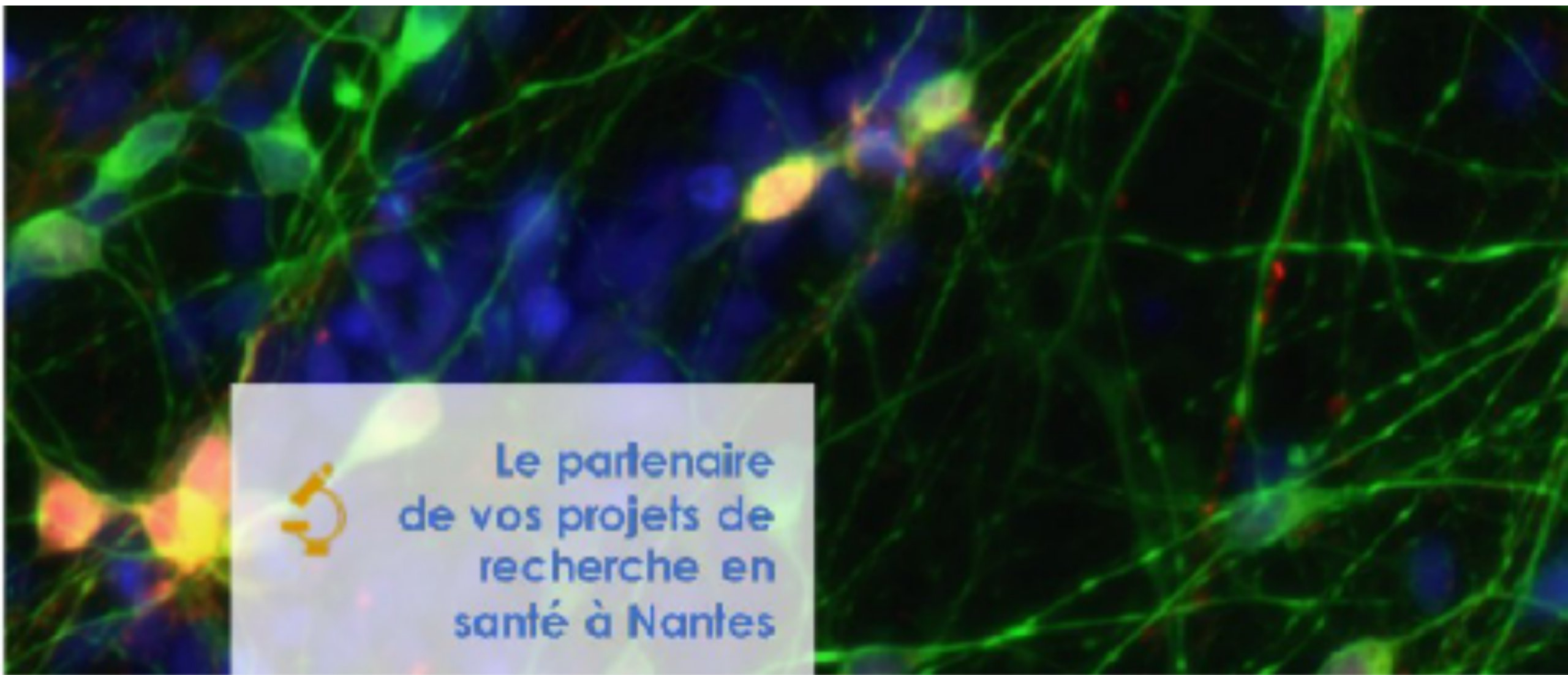


Gestion et stockage de vos données de recherche : quels enjeux et quels moyens ?



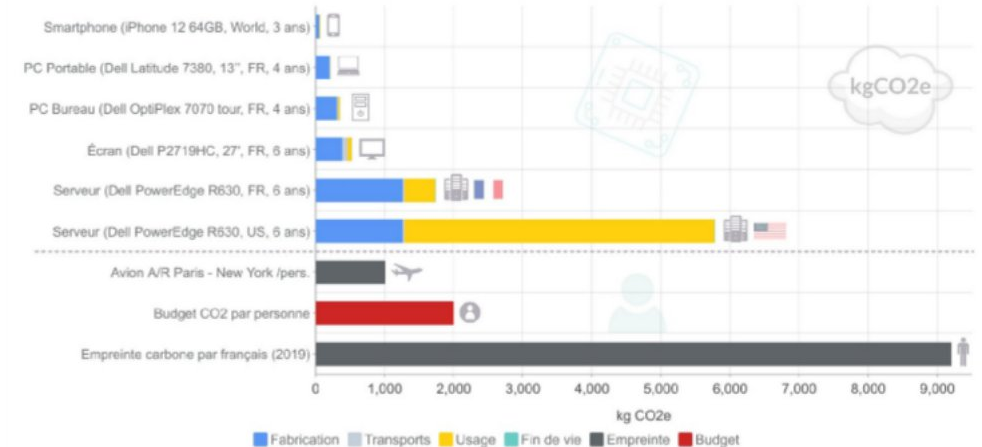
Actions passées et en cours

- Séminaire du 10 Novembre 2021 :
 - [« Le paysage des infrastructures de calcul et de stockage au sein des dynamiques régionales et nationales qui se mettent en place »](#)
 - *Yann Capdeville (LPG), Audrey Bihouée (plateforme BiRD), Pierre-Antoine Gourraud (CHU Nantes) et Richard Redon (cluster SysMics)*
- Séminaire SFR, DELPHI et SysMics du 15 Mars 2023 :
 - [Infratructures numériques pour l'imagerie et l'IA pour la recherche en Santé-Biologie à Nantes](#)
 - *Pierre-Antoine Gourraud, Geoffrey Desvaux, Audrey Bihouée, Alban Gaignard, Perrine Paul-Gilloteaux, Heidi Derrien, Yann Capdeville, Delphine Loussouarn, Raphaël Bourgade, Charles Lepine*
- Séminaire ITX du 17 Novembre 2023
 - [Introduction aux impacts environnementaux du numérique](#)
 - *David Benaben, Ingénieur informaticien. INRAE, Univ. Bordeaux, UMR BFP, Villenave d'Ornon.*
- Page Intranet SFR Bonamy « [Gestion des données numériques](#) »
 - dans chapitre « Science Ouverte », rubrique « Recherche Responsable ».
- Digital Clean Up Day 16 mars 2024
- Séminaire SFR 23 janvier 2024 [Gestion et stockage de vos données de recherche](#)

Enjeux

- La donnée numérique est au cœur de nos recherches.
- Un volume à potentielle croissance infinie qui ne peut pas profiter de ressources infinies : apprendre à supprimer
- Une recherche éco-responsable :
2020 : entre 1.8% et 2.8% produit par le numérique (vs 1.9 pour aviation civile)
[GREENER principles for environmentally sustainable computational science](https://doi.org/10.1038/s43588-023-00461-y)
<https://doi.org/10.1038/s43588-023-00461-y>
Mutualiser pour réduire l'empreinte
Réutiliser plutôt que calculer
- Un enjeu de Science Ouverte (partage des données pour réutilisation) : stocker et partager oui mais de manière pertinente.

Ordre de grandeur émission GES (kg)



Séminaire Institut du Thorax, David Benaben, 17/11/23



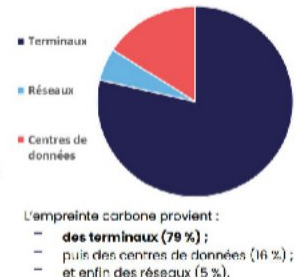
Impacts environnementaux du numérique en France

- L'empreinte carbone du numérique en France : **17 Mt CO2 eq. soit 2,5 % de l'empreinte nationale**

- **10 % de la consommation électrique française** soit 48,7 TWh par an

407 kg de CO2 : c'est l'empreinte GES du numérique d'un-e français-e. C'est déjà 25% de son objectif de bilan carbone 2050

Mais l'empreinte d'ici 2050 en France pourrait tripler si aucune action.



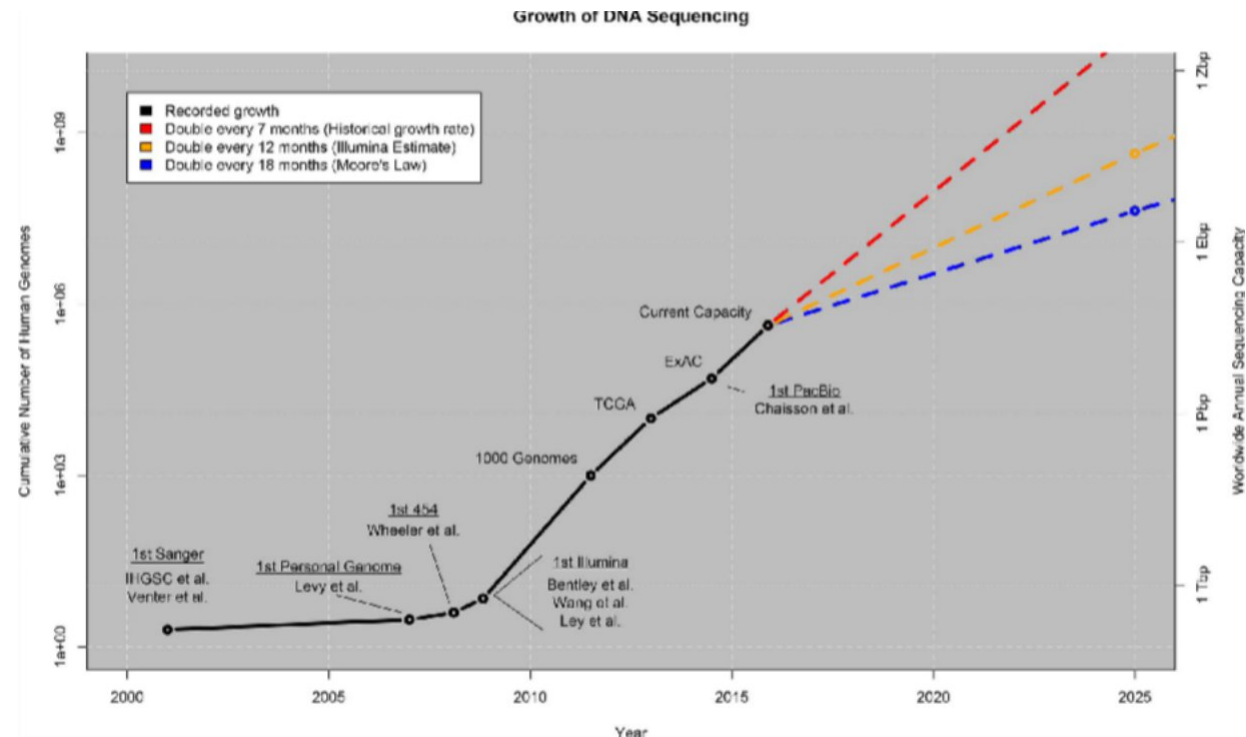
ANF EcoInfo ,ADEME, Julie Mayer, 21/11/23

Problématique en 5 V :

- Volume de données à traiter conséquent et croissant
 - Variété d'informations importante (dans sa nature et sa structure)
 - Vitesse dans la fréquence d'acquisition et de mise à jour de cette information
 - Valeur de la donnée
 - Véracité de l'information
- + la sensibilité de l'information

La production de données en Sciences de la Vie

- Un séquenceur de type NovaSeq peut générer 6 TB en 2 jours
 - Un microscope automatisé ou un feuille de lumière peut générer 1TB par jour
 - Une analyse classique multiplie par 1.5 X la taille des données d'entrée en fichier temporaire.
-
- Le rythme de croisière estimé à la SFR, d'après une enquête :
 - en 2021 - 300 TB/an
 - en 2026 - 600 TB/an

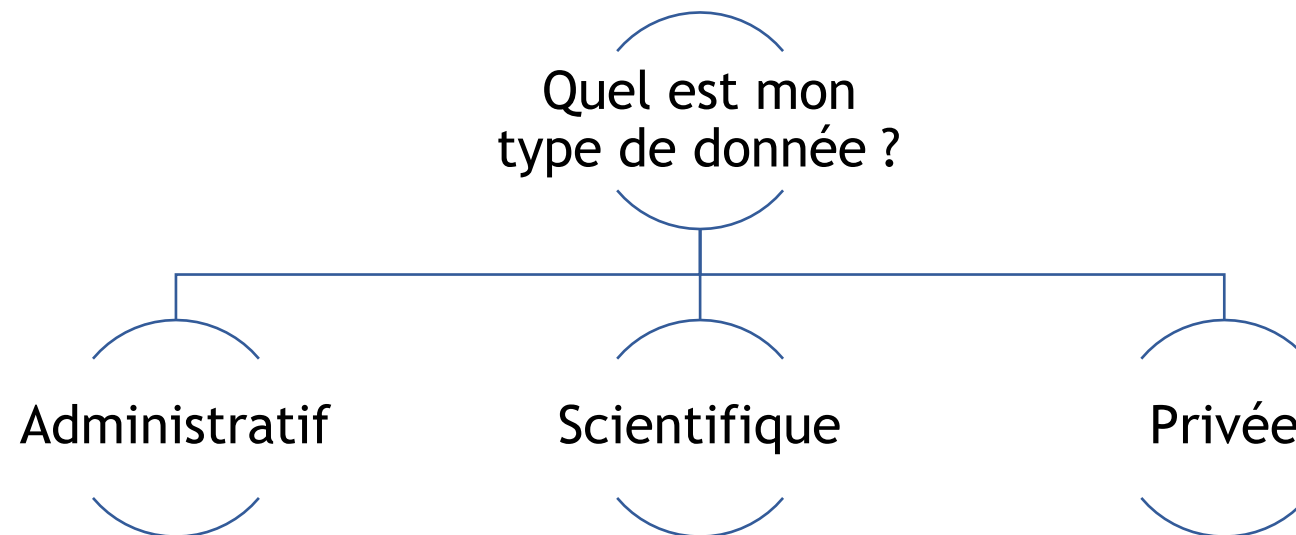
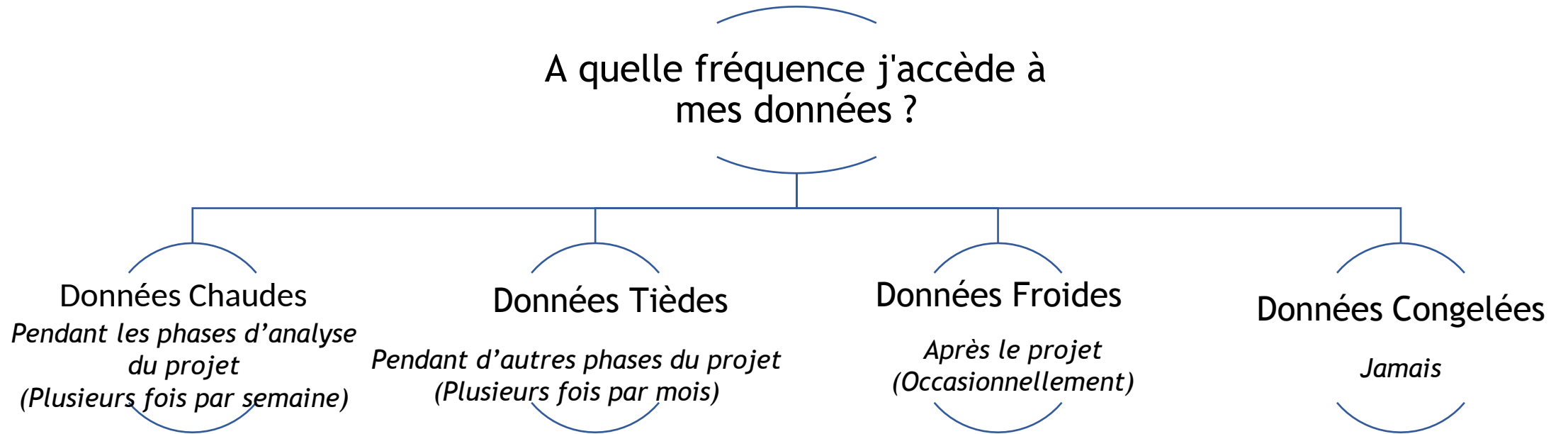


Zachary D.S. *et al.*, Big Data: Astronomical or Genomical?, 2015

Choix du stockage

- Nécessaire d'identifier la **nature** des données et l'**usage** qui en sera fait pour effectuer un choix adapté d'espace de stockage.
- Le choix du stockage peut en effet varier en fonction de plusieurs critères :
 - l'étape du cycle de vie du projet de recherche,
 - le niveau de sécurisation,
 - le protocole d'accès aux données,
 - la performance,
 - le lieu de stockage et la connectivité au réseau du site de production des données,
 - la volumétrie,
 - le niveau de sensibilité des données (cf. fiche données sensibles),
 - la nature structurée ou non des données,
 - les collaborations internes/externes souhaitées aux différentes étapes (collecte, traitement/analyse, publication),
 - etc.

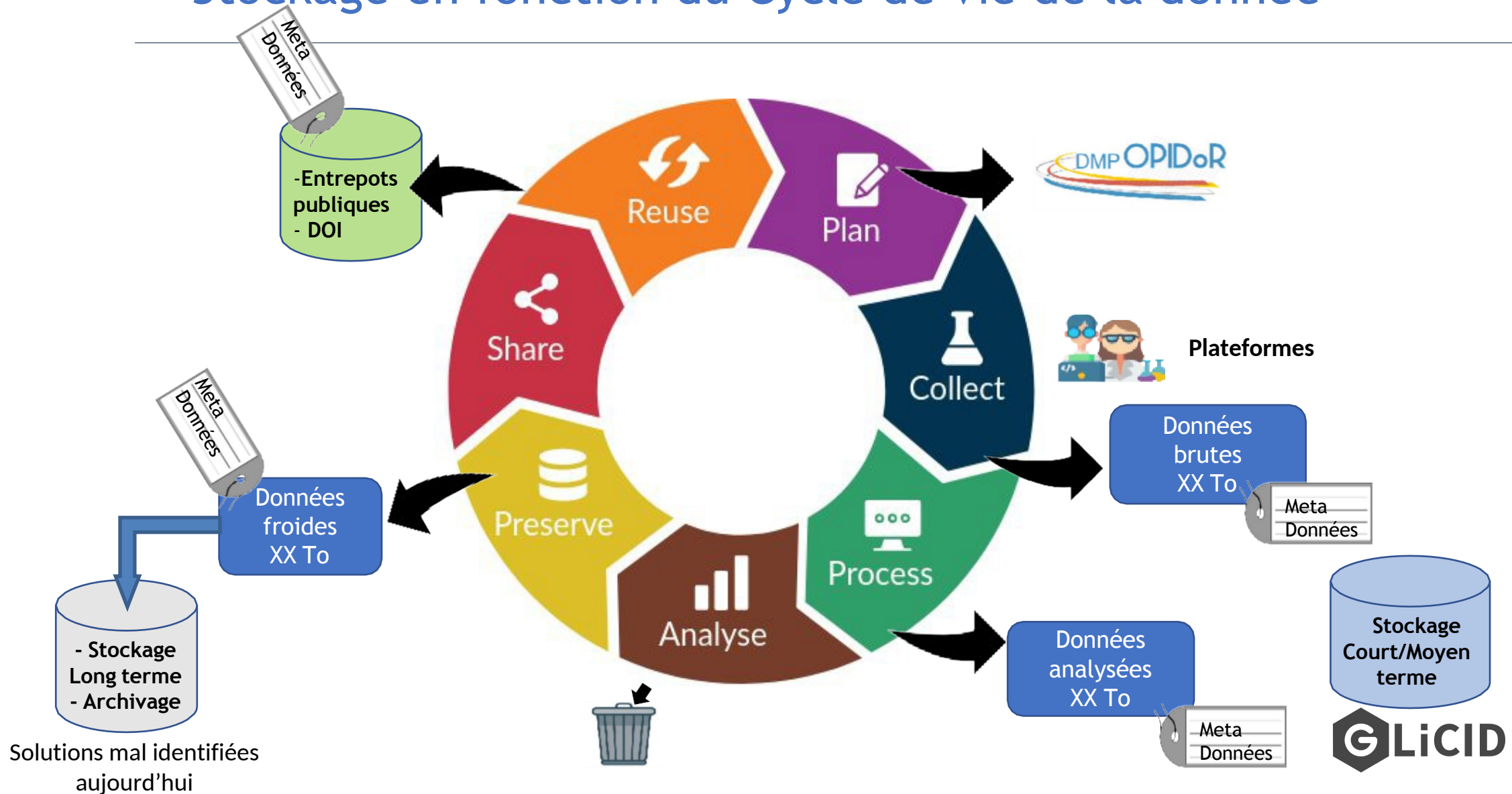
Typologie de mes données



Stockage en fonction de la typologie de mes données

Typologie	Chaudes 	Tièdes 	Froides 	Congelées
Bureautique / administrative/ Articles <i>(doc, xls, ppt, pdf...)</i>	– SNPS : Cronos – UNCloud – INSERM		Solutions d'archivage locales(?) CINES Hal	Détruites
Scientifique <i>(Fichiers issus d'appareil de mesures scientifiques et de leur analyse)</i>	– <i>Ne nécessitant pas de calcul HPC :</i> -> NU/INSERM/SIEN – <i>Nécessitant du calcul HPC</i> -> GLiCID		Déposées sur un entrepôt adapté et/ou publiées	Détruites
Privée <i>(Photos de vacances, musique, etc...)</i>	Domicile Cloud privé		Domicile Cloud privé	Détruites

Stockage en fonction du Cycle de vie de la donnée



Le stockage est nécessaire tout au long du cycle de vie de la donnée, mais les modalités seront différentes

Programme (1/2)

- | | |
|-------------|--|
| 9h15-9h30 | L'atelier de la donnée OREADE : pour se faire accompagner à Nantes Université (PGD, partage de données, formation, etc.)
Pierre François, Service Données de la Recherche, Bibliothèques Universitaires, Nantes Université |
| 9h30-9h45 | Centre de référence Thématique Institut Français de Bioinformatique
Jean François Dufayard, CIRAD et IFB, Montpellier |
| 9h45-10h00 | Numérique responsable
Melisande Febvre, DSIN, Nantes Université |
| 10h00-10h15 | Enjeux liés aux données de recherche en santé
Richard Redon, l'institut du Thorax, SysMics, IFB |
| 10h15-11h00 | Echanges et Discussions - Pause café |

Programme (2/2)

- | | |
|-------------|--|
| 11h00-11h15 | La gestion de vos données de recherche ne nécessitant pas de calcul intensif
Patrice Gonnord, Service Numérique du Pôle Santé (SNPS NU),
Arnaud Abélard, Direction des Systèmes d'Information et du Numérique (DSIN NU, SIEN) |
| 11h15-11h30 | La gestion de vos données de recherche nécessitant du calcul intensif
Audrey Bihouée, Mésocentre régional GLiCID/ Plateforme BiRD |
| 11h30-11h45 | Solutions proposées à l'INSERM
Stéphane Cesbron, DSI / DR Grand Ouest Inserm |
| 11h45-12h00 | Collecte et conservation patrimoniale des données de la recherche
Tulia Klavzer, Département Archives et Patrimoines, Bibliothèques Universitaires,
Nantes Université |
| 12h00-13h00 | Echanges et discussions autour de cas concrets |

Discussion - Table ronde

Exemple d'une demande reçue le 23/09/2025

Bonjour Audrey,

J'ai discuté avec *** sur la possibilité de disposer d'un lieu de stockage de données du laboratoire. Il m'a conseillé de me rapprocher de toi pour en discuter.

Nous avons un certain nombre de données numériques que nous devons stocker **un peu longtemps** (images TEP, IRM, images anapath numérisées, image de coupes radioactives, etc.

Comment peut-on procéder pour disposer d'un lieu de stockage au sein de Nantes université?

Merci par avance,
Bonne journée,

Discussion - Table ronde

Exemple d'une demande reçue le 23/09/2025

Nous avons un certain nombre de données numériques que nous devons stocker **un peu longtemps** (images TEP, IRM, images anapath numérisées, image de coupes radioactives, etc.).

Comment peut-on procéder pour disposer d'un **lieu de stockage** au sein de Nantes université?

Quelle échelle
de temps
En mois ?
En Années ?

Quel volume ?

Quel usage ?
– Archivage ?
– Fréquence d'accès
– Traitement ?

Discussion - Table ronde

Exemple d'une demande reçue le 23/09/2025

Nous avons un certain nombre de données numériques que nous devons stocker **un peu longtemps** (images TEP, IRM, images anapath numérisées, image de coupes radioactives, etc.).

Comment peut-on procéder pour disposer d'un **lieu de stockage** au sein de Nantes université?

Quelle échelle de temps
En mois ?
En Années ?

5 à 10 ans

Quel volume ?

5 à 10 To

Quel usage ?

- Archivage ?
- Fréquence d'accès
- Traitement ?

Une partie avec un accès non fréquent
Une partie avec un accès plus fréquent quand expérience en cours