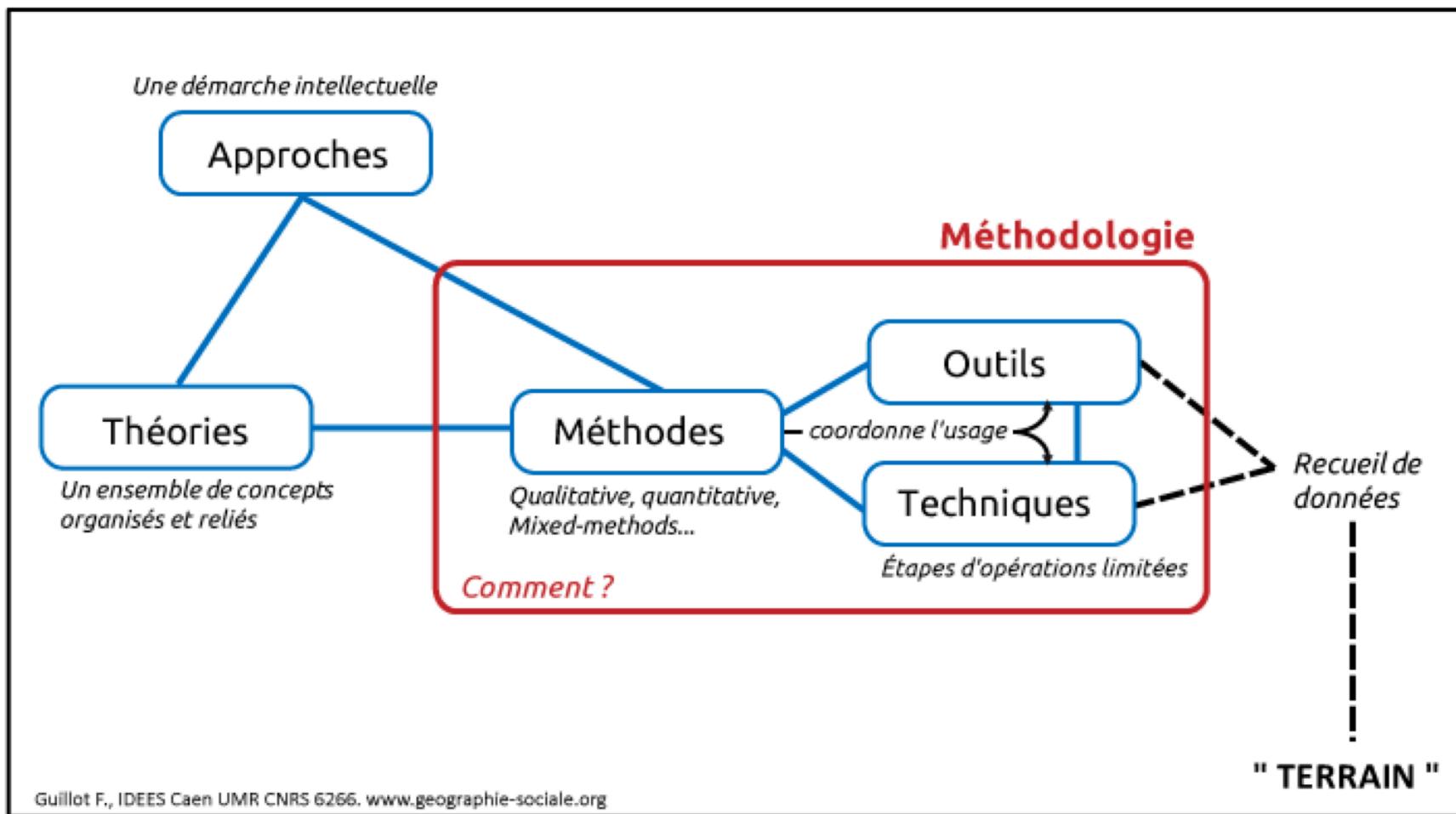
Partage de données

Méthodologie

- « Production de données » ne peut être séparé d'une « réflexion méthodologique » sur les données et sur leur production
- **Méthodologie :**
 - recours à un matériel informationnel collecté par des observations sur le terrain
 - méthodologie peut être définie comme l'ensemble des méthodes, des procédés et des règles permettant de choisir les outils et techniques adaptés à une analyse des données.
 - Elle permet au chercheur de contrôler la **qualité** de ses recherches et de répondre à ses **objectifs**.

Partage de données

Méthodologie



Partage de données

Notion de Qualité

- Définition : évaluation objective fondée sur des critères prédéfinis, liés eux-mêmes à un **besoin**, à un **objectif**.
 - Ex : données d'une enquête sont de « bonne qualité » n'a pas grand sens en soi : tout dépend de ce que l'on veut en faire, autrement dit de l'utilisation à venir.
- Une telle approche doit « *naturellement* » pousser à objectiver le travail de production, à se poser dès le départ la question de l'**usage** des données produites, mais aussi, à repenser leur **organisation**.

Partage de données

Notion de Qualité pour la mise à disposition des données

- **Pertinence** : comment l'information répond aux besoins réels des utilisateurs.
- **Exactitude** : la mesure dans laquelle l'information décrit correctement le phénomène qu'elle devrait évaluer.
- **Accessibilité** : qui renvoie à la facilité avec laquelle on peut l'obtenir auprès du producteur.
- **Interprétabilité** : qui se caractérise par la disponibilité d'informations supplémentaires nécessaires à son interprétation.
- **Cohérence** : la vérification se fait par l'existence d'informations annexes au sujet étudié (dans l'espace et dans le temps)

Partage de données

TGIR PROGEDO

Cadre juridique et institutionnel favorable



- Acteur central de la politique nationale sur les données en sciences sociales.
- Soutient la réalisation des grandes enquêtes européennes et la mise à disposition des données françaises.

Partage de données

TGIR PROGEDO

- CESSDA Fr (- Consortium of European Social Science Data Archives) – Réseau Quételet :
 - Enquêtes académiques réalisées par des chercheurs
 - Statistiques publiques collectées par l'INSEE, les services ministériels, les organismes chargée d'une mission de service public
 - **Partenaires**
 - ADISP, CDSP, INED : diffusion gratuite des données agrégées
 - CASD : garantit confidentialité, sécurité élevée pour pour travailler sur des données très détaillées
- ICPSR : Base internationales dédiée aux sciences sociales comportementales .
- SHARE, ESS, GGP : grandes enquêtes européennes

Partage de données

Réutilisation

- « Partage » implique une acceptation de l'éventualité de la réutilisation des données produites – « ouvrir » « mettre à disposition » ses données.
- Pratiques personnelles propres à chaque chercheur
 - Loin d'un gestion réfléchi et scientifique des données à l'échelle d'une structure / communauté
- Motivations pour le chercheur
 - critères d'acceptation des projets, articles, communications(H2020) : capacité à produire et préserver ses données dans une norme et format accessible, documentée et réutilisable
 - Fonctionnement normal d'une équipe : départ / arrivée projet etc ...

Partage de données

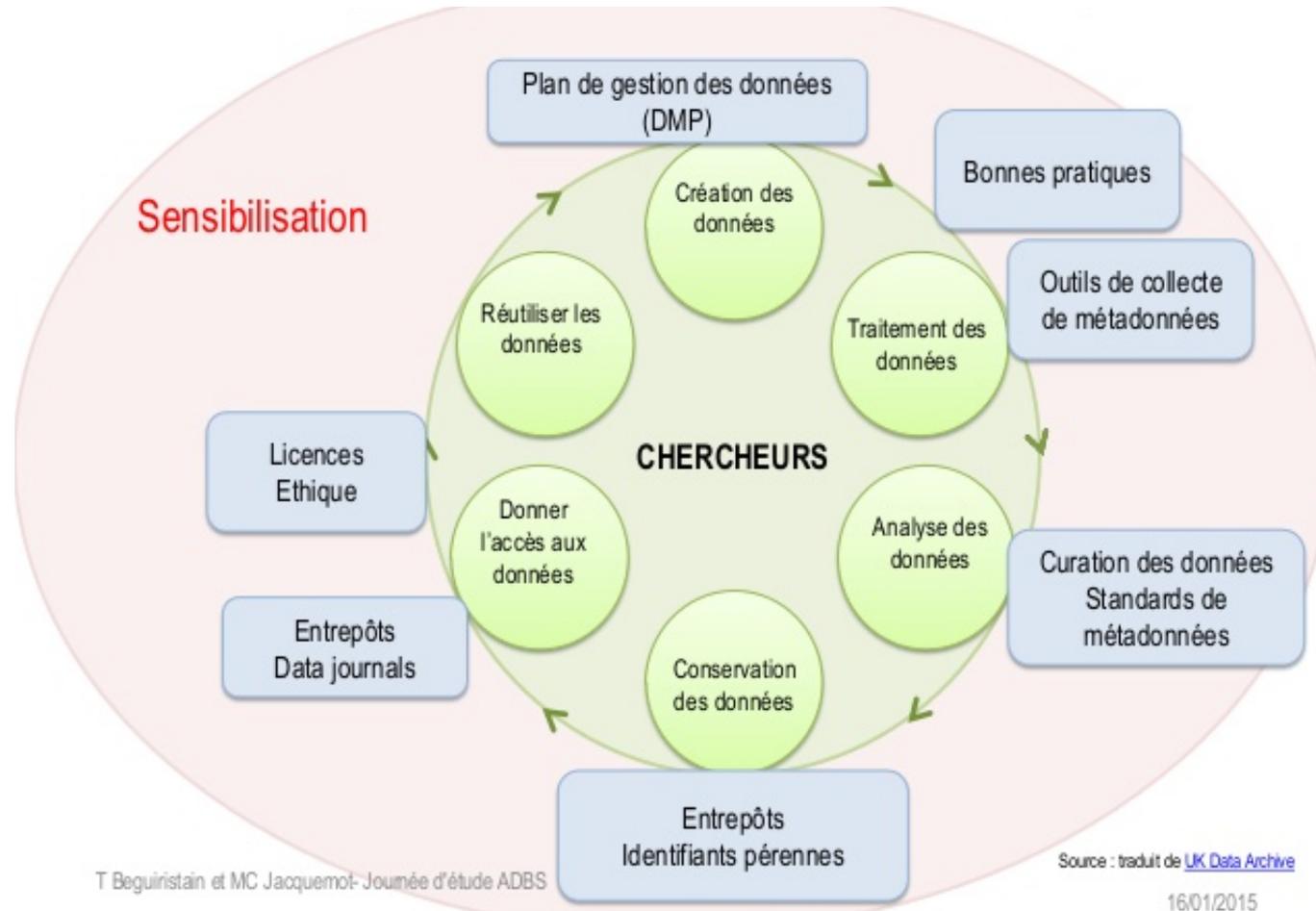
Réutilisation

- Permettre à d'autres de réutiliser, répliquer
 - **Par qui? Pour quoi ? Par d'autres chercheurs, par la veille, les outils du datamining, pour l'innovation....**
- **Conditions de réutilisation**
 - Fiabilité : provenance et qualité et intégrité
 - Accessibilité : cadre juridique (Quelle licence pour quelle diffusion?)
 - Interopérabilité : Quel codage? Quel format?
 - Etre documenté de manière adéquate
 - Éthique : science et société
 - Respecter les règles de confidentialité : données personnelles, etc.
- ***Aider à trouver la bonne formule , dans son contexte, pour son projet, pour ses données***

Cadre méthodologique et technologique

- **Responsabilité** du chercheur :
 - Soucieux de la qualité des données qu'il offre. Donnée = information
- Apprendre les enjeux :
 - Sensibilisation
 - Formation des jeunes chercheurs et doctorants à la gestion et valorisation des données
- S'entourer de compétences
 - Statisticien avec une **culture des données**
 - Documentaliste : Documentation, interopérabilité
 - Informaticien
 - Service juridique : accessibilité
- Accompagnement tout au long du processus de recherche

Cadre méthodologique et technologique



T Beguiristain et MC Jacquemot- Journée d'étude ADBS

Source : traduit de [UK Data Archive](#)

16/01/2015

Institut des Sciences de Homme - ISH

- 21 Laboratoires associés
 - lieux de production et élaboration des données de la recherche
 - Disciplines de recherche : *Anthropologie, Archéologie, Cultures, Droits, Economies, Ergonomie, Géographie, Gestion, Histoire, Information, Informatique, Langage, linguistique, Science politiques, Sociétés, Sociologie ...*
 - Éléments structurants des projets et lieu d'innovation
- se positionne au plus près de la construction des projets de recherche favorisant les collaborations et aide à l'émergence de nouveaux projets(incubateur)
- collabore avec LABEX ASLAN et IMU présents sur site vecteur de l'interdisciplinarité
- accompagne les équipes de recherche des laboratoires dans la réalisation de projets interdisciplinaires sur l'ensemble du travail jusqu'à la valorisation
- Formation avec rapprochement Ecoles Doctorales

ISH – PANELS et PAGES

- Plateformes pour la recherche en sciences humaines et sociales
- PANELS données statistiques et PAGES données textuelles :
 - Environnement scientifique et technique
 - Cadre méthodologique et technologique
- PAGES et PANELS : relais et facilitent l'utilisation des données et services mis à disposition par les TGIRs.

ISH - Sociétés à l'ère du numérique

PANELS

- Destinée aux méthodologies et aux outils dédiés à l'observation (par l'enquête, la veille ou le moissonnage de traces numériques), le traitement, l'analyse statistique et la modélisation.
- Diffusion et la promotion des données du réseau, et toutes les propositions émanant de la TGIR PROGEDO.
- Formations aux utilisateurs (Formations statistiques aux méthodes et outils logiciels, en groupe ou individuelles).
- Membre de l'**ICPSR**, la **plus grande archive de données numérique au monde en sciences sociales comportementales**.

ISH - Sociétés à l'ère du numérique

- Contexte de l'ISH et de ses laboratoires
 - Interdisciplinaire
 - propice à *“la réflexion collective sur les transformations induites par les technologies numériques dans les domaines des sciences sociales”*.

Numérique et nouveaux défis

- Nouveaux matériaux d'enquêtes bruts ou agrégés accessibles aux chercheurs :
 - Divers dispositifs numériques : : sites web, forums de discussion, plateformes interactives, réseaux sociaux ...
 - Données produites par l'activité humaine ou données du monde social—traces numériques, échanges textuels ou visuels, contenus multimédias
 - Données diffusées en open data
 - Données rendus accessibles par les banques de données
 - Données modélisées selon le standard RDF du web sémantique
 - ...
- Nouveaux terrains d'investigation :
 - phénomènes sociaux, culturels, politiques, économiques ...
- Objet de recherche complexe : Internet
 - Web ouvert, vaste, hétérogène et en évolution, en relation permanente avec les différentes sphères de la vie sociale

Numérique et nouveaux défis

Numérique comme instrument d'analyse

- Nouveaux dispositifs d'investigation scientifiques :
 - observation et analyse du web. ex: cartographie, analyse des réseaux sociaux,
- Nouvelles méthodes et nouveaux outils :
 - Ouverture des logiciels / codes,
 - Collecte Automatique de Données via Crawlers , API / service web
 - Datamining pour extraire l'information

Numérique et nouveaux défis

Numérique comme objet d'étude

Le numérique donne naissance à des nouvelles formes d'information ou de création.

- Qu'apportent les formes spécifiques du journalisme numérique (web documentaires, data journalisme...)
- Le recul nécessaire à la compréhension d'un événement est-il possible quand le numérique donne la possibilité d'une information en temps réel?

Numérique comme outil d'enrichissement

Les outils numériques enrichissent les modalités possibles de lecture et d'écriture.

- Web sémantique - LOD
- Comment formuler des hypothèses de lecture en usant du numérique (associer un son, un mouvement ou une image à un texte...)?

LES SCIENCES SOCIALES ET LEURS DE DONNÉES

Merci.

sonia.guerin-hamdi@ish-lyon.cnrs.fr

« Humanités Numériques et données ouvertes »
24-25 mai 2016, Lyon

