

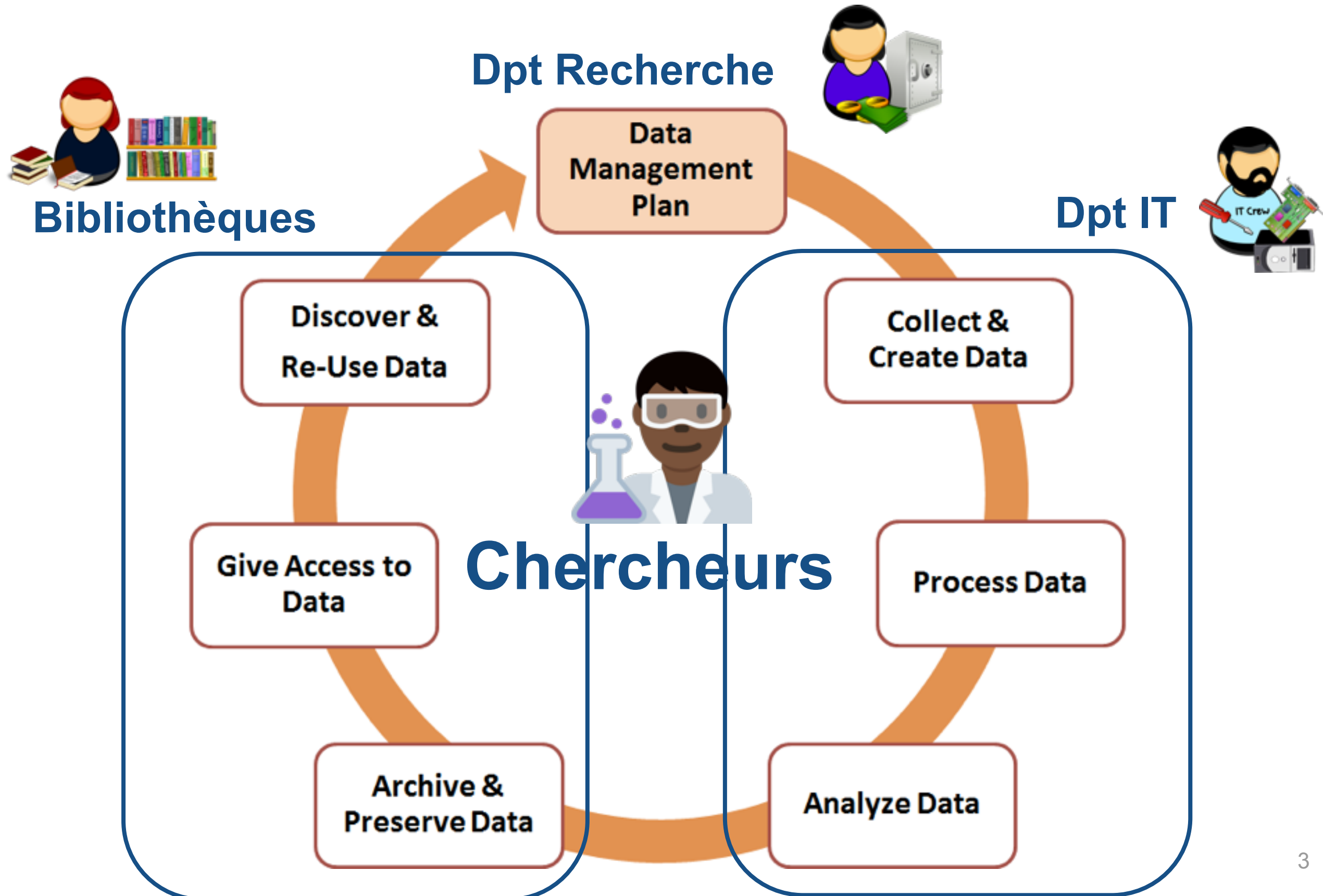
# Gestion des données de la recherche

ArchiRès - 24 janvier 2019

Anthony Leroy  
Systèmes & Innovation (DBIS-ULB)



# Le GT GDR initie le dialogue essentiel entre les principaux acteurs de la GDR





# Composition du groupe de travail GDR

## Chercheurs



Jean-Marie  
Vanderwinden  
(Santé)



Olivier  
Klein  
(SH)



Quentin  
Delhaye  
(EPB)



Thierry  
Massart  
(Sciences)

## Conseiller vie privée et sécurité



Olivier  
Markowitch

## Dépt Recherche



Rachel  
Leproult

## SISC (relais DI)



Raphaël  
Leplae

## Dépt des Bibliothèques et Info Scientifique



Benoît  
Pauwels



Françoise  
Vandooren



Anthony  
Leroy

# Objectifs du groupe de travail



faire état des attentes et des besoins concrets des chercheurs **(de l'ULB)**



analyser la littérature existante et les expériences d'autres institutions



formuler des lignes directrices et des propositions concrètes quant aux infrastructures et services à proposer aux chercheurs

# Le GT a scindé la problématique en 7 thèmes principaux

**Contraintes légales  
Exigences des bailleurs de fonds**

**Besoins des chercheurs  
Politique institutionnelle**

**Plan de gestion des données**

**Inventaire des données de  
recherche**

**Gestion quotidienne: collecte,  
stockage, traitement**

**Dépôt, accès et préservation**

**Support aux chercheurs**

# Deux sources d'information principales



**Sources d'inspiration**  
(littérature, collaboration)



**Enquête**  
(auprès des responsables d'UR)

# Deux sources d'information principales



**Sources d'inspiration**  
(littérature, collaboration)



**Enquête**  
(auprès des responsables d'UR)



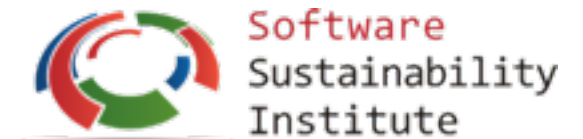
# Collaborations et sources d'inspiration



Stanford University



e-infrastructures  
austria



+130 publications analysées



# Deux sources d'information principales



**Sources d'inspiration**  
(littérature, collaboration)



**Enquête**  
(auprès des responsables d'UR)

# Enquête sur les pratiques des chercheurs en matière de GDR



lancée mardi 2 mai 2017



accueil très positif



nombreuses remarques proches du terrain



+50 % des répondants acceptent d'être recontactés

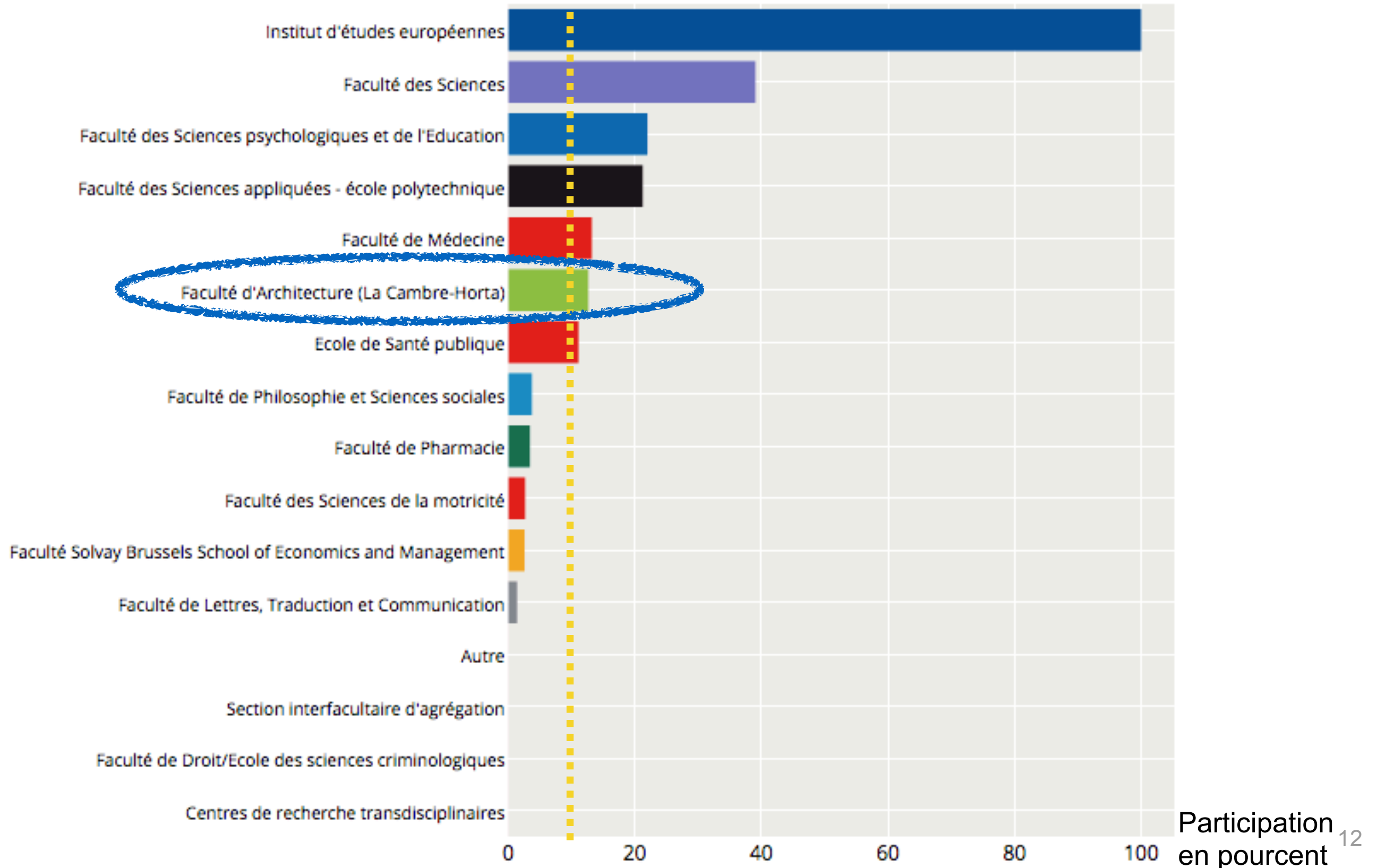


161 réponses (125 individuelles, 36 globales)



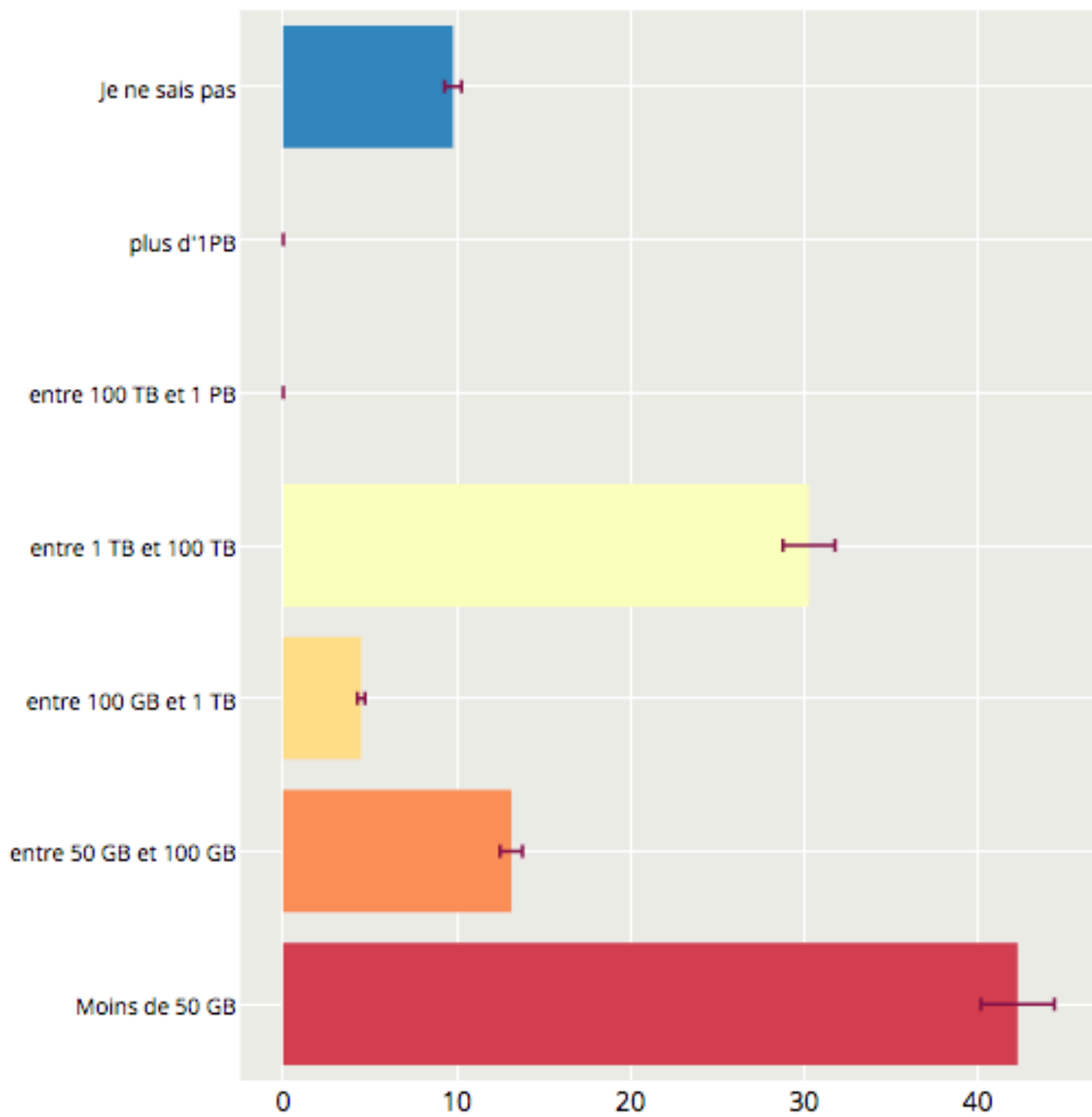
concerne les usages de **+620** chercheurs (+15%)

# La participation dépasse les 10% pour 7 facultés

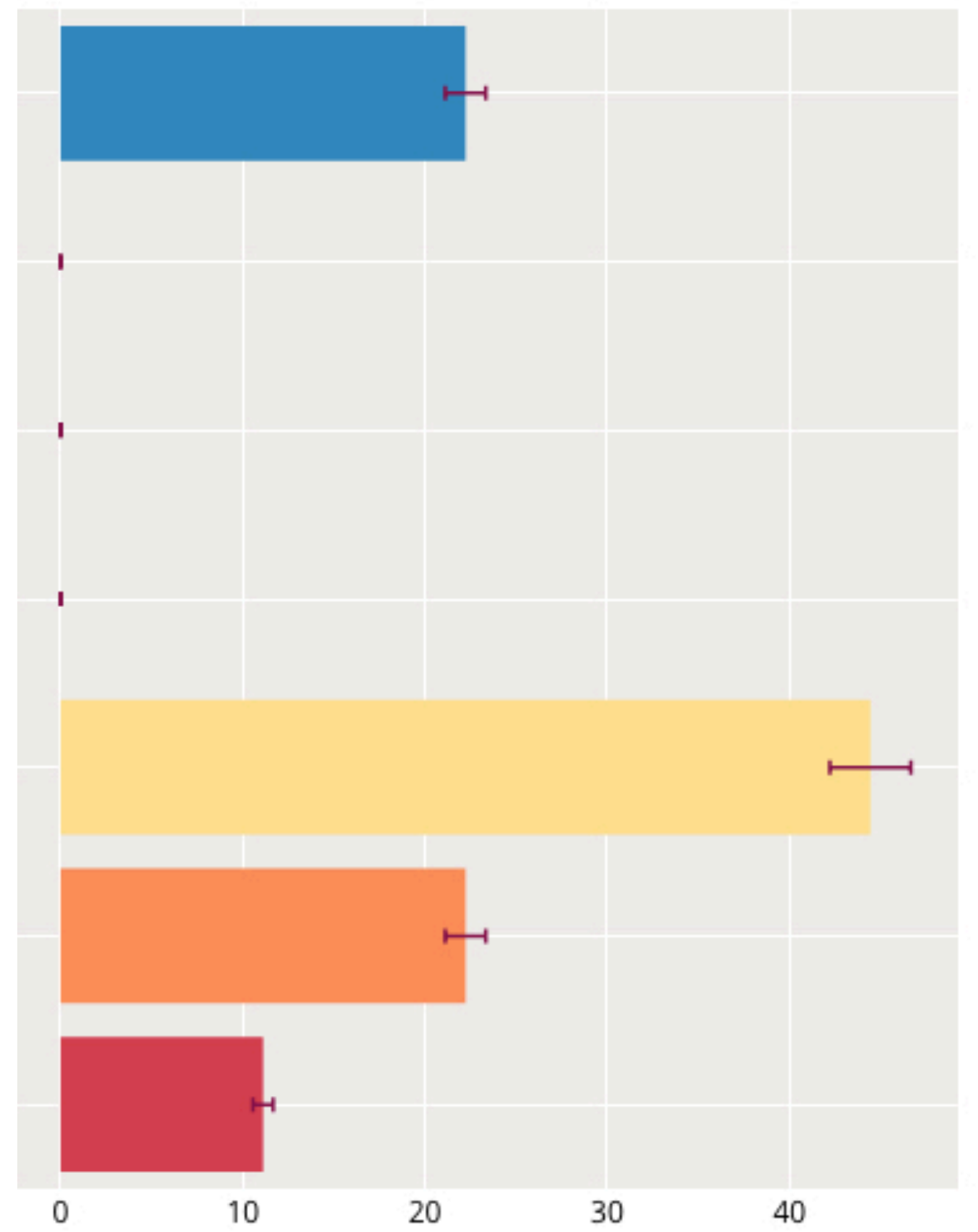


# Long tail or big data?

## A l'ULB



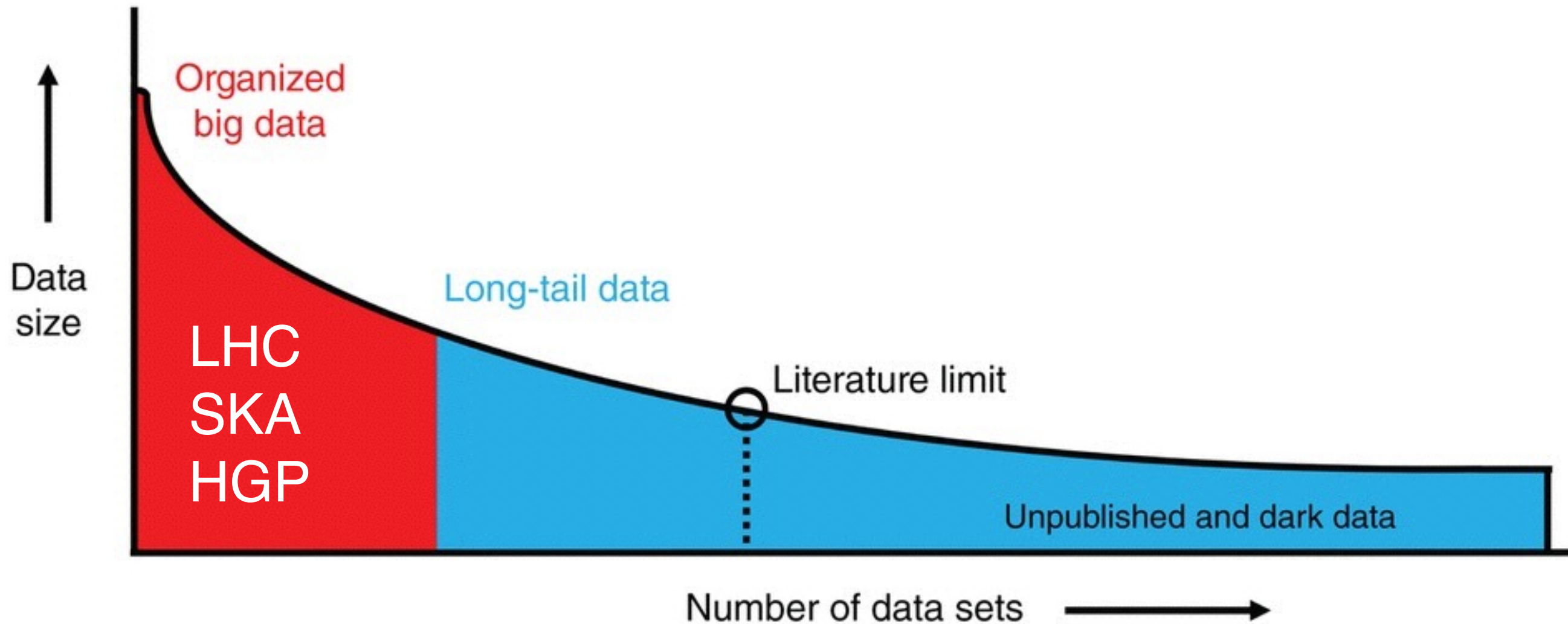
## En architecture



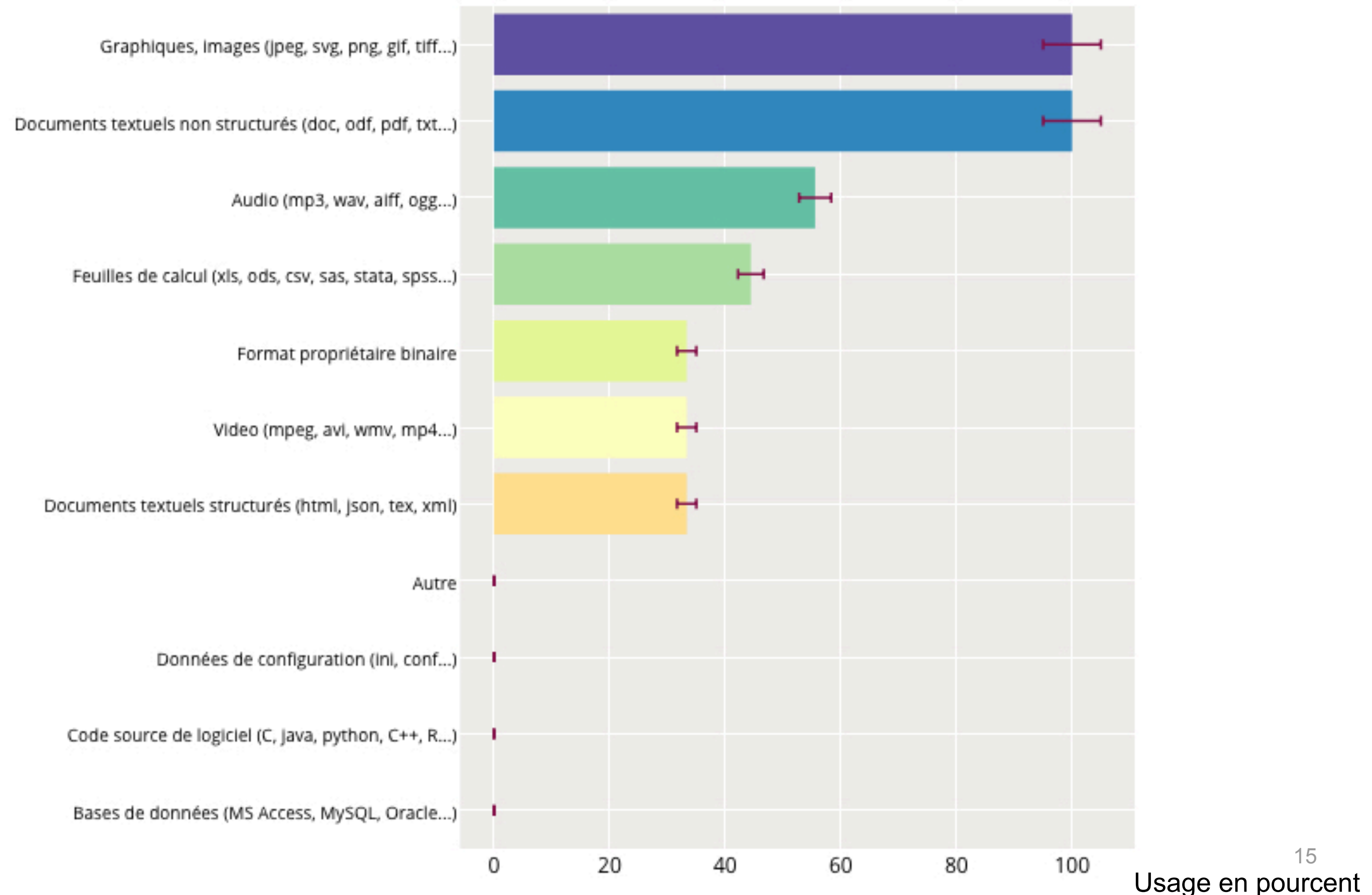
Usage en pourcent



Les projets “big data” disposent d’infrastructures spécifiques de gestion des données de recherche, mais pas le long tail

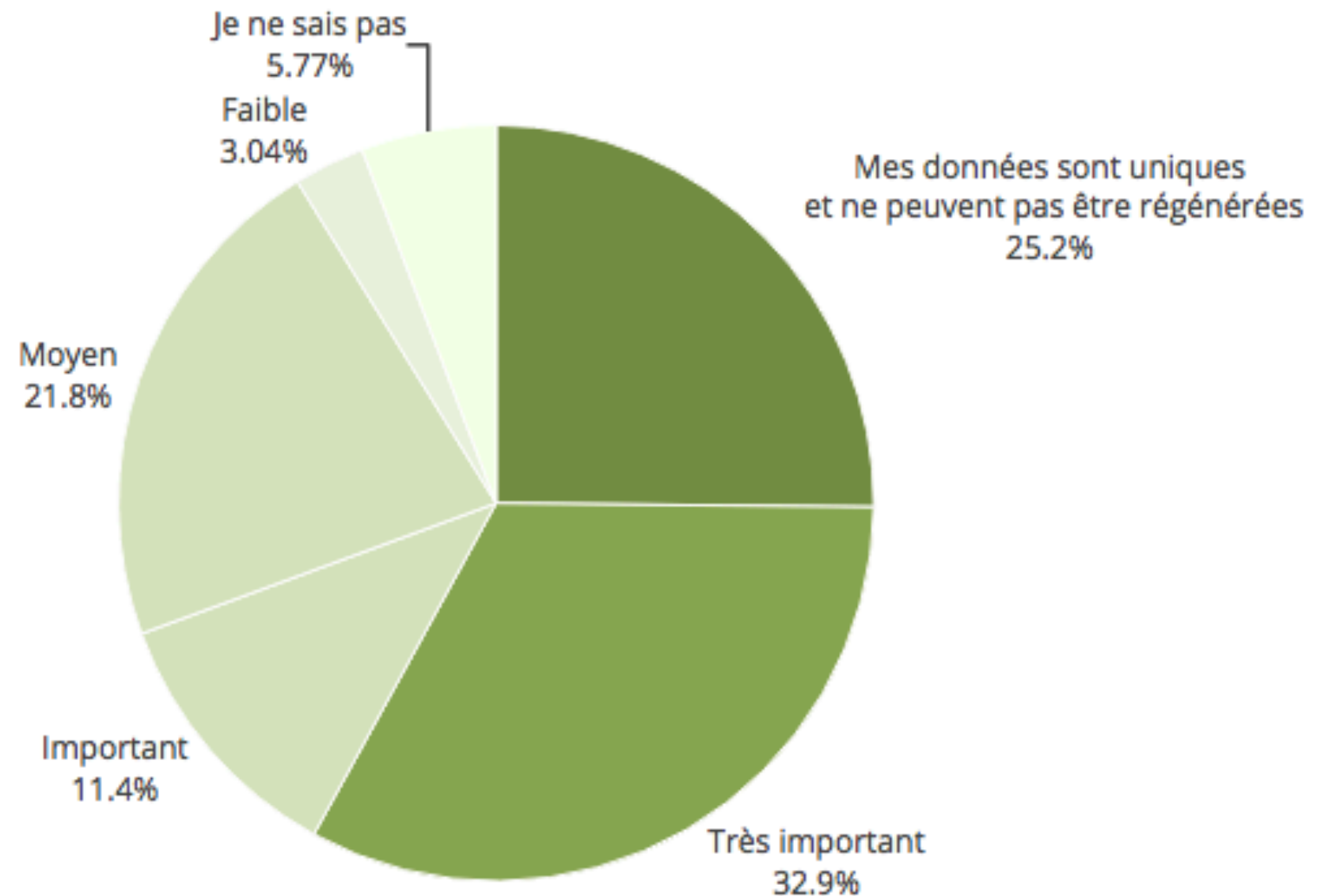


# Types de formats exploités en Archi.



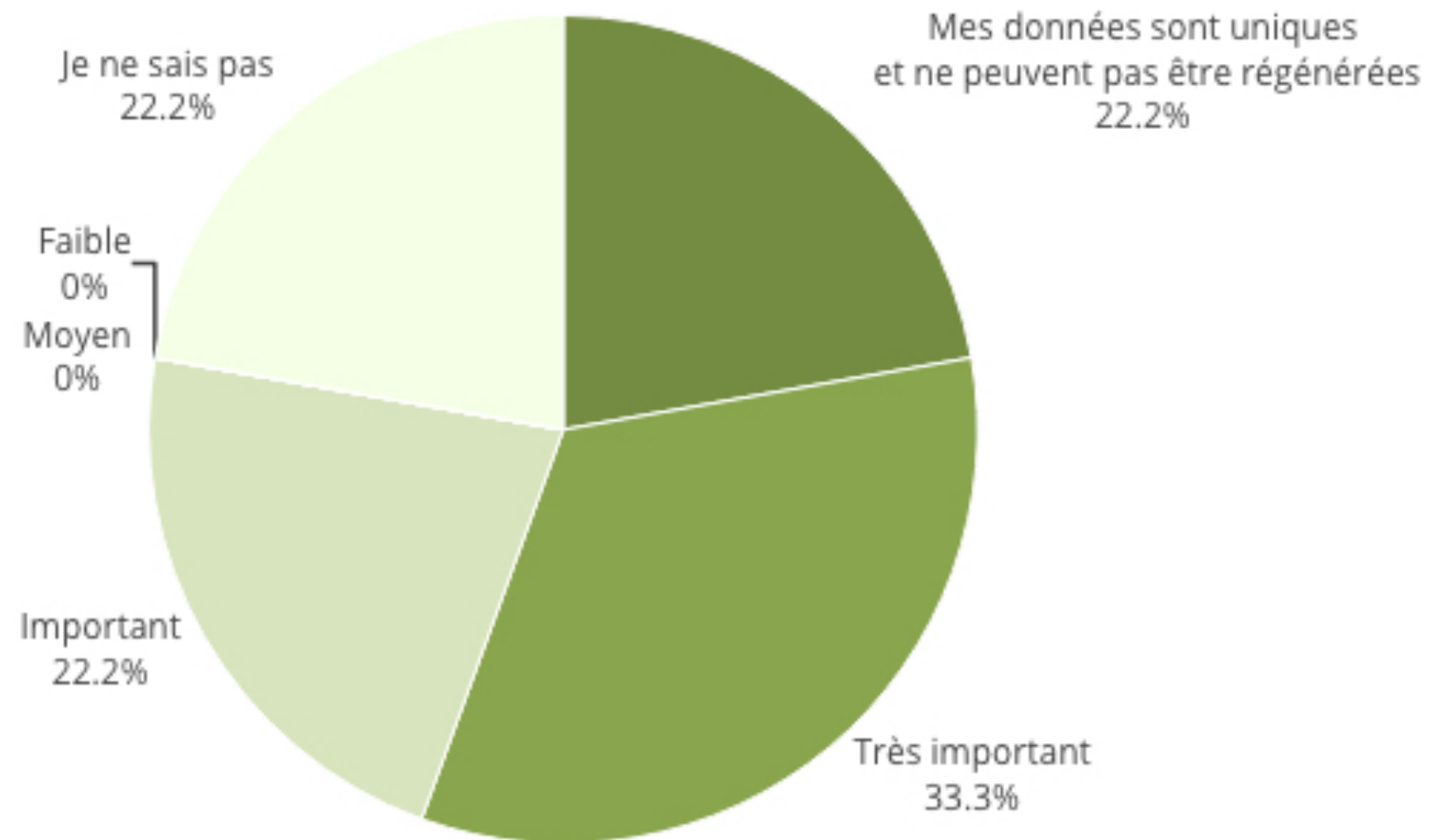


# Les données générées à l'ULB sont précieuses



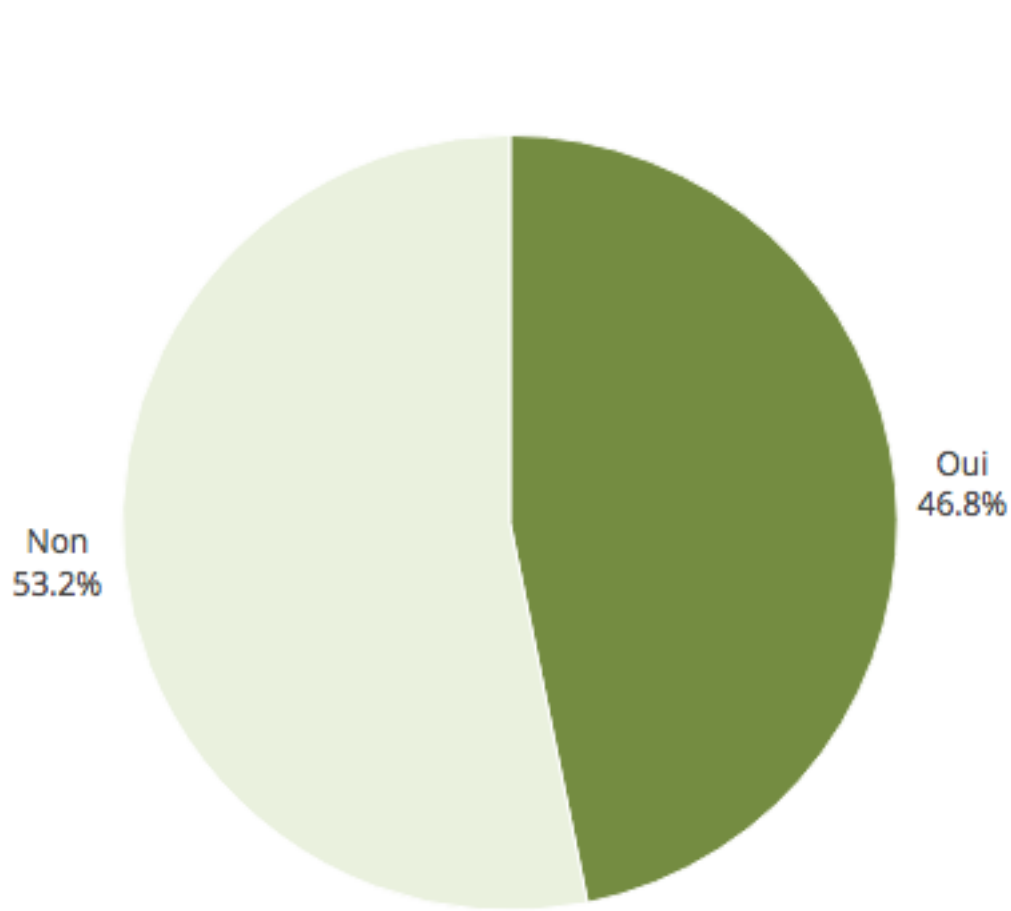
Une majorité des répondants estime que leurs données sont uniques et ne peuvent être régénérées ou qu'un effort/coût très important serait nécessaire pour les régénérer

# en Architecture aussi...

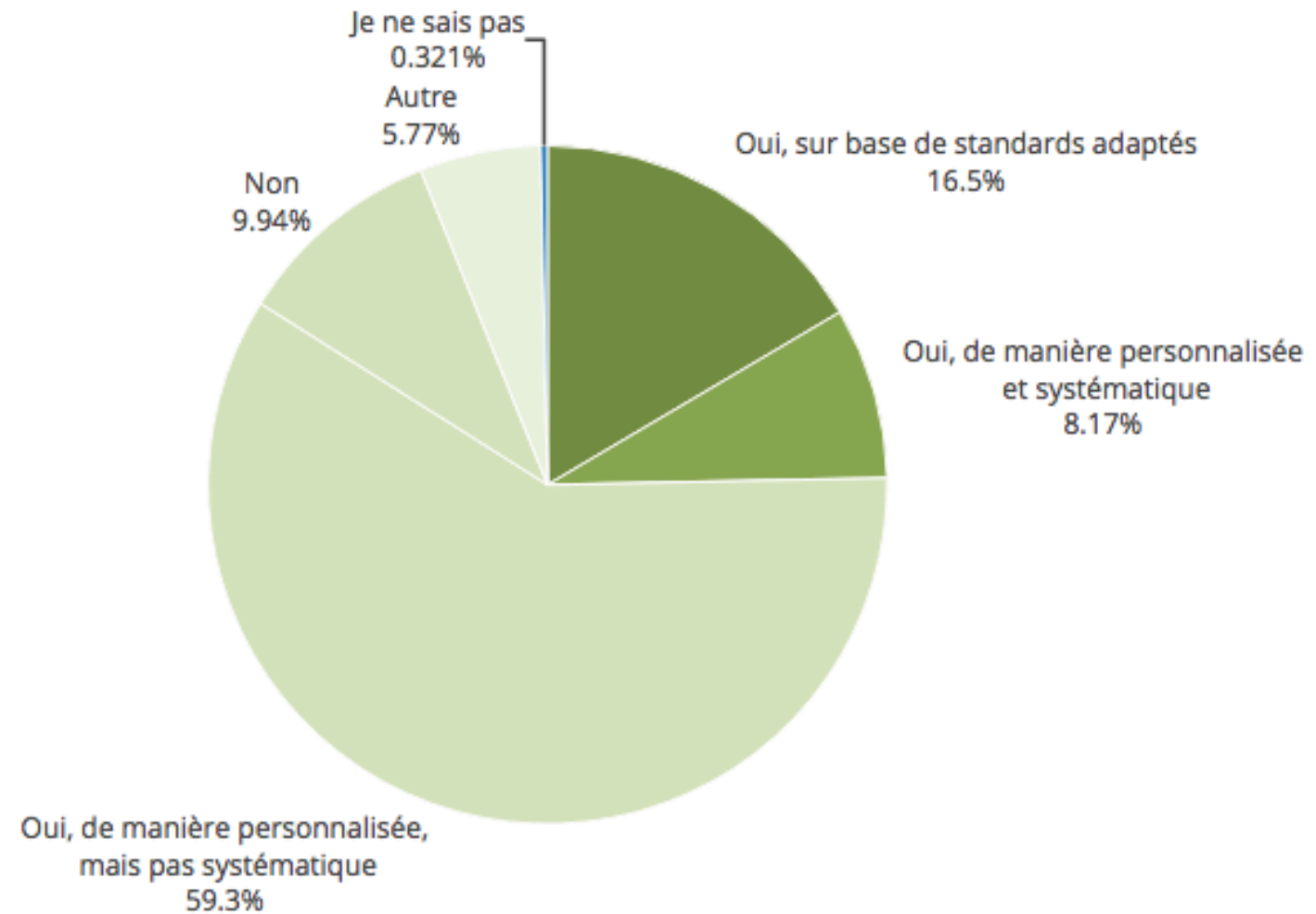




# Certaines données générées à l'ULB sont **vulnérables**



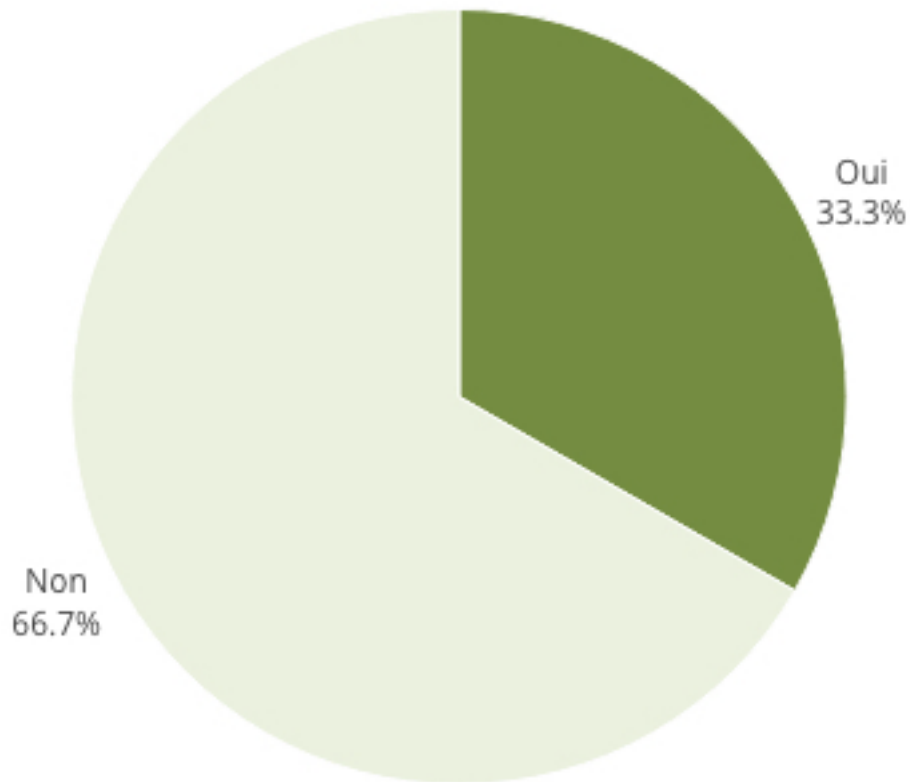
Près de 47% des répondants ont déjà subi une **perte de données**



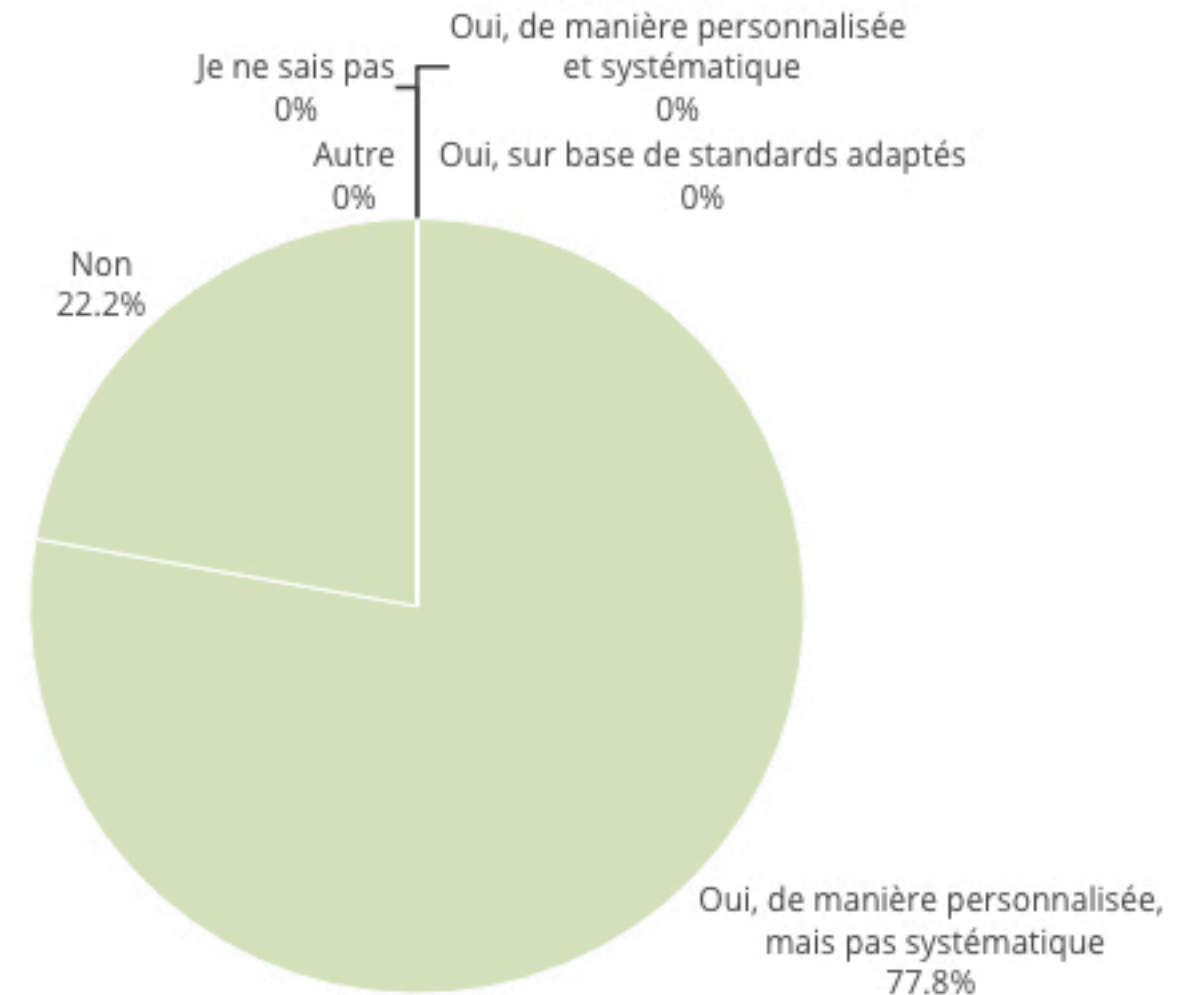
Seuls 16.5% des répondants indiquent **documenter** leurs données de recherche sur **base de standards adaptés**.



# en fac. d'architecture aussi...



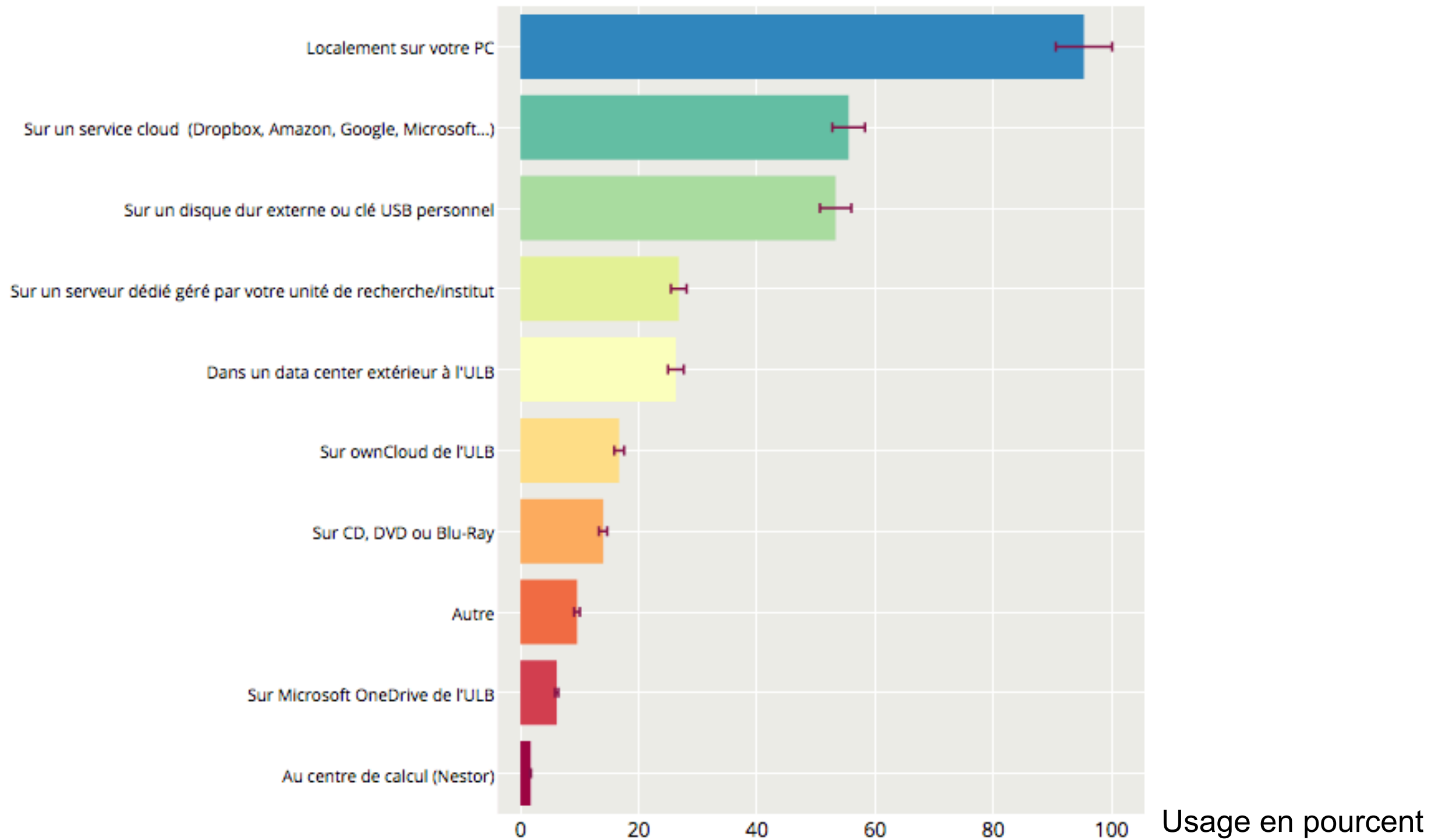
Près de 1 répondants sur 3 ont déjà subi une **perte de données**



0% des répondants indiquent **documenter** leurs données de recherche sur **base de standards adaptés**.

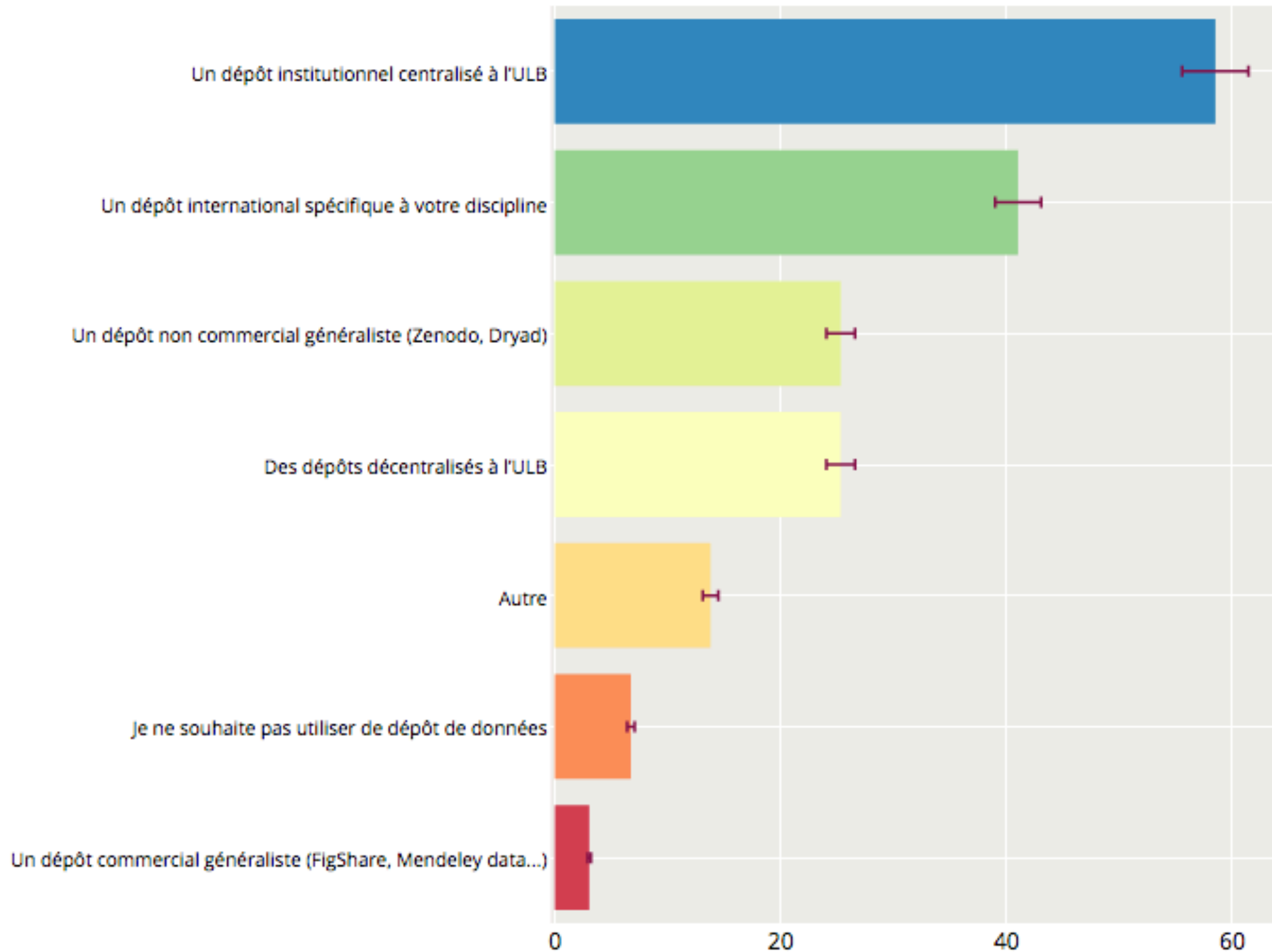


# Les chercheurs générant moins de 100GB/an stockent leurs données principalement sur un service cloud ou un disque dur externe



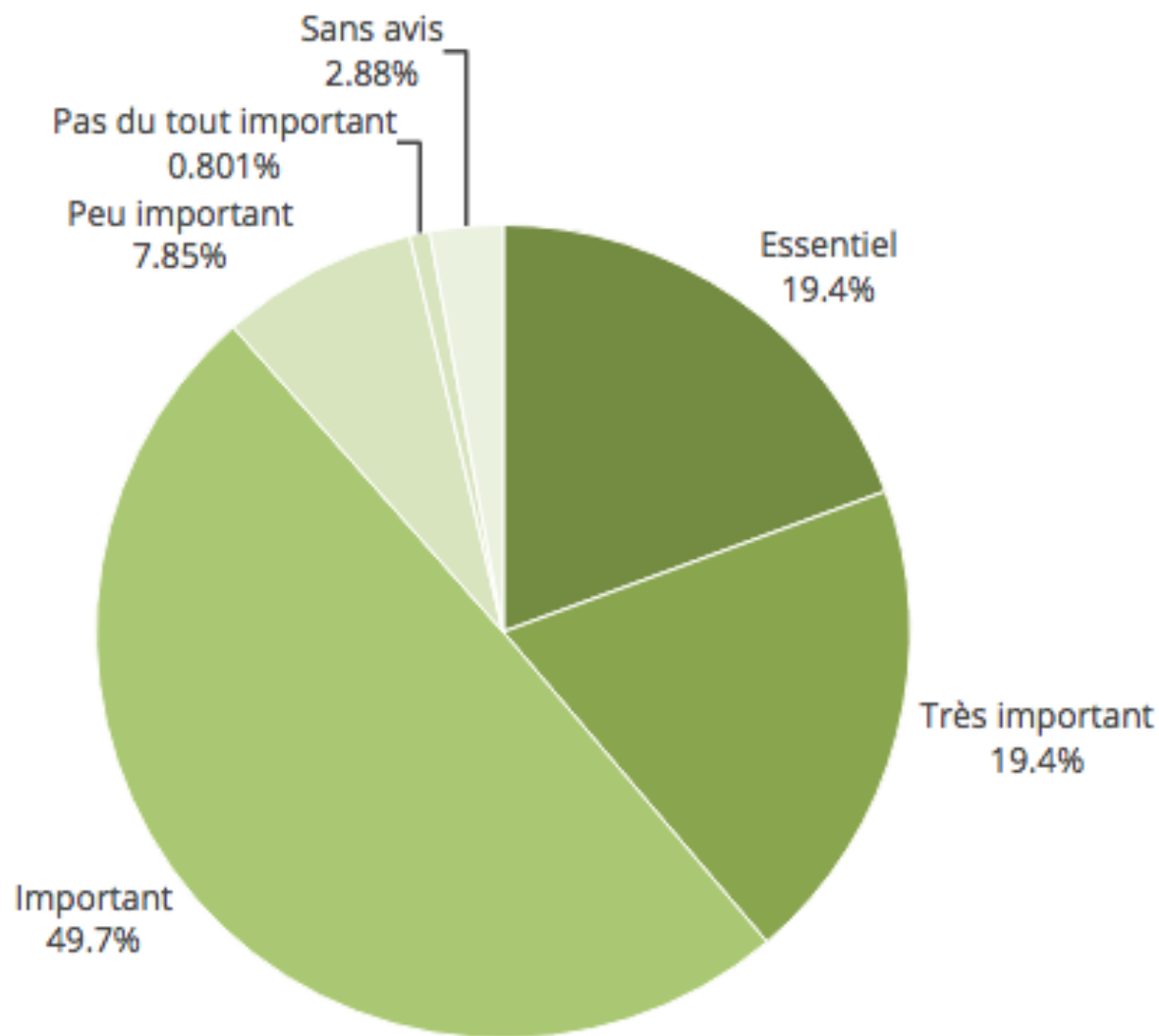
Plus de 86% des répondants indiquent qu'ils utiliseraient une infrastructure de stockage institutionnelle si elle était disponible à l'ULB

# Une large majorité des répondants utiliserait un dépôt institutionnel de données de recherche s'il était disponible à l'ULB

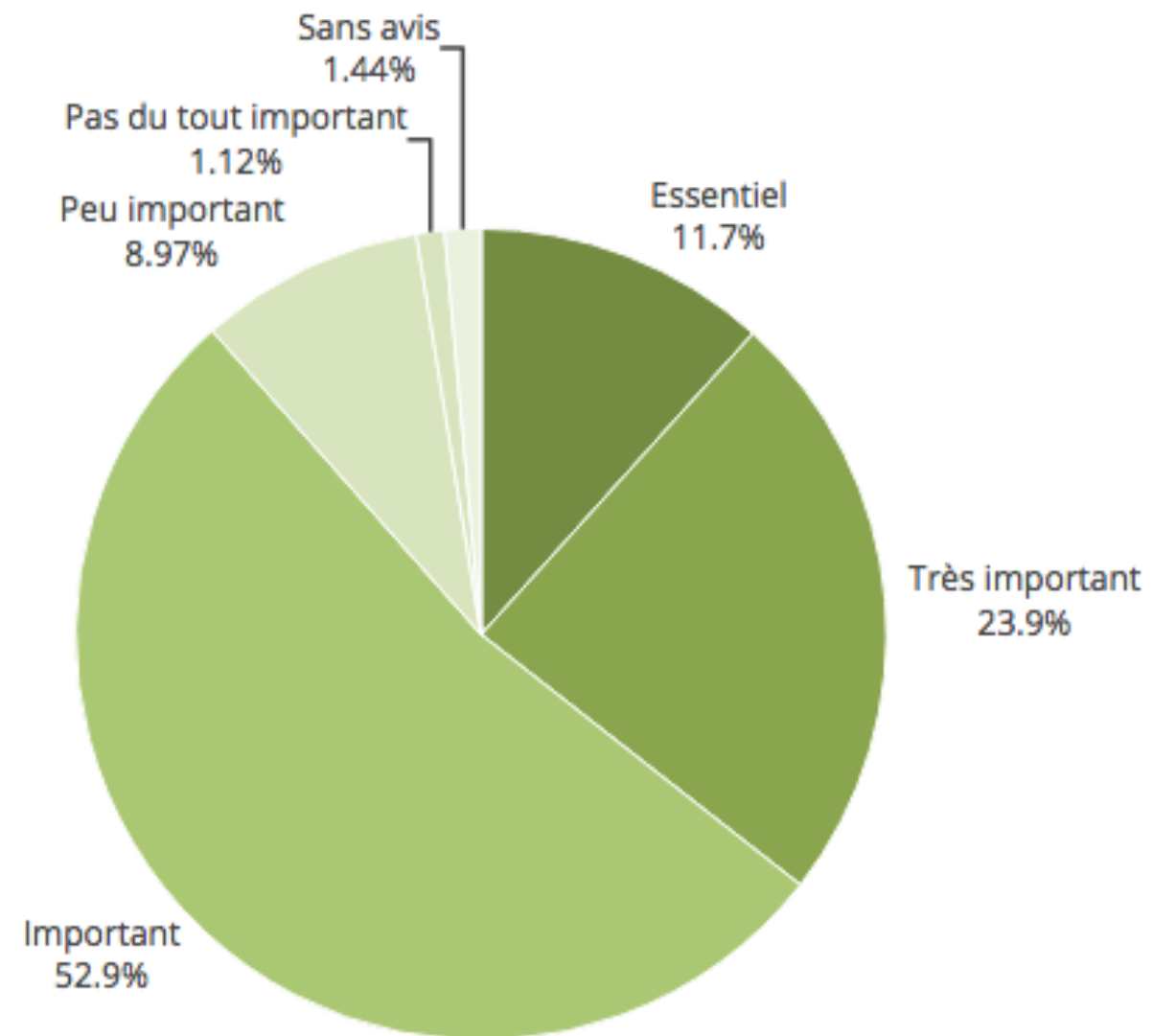




# Les répondants à l'enquête sont demandeurs de **lignes directrices**



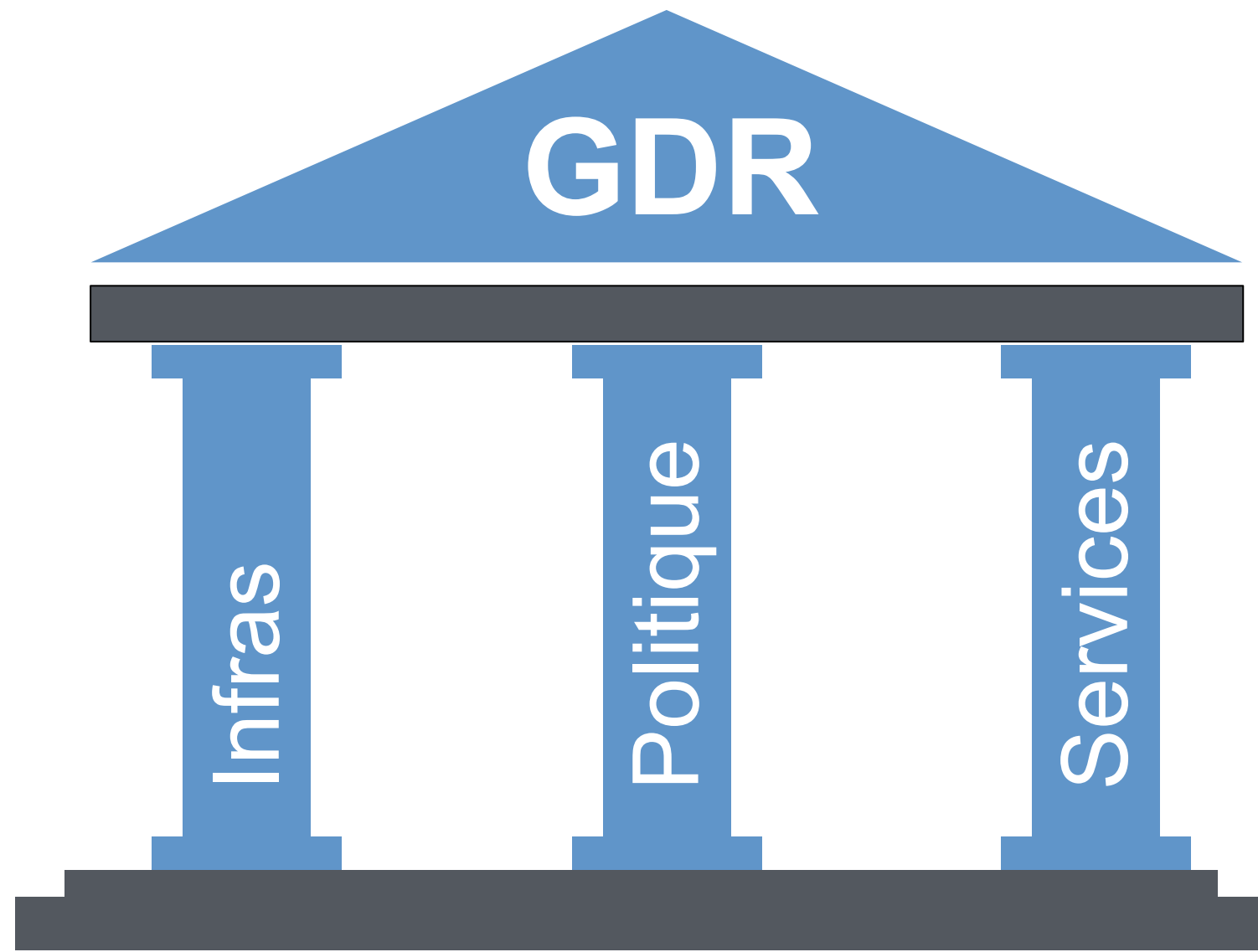
L'ULB doit-elle inclure explicitement la GDR dans les **responsabilités** du chercheur ?



L'ULB doit-elle fournir des **recommandations** et des **conseils** sur la gestion des données ?

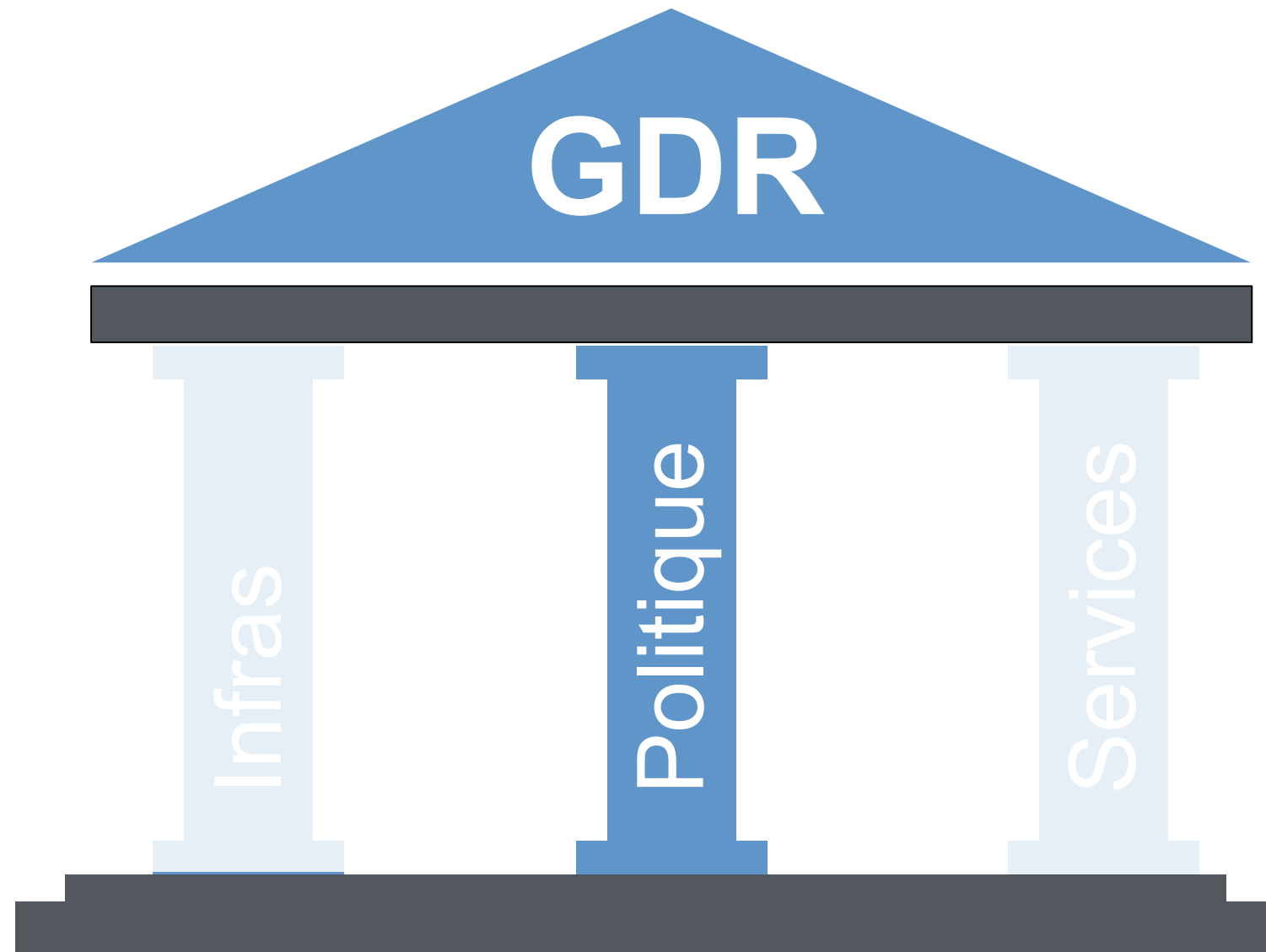
# Principales recommandations

# Une GDR efficace s'appuie sur 3 piliers principaux






# Le pilier central est la politique de l'institution en matière de GDR





# L'ULB devrait se doter d'une politique institutionnelle de GDR



contexte politique favorable (GDPR, EOSC, décret OA)



lignes directrices générales, succinctes et claires



ambitieuse



associée à une feuille de route

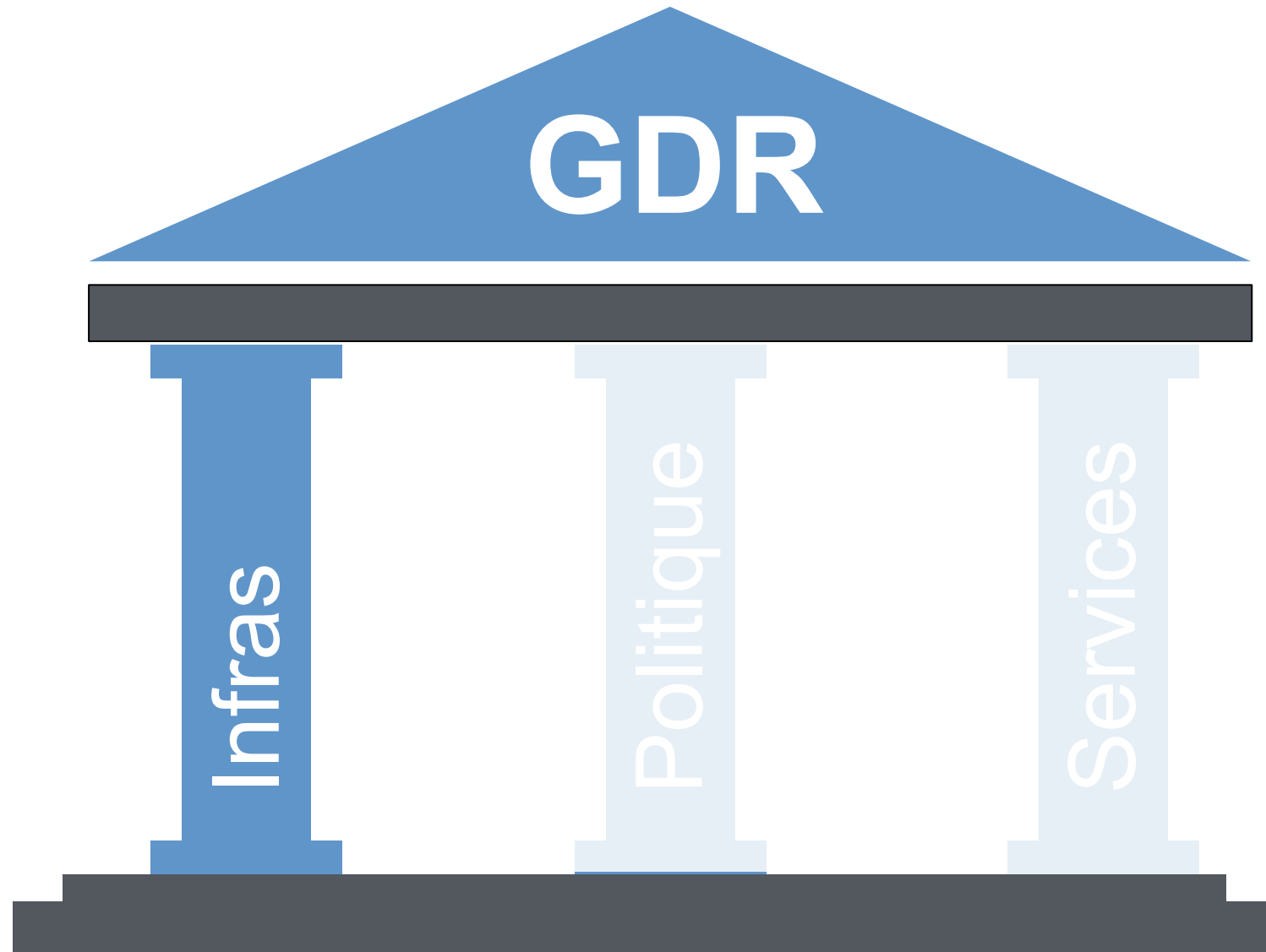


définition claire des rôles et responsabilités



encourageant l'ouverture des données tant que possible

# Le deuxième pilier : les infrastructures





# L'ULB devrait consentir des investissements importants dans la mise en place **d'infrastructures** de GDR essentielles pour les chercheurs



## **Réseau**

important déploiement en cours



## **Stockage**

solution de stockage évolutive, simple d'utilisation, multiplateforme, centralisée (GPDR), interopérable avec EOSC et autres institutions, à coût compétitif



## **Dépôt de données**

solution destinée à garantir l'accès et la préservation des données sur le très long terme  
répertorier les jeux de données produits à l'ULB dans le dépôt institutionnel, permettre la citation et le lien avec les publications



Une solution **institutionnelle** de **stockage** et de **partage** des données répondrait à l'un des besoins communs fondamentaux des chercheurs



besoin universel, indépendant du domaine de recherche



critique au vu des usages actuels et des risques de perte



essentiel pour la conformité au GDPR



avantage pour la rédaction du plan de gestion (DMP)



# Le GT recommande NextCloud



Solution hébergée à l'ULB



Interopérable avec les solutions de stockage déjà utilisées à l'ULB



Simple d'utilisation, client multi-plateforme



Point central de gestion d'accès aux données (GDPR)



Multiples couches de sécurité

(ISO 27001, cryptage de bout en bout, authentification à deux facteurs... )



Expertise de référence à l'ULB, adoption en cours par les autres universités de la FWB (Federation)



# Outils de **collecte** et de **traitement** des données: un paysage très hétérogène



Fortement dépendant du domaine de recherche



Privilégier les solutions ouvertes et interopérables



Eviter le “vendor lock-in”



Adopter des formats ouverts pour la soumission en dépôt (accès et préservation)



# Le GT propose de tester des prototypes pour les outils généralistes



GitLab

Logiciel de versionnage de code source

Installation institutionnelle possible

Standard reconnu



eLabFTW

ELN open source



SCI NOTE

Installation institutionnelle possible



OSF

Installation locale possible

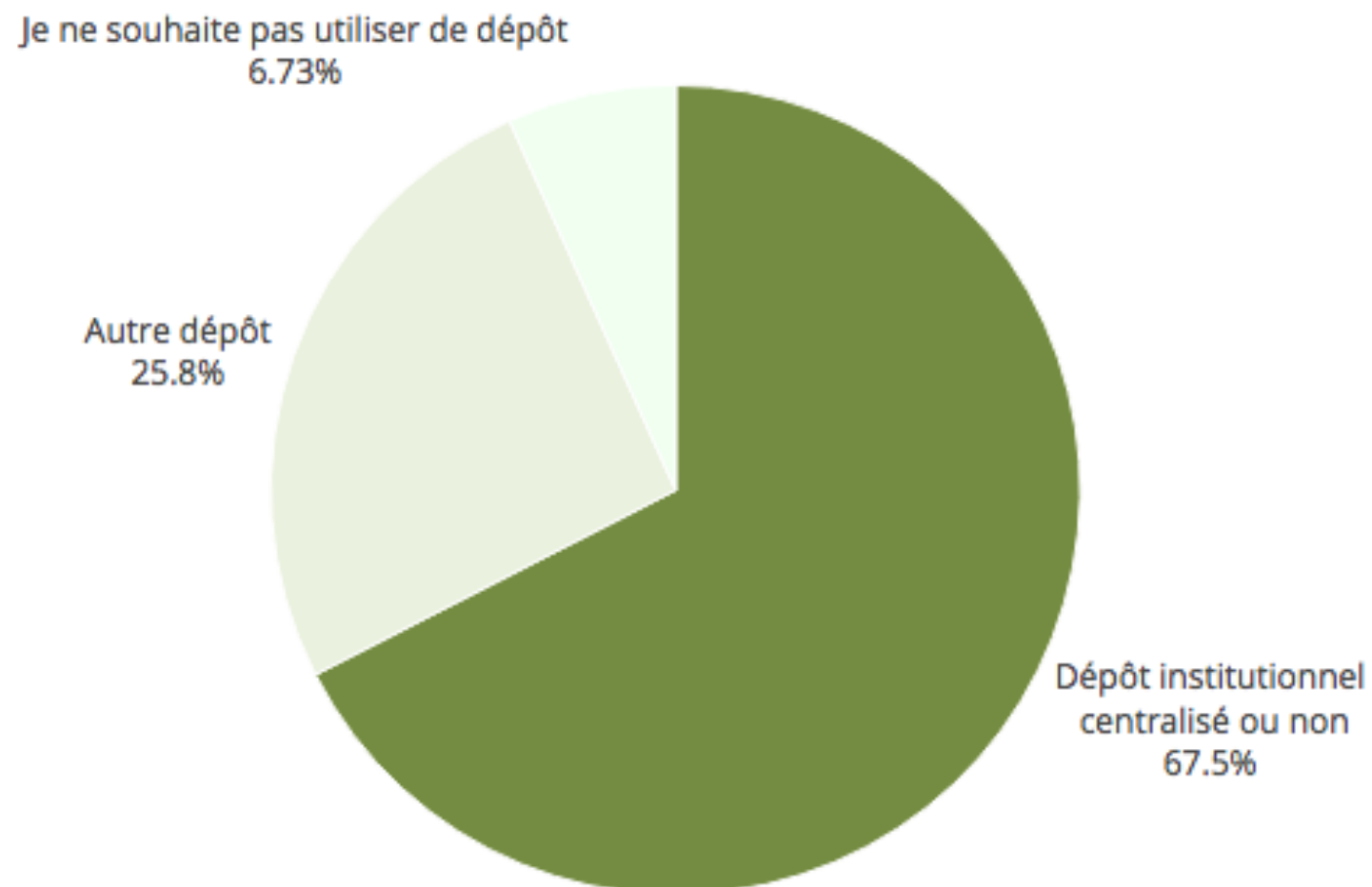


# Dépôt de données : accès et préservation

## Au minimum

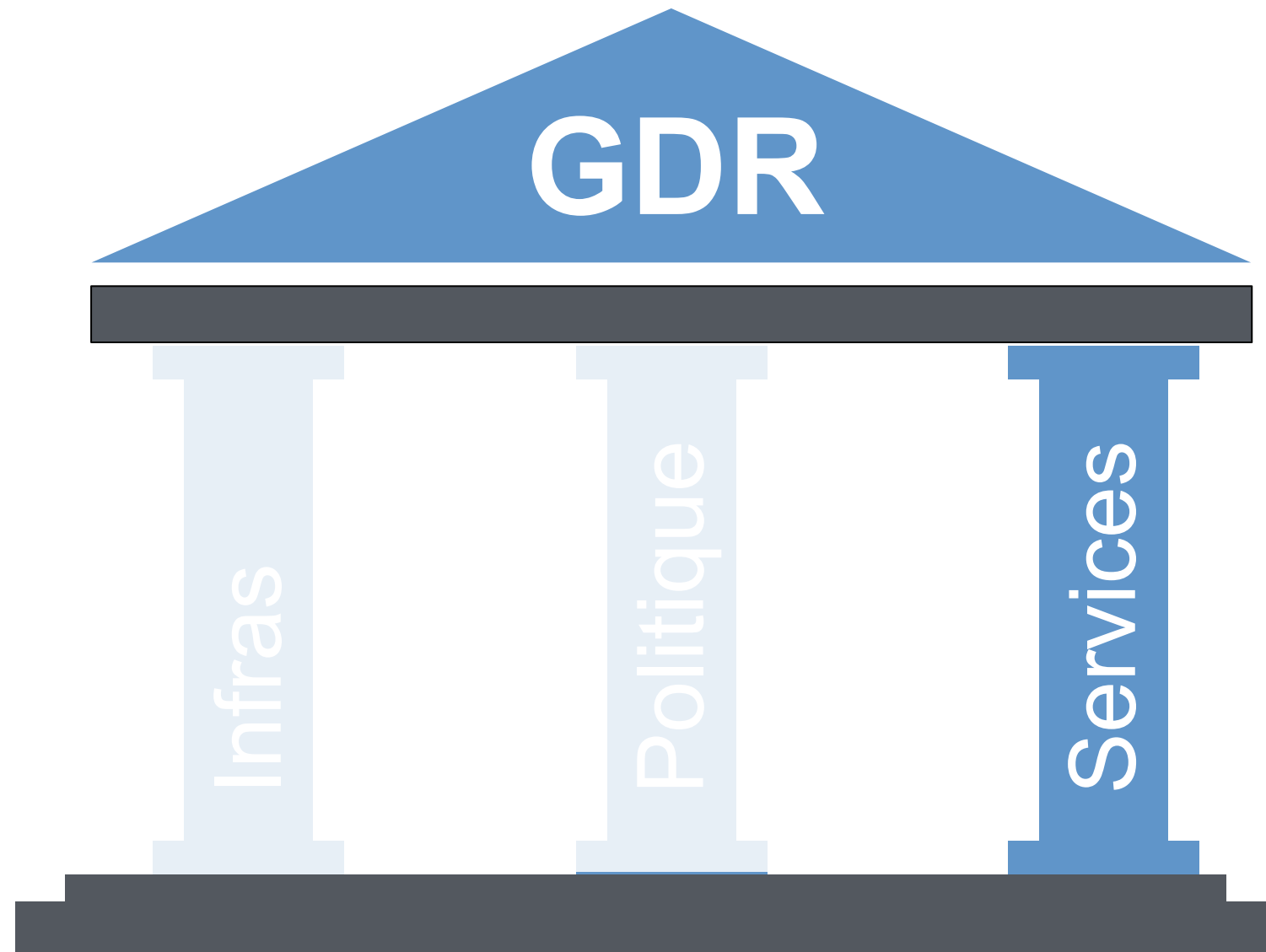
- répertorier les jeux de données produits à l'ULB dans le dépôt institutionnel
- permettre la citation et le lien avec les publications

## Intérêt pour un dépôt institutionnel de données de recherche





# Le troisième pilier : les services et le support aux chercheurs





L'ULB devrait consentir des investissements importants dans la mise en place **des services** de GDR essentiels pour les chercheurs



répondre aux besoins immédiats de création de DMP en proposant l'outil [dmponline.be](http://dmponline.be) en accompagnant dans la préparation du plan



développer un programme de formation doctorale à la GDR



service de support de premier niveau à destination des chercheurs  
site web de référence en matière de GDR (DI, DR, DPO, DBIS)

La mise en oeuvre des recommandations formulées permettra d'implémenter les infrastructures, les services et le soutien essentiels à nos chercheurs







**Merci !**