Choisir un entrepôt adapté à ses données SHS

Les plateformes de stockage de données

Joachim Schöpfel, Université de Lille

Université de Lille 3 Sciences humaines et sociales

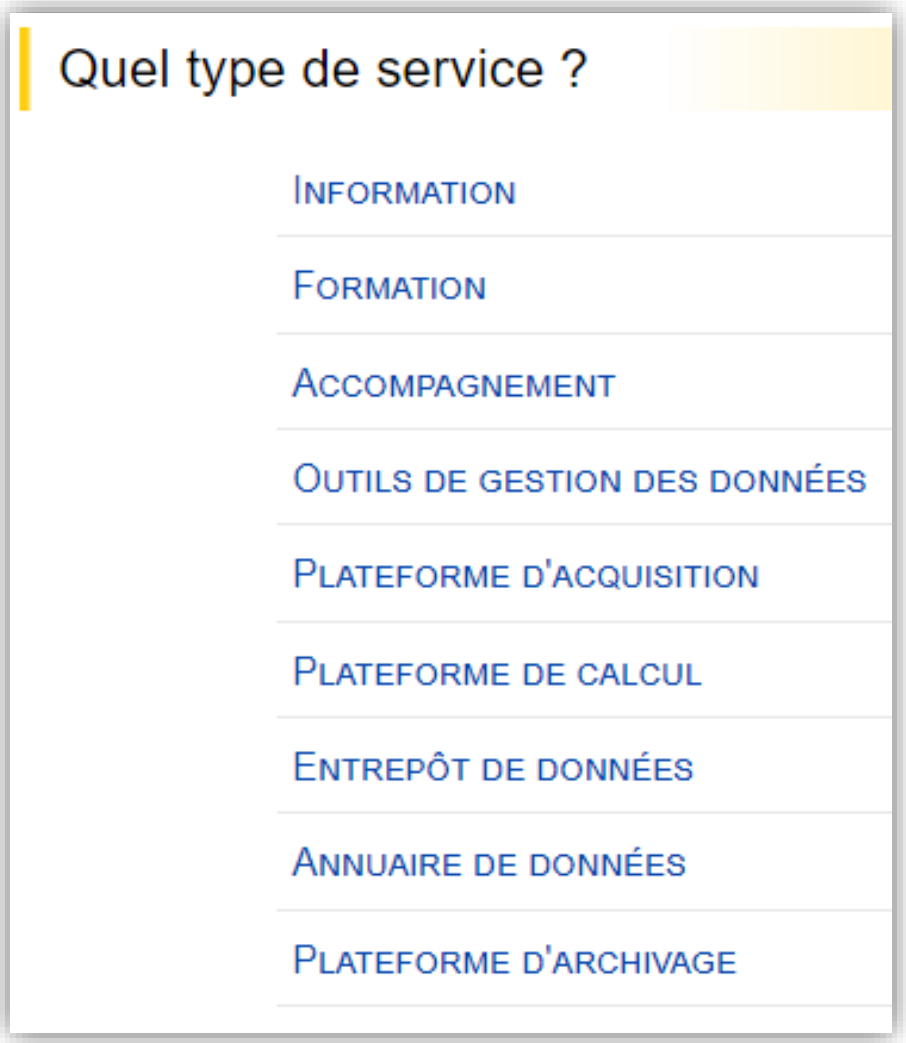
GériCo

MESHS Lille Nord de France

Région Hauts-de-France

Au préalable…

- Pas seulement « entrepôt »
  - Mais toute forme de plateforme
- Pas seulement « stockage »
  - Mais aussi gestion, utilisation, partage, *review*, suppression
- Pas seulement « préservation »
  - Mais aussi exposition, publication, réutilisation



Source : Cat OPIDoR

A double titre…

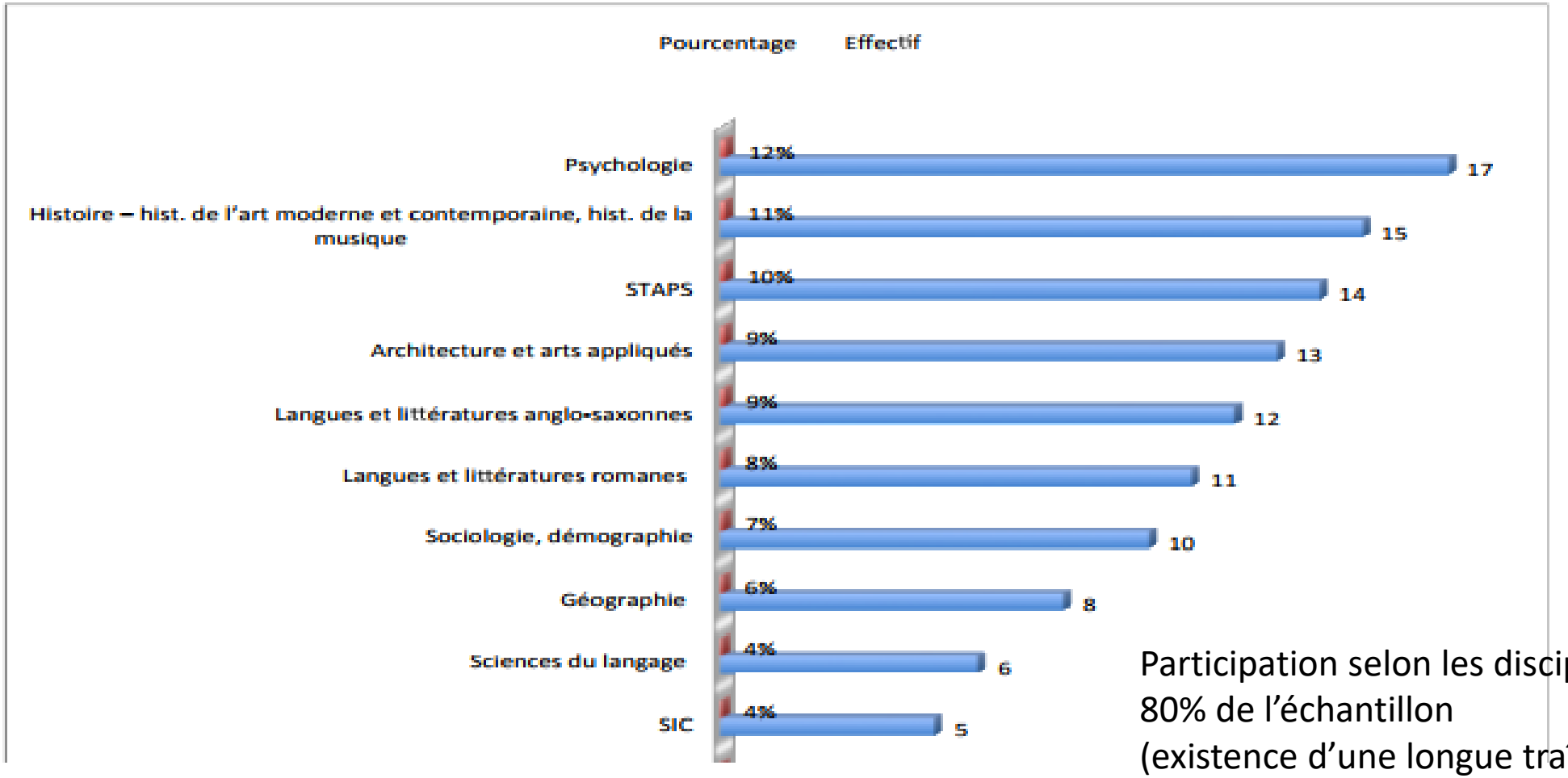
Support à la recherche

- Conseil, assistance
- Formation
- Gestion
- Administration
- Pilotage
- Conception (architecture)

Production

- Données documentaires
- *Text and data mining*
- Etudes
- Projets de recherche

Disciplines, communautés



Participation selon les disciplines CNU
80% de l'échantillon
(existence d'une longue traîne)

Source : Serres et al. (2017)

Types de données

Sources et résultats
Lien avec disciplines et instruments
Catégories +/- larges
Données +/- structurées
Données « chaudes » ou « froides »

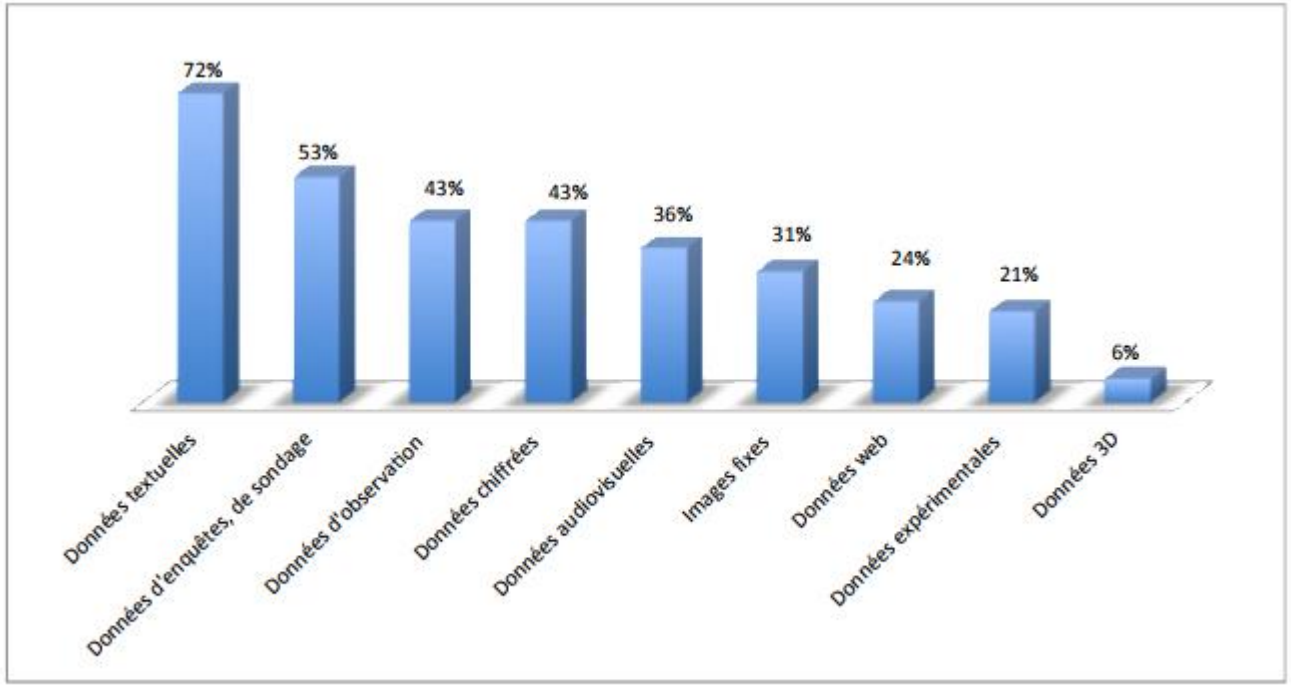


Figure 13 : Catégories des données sources (N = 127)

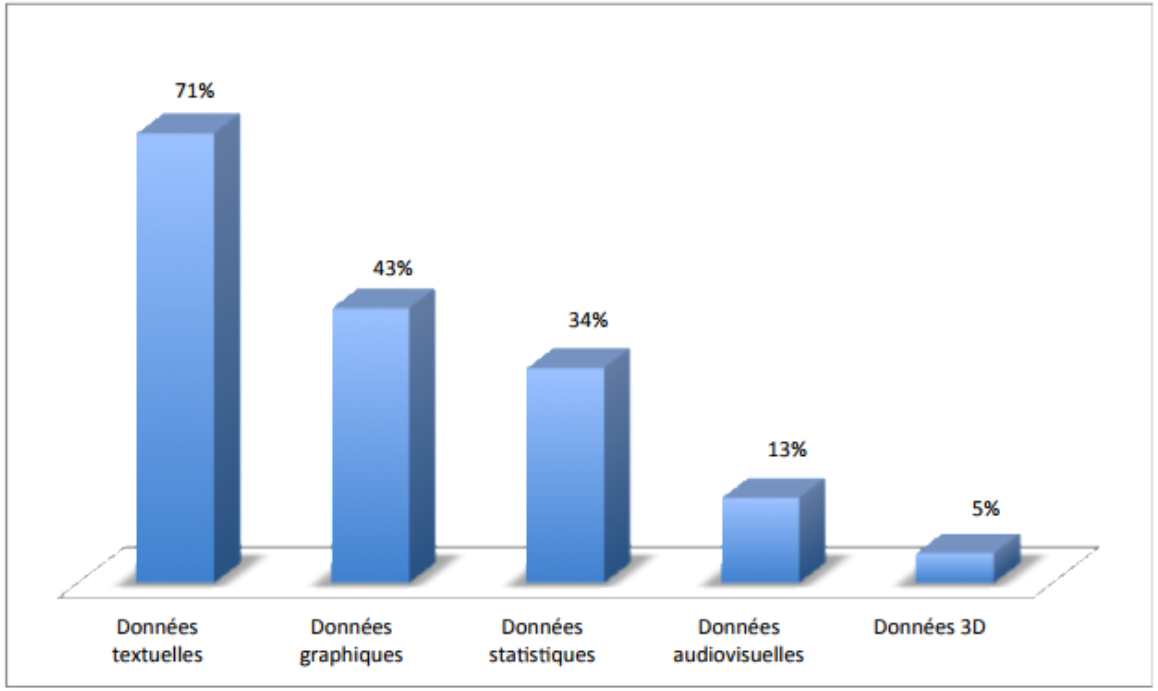
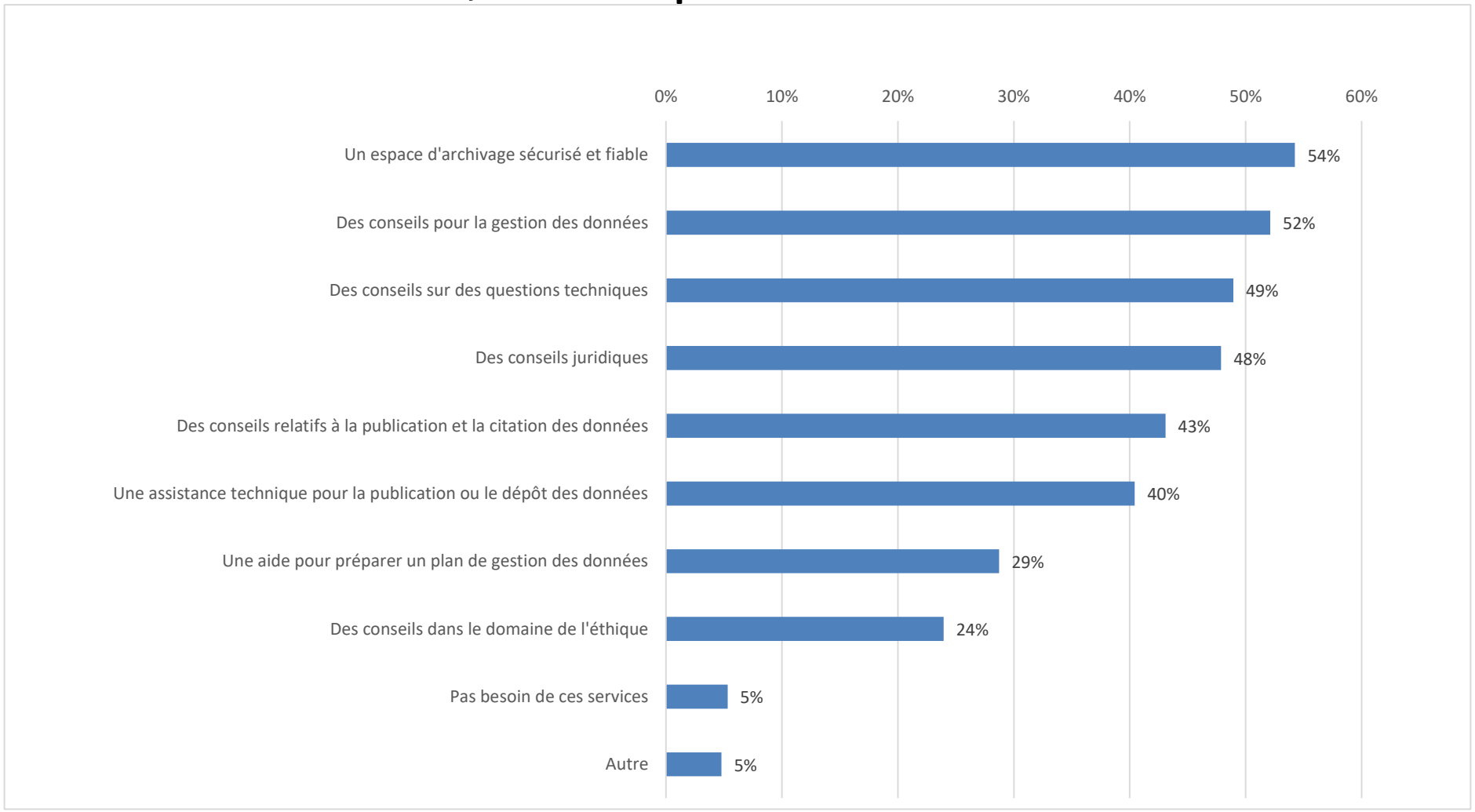


Figure 17 : Catégories de données produites (N = 127)

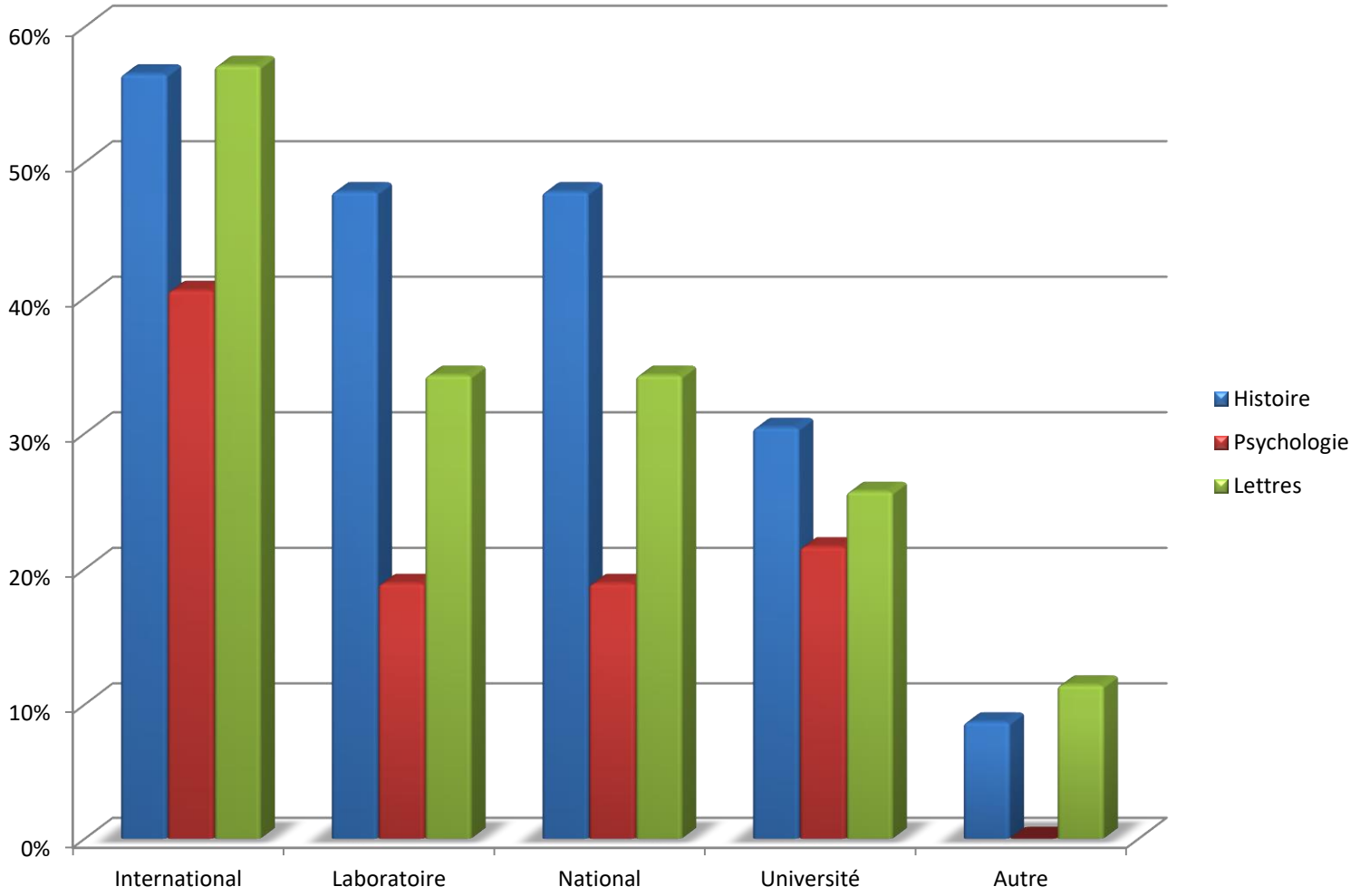
Source : Serres et al. (2017)

Avant de choisir, comprendre les besoins



Source : Prost & Schöpfel (2015)

Type de plateforme préféré



Source : Prost & Schöpfel (2015)

Support OA préféré

Contraste entre préférence
et pratique réelle (ex TGIR)

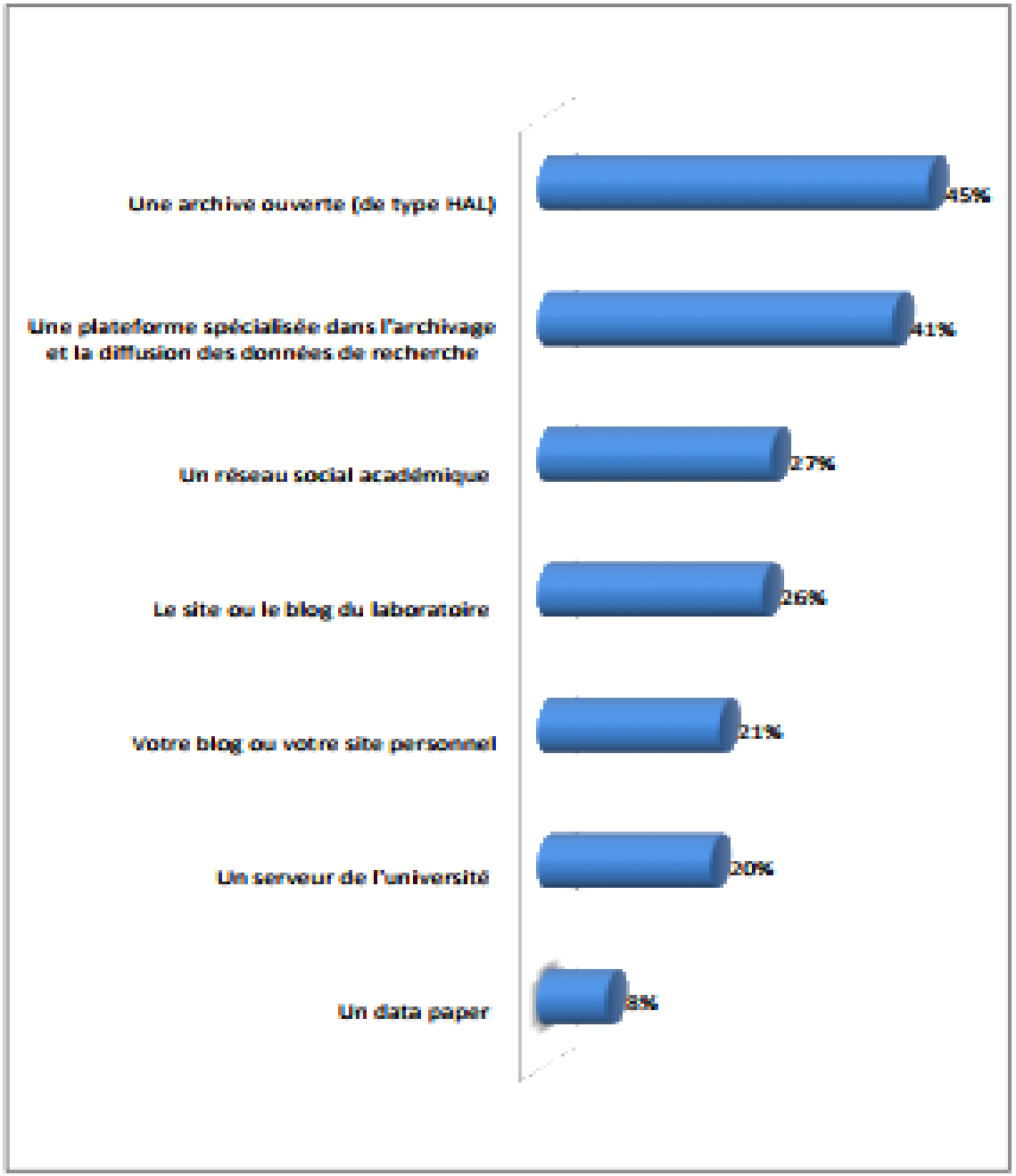


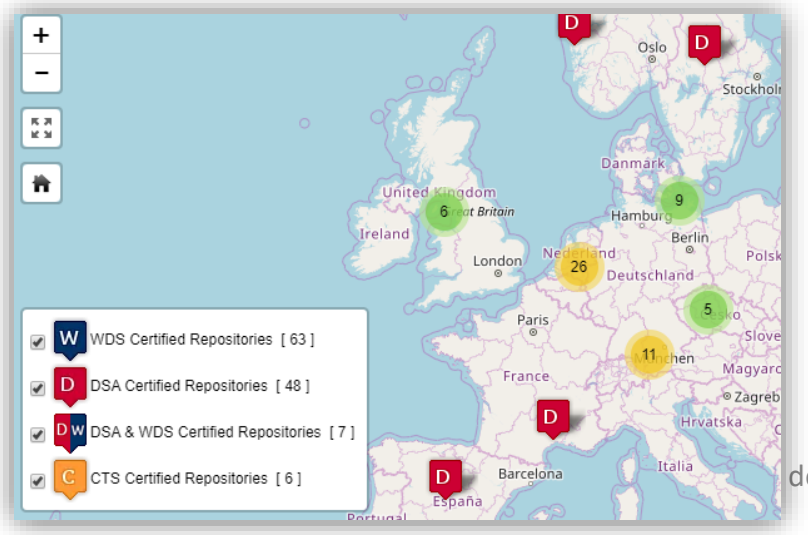
Figure 44 : Support à privilégier pour un dépôt en libre accès (N = 96)

Source : Serres et al. (2017)

Critères de qualité (1)

Certification, référencement

- CLARIN B centre label
- CoreTrustSeal
  - Data Seal of Approval (DSA)
  - ICSU World Data System (WDS)
- re3data, Cat OPIDoR



Conformité avec recommandations

DataCite Metadata Schema

FAIR Guiding Principles

- *Findability*
  - Métadonnées
- *Accessibility*
  - Ouverture
- *Interoperability*
  - Standards
- *Reusability*
  - Documentation

Critères de qualité (2)

Fonctionnalités

- Dépôt (facilité, authentification…)
- Diffusion (exposition, licences, *data papers*…)
- Gestion (confidentialité, partage…)
- Manipulation (versionnage, suppression, validation…)
- Curation (métadonnées, documenation…)
- Stockage (y compris pour *peer review*)
- Préservation (logiciels, cryptage…)
- Sécurité SI/données

Gouvernance, statut, modèle économique

- Public vs privé
  - figshare
- Infrastructure nationale vs service international
  - Zenodo
- Organisme français ou pas
  - Huma-Num
- Gratuit vs payant
  - DRYAD

Entrepôts de données en France (1)

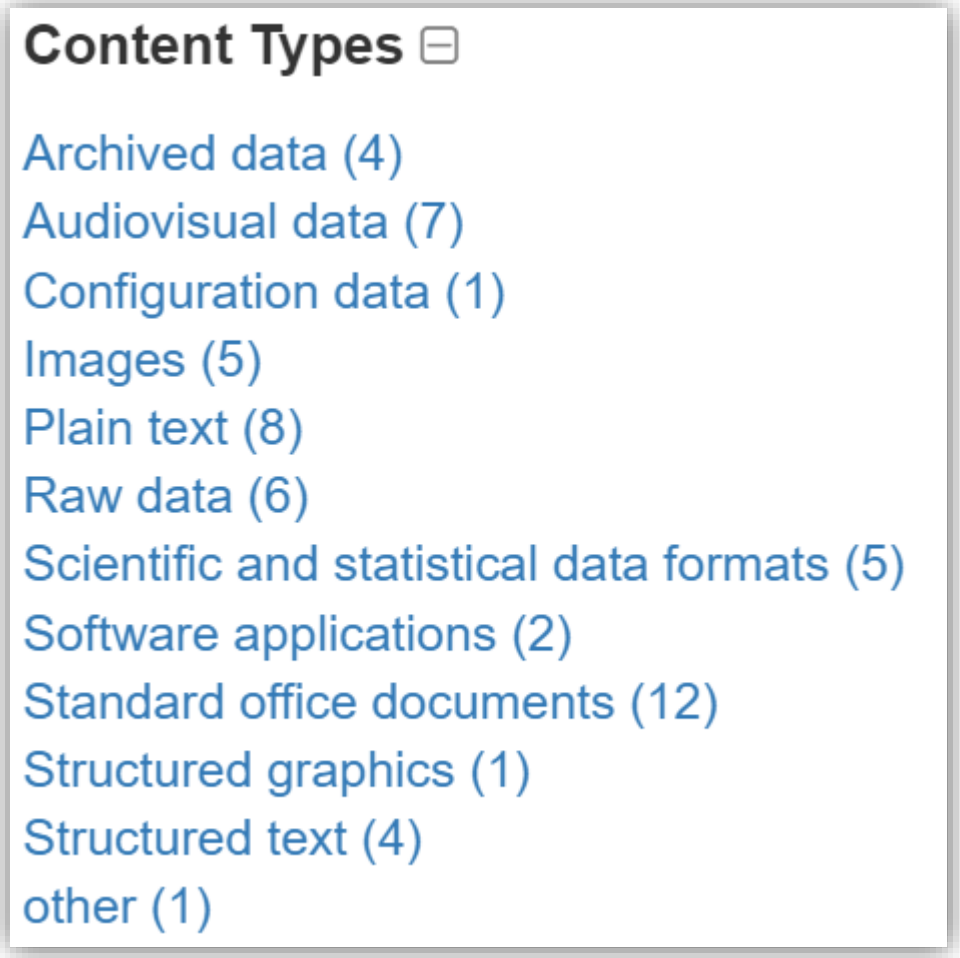
Cat OPIDoR -SHS

- 50 services dont 12 entrepôts
- 6 disciplinaires
  - Linguistique
  - Archéologie
  - Anthropologie
- 6 autres (thématique, types de données…)
  - Enquêtes
  - Images
  - Bassin méditerranéen

Entrepôts de données en France (2)

re3data -SHS

- 15 entrepôts
  - Dont plus de la moitié à caractère international
- 12 disciplinaires
  - Dont NAKALA
- 3 institutionnels
  - Dont CINES
- 4 autres
  - Dont CINES
- Pb indexation



Entrepôts de données SHS (2)

Metadata standards

ABCD - Access to Biological Collection Data (1)
CF (Climate and Forecast) Metadata Conventions (1)
DCAT - Data Catalog Vocabulary (1)
DDI - Data Documentation Initiative (68)
DIF - Directory Interchange Format (1)
Darwin Core (2)
DataCite Metadata Schema (24)
Dublin Core (50)
EML - Ecological Metadata Language (3)
FGDC/CSDGM - Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata (8)
ISO 19115 (8)
OAI-ORE - Open Archives Initiative Object Reuse and Exchange (5)
PROV (1)
RDF Data Cube Vocabulary (3)
Repository-Developed Metadata Schemas (7)
SDMX - Statistical Data and Metadata Exchange (1)
other (12)

F

Data access

A

closed (55)
embargoed (31)
open (287)
restricted (242)

API

I

FTP (13)
NetCDF (1)
OAI-PMH (56)
OpenDAP (1)
REST (38)
SOAP (3)
SPARQL (3)
SWORD (12)
other (35)

Data licenses

Apache License 2.0 (7)
BSD (10)
CC (122)
CC0 (19)
Copyrights (135)
ODC (13)
OGL (5)
OGLC (2)
Public Domain (25)
RL (4)
other (216)

R ?

Certificates

CLARIN certificate B (22)
CoreTrustSeal (4)
DINI Certificate (1)
DSA (41)
RatSWD (25)
WDS (5)
other (3)

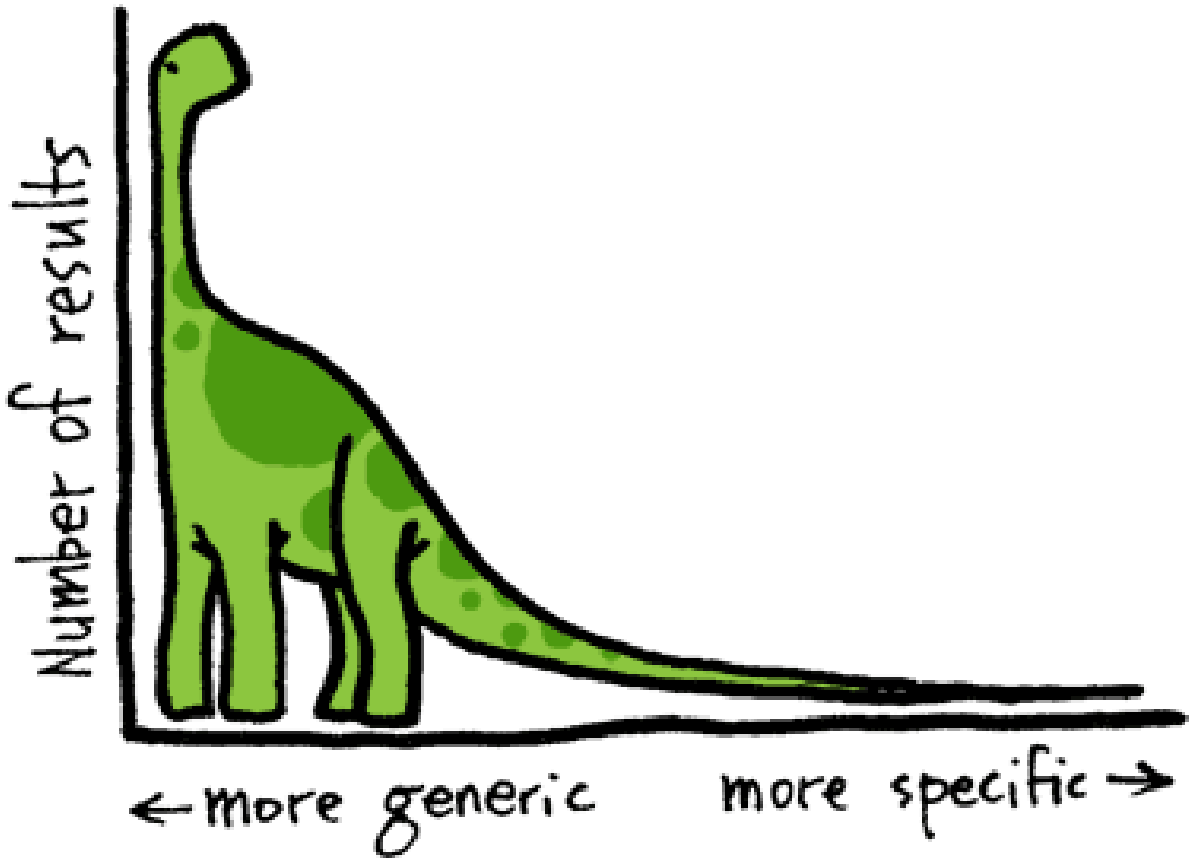
Software

CKAN (8)
DSpace (12)
DataVerse (14)
EPrints (1)
Fedora (18)
MySQL (5)
Nesstar (15)
dLibra (1)
eSciDoc (1)
other (55)
unknown (170)

Demandes

- Espace serveur pour machine virtuelle
- Conservation d'une solution ad hoc
- Partage de données confidentielles
  - En particulier, avec d'autres organismes (hôpitaux…)
- Exposition
- Pour publication (mais peu en SHS, pour l'instant)
- Pour *data paper*
- …

Quid de la longue traîne ?



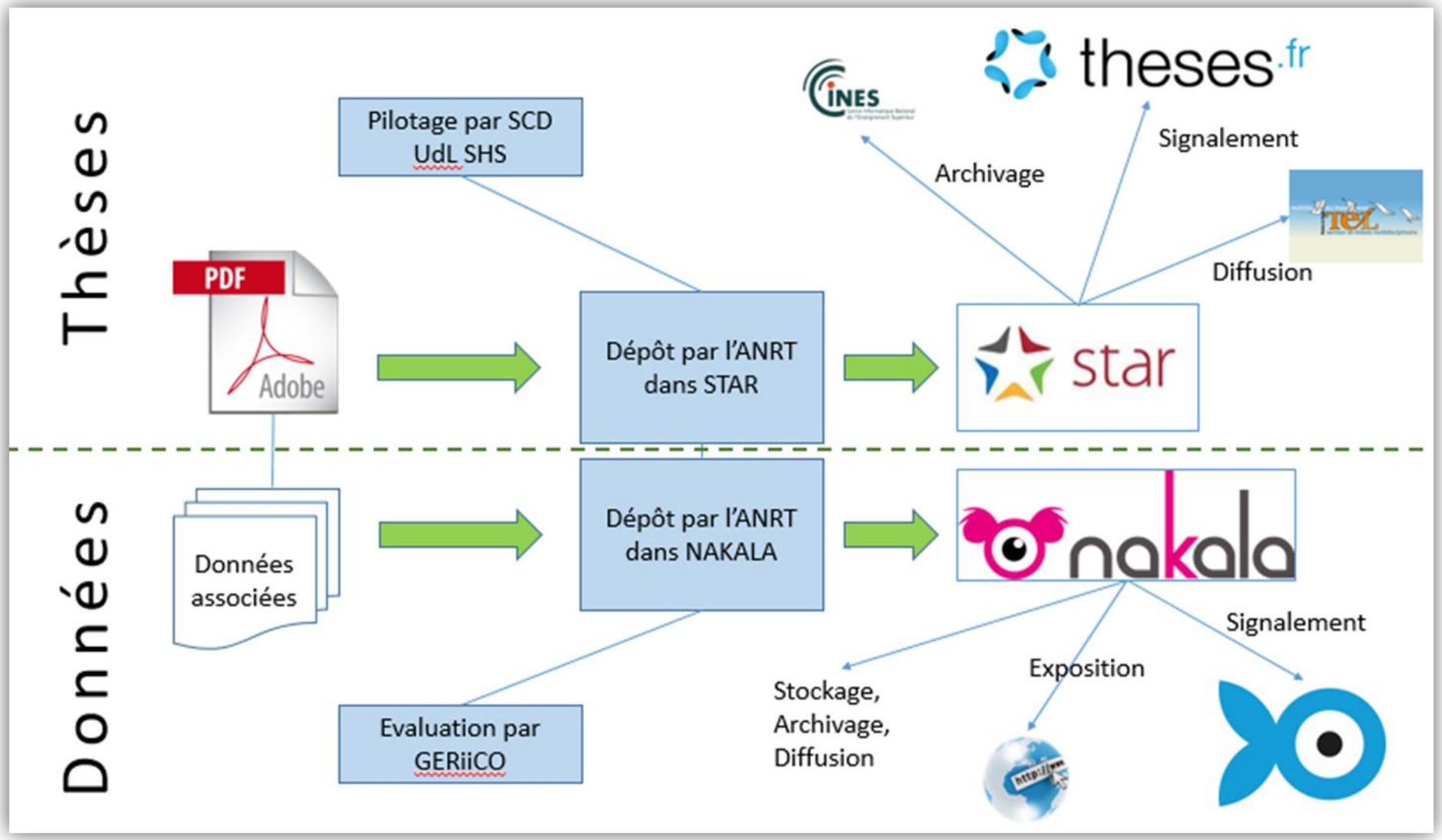
Source : www.sitepronews.com

- Données orphelines
- *Small data*
- Données non structurées
- Solutions transitoires
- …

« *Pour les données de recherche, il existe 100 infrastructures en France, mais il n'y a pas de structure banalisée qui serait une infrastructure par défaut avec vocation à disparaitre quand toutes les disciplines seront pourvues (…) L'objectif pour Marin Dacos est que les communautés de la longue traîne soient représentées.* »

(source : CORIST 19 octobre 2017)

Projet lillois



Références

- Kindling, M. et al., 2017. The landscape of research data repositories in 2015: A re3data analysis. *D-Lib Magazine* 23 (3/4). http://www.dlib.org/dlib/march17/kindling/03kindling.html
- Prost, H., Schöpfel, J., 2015. *Les données de la recherche en SHS. Une enquête à l'Université de lille 3*. Rapport final. Université de Lille 3, Villeneuve d'Ascq. http://hal.univ-lille3.fr/hal-01198379v1
- Schöpfel, J., Prost, H., 2016. Research data management in social sciences and humanities: A survey at the University of Lille 3 (france). *LIBREAS. Library Ideas* 29, 98-112. http://hal.univ-lille3.fr/hal-01395816
- Serres, A. et al., 2017. *Données de la recherche en SHS. pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2*. Rapport. Université Rennes 2. https://hal.archives-ouvertes.fr/hal-01635186

Merci !

joachim.schopfel@univ-lille3.fr