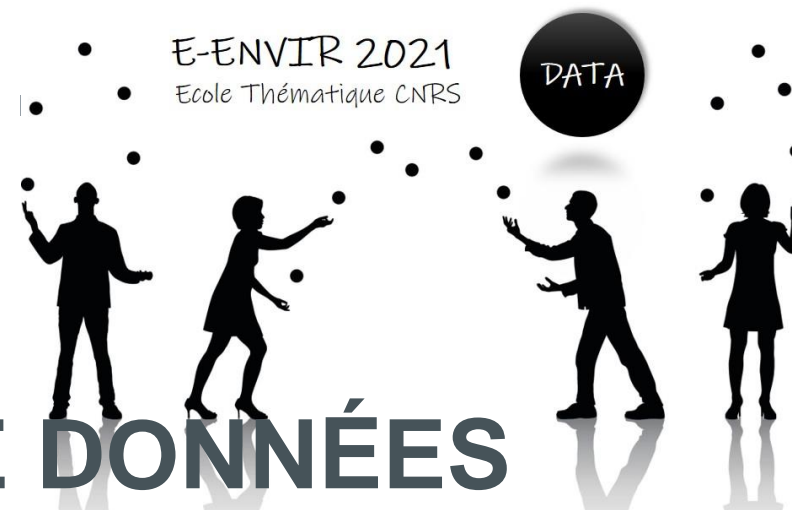


TD - CARACTÉRISATION ET PRÉSENTATION DE MON JEU DE DONNÉES

ANIMATION : AMANDINE HÉNON ET EMILIE LERIGOLEUR



Source : issue du film de Julio Pot "Who are you ?"



Source :Cenis

OBJECTIFS DE LA SESSION

- Découvrir les résultats du questionnaire en ligne
 - Apprendre à se connaître
 - Découvrir la diversité des jeux de données à “FAIRiser”
- Découvrir les jeux de données exemples utilisés durant tout l'École Thématique

Enquête e-Envir 2021

Cette enquête vise à situer en amont de l'école thématique vos connaissances sur la gestion des données et à caractériser le jeu de données avec lequel vous allez travailler durant l'École Thématique.

Merci de répondre autant que possible à l'ensemble des questions et d'utiliser avec parcimonie les réponses "non concerné" et "je ne sais pas". Si votre jeu de données est complexe et satisfait plusieurs réponses, tenter toutefois d'indiquer la réponse qui soit "dominante". Les résultats de cette enquête resteront anonymes et présentés de manière synthétique en introduction de l'E.T.

Temps de remplissage : 20min

Vous pouvez à tout moment arrêter et reprendre ultérieurement l'enregistrement

Il y a 37 questions dans ce questionnaire.

Suivant

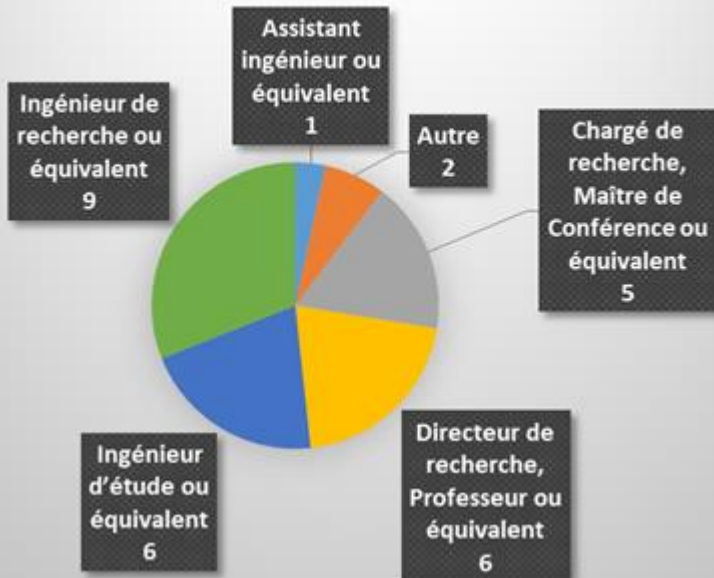
L'ENQUÊTE EN BREF

- 29 répondants
- Structure du questionnaire :
 - La première partie se focalise sur “vous” : votre profil, votre positionnement dans le paysage de la recherche et votre relation à la donnée. Connaître également vos connaissances des principes FAIR et de l’Open Science.
 - La seconde partie permet d’en savoir plus sur vos jeux de données : discipline, thématique, type de données, lien avec un projet, des partenaires, volumétrie, format...

Nous allons ici analyser l’hétérogénéité de vos réponses afin de dresser à la fois le profil des répondants mais aussi des données qui nous accompagneront durant cette École Thématique.

PROFIL DES RÉPONDANTS

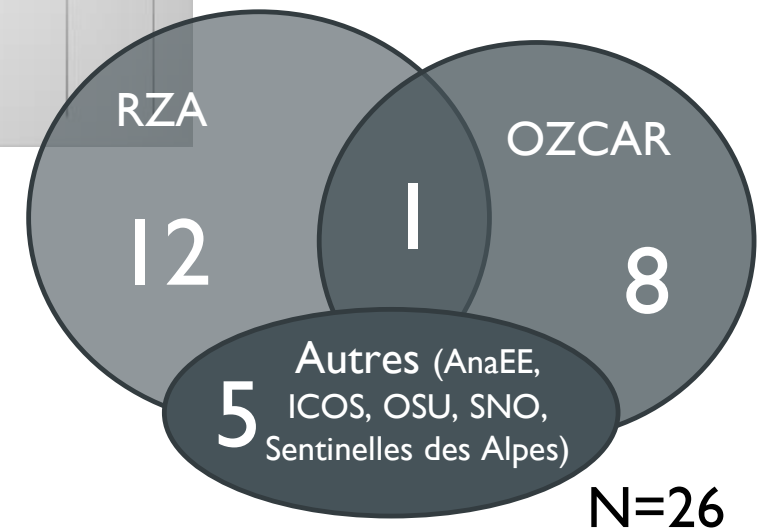
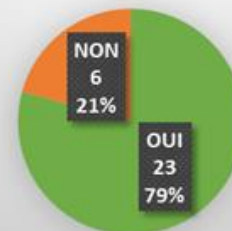
Fonction des répondants



Quel est votre rôle dans cette IR ?

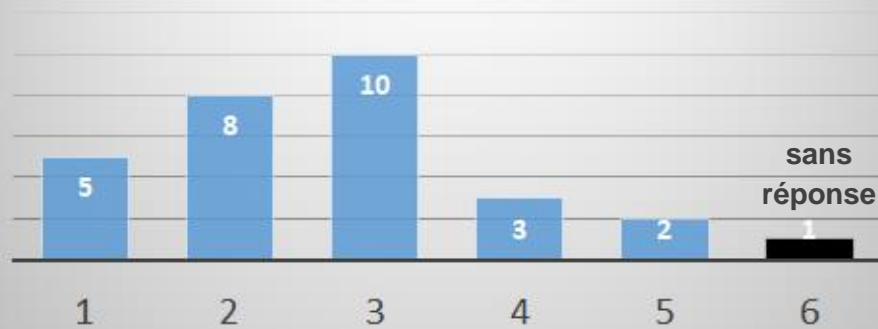


Vos recherches et données produites concernent-elles une infrastructure de recherche (IR) ?



NIVEAU DE CONNAISSANCE ET MOTIVATION ⇔ “FAIR”

Quel est votre niveau de connaissance sur la FAIRisation des données ?



Comment évaluez-vous votre motivation à produire des données FAIR et ouvertes autant que nécessaire à compter d'aujourd'hui et dans un futur proche ?



faible  forte

Quel est votre niveau de connaissance sur l'ouverture des données ?



aucune connaissance  expert ou très bonne maîtrise des connaissances

Vos remarques

La FAIRisation comme moyen de :

- banqueriser ses données, les promouvoir.
- préserver ses données pour une réutilisation.
- anticiper la bonne gestion des données le long du cycle de vie.
- impliquer la majorité des acteurs de la donnée.

Cependant, plusieurs personnes pointent la complexité du processus, les contraintes et ne semblent pas convaincues.

VOS JEUX DE DONNÉES

Quel type de données ?

Données spatiales, temporelles ou spatio-temporelles
Données qualitatives
Données quantitatives
Images, photos
Données d'occurrence
Enquête
Inventaire
Archive
Mesures
Données issues de capteurs
Analyses
Ressources pédagogiques

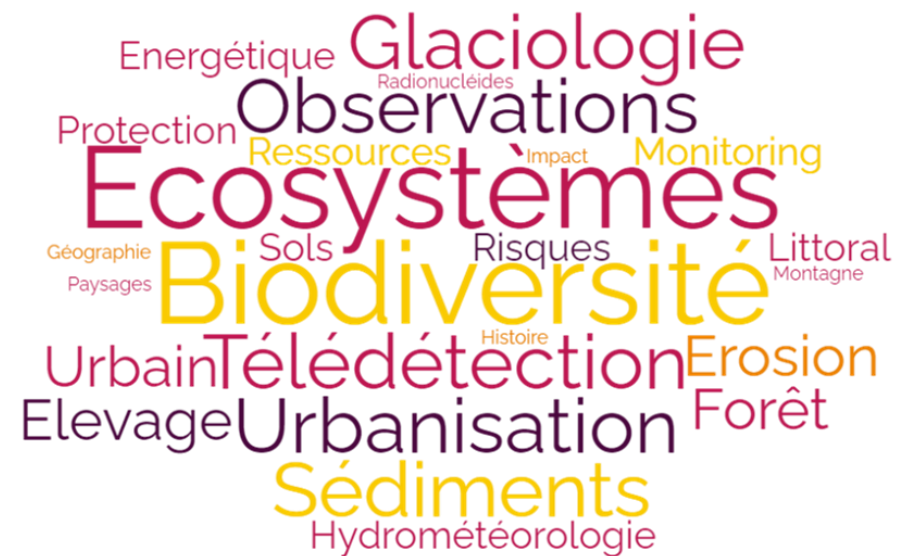


VOS JEUX DE DONNÉES

Les disciplines

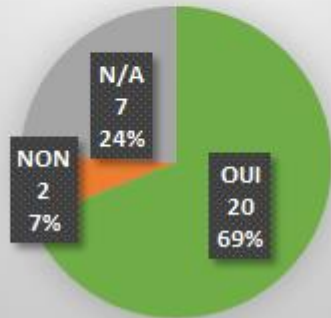


Les thématiques

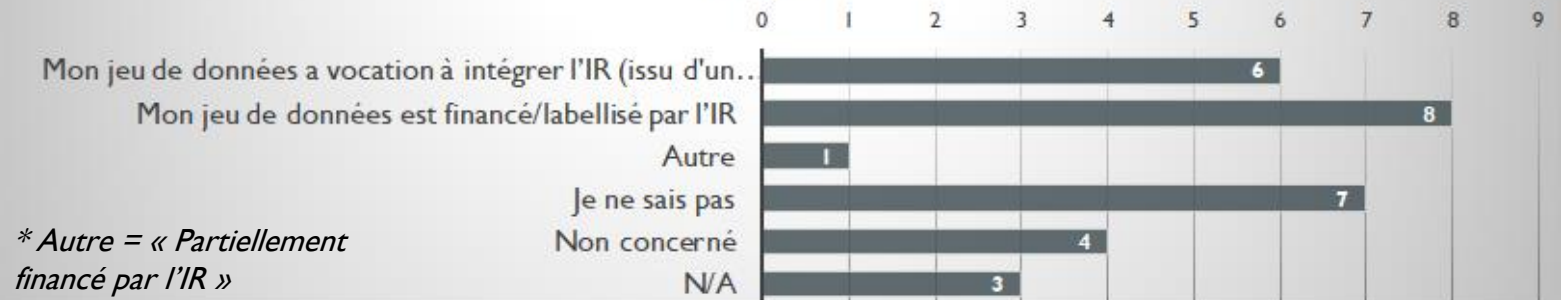


QUELS LIENS AVEC UN PROJET ET/OU UNE INFRASTRUCTURE ?

Ce jeu de données est-il en lien avec un ou plusieurs projet(s) de recherche ?



Quel est le positionnement de votre jeu de données par rapport aux IR ?

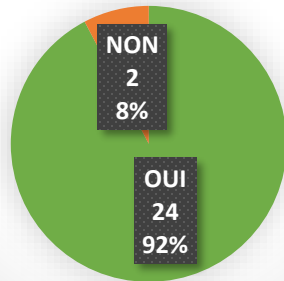


* Autre = « Partiellement financé par l'IR »

- Les projets sont quasi exclusivement liés à un projet de recherche à l'échelle :
 - Local : projets financés par ZA, LabEx...
 - Régional ou inter-régional : projets financés par les Régions, SOAHAL - Système d'observation de la rivière Allier et de ses annexes hydrauliques, Observatoire hydro-sédimentaire du Louroux, INTERREG...
 - National : ANR, programme du PIREN-Seine...
 - International : projets H2020 (N4C, HyMeX) ou le projet Global Precipitation Mission, projet IPEV...

QUELS LIENS AVEC VOS COLLABORATEURS ?

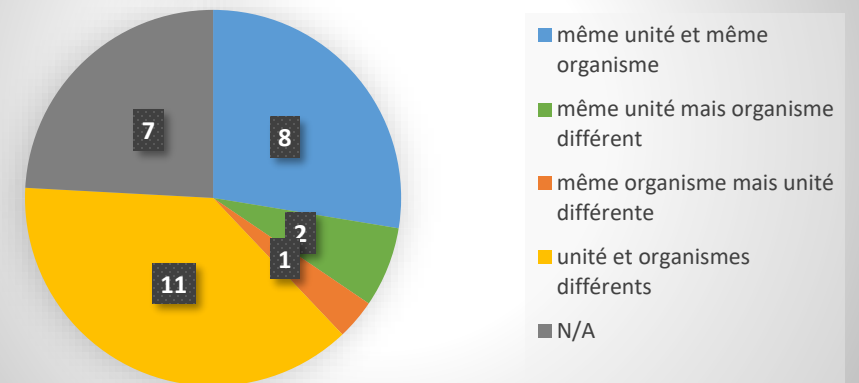
Y a-t-il d'autres personnes associées à la production de ce jeu de données ?



Ont-elles le même statut que vous ?

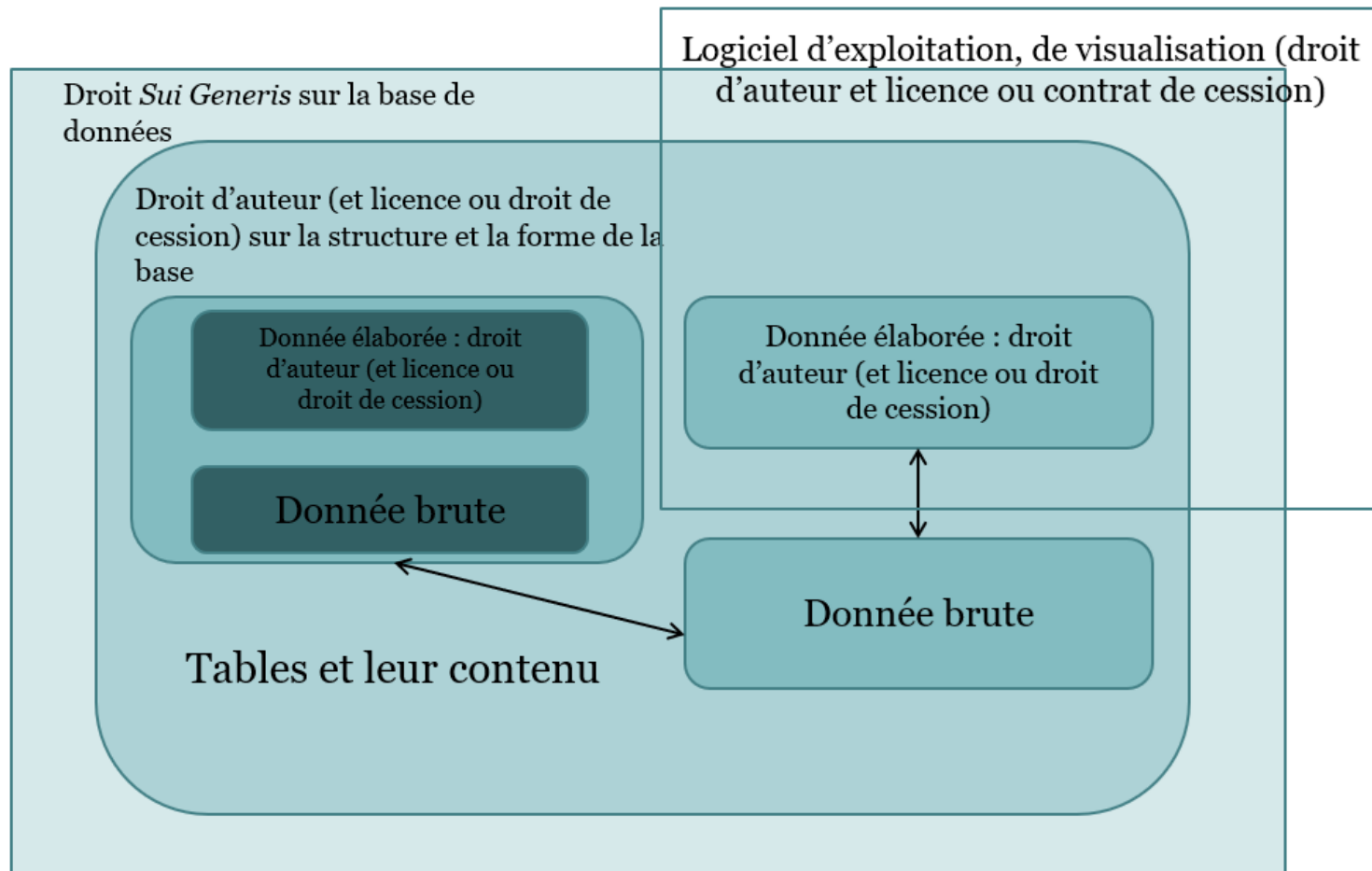


Font-elles partie de la même unité et du même organisme que vous ?



- La collaboration avec divers personnels et organismes représente une majorité dans les projets cités.
- Ces différences sont à prendre en considération lors de la mise en place d'un projet car elles ont des conséquences sur le plan juridique.

Focus Propriété des données : Le droit applicable aux données



- Veiller aux droits d'auteur de chacun
- Encadrer les droits de diffusion avec des licences
- Etablir un document, une charte en cas de collaboration
- Ne pas hésiter à faire appel au service juridique du CNRS ou de votre établissement pour éclaircir vos doutes

Focus Propriété des données : **Le droit d'auteur**

Exception du droit d'auteur pour les ITA : restreint !

Cette restriction ne remet pas en cause le droit de citation.

PROPRIÉTÉ INTELLECTUELLE

PROPRIÉTÉ INDUSTRIELLE

BREVETS

MARQUES –SIGNES
DISTINCTIFS

DESSINS ET MODELES

PROPRIÉTÉ LITTÉRAIRE ET
ARTISTIQUE

LOGICIELS

BASES DE
DONNEES

DROIT
D'AUTEUR
et
DROITS
VOISINS

Droit d'auteur sur son œuvre

Droits
patrimoniaux

Droit de représentation
Droit de reproduction
Durée : vie de l'auteur +
70 ans après sa mort
**Droits cessibles à un
tiers**

Droits extra-
patrimoniaux
OU « droit moral »

Droit de divulgation
Droit à la paternité de l'œuvre
Droit au respect de l'œuvre
Droit de retrait et de repentir
**Droit imprescriptible, inaliénable,
perpétuel**

Sources : DAUMAS-LADOUCE, M.-F. et UZAN, J., Propriété intellectuelle et valorisation

VOS JEUX DE DONNÉES

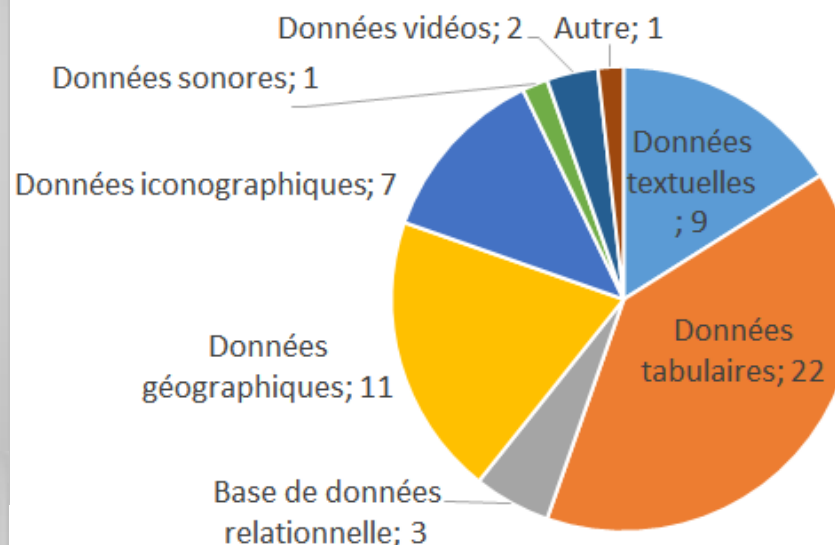
Quelle est la dimension de votre jeu de données ?



Les données que vous souhaitez partager sont :



Quel est le format de vos données ?



La taille de vos fichiers

- Une grande majorité entre quelques Ko et quelques centaines de Mo (18 réponses)
- Quelques personnes de l'ordre du Go (4 réponses)
- Une personne parle de To

Les extensions :

- Données textuelles : .txt et .csv (4), .pdf, et .scorm
- Données tabulaires : .xls(x) (7), .csv (4), .txt (2), .odt
- Bases de données relationnelles : PostgreSQL, .accdb
- Données géographiques : .shp (7), vecteur
- Données iconographiques : .jpeg, .jpg, .png, .geotiff, .bmp
- Données vidéo : .mp4

VOS JEUX DE DONNÉES : exemples

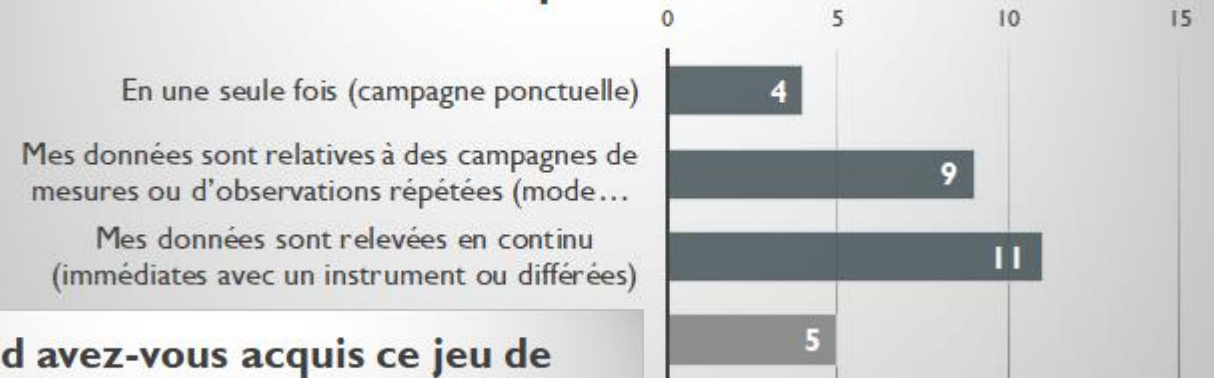
- Chroniques de données de **débit, précipitations, température, concentrations chimiques, turbidité**, etc.... (Observatoire de l'Eau et de l'Érosion aux Antilles, format CSV ou XLSX)
- Données **pluie-débit** sur un petit bassin versant urbain (chroniques temporelles)
- Jeu de données **hydrométéorologiques** ou variés issu d'un projet de recherche sur les basses eaux (HydroPop) combinant des données **d'enquêtes**, des données hydrométéorologiques variées...
- Jeux de données géographiques et non géographiques en **géophysique** et en **SHS** récoltés sur un site (PIREN et ZA Seine)
- Jeu de données issues du projet ECONECT : **séries temporelles** issues de systèmes sentinelles de l'environnement (**capteurs physico-chimiques** en milieu aquatique, données de **ruches connectées...**)
- Jeu de données issu d'expérimentation en **chambre climatique** : données de différentes instrumentations sur des projets de courte durée (2 ans max.)
- Jeu de donnée associant des champs de saturation (**teneur en eau**) dans le sol (BV Strengbach, OHGE)
- Données **géochimiques**
- Analyse de **radioactivité** dans des sédiments de sources

INSTRUMENTATION ?

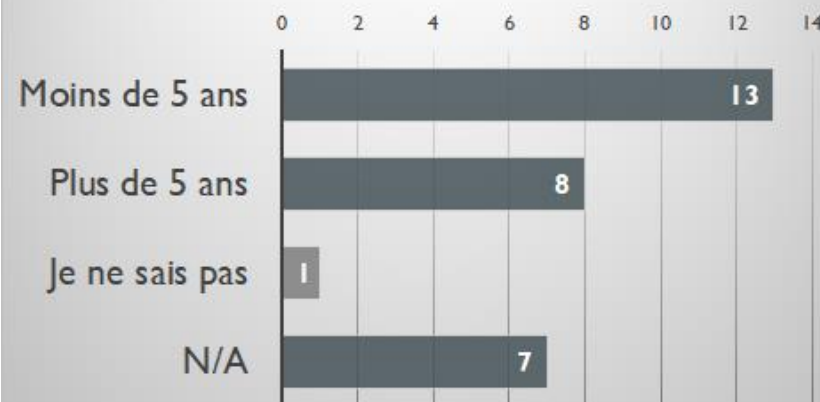
Quels instruments ?

- ▶ Des sondes et des capteurs en grande majorité (température, météo, pression, pluviomètre, piézomètre...)
- ▶ Mais aussi un appareil photo numérique avec prise de vues automatique, un drone, des appareils d'analyses biologiques (spectromètre, cytomètre)

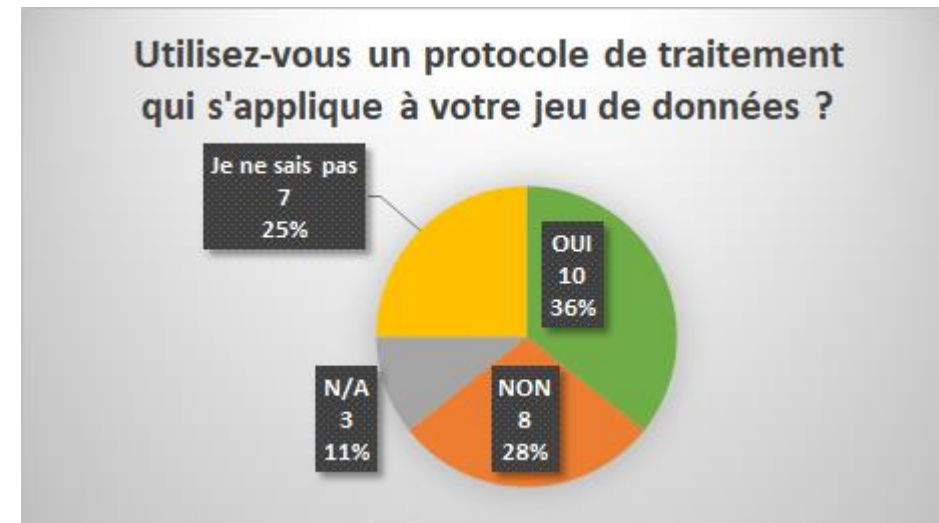
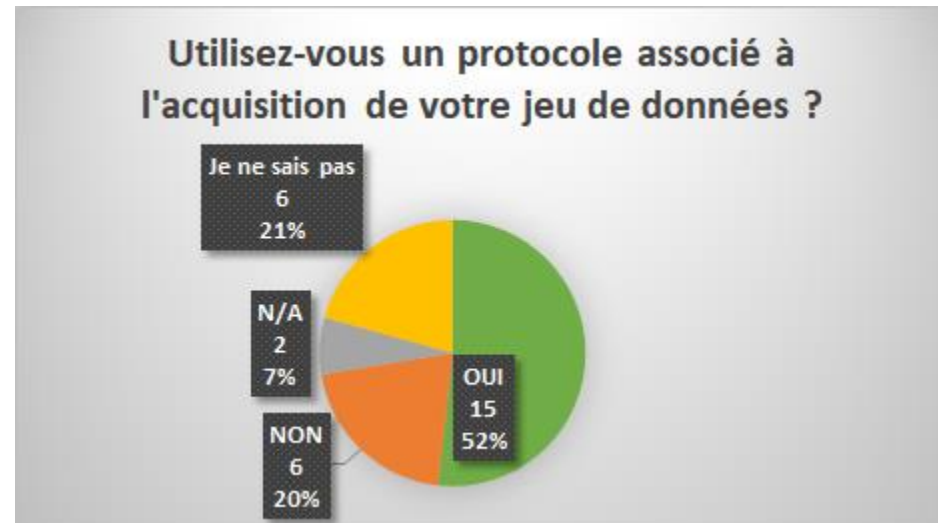
A quelle fréquence ces données ont-elles été acquises ?



Quand avez-vous acquis ce jeu de données ?

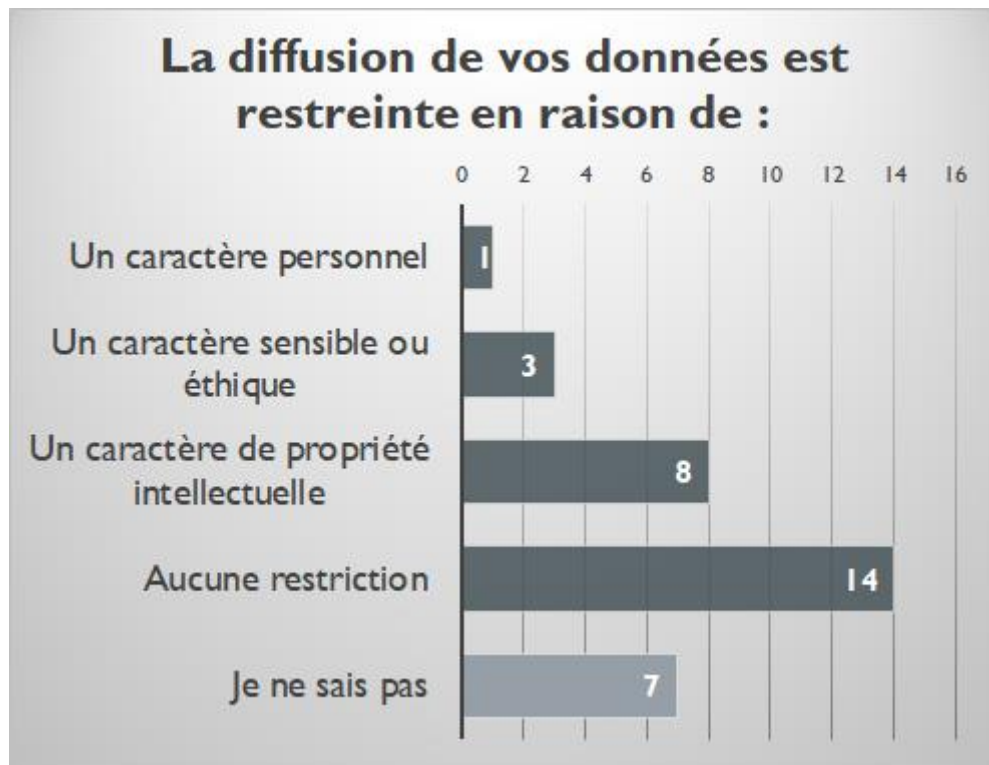


PROTOCOLES ?



- Utilisation globale d'un protocole de collecte et d'un protocole de traitement... mais aussi :
 - absence de protocoles (rédigés ?) pour environ 1/5^{ème} des répondants !
 - environ 1/5^{ème} ne savent pas !!!!!
- Nouvelles questions :
 - sous quelle forme existent ces protocoles (document qualité, script...) ?
 - sont-ils diffusables ?

QUELLES RESTRICTIONS À LA DIFFUSION ?



Pouvez-vous nous en dire plus sur ces restrictions ?

Focus RGPD

Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE



PROTECTION DES DONNÉES PERSONNELLES

Qu'est-ce qu'une donnée personnelle ?

Toute information relative à un particulier identifié ou identifiable, directement ou indirectement, grâce à un identifiant ou à un ou plusieurs éléments propres à son identité

Par exemple :



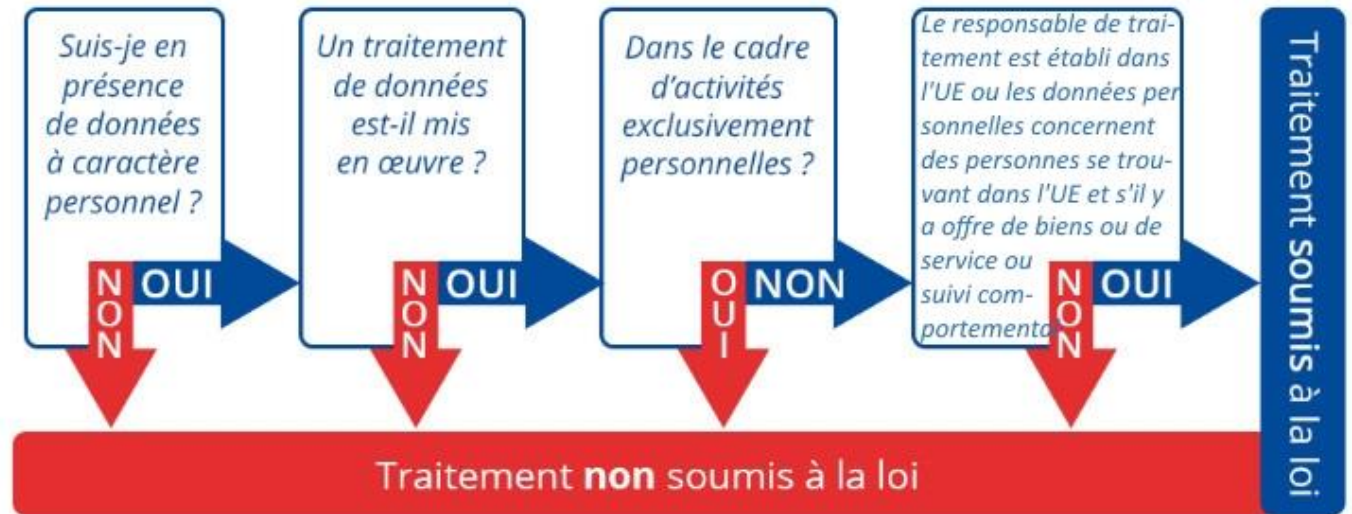
Source : Règlement général sur la protection des données (RGPD) du 27 avril 2016

vie-publique.fr | Paris 2019

<https://www.vie-publique.fr/eclairage/19588-rgpd-reglement-general-sur-la-protection-des-donnees-de-quoi-sagit-il>



Objectif = donner aux citoyens de l'union européenne davantage de contrôle et de visibilité sur leur vie privée



* que le traitement ait lieu ou non dans l'UE

Sources : CNIL et Intranet CNRS, cellule Protection des Données

Focus RGPD

Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE

Qui est responsable du traitement des données personnelles à l'échelle d'une unité de recherche ? → le DU !
(possibles co-responsables si plusieurs unités)

Qui doit s'assurer de la conformité du traitement ? → le DU !

Quel est le délégué à la protection des données vers qui on doit s'adresser ? → c'est le délégué de l'employeur du DU : il existe une cellule nationale déléguée à la protection des données (DPD)

Au CNRS : contacter la cellule DPD du CNRS en fournissant :

- le formulaire de déclaration et d'inscription au registre tenu par la cellule DPD du CNRS (téléchargeable depuis le site intranet CNRS de la DPD)
- la présentation du projet
- les éventuels questionnaires et guides d'entretien
- les éventuels avis du comité de protection des personnes ou du comité d'éthique

→ En retour la cellule DPD-CNRS fournit un certificat attestant de l'inscription de cette recherche dans le registre du labo concerné qu'elle met à jour pour chaque labo.

Source : [article Journal CNRS mai 2020](#)

Quelques infos / astuces...

- Se poser la question : pourquoi on collecte cette donnée ? Est-elle réellement nécessaire ou est-ce une habitude de collecte ?
- Enlever le plus d'infos possibles pour qu'on ne puisse pas identifier la personne (ex : date naissance, géolocalisation)
- Fixer/limiter la durée de conservation des données d'identification
- Anonymiser un maximum !

VOS OBJECTIFS DURANT L'ÉCOLE THÉMATIQUE

Le classement de vos objectifs concernant votre jeu de données

- 1. Rendre mon jeu de données le plus interopérable possible (standards de données...)**
- 2. Comprendre et commencer à écrire un Plan de Gestion de données**
- 3. Savoir décrire mon jeu de données pour qu'il soit trouvable**
- 4. Savoir diffuser mon jeu de données dans un entrepôt adapté : DOI**
- 5. Savoir valoriser mon jeu de données grâce à un data paper**

BILAN - DISCUSSION

Une grande hétérogénéité dans les disciplines et les thématiques de recherche, mais beaucoup de points communs :

- Données en rapport avec une I.R.
- Une diversité d'acteurs et d'organismes au sein d'un même projet
- Beaucoup de données curées et harmonisées
- De divers types et sous différents formats
- Utilisant souvent des instrumentations

Mais vous êtes tous animés par une volonté de produire des données plus FAIR, de mieux connaître les Plans de Gestion de Données et de rendre les données plus interopérables.

LIENS UTILES

Aspects juridiques

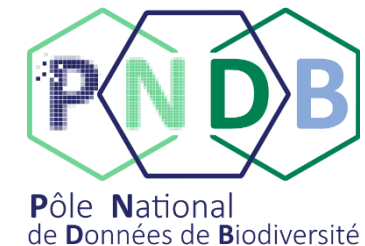
- A qui appartiennent les données (L. Maurel) : <https://mate-shs.cnrs.fr/actions/tutomate/tuto25-propriete-donnees-lionel-maurel/>
- A qui appartiennent les données de Recherche ? (L. Maurel) : <https://mate-shs.cnrs.fr/wp-content/uploads/2020/09/tuto25-mate-Donn%C3%A9es-de-recherche.pdf>
- Logigramme sur l'appartenance des données (Institut Pasteur) : <https://www.pasteur.fr/fr/file/20707/download>
- Résultats de recherche et open data : le cadre juridique (N. Gandon) : https://fredoc2018.sciencesconf.org/data/pages/Open_data_cadre_juridique_N_Gandon.pdf
- La diffusion numérique des données en SHS - Guide des bonnes pratiques éthiques et juridiques (N. Ginouvès) : <https://hal-amu.archives-ouvertes.fr/hal-01903040>
- Questions juridiques autour de l'ouverture des données (N. Le Ba) : http://rdbd.cnrs.fr/IMG/pdf/20181107_questions_juridiques_autour_de_l_ouverture_des_donnees_rdbd.pdf?515/adba03a95aa28dd17a08e0c052414e099ad26347

REMERCIEMENT

cnrs

- CNRS soutien financier
- R. OMNES (conseillère formation, DR10) suivi de la formation
- P. RUPP (ICUBE), C. DIDIERJEAN (EDYTEM) gestion administrative et financière
- Le comité organisateur : BED/RZA, OZCAR/THEIA, UMS BBEES, PNDB, avec la participation de la MISHA
- Centre Saint-Thomas, hébergement

- Retrouvez les informations sur : <https://e-envir-21.sciencesconf.org/>



Maison Interuniversitaire des
Sciences de l'Homme
- Alsace -

