

# Des données chaudes, tièdes, froides

Café Science Ouverte PAMIR

—

23 mai 2025

Serge Cohen, CNRS



# Un parallèle culinaire... ?

Les données, ça se consomme chaud, mais ça se conserve mieux au frais...

Ou bien, comme un matériau, la donnée se consolide au fur et à mesure qu'elle refroidit ?

# Au programme

- Le cycle de vie de la donnée
- Stockage vs. accès
- Des grilles de lecture qui se recouvrent partiellement

# Le cycle de vie de la donnée

| Temps              | T0                                                                   |                                         | T1                                                                                                                                                  | T2                                                     | T3         | T4                                                                                                                                                                               | T5                      |             |
|--------------------|----------------------------------------------------------------------|-----------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------|-------------|
| Actions            | Application métier :<br>- choix<br>- implémentation<br>- paramétrage | Planification de la gestion des données | Acquisition des données :<br>- création (manuelle/automatique (avec contrôles)<br>- captation (autre application/base de données interne/Open data) | Contrôle qualité "métier"<br>- manuel<br>- automatique | Validation | Open Data :<br>- analyse<br>- sélection<br>- corrections<br>- enrichissements<br>- conversion<br>- contrôle qualité<br>- publication (ODbL / Creative commons / Licence ouverte) | Tri archives            |             |
|                    |                                                                      |                                         | Modifications<br>Traitements<br>Corrections<br>Enrichissement                                                                                       |                                                        |            | Export vers d'autres applications                                                                                                                                                | Conservation définitive | Destruction |
| Cadre d'action     | Gouvernance de l'information                                         |                                         | Vie de la donnée (besoins métiers)                                                                                                                  |                                                        |            | Réutilisation                                                                                                                                                                    | Archives : sort final   |             |
| Acteurs principaux | Décideurs / Direction des systèmes d'information                     |                                         | Services producteurs                                                                                                                                |                                                        |            | Techniciens : Open Data, Archives...                                                                                                                                             |                         |             |



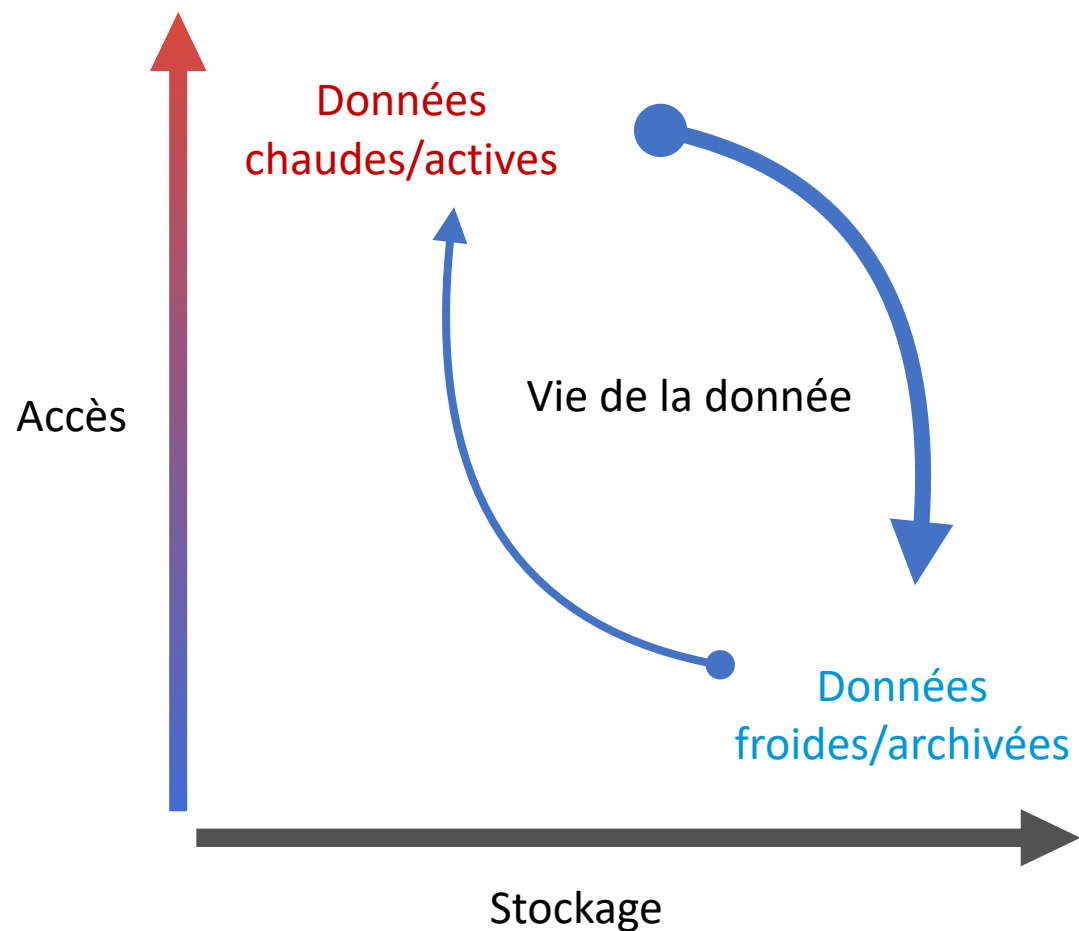
# Que serait la *température d'une donnée* ?

- Positionnement de la donnée dans son cycle de vie (est-elle plus proche de l'acquisition ou de l'archivage) ?
  - ➔ Concept permettant d'inscrire la donnée dans une séquence temporelle (séquence à *sens unique*)
  - ➔ Simplification qui limite certains aspects opérationnels/techniques

# La donnée : où la stocker, comment y accéder...

- Stockage et accès sont des aspects duaux de la donnée
  - ➔ Il ne peut pas y avoir d'accès sans stockage qui le rend possible
  - ➔ Un stockage sans possibilité d'accès est inutile
- Qui ont des caractérisations orthogonales...
  - ➔ Le stockage est caractérisé par sa capacité/son volume (Go/To/...)
  - ➔ L'accès est caractérisé à la fois par son débit (Go/s...) et sa latence (ms...) (on parle aussi de temps d'accès)

# Stockage vs. accès : un équilibre des coûts ?



Un parallèle avec une bibliothèque...

Livre accessible en salle de lecture

Livre en réserve, sur demande

Livre en grande réserve/  
prêt inter-bibliothèques

# La donnée : où la stocker, comment y accéder...

- Stockage et accès sont des aspects duals de la donnée
  - ➔ Il ne peut pas y avoir d'accès sans stockage qui le rend possible
  - ➔ Un stockage sans possibilité d'accès est inutile
- Qui ont des caractérisations orthogonales...
  - ➔ Le stockage est caractérisé par sa capacité/son volume (Go/To/...)
  - ➔ L'accès est caractérisé à la fois par son débit (Go/s...) et sa latence (ms...) (on parle aussi de temps d'accès)
- Avec des caractéristiques complémentaires :
  - ➔ Disponibilité (minimiser l'incertitude sur le temps d'accès à une donnée)
  - ➔ Sécurité (minimiser le risque de perdre définitivement une donnée)
  - ➔ Pérennité (conserver la possibilité **d'utiliser** la donnée dans le futur)

# Une combinaison de grilles de lecture

- Capacité : volume de données qui peuvent être stockées par le système
- Accès : vitesse/débit du transfert de la donnée du stockage vers le système l'exploitant (la redondance peut permettre la multiplication du débit par parallélisation)
- Accès : latence/temps d'accès, temps entre une requête et le début de sa réalisation
- Disponibilité : mise en place de redondance (tant du stockage que de l'accès) permettant de limiter le risque de blocage
- Sécurité : mise en place de redondance (principalement sur le stockage) permettant de limiter le risque de perte définitive
  - ➔ Cette redondance peut prendre la forme de *sauvegardes* (hors ligne)
- Durée de vie : archivage des données, mise en place de système de stockage à (très-)haute capacité mais à accès réduit; avec pour but de conserver la données sur le temps long
  - ➔ L'archivage vise à garder la donnée dans son *format normal* sur le temps long
  - ➔ L'archivage pérenne complète le dispositif en cherchant à assurer la capacité à exploiter la donnée dans le futur (cela implique également une activité de conservation des moyens d'accès et de traitement de la donnée)
- Catalogage/méta-donnée

# Questions ?