



DONNÉES DE RECHERCHE EN SHS

Pratiques, représentations et attentes des chercheurs : une enquête à l'université Rennes 2

Journée d'étude MaDICS-ADOC « Variété des données SHS »
Conférence INFORSID
Nantes, 28 mai 2018



Maison
DES
Sciences
de l'Homme
EN
Bretagne

Morgane Mignon (Maison des Sciences de l'Homme en Bretagne)

L' ENQUÊTE : PAR QUI ET SUR QUI ?



Maison ^{DES}
Sciences
de l'Homme
EN Bretagne



- **Plusieurs acteurs**
 - Co-organisateur : URFIST de Rennes, MSHB et SCD Rennes 2
 - Soutien des instances de l'université Rennes 2
- **Périmètre**
 - Université Rennes 2 : les SHS
 - Les enseignants-chercheurs, chercheurs et ingénieurs
- **3 objectifs principaux**
 - Établir un état des lieux
 - Élaborer une offre de services
 - Alimenter la recherche sur les pratiques informationnelles

L'ENQUÊTE : QUELLE MÉTHODOLOGIE ?

- **L'enquête quantitative : questionnaire statistique en ligne (sous LimeSurvey) :**
 - **Sources du questionnaire :**
 - [Les données de recherche en SHS. Une enquête à l'Université de Lille 3](#)
et deux autres enquêtes
 - retravaillées
 - **Un long processus d'élaboration (depuis mai 2016)**
 - ⇒ lancement le 6 février 2017
 - **32 questions structurées autour de cinq grands axes :**
 - Informations personnelles
 - Typologie des données de recherche
 - Pratiques de stockage et d'archivage
 - Pratiques de partage et de diffusion
 - Besoins

L'ENQUÊTE : QUELLE MÉTHODOLOGIE ?

- **L'enquête qualitative : des entretiens** : pour prolonger et préciser les réponses du questionnaire
 - **Entretiens enregistrés, d'une demi-heure environ**
 - Chercheurs volontaires
 - VP, Directeurs de labos volontaires
 - **Scénario d'entretien selon 2 modèles** :
 - Axé sur les pratiques individuelles et les représentations pour les chercheurs + retour sur le questionnaire
 - Davantage orienté sur le positionnement par rapport à une politique des données de recherche et les besoins d'un labo, pour les Directeurs de labo, les VP.
 - **Total de 21 entretiens** :

Durée totale :
12:01:01
Parties thématiques :
11:59:17 (99,8%)
Nombre d'extraits :
267





LES RÉSULTATS

Partie 1

Informations personnelles

INFORMATIONS SOCIO-DÉMOGRAPHIQUES

- Tranche d'âge
- Statut
- Responsabilités actuelles
 - Éducatives
 - Scientifiques
 - Encadrement de thèses
 - Politiques
 - Éditoriales
 - Administratives)
- Laboratoire d'appartenance
- Section CNU

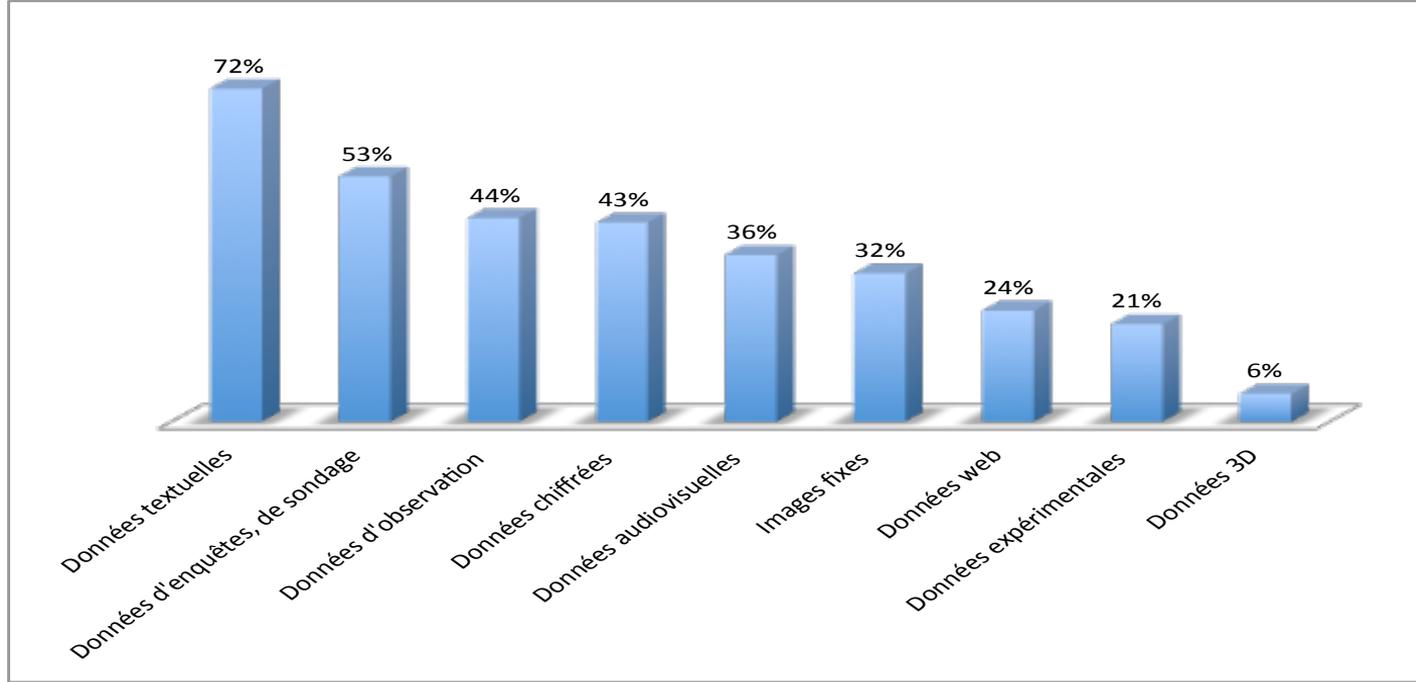


LES RÉSULTATS

Partie 2

Typologie des données

QUELLES CATÉGORIES DE DONNÉES SOURCES ?



QUELLES CATÉGORIES DE DONNÉES SOURCES ?

« du matériau langagier [extrait d'] un terrain ethnographique [investi] sur la durée, sur la profondeur, avec une compréhension très contextualisée de la situation ; des conversations ; du témoignage ethnographique ; des entretiens ; des biographies langagières orales ou écrites ; des documents d'archive ; des documents audiovisuels ; des émissions radiophoniques; des corpus de presse; des programmes, les discours des inspecteurs généraux, des manuels ; des corpus de dossiers d'aide à l'enfance ; des corpus de textes de loi ; des échanges avec les artistes ; des interviews avec des collègues ; des relevés GPS ; des données d'enquête ; des images ; des textes ; des corpus de textes ; des corpus visuels ; des adresses ; des ouvrages ; des articles ; des facsimiles d'ouvrages ; des documents ; des brochures ; des photos ; des fragments de pierres, des inscriptions ; de la donnée géographique numérique ; des cartes papier ; des données de type raster, matriciel, tout ce qui est image, photographie aérienne, image satellite ; de la donnée vectorielle, avec une composante géométrique (la donnée c'est de l'objet) ; de la donnée attributaire (ce qui vient décrire l'objet ; des tables) ; des informations informelles glanées lors de colloques ; des données récupérées sur Internet ; des SMS ; des enregistrements de webcams ; des vidéos (youtube ou autres plateformes) ; des pages de forums Internet, de discussion, sites web, ; des pages de tchat SOS ; des pages ou échanges de réseaux sociaux, comptes twitter, Facebook, ; des PDF d'œuvres sur Internet ; des index et listes ; des photos aériennes. ; des données numériques, statistiques ; des cartes ; des données archéologiques ; des captures d'écran ; des traces GPS ; des trajectoires de marche ; des indicateurs liés à de petites tâches cognitives ; des données territorialisées... »

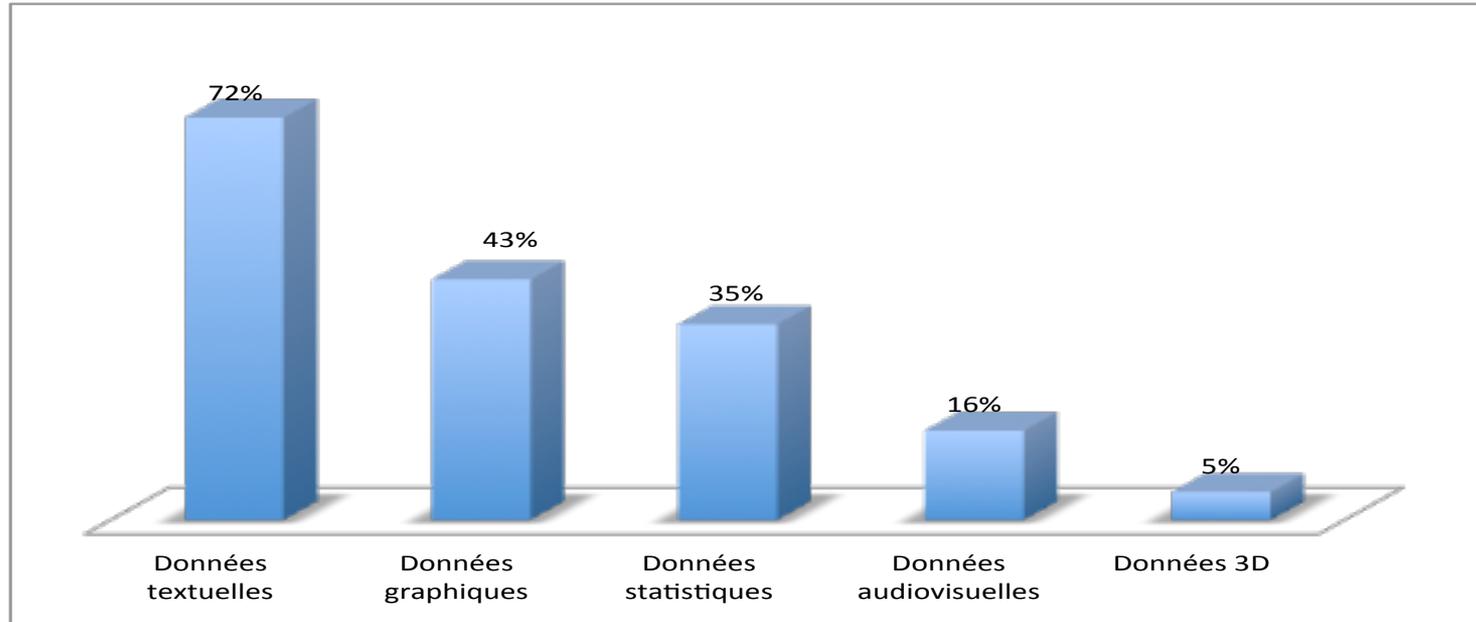
TYPES DE DONNÉES ET DISCIPLINES

« La donnée que je manipule est diverse »

« Après, il y a des pratiques qui sont différentes selon les disciplines. Principalement en géographie et sociologie. Et selon les disciplines, on mobilise des matériaux qui ne sont pas forcément les mêmes, et avec des niveaux d'appropriation des outils numériques variables d'un collègue à l'autre. »

« Ce n'est pas tant une question de discipline, mais de paradigme de recherche avec des méthodes qui ne sont pas les mêmes. »

QUELLES CATÉGORIES DE DONNÉES PRODUITES ?



QUELLES CATÉGORIES DE DONNÉES PRODUITES ?

«des données textuelles; des bases de données ; des enregistrements audio ou vidéo d'entretiens ; des notes sur des entretiens ; des questionnaires ; des formulaires papier remplis ; des tableaux Excel ; des fichiers Word ; plusieurs versions de textes ; des fiches à partir de documents d'archive ; des sites ou plateformes internet ; des retranscriptions d'entretiens ; des analyses statistiques lexicales ; des dossiers PDF, des textes, essentiellement des textes ; des livres ; des revues ; des données Excel ; des données chiffrées ; des données qualitatives ; des études de cas cliniques ; des logiciels ; des tâches ; des petits programmes ; de petites épreuves cognitives; des tests de raisonnement; des fichiers texte avec reconnaissance de caractères, des transcriptions ; des documents numérisés ; des photographies de documents originaux ; des enregistrements vidéo ; des rushs ; des documentaires ; des catalogues prosopographiques ; des arbres généalogiques ; de recensements d'archives ; des campagnes de photos ; des fiches thématiques ; des bases de données géographiques ; des traces GPS ; des indicateurs ; des données biologiques, des dosages ; des électrocardiogrammes ; des consentement de participation (papier) ; des protocoles ; des fichiers Latex ; des fichiers Matlab ; des fichiers partagés ; des bibliographies ; des fichiers statistiques ; des nuages de points, des captures de mouvement ; des photographies aériennes ; des cartes ; des croquis ; des enregistrements sonores suivis sur plusieurs années ; des photographies suivies sur plusieurs années ; des corpus multilingues ; des carnets papier... »

VERS UNE DÉFINITION DES DONNÉES DE RECHERCHE

- Les données sont une construction

« Le problème, c'est que l'on parle de "données" mais en fait, pour moi, ce ne sont pas des données, ce sont des construits. [...] Le mot "donnée", concrètement cela m'embête. Cela veut dire que les choses sont données. Si vous avez construit tout un appareil méthodologique qui permet de prélever des éléments, en aucun cas c'est donné, mais c'est construit, prélevé... Donc cela pose problème déjà, parce que les "données", on a l'impression que c'est quelque chose qui est donné, physique, factuel, que les choses sont comme ça, en fait non. »

- C'est le travail de recherche qui confère une valeur scientifique aux données

« Ma collecte de données, ce sont [...] tous les échanges avec les artistes, mais qui ne sont pas méthodiques, ce ne sont pas des formes d'enquêtes ou d'interviews, mais c'est divers types d'échanges à diverses occasions qui peuvent avoir l'air de rien du tout mais qui sont des données sur lesquelles je m'appuie réellement. »

VERS UNE DÉFINITION DES DONNÉES DE RECHERCHE

- Rapport mutuel d'influence entre les données et les pratiques de recherche

« Et puis après effectivement, tous ces logiciels comme Gephi, Paraview [...] cela m'a sorti de ce qu'on a appelé la lecture de près pour aller vers d'autres pratiques de géographe, vers la spatialisation des données. Donc cela a bouleversé ma façon de travailler, cela a radicalisé ma façon de travailler. »

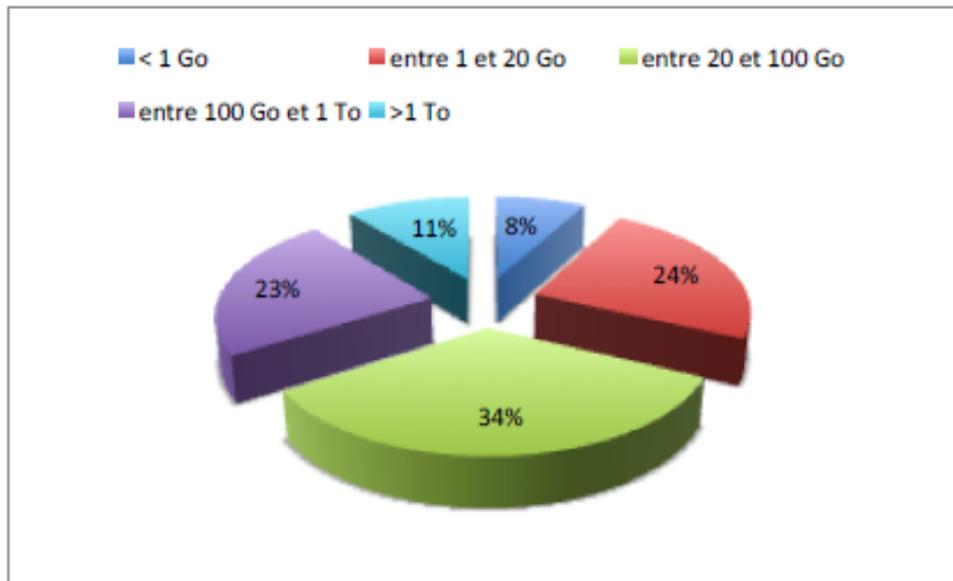


LES RÉSULTATS

Partie 3

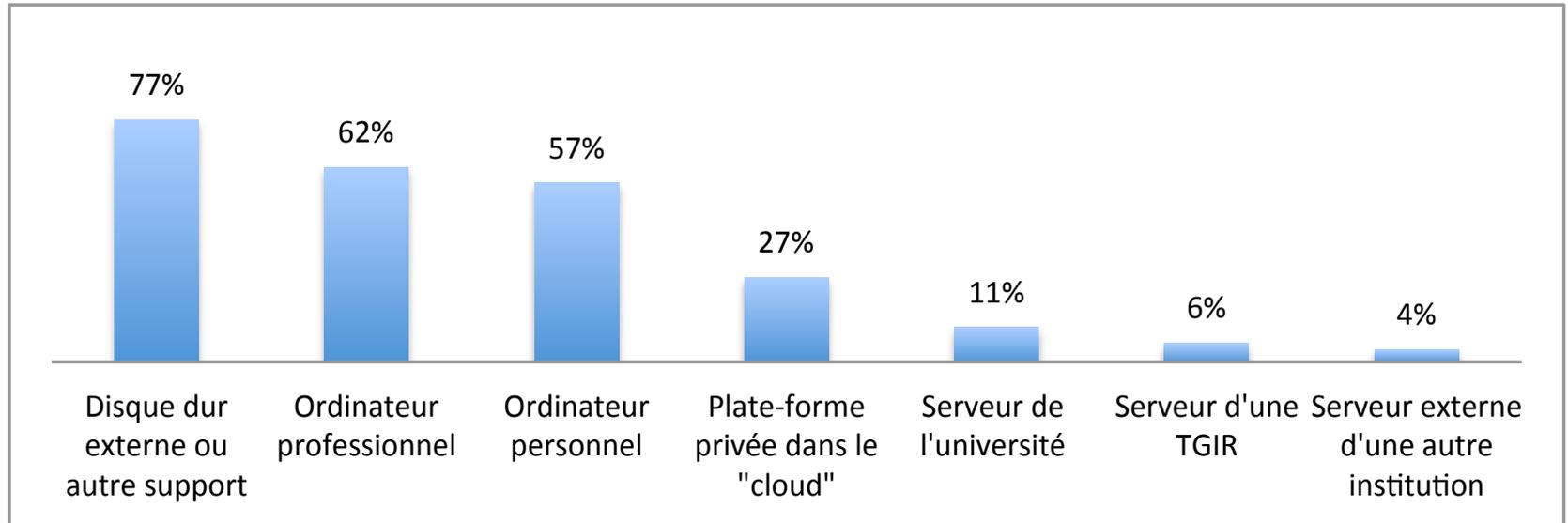
Stockage et archivage

VOLUME DES DONNÉES



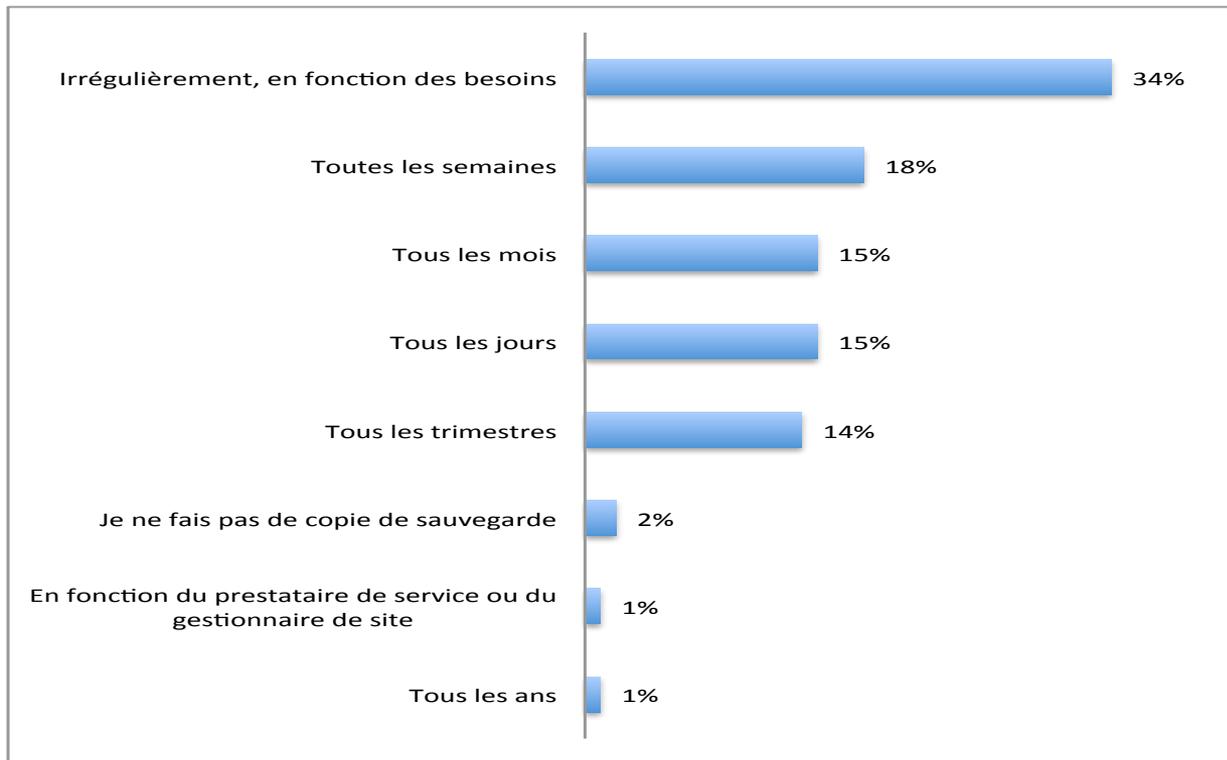
**Espace de stockage
requis par les chercheurs
(N = 95)**

OÙ SONT STOCKÉES LES DONNÉES ?

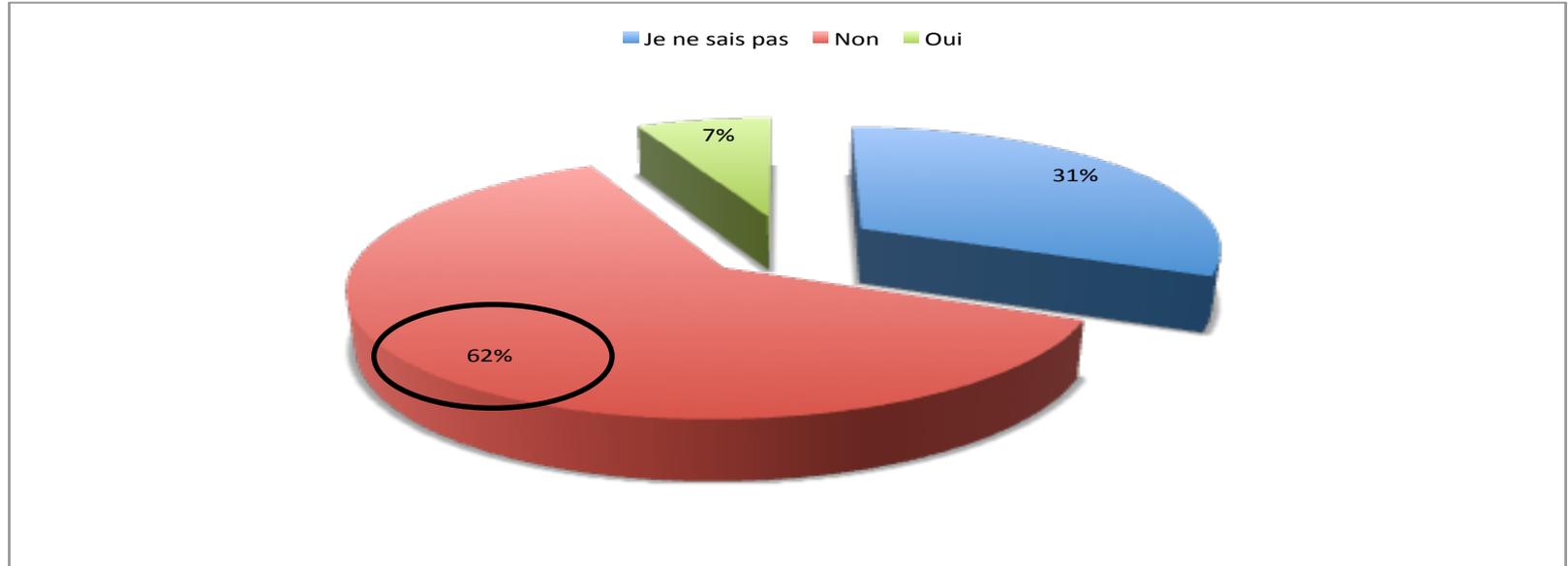


Lieux de stockage (N =122)

SAUVEGARDE DES DONNÉES : LES PRATIQUES



ARCHIVAGE DES DONNÉES : LES PRATIQUES



Y a-t-il une politique d'archivage dans le laboratoire ? N = 122

MÉTADONNÉES : LES PRATIQUES

Métadonnées	Nombre de réponses	Pourcentage
N'attribue pas d'informations supplémentaires	57	50%
Des informations scientifiques	55	48%
Des informations administratives	29	25%
Des informations techniques	14	12%

Types de métadonnées attribuées (N = 121)

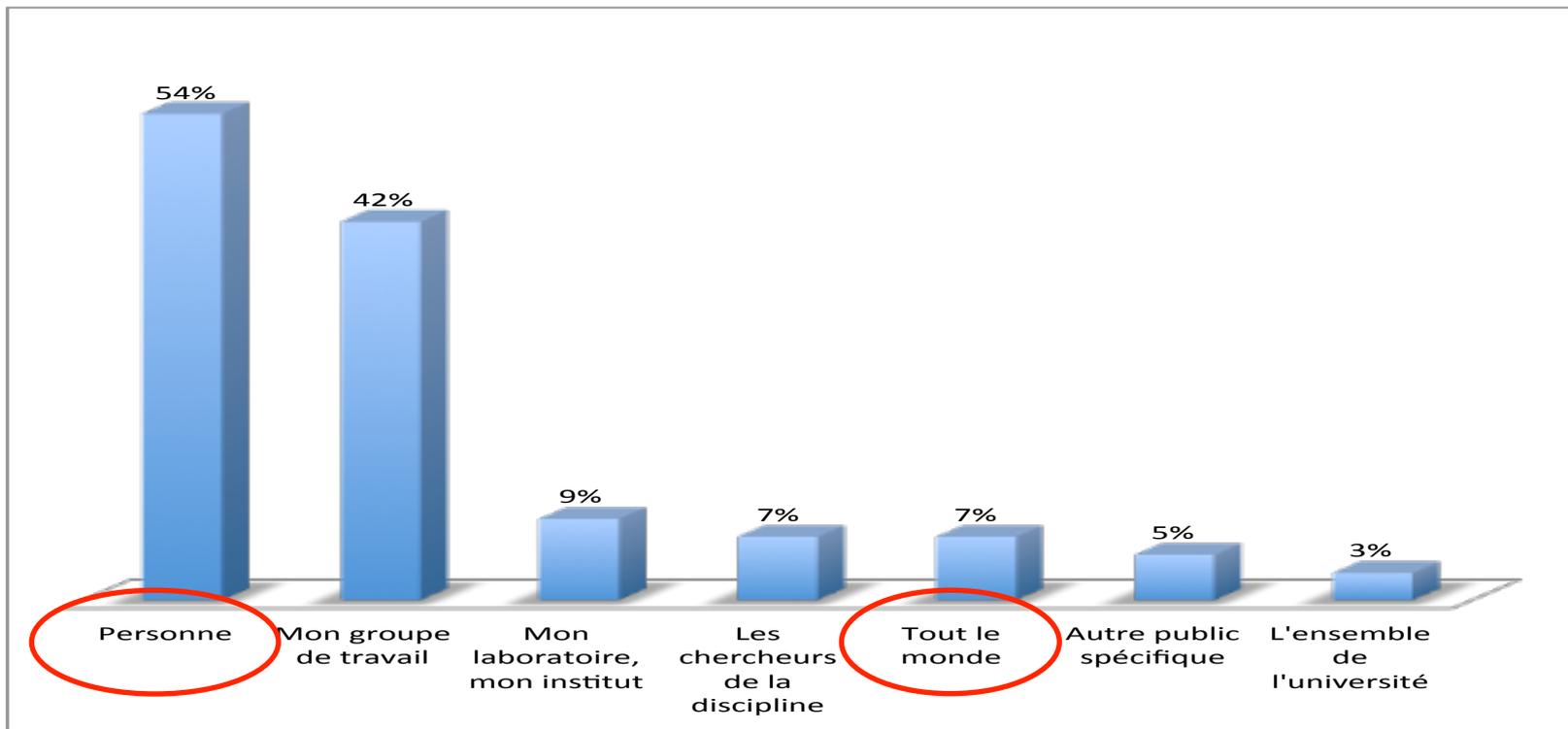


LES RÉSULTATS

Partie 4

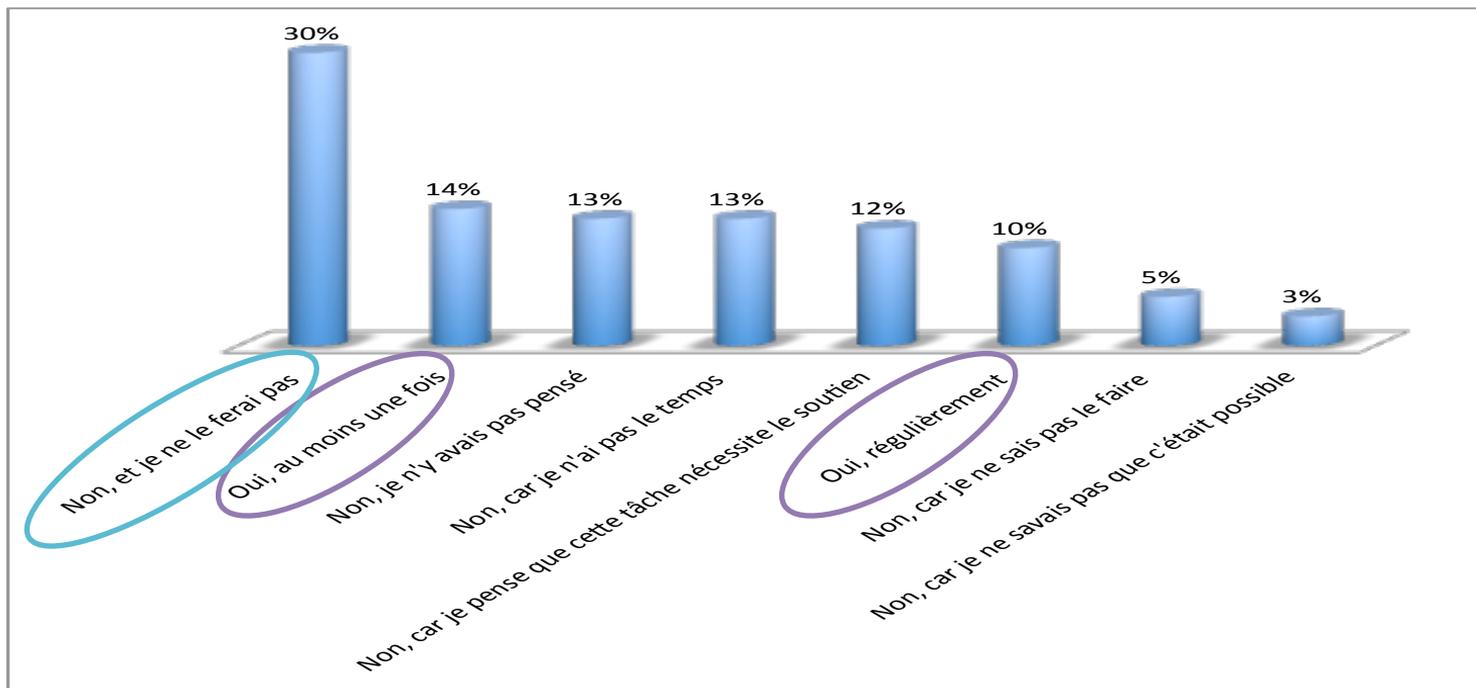
Partage et diffusion des données

PARTAGE DES DONNÉES : LA RÉALITÉ



« Qui peut accéder à vos données ? » N = 121

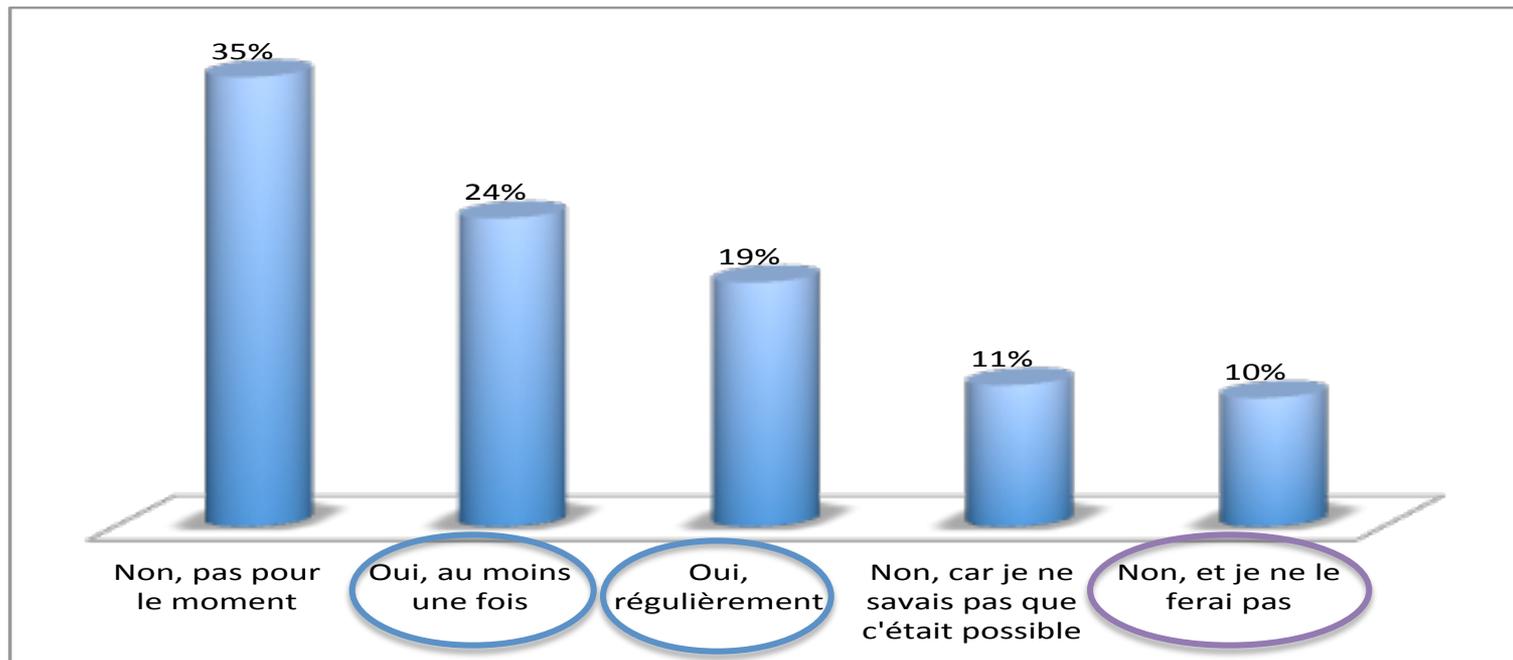
PARTAGE DES DONNÉES : LES PRATIQUES



« Vous-même, avez-vous déjà stocké vos données en libre accès sur le web ? »

N = 116

PARTAGE DES DONNÉES : LES PRATIQUES



« Avez-vous déjà téléchargé les données d'autres chercheurs ? »

N = 121

PARTAGE DES DONNÉES : LES REPRÉSENTATIONS

	Nb de réponses	%
Une meilleure visibilité de vos travaux de recherche	70	65%
L'occasion de nouveaux contacts, de nouvelles coopérations scientifiques	58	54%
Le souhait de voir se développer d'autres recherches à partir de vos jeux de données	43	40%
L'adhésion aux valeurs du Libre Accès aux résultats de la recherche scientifique	43	40%
Une obligation faite par le financeur de votre recherche	35	33%
La reconnaissance de la communauté scientifique	34	32%
Le souci de pouvoir faire valider vos résultats à partir des jeux de données	12	11%
De possible retombées financières	8	7%

Facteurs d'incitation au libre accès de ses données de recherche

N = 107

PARTAGE DES DONNÉES : LES REPRÉSENTATIONS

**Facteurs de frein
au libre accès
de ses données
de recherche
N = 108**

	Nb de réponses	Pourcentage
Des freins juridiques liés à la nature de vos données	58	54%
Le manque de temps	55	51%
La crainte du plagiat	40	37%
Des raisons scientifiques	29	27%
Le manque de connaissances dans la description des jeux de données	28	26%
Le manque de compétences	20	19%
Des raisons personnelles	13	12%
Des raisons économiques	8	7%
Des raisons institutionnelles	3	3%

PARTAGE DES DONNÉES : LES SOUHAITS

	Nombre de réponses	Pourcentage
Une archive ouverte (de type HAL)	43	45%
Une plateforme spécialisée dans l'archivage et la diffusion des données de recherche	39	41%
Un réseau social académique	26	27%
Le site ou le blog du laboratoire	25	26%
Votre blog ou votre site personnel	20	21%
Un serveur de l'université	19	20%
Un data paper	8	8%

Si vous souhaitez diffuser vos données en libre accès, quel est (ou serait) votre support principal ? N = 96

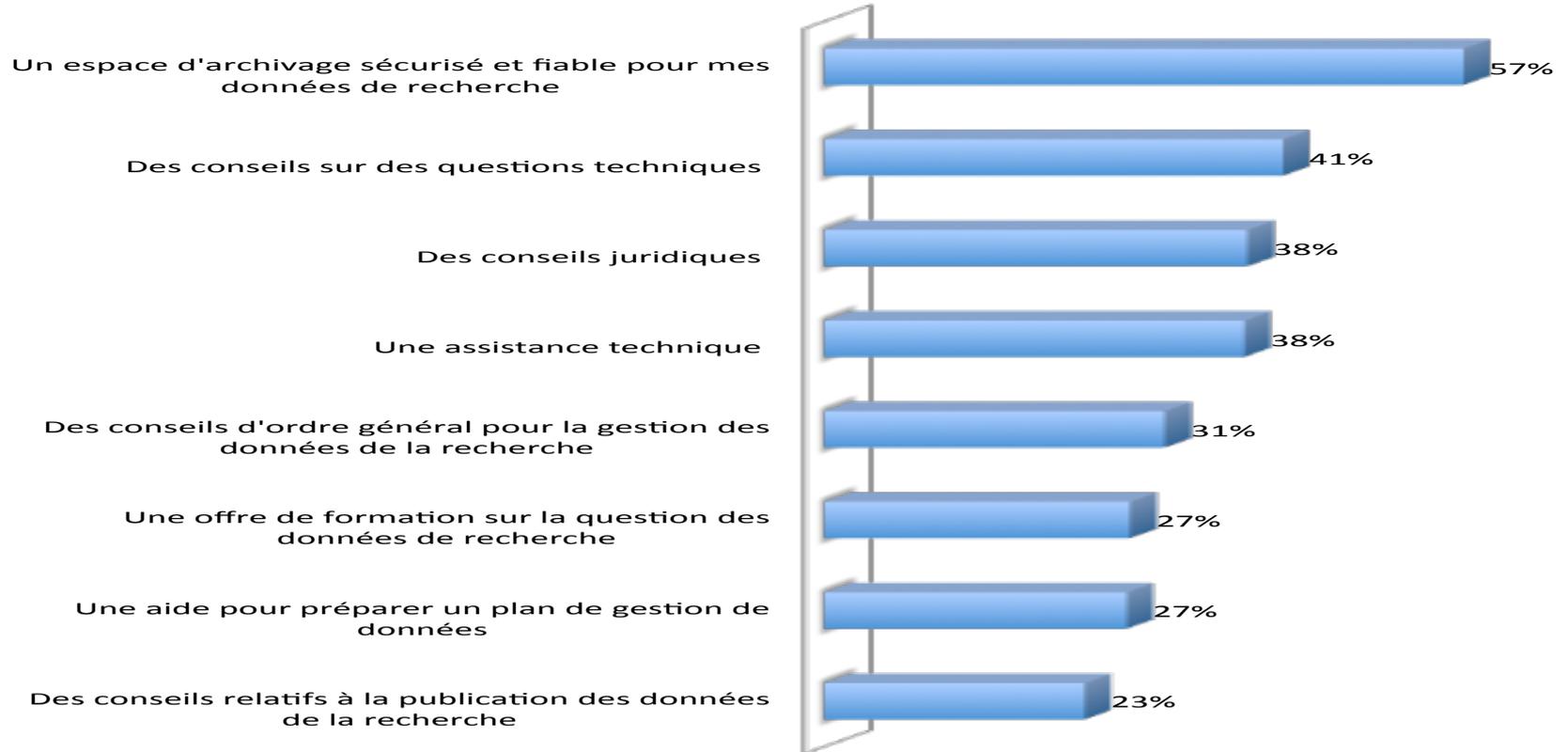


LES RÉSULTATS

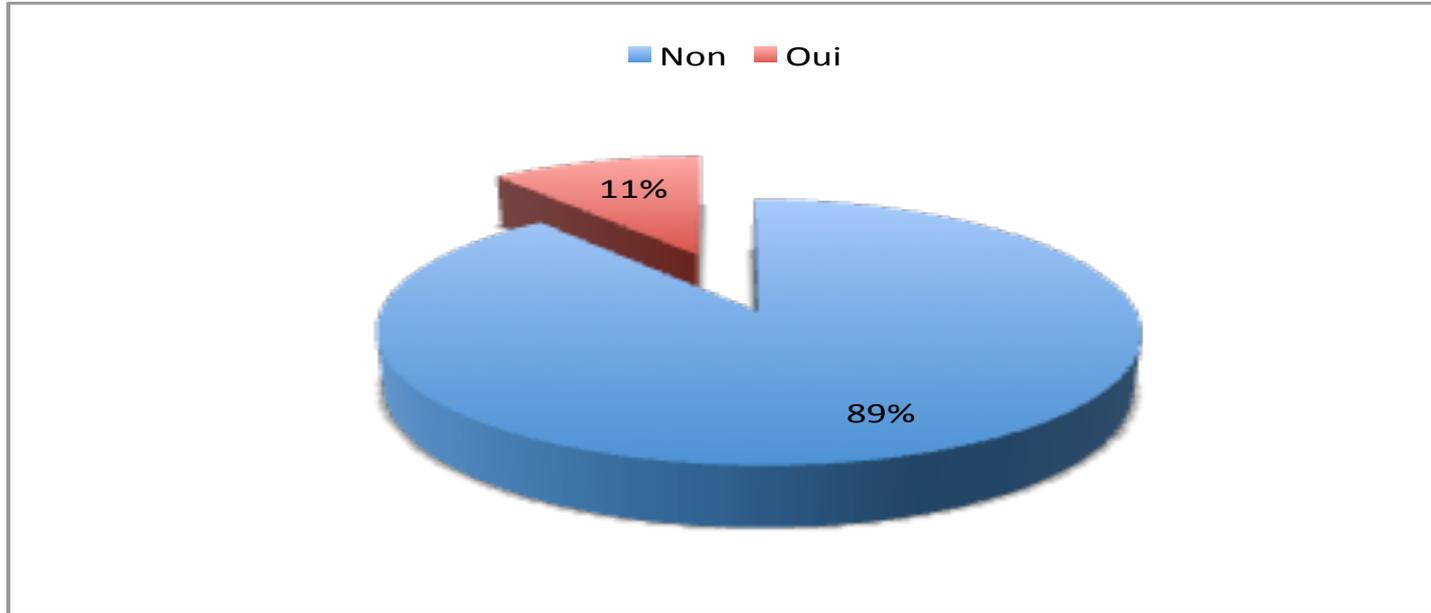
Partie 5

Besoins et attentes des chercheurs

BESOINS ET ATTENTES



PLAN DE GESTION DES DONNÉES : (MÉ)CONNAISSANCE



Savez-vous ce que c'est qu'un plan de gestion de données ?

N = 96



CONCLUSIONS ET PROPOSITIONS

EN CONCLUSION : QUELLES LEÇONS ? SUR LE FOND

→ Confirmation de la complexité des données de recherche en SHS

- Difficultés des définitions
- Importance des aspects épistémologiques en SHS : « données ou obtenues »
- Diversité disciplinaire, des domaines de recherche, importance du niveau micro
 - Impossible d'élaborer une politique générale et unique pour tous

→ Identification de services à mettre en œuvre

- Espaces de stockage à la fois sécurisés et collaboratifs
- Accompagnement en terme de gestion des données

EN CONCLUSION : 6 PROPOSITIONS

- 1/ Créer un **groupe de travail** chargé de la réflexion sur une politique des données de recherche
- 2/ Développer des **services mutualisés** pour la sensibilisation et la formation
- 3/ Offrir un **service juridique** spécialisé
- 4/ Mieux adapter les espaces internes de **stockage des données** de recherche aux besoins individuels et collectifs des chercheurs
- 5/ **Sensibiliser à la description** des jeux de données et à la rédaction des Plans de Gestion de Données
- 6/ Sensibiliser aux **solutions institutionnelles** de partage et d'archivage des jeux de données

EN CONCLUSION :

RECOMMANDATIONS GÉNÉRALES

- **Partir de la complexité de la question des données**
- **Adapter les discours aux préoccupations des chercheurs (pas de "discours d'injonction" sur le partage des données)**
- **Décider du "bon niveau" d'une politique des DR :**
 - Une seule politique d'établissement impossible et inefficace
 - Déterminer ce qui relève de chaque niveau de décision :
 - les communautés de recherche
 - les unités de recherche
 - l'établissement
 - les différents établissements de tutelle : pour les UMR

LES CHERCHEURS ET LEURS DONNÉES EN SHS : QUELLES PRATIQUES, QUELLE(S) POLITIQUE(S) ?



Journée d'étude

Vendredi 15 juin 2018
Maison des Sciences de l'Homme en Bretagne (MSHB)
Amphithéâtre Robert Castel
2 avenue Gaston Berger, Rennes

Pour plus d'information :
<https://jedonnees-shs.sciencesconf.org/>

<https://jedonnees-shs.sciencesconf.org/>

Rapport de l'enquête *Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'université Rennes 2*

En libre accès sur Hal :

<https://hal.archives-ouvertes.fr/hal-01635186>

- Rapport de synthèse
- Annexes (avec tous les tris à plat de l'enquête statistique) données brutes des croisements statistiques
- Visualisation des données sur [Tableau public](#)

Présentation de l'enquête

- Journée d'étude le 15/06 co-organisée par l'URFIST de Rennes, la MSHB, le SCD Rennes 2 et la MSH Ange Guépin
- Avec le soutien de l'université Rennes 2, du Réseau national des MSH et du GIS Réseau des URFIST

CONTACTS

URFIST de Rennes

alexandre.serres@univ-rennes2.fr

marie-laure.malingre@univ-rennes2.fr

SCD Rennes 2

cecile.pierre@univ-rennes2.fr

didier.collet@univ-rennes2.fr

MSHB

morgane.mignon@mshb.fr