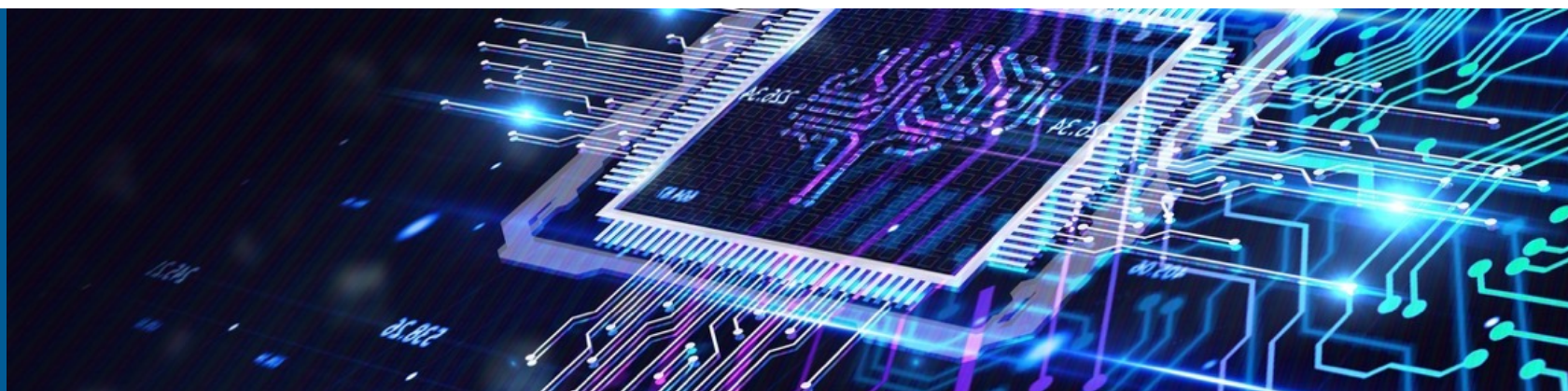




CSC

ICT Solutions for
Brilliant Minds



Partager les données de la recherche. Perspectives Européennes

Damien Lecarpentier, CSC-IT Center for Science, Finland

Assises nationales des données de la recherche, 26-27 Novembre 2024

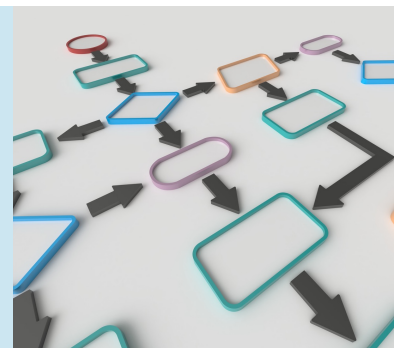


Notre domaine d'expertise comprend notamment les infrastructures de recherche, l'interopérabilité et la transformation digitale



Chiffre d'Affaires en 2023

90 M€



L'un des datacenters les plus écologiques au monde à Kajaani.

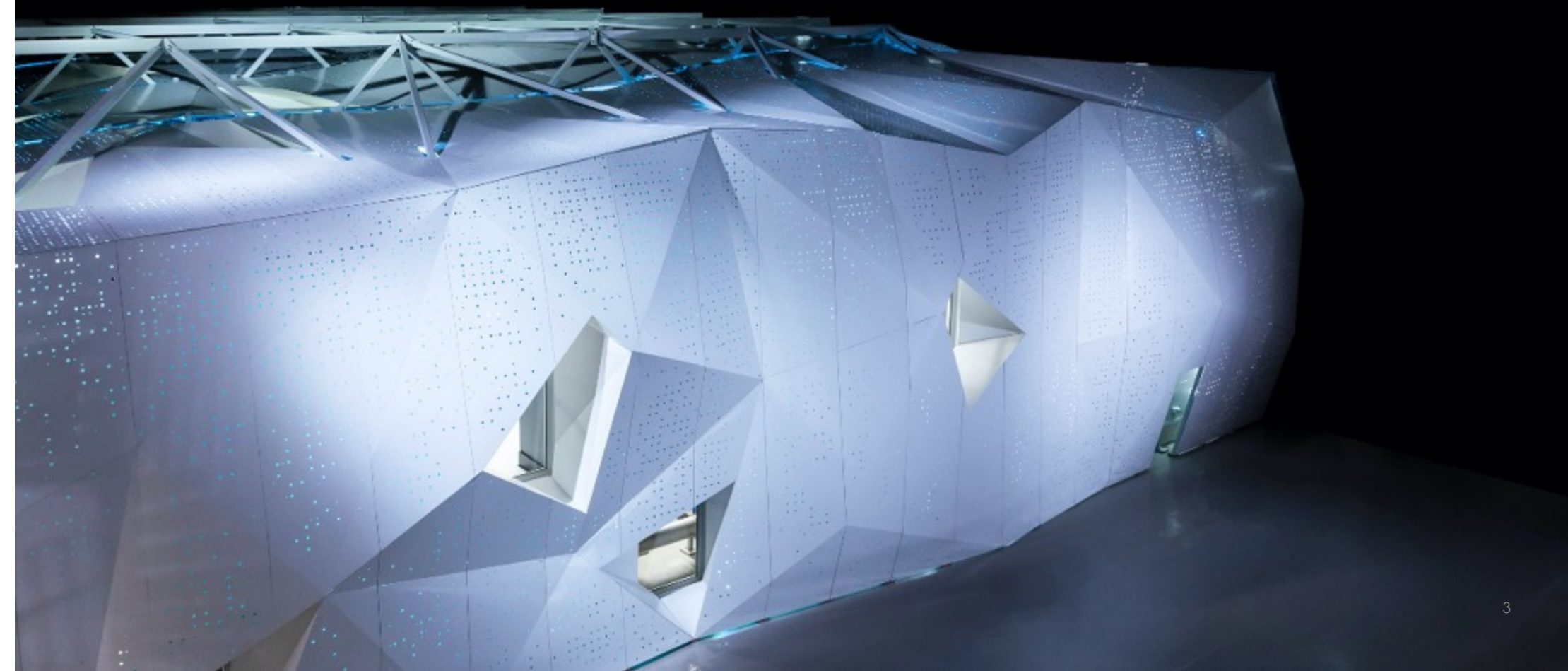


Une entreprise publique à but non lucratif, détenue à **70 %** par l'État finlandais et à **30 %** par des établissements d'enseignement supérieur finlandais, et chargée de missions spécifiques.

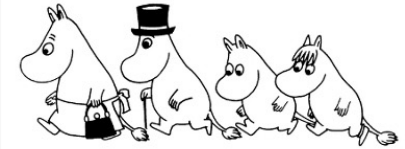


Approx.
660
Employés en
2023

LUMI



CSC: Un acteur clé de la Science Ouverte en Finlande



Open
Science

Fairdata.fi



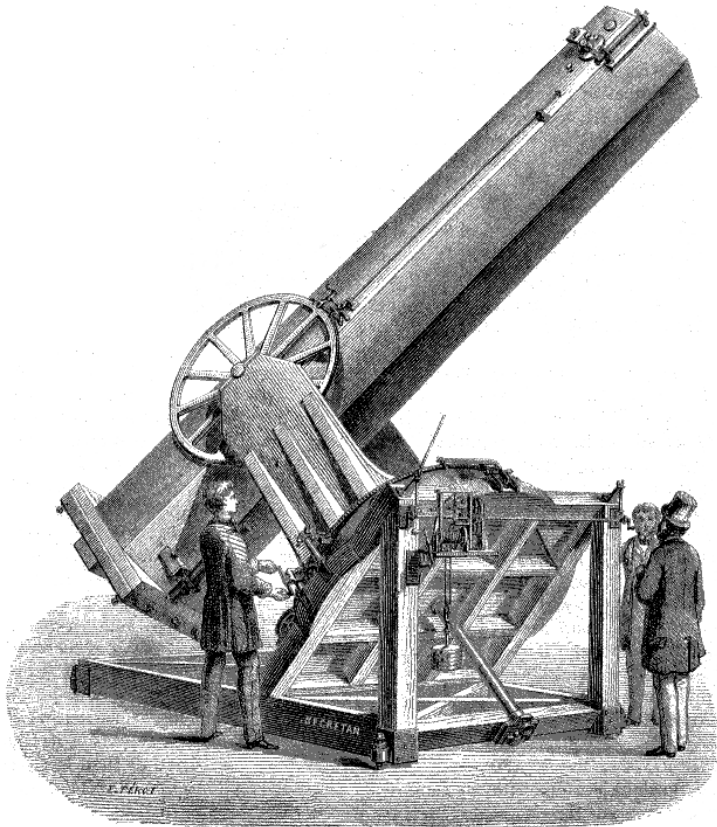
EOSC 
Finnish Forum

- **Membre du Groupe de Pilotage National pour la Science Ouverte et la Recherche**, au sein de l'Initiative Nationale pour la Science Ouverte.
- **Opérateur de la plateforme et des services Fairdata.fi**. Les services Fairdata font partie des services de préservation numérique offerts par le ministère de l'Éducation et de la Culture de Finlande.
- **Membre actif de l'Association EOSC et coordinateur du Forum EOSC en Finlande**, en collaboration avec le Ministère de l'Éducation et de la Culture.

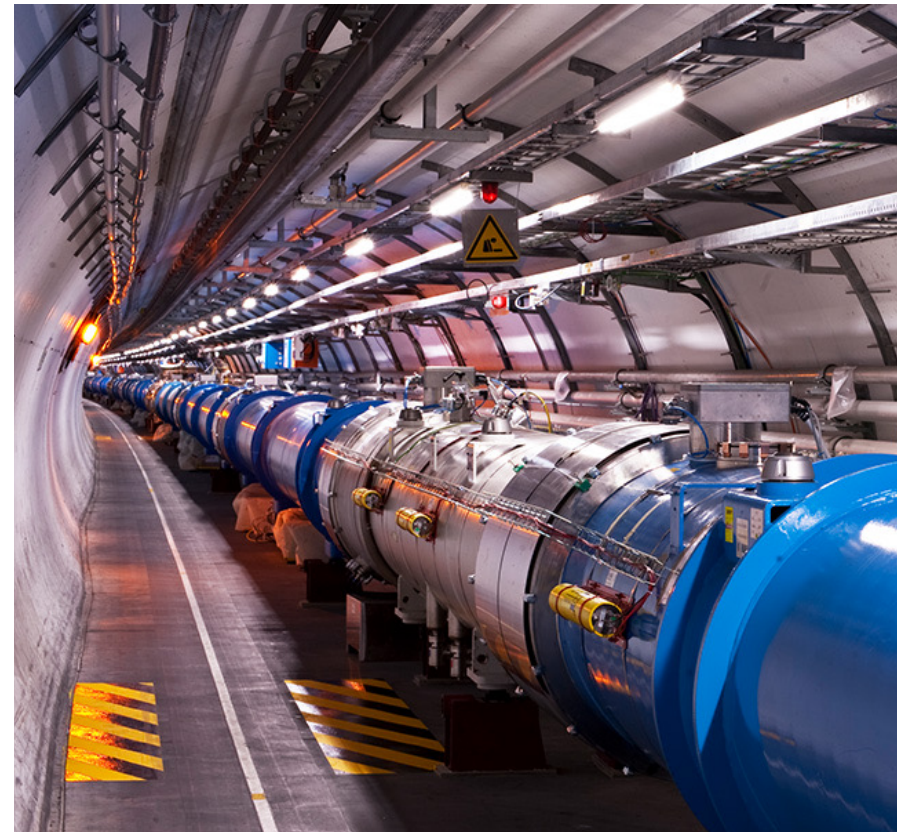
Partager les données de la recherche: une préoccupation récente?



Partager les données collectées par les instruments



Télescope de Foucault, By Unknown author - Public Domain



LHC, By Unknown author - Wikimedia Commons

Pourquoi partager les données?



1. Amélioration de la qualité et de la transparence de la recherche: vérification plus précise des données, plus grande confiance dans la science.



2. Efficacité et réduction des coûts: réduction des risques et des coûts de duplication dans la collecte et la production des données, mise en commun des données et donc réutilisation.



3. Collaboration et accélération de la connaissance: favoriser un engagement plus large des autres chercheurs mais aussi des citoyens, des gouvernements et des organisations, faciliter la recherche multidisciplinaire pour répondre aux défis sociétaux complexes, accélérer l'innovation en permettant aux chercheurs de différents domaines de collaborer et de combiner leurs connaissances.



4. Valorisation et Reconnaissance: pour les chercheurs, une plus grande découvrabilité de leur travail, valoriser ce qui est parfois considéré comme un travail invisible (la collecte et la production de données).



5. Équité et Inclusion: réduire les inégalités en rendant les ressources de recherche accessibles à un plus grand nombre de personnes, y compris celles des pays en développement.

Quelques textes fondateurs pour le partage des données

- **OCDE (2004) Déclaration sur l'accès aux données de la recherche financée par des fonds publics.**
Recommande le partage des données pour promouvoir la recherche scientifique et l'innovation, valoriser les investissements publics, et renforcer la participation des pays en développement au système scientifique mondial.
- **UNESCO (2013): Principes directeurs pour le développement et la promotion du libre accès.**
Reconnaît l'extension du concept du libre accès aux données de la recherche.
- **FORCE 11 (2014): Joint Declaration of Data Citation Principles.** Etend aux données de la recherche les exigences concernant les citations et encourage les pratiques de citation qui sont à la fois compréhensibles pour l'homme et applicables à la machine.



Mari Helin, Unsplash

Les grandes orientations de la CE

- **Riding the Wave (2010):** Identifie les avantages et les coûts de l'accélération du développement d'une infrastructure électronique pour les données scientifiques.
- **Recommandation de la Commission sur l'accès aux informations scientifiques et à leur conservation (2012).** vise à promouvoir l'accès ouvert aux données de la recherche et leur préservation. Encourage les États membres à mettre en place des politiques et des infrastructures pour faciliter le partage des données de la recherche.
- **European Open Science Cloud: Réaliser l'agenda européen pour la science ouverte (2016).** Détaille la vision de la CE pour le Cloud européen de la science ouverte.
- **Recommandation de la Commission sur l'accès aux données de la recherche et leur préservation (2018).** Introduit des exigences plus strictes pour la gestion et le partage des données de recherche, y compris la nécessité de plans de gestion des données (DMP) et l'adoption des principes FAIR, et souligne l'importance de la préservation des données.
- **Stratégie européenne pour les données (2020):** vise à créer un espace européen des données où les données peuvent circuler librement entre les secteurs et les frontières.



Les programmes de financement Horizon 2020 & Horizon Europe



2017 - Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020

Pilote de libre accès aux données de la recherche (Open Research Data Pilot) : invite les participants à rendre gratuitement accessible les données issues de la recherche financée par la CE.

Plan de Gestion des Données (Data Management Plan) – invite les participants au pilote à établir un plan de la gestion des données tout au long du cycle de vie du projet.

2021-2027 – Horizon Europe: Le libre accès aux données et le PGD deviennent obligatoires et font partie intégrante de l'évaluation des propositions.

En plus d'être ouvertes, les participants sont incités à rendre leurs données « FAIR ».



Les plans nationaux



2017: National Programme Open Science (NPOS)



2020: Declaration for Open Science and Research



2023: Action Plan for Open Science



2018: Plan National pour la Science Ouverte



2021: National strategy for data management based on the FAIR principles



2020: Second Plan National pour la Science Ouverte

Dans certains pays, la mise en oeuvre de la politique de partage des données de la recherche se fait au niveau régional ou institutionnel.



2019: National Framework on the Transition to an Open Research Environment



2022: Italian National Plan for Open Science



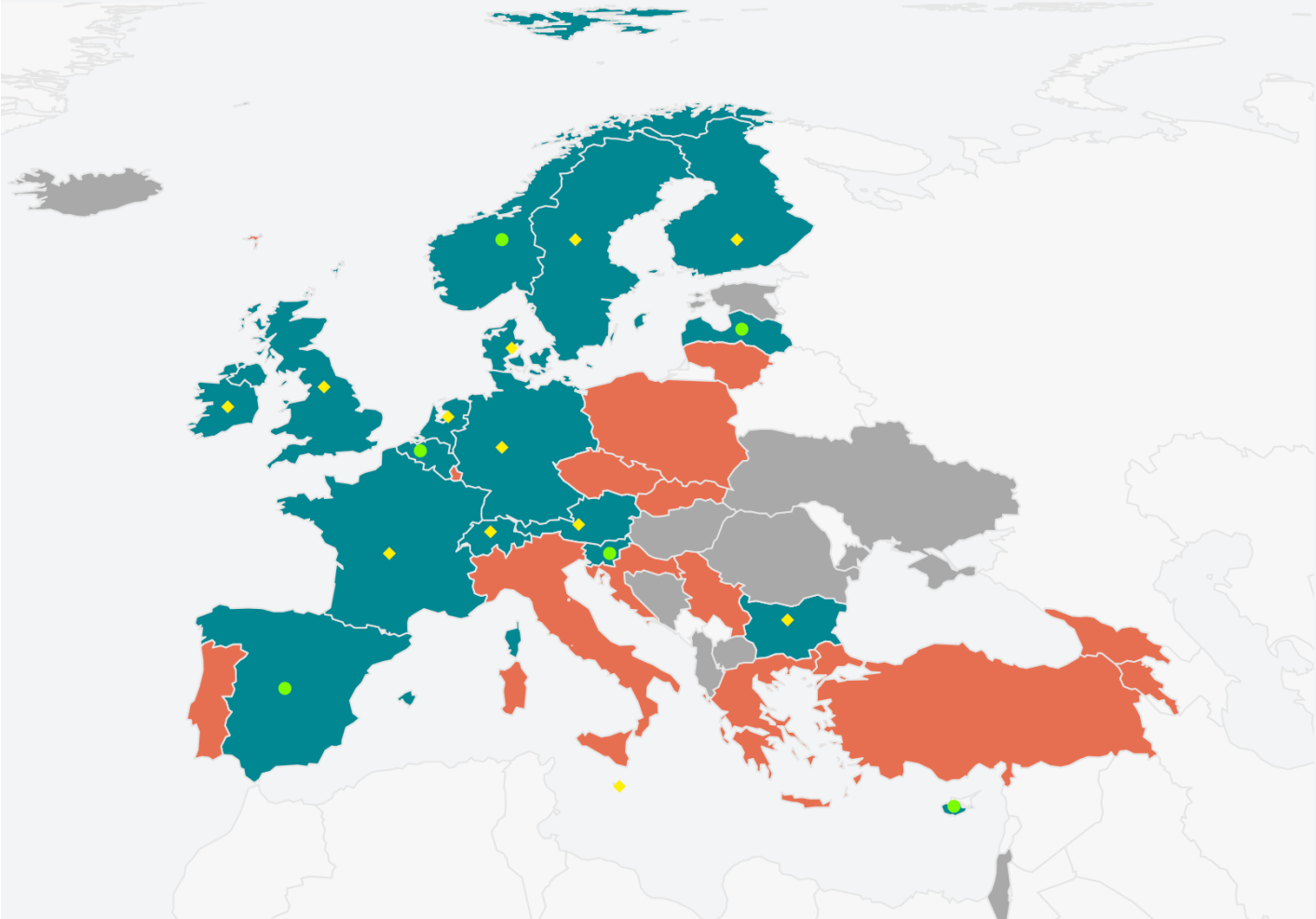
2023: National Strategy for Open Science 2023-2027

2022: National Action Plan for Open Research

National policy on FAIR data

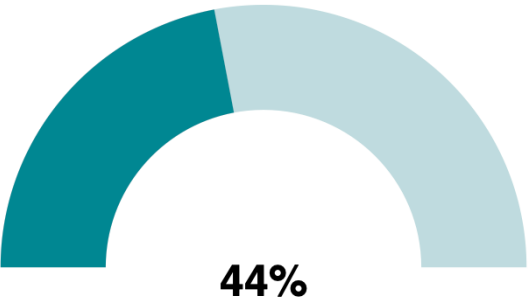


- Has national policy
- Does not have national policy
- Awaiting Data
- National policy is mandatory
- ◆ National policy is not mandatory



18 out of 41
countries have a national policy on FAIR data

Percentage of countries with a national policy on FAIR data



Des principes à la réalité

- **La proportion de chercheurs qui stocke des données dans des archives de données de recherche est encore relativement faible.** Le stockage des données dans des supports physiques (personnels et institutionnels) reste beaucoup plus répandu (57%). 40 % des chercheurs ont parfois stocké des données dans des archives de données de recherche, tandis que 22 % des répondants l'ont fait pendant l'activité de recherche en cours/la plus récente (avec quelques variations selon le type de données) (EC, 2022).
- **Une mise en œuvre du partage des données plus difficile que l'accès libre aux publications dans les universités.** Selon une étude de l'EUA, l'écart entre l'importance stratégique et la mise en œuvre est plus grand dans le partage des données que dans l'accès aux publications, avec 64 % des institutions trouvant que la mise en œuvre de l'ouverture des publications est élevée, contre seulement 16 % pour le partage des données (EUA, 2021).



Le rôle des acteurs de la recherche dans le partage des données



Mobiliser les acteurs



Facteur clé pour la mise en oeuvre des politiques

La mobilisation des acteurs de la recherche est un facteur déterminant de la mise en oeuvre des politiques de science ouverte et de partage des données.



Inclusion de tous les acteurs concernés.

Au-delà des seuls chercheurs, il est importants de convier tous les acteurs concernés (agences de financement, fournisseurs d'infrastructures et de services, etc.).

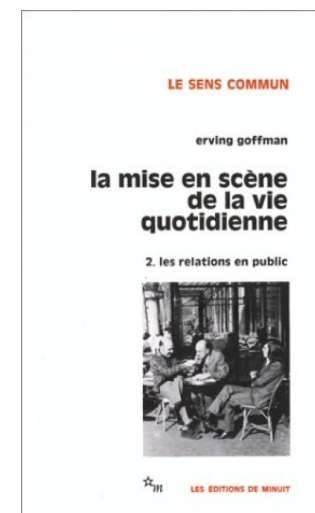
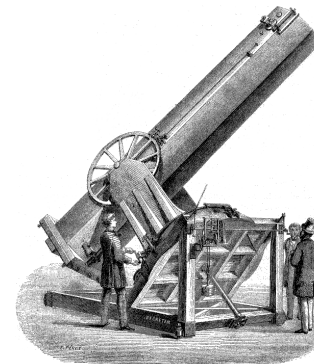
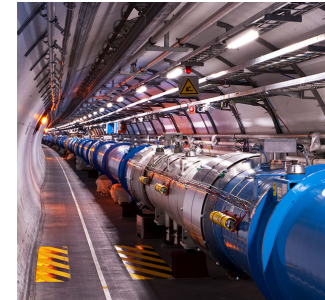


Collaboration ciblée avec les chercheurs et universitaires

Pour assurer l'acceptation et l'intégration des politiques à la pratique scientifique, une étroite collaboration avec les chercheurs est nécessaire, en tenant compte des particularités des régions et des disciplines.

Des cultures de partage différentes?

- **Facteurs historiques et collaborations internationales.** Des disciplines comme la génomique, l'astrophysique ou la physique des particules ont une culture de partage des données ancrée grâce à des collaborations qui nécessitent le partage des données pour coordonner les efforts.
- **Complexité des données dans certaines disciplines.** Dans les sciences humaines et sociales, les données sont souvent qualitatives (e.g. récits personnels, enquêtes de terrain) rendant leur standardisation et partage plus difficile.
- **Ressources limitées.** Le manque de ressources et les coûts élevés constituent des obstacles majeurs à la mise en oeuvre de systèmes de partage de données efficaces dans certaines disciplines.



La question de la gouvernance

- **Quel rôle pour l'État?** La problématique de la gouvernance renvoie aux interactions entre l'État et la société et aux modes de coordination susceptibles de rendre possible l'action publique.
- En matière de politique de science ouverte et de partage des données, certains avancent l'idée que l'État a **un rôle minimal** à jouer, celui de facilitateur pour amener au compromis ou au consensus.
- Fabriquer des politiques publiques revient alors pour l'État à **organiser un échange égalitaire entre des partenaires** aux intérêts divergents et aux logiques d'action différentes et permettre ainsi la réalisation de gains mutuels.



iStock, credit: warchi

L'exemple de la Finlande

- En 2018, le ministère de l'Éducation et de la Culture finlandais a **transféré la coordination** de la science ouverte à la communauté de recherche elle-même.
- La responsabilité de la coordination a été confiée à la Fédération des sociétés savantes finlandaises.
- Le ministère **soutient** le travail de coordination **mais ne dirige pas** son contenu.

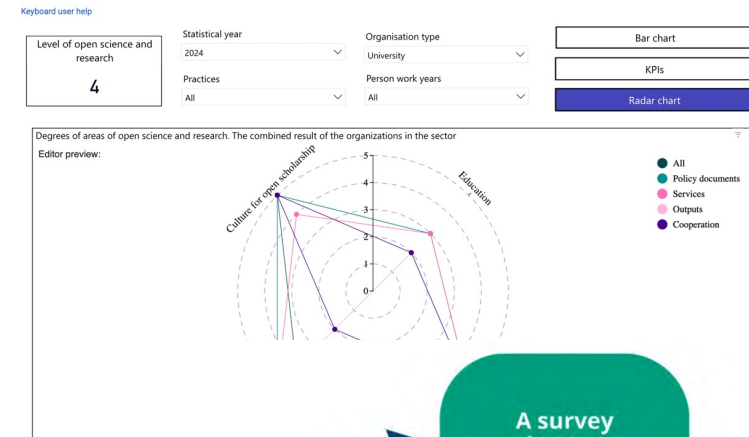
The Steering Group 2023–24

First name	Surname	Appointing organisation	Affiliation
Ella	Bingham	Finn-ARMA	Aalto University
Jaana	Erkkilä-Hill	Unifi	The University of the Arts
Tuija	Gadd	Researcher (Tulanet)	Finnish Food Authority
Minna	Harjuniemi	FUCIO and AAPA	University of Helsinki
Jari	Hämäläinen	Unifi	LUT University
Janne	Kanner	CSC	CSC
Henrik	Kunttu	Unifi	Jyväskylä University
Helena	Laaksonen	Finnish Social Science Data Archive	Finnish Social Science Data Archive
Leo	Lahti	Researcher (Unifi)	Turku University
Hanna	Lahtinen	Arene	Laurea University of Applied Sciences
Riitta	Lähdemäki	FUN	Tampere University
Riitta	Maijala	Academy of Finland	Academy of Finland
Sanna	Marttinen	Tulanet	Tulanet
Liisa	Maunuksela	Tulanet	Finnish Food Authority
Maija	Ojanen-Saloranta	Tulanet	VTT Technical Research Centre of Finland
Emilia	Palonen	TSV	Finnish Political Science Association
Taina	Pihlajaniemi	Unifi	University of Oulu
Selliina	Päällysaho	Researcher (Arene)	Seinäjoki University of Applied Sciences
Juhani	Soini	Arene	Turku University of Applied Sciences
Arja	Tuuliniemi	The National Library of Finland	The National Library of Finland

L'exemple de la Finlande

- **Le suivi national (monitoring) des politiques est organisé par la communauté de recherche finlandaise et la soutien de plusieurs manières:** Il offre une vue d'ensemble de l'état actuel des progrès de la communauté de recherche finlandaise. Il permet aux organisations d'évaluer leurs propres progrès (auto-évaluation) et de se comparer aux autres notamment par rapport aux benchmarks européens.
- **Le suivi n'est pas conçu pour être utilisé dans le modèle de financement national des universités par le Ministère de l'Éducation et de la Culture** pour l'enseignement supérieur et d'autres ministères pour les instituts de recherche. Cependant, les organisations de recherche ont la possibilité de choisir des indicateurs de science ouverte et de recherche dans leurs négociations bilatérales avec le ministère qui les finance.

Levels and degrees of openness



Le Rôle des Infrastructures



Quelles Infrastructures?

- Les infrastructures de la science ouverte désignent les infrastructures de recherche partagées qui sont nécessaires pour soutenir la science ouverte et répondre aux besoins des différentes communautés. Ces infrastructures peuvent être virtuelles ou physiques (UNESCO, 2021)
- Elles comprennent notamment:
 - les équipements scientifiques et instruments de base;
 - les collections, les revues et les plates-formes de publication en libre accès;
 - les systèmes de bibliométrie et de scientométrie pour évaluer l'impact scientifique, les réseaux de citation et la productivité académique;
 - les dépôts, les archives pour les données scientifiques;
 - les infrastructures informatiques et de manipulation des données ouvertes, etc.



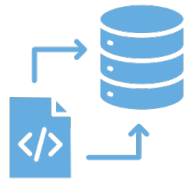
Des infrastructures pas toujours à la hauteur?

- **Conformité limitée aux exigences d'Horizon** : Une étude de l'Agence Executive du Conseil Européen de la Recherche (ERCEA) de 2021 portant sur 241 dépôts montre que de nombreux entrepôts de données ne respectent pas les exigences du programme Horizon en matière de science ouverte, notamment en matière de métadonnées (HAL est parmi les dépôts les mieux classés).
- **Conformité limitée aux exigences FAIR** : Les résultats d'une étude menée dans les pays nordiques par le projet EOSC-Nordic (2020) montrent que, pour un échantillon de 1018 objets numériques provenant de près de 100 dépôts, 53 % obtiennent un score inférieur à 0,33 (un tiers des 22 indicateurs de maturité) et 31 % un score entre 0,33 et 0,5. Ainsi, 84 % des ensembles de données passent moins de 50 % des indicateurs de conformité aux principes FAIR. Seulement 16 % des dépôts sont considérés comme "matures" (au-dessus de 0,5).



iStock

Quels échelons prioriser?



Entrepôts génériques ou disciplinaires?

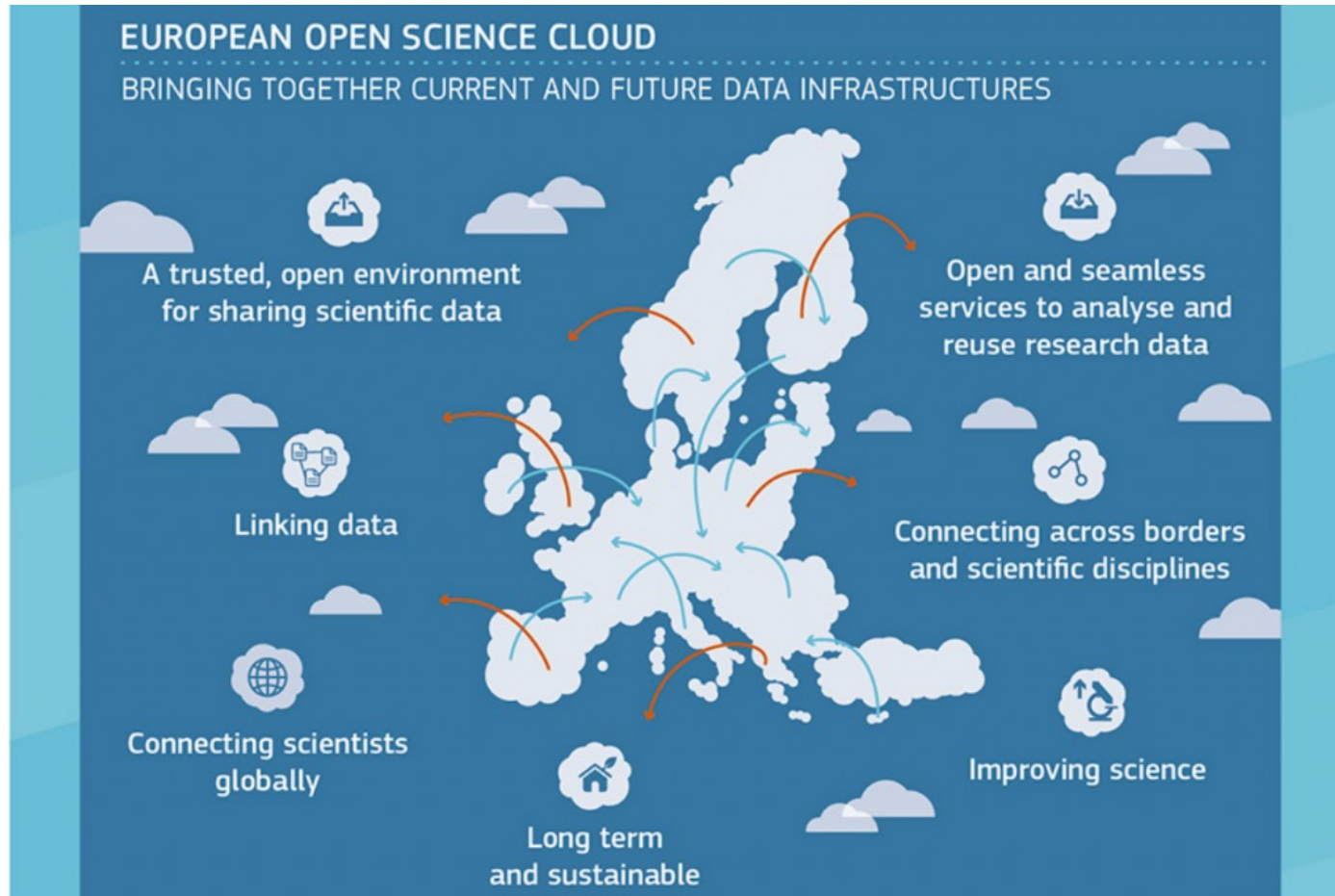
Investir dans des entrepôts génériques permet une centralisation et une standardisation, mais peut ne pas répondre aux besoins spécifiques de certaines disciplines qui ont leurs propres dépôts. D'un autre côté, financer des infrastructures disciplinaires permet de mieux adapter les solutions aux besoins de chaque domaine, mais peut entraîner une fragmentation des ressources et des efforts, ainsi qu'une duplication des coûts.



Infrastructure européenne ou nationale?

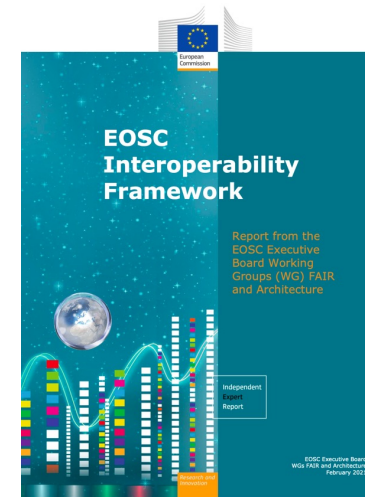
Investir dans des infrastructures européennes comme l'EOSC offre des avantages en termes de centralisation, d'économies d'échelle, d'accès élargi et de collaboration internationale, mais peut poser des défis en matière de complexité de gestion, d'adaptation aux besoins locaux et de financement.

Le Cloud Européen pour la Science Ouverte (EOSC)

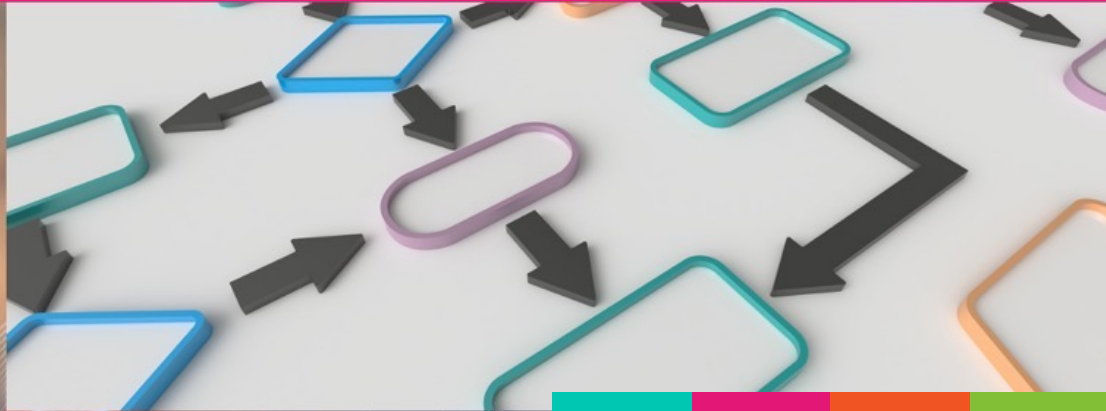


Favoriser l'accès aux ressources: la question de l'interopérabilité

- **Cadre d'interopérabilité EOSC.** Le EOSC Interoperability Framework vise à standardiser et à coordonner les infrastructures européennes, permettant une intégration harmonieuse des données de recherche à travers les disciplines et les frontières.
- **Projet Fidelis (2025-2027) pour les entrepôts de données.** FIDELIS entend développer un réseau d'entrepôts numériques de confiance en Europe, avec l'objectif de faciliter une fédération de Trustworthy Digital Repositories, et de garantir l'interopérabilité entre les dépôts.
- **Objectifs clés:** Harmonisation des normes, des API et des solutions pour faciliter l'accès aux ressources, favoriser l'interopérabilité et permettre des services de recherche fédérée inter-dépôts.



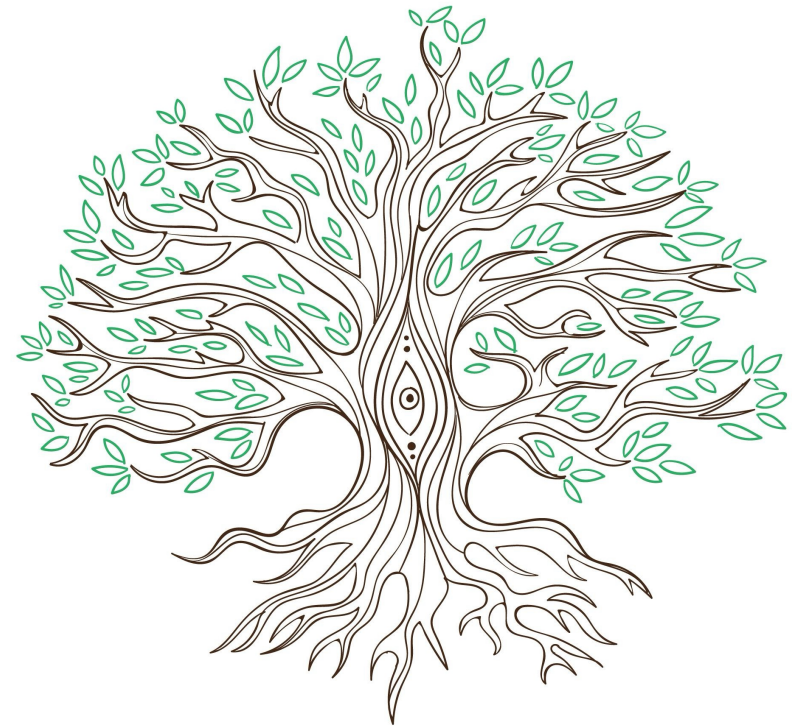
Financer le partage des données



L'importance des financements pérennes pour la science ouverte



- **Infrastructures durables.** Les infrastructures de science ouverte nécessitent des financements stables, en particulier les dépôts et les archives qui s'inscrivent nécessairement sur le long terme. Cependant les programmes public sont souvent limités à quelques années, compliquant le maintien (et parfois l'adoption) de ces infrastructures.
- **Formation des acteurs et sensibilisation.** Investir dans la formation est cruciale pour que l'ensemble des acteurs adoptent les principes de la science ouverte.
- **Collaboration internationale.** Financer les collaborations internationales et de projets de recherche à grande échelle pour favoriser l'échange de connaissances et de ressources.



Freepik



Merci !

Damien Lecarpentier
Director, International Collaborations
& Partnerships



facebook.com/CSCfi



twitter.com/CSCfi



linkedin.com/company/csc--it-center-for-science



github.com/CSCfi

Références bibliographiques

- Berkowitz, H. et Delacour, H. (2022). **Ouvrir les données de la recherche : Quelles implications pour les sciences sociales ?** *M@n@gement*, Vol. 25(4), 1-31.
- Data Citation Synthesis Group: **Joint Declaration of Data Citation Principles**. Martone M. (ed.) San Diego CA: FORCE11; 2014 <https://doi.org/10.25490/a97f-egyk>
- European Union (2010). **Riding the wave. How Europe can gain from the rising tide of scientific data**. Final report of the High level Expert Group on Scientific Data A submission to the European Commission.
- European Commission (2012). **Recommandation de la Commission sur l'accès aux informations scientifiques et à leur conservation**.
- European Commission (2016). **European Open Science Cloud: Réaliser l'agenda européen pour la science ouverte**. Publications Office, <https://data.europa.eu/doi/10.2777/940154>.
- European Commission (2018). **Recommandation de la Commission sur l'accès aux données de la recherche et leur préservation**.
- European Commission (2020). **Stratégie européenne pour les données**. Communication de la Commission Européenne.
- Jaunsen A., Kleemola M., Alaterä T., Lehvaslaiho H., Hasan A., Nordling J., & Assinen P.. (2020). **D4.1 An assessment of FAIR-uptake among regional digital repositories (1.0)**. Zenodo. <https://doi.org/10.5281/zenodo.4045402>
- Longley Arthur P., Hearn L., (2021). **Toward Open Research: A Narrative Review of the Challenges and Opportunities for Open Humanities**, *Journal of Communication*, Volume 71, Issue 5, October 2021, Pages 827–853, <https://doi.org/10.1093/joc/jqab028>

Références bibliographiques

- Morais R., Saenen B, Garbuglia F, Berghmans F, Gaillard V (2021). **From principles to practices: Open Science at Europe's universities. 2020-2021** EUA Open Science Survey results. Brussels & Geneva, European University Association.
- Lazzeri, E. (2024). **Update of the Study on the readiness of research data and literature repositories to facilitate compliance with the Open Science Horizon Europe MGA requirements (1.0)**. Zenodo. <https://doi.org/10.5281/zenodo.13919643>
- Open Science Coordination in Finland, Federation of Finnish Learned Societies (2022). **Monitoring Model for Open Science and Research. Principles and Practices**. <https://doi.org/10.23847/tsv.238>
- OCDE (2004). **Déclaration sur l'accès aux données de la recherche financée par des fonds publics**, OECD/LEGAL/0321
- OECD (2007), **OECD Principles and Guidelines for Access to Research Data from Public Funding**, OECD Publishing, Paris, <https://doi.org/10.1787/9789264034020-en-fr>.
- OECD (2015), "**Making Open Science a Reality**", *OECD Science, Technology and Industry Policy Papers*, No. 25, OECD Publishing, Paris, <https://doi.org/10.1787/5jrs2f963zs1-en>.
- Spichtinger D. (2024). **The Role of Repositories in Horizon 2020 and Horizon Europe Open Access and Data Management Requirements. A Comparative Perspective**, in *Handbuch Repositorien management*, Hg. v. Blumesberger et al., 2024, S. 353–367 <https://doi.org/10.25364/978390337423219>
- UNESCO (2013). **Principes directeurs pour le développement et la promotion du libre accès**. Publications de l'UNESCO.
- Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* **The FAIR Guiding Principles for scientific data management and stewardship**. *Sci Data* **3**, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>