SOLEIL et Science ouverte

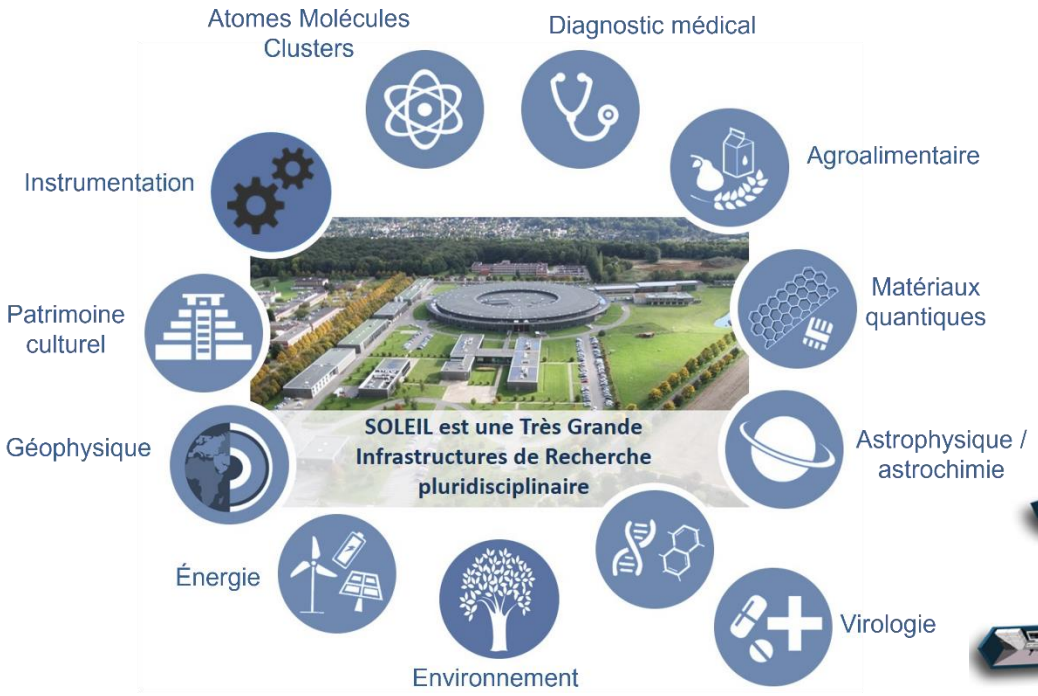
Présenté par Brigitte Gagey

Synchrotron SOLEIL - Stratégie des Systèmes d'Information et Qualité

11 mai 2022

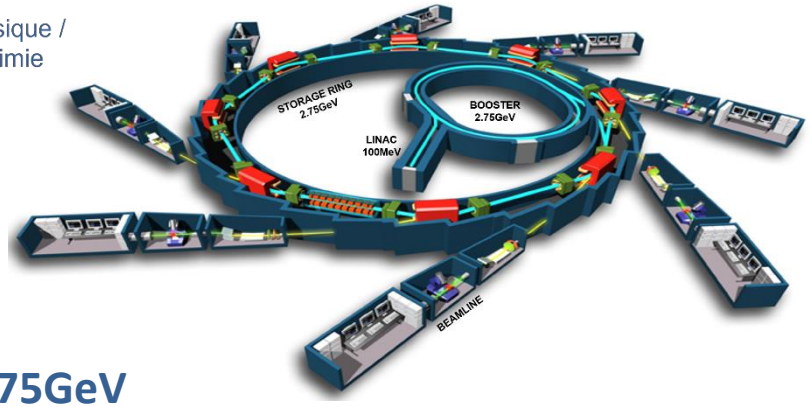
Le Synchrotron SOLEIL, une TGIR pluridisciplinaire

Un instrument de haute technologie au service de la recherche, et des grands enjeux nationaux et de l'industrie



- **Ouvert aux utilisateurs en 2008**
- **Budget total de construction 515 M€**
- **Budget annuel de fonctionnement : 63 M€**
- **~450 personnes**

72% 28%



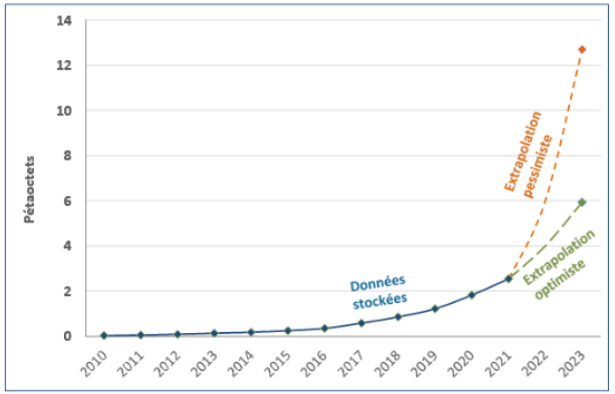
- **Anneau de stockage 354m, 2.75GeV**
- **29 Lignes de lumière couvrant 9 ordres de grandeur en énergie**

~ 700 projets d'expérience par an,

~ 3.000 utilisateurs (individus) par an, provenant de ~ 900 laboratoires.

Plus de 12.000 utilisateurs depuis 2008

4 pétaoctets de données stockés au 01/05/22



➔2021 : données stockées

2021➔ : projection basée sur une enquête auprès des scientifiques de SOLEIL

Science ouverte et les installations Photons et Neutrons : une longue histoire

pandata europe (INFRA-2010-3.3) ➔ un modèle de politique de données (https://doi.org/10.5281/zenodo.3738497) posant des principes communs, adopté progressivement en fonction du contexte de chaque installation

pandata europe | pandata ODI | SINE2020 | EUCALL | CALIPSOplus | LEAPS INNOVATION | DAPHNE 4NFDI

2010 **2015** **2018** **2021**

European Open Science Cloud (EOSC) — EOSC Future

2019	2020		2021	2022
Cluster of 6 ESFRI PaN Sources + EGI	panosc (photon and neutron open science cloud)	&	ExPaNDS (European Open Science Cloud Photon and Neutron Data Services)	10 PaN national RIs +EGI

Policies **Data Access** **Data Management Plans**

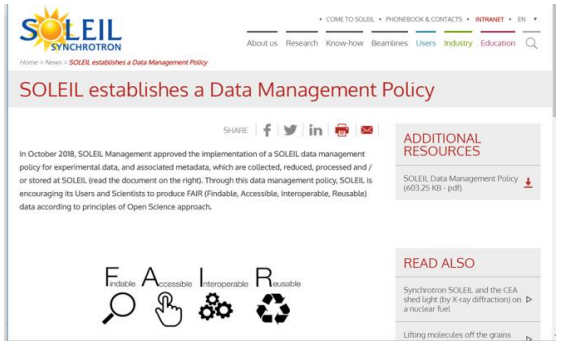
Software Catalog **Remote Analysis**

AAI **Training**

Politique des Données établie en concertation avec les Utilisateurs et Chercheurs, avec le soutien d'experts CNRS en gestion des données de la recherche. Publiée en octobre 2018

Encourager et aider Utilisateurs et Chercheurs à produire des données **FAIR**
Faciles à trouver, Accessibles, Interopérables, Réutilisables

« Des données ouvertes autant que possible, fermées autant que nécessaire »



Applicable aux données d'exp. de la recherche publique

→ 5 axes :

1. Qualification des Données
- Plan de Gestion des Données (DMP)
- Métadonnées

3. Préservation des Données
- Données brutes et métadonnées conservées jusqu'à 10 ans à terme
- Infrastructures mutualisées de stockage et de calcul

Mise en œuvre de la **politique de données** à **SOLEIL**

2. Ouverture des Données
- Standard de format de données
- Identifiant persistant unique (DOI)
- Catalogue de données en ligne
- Embargo de 3+1+1 ans
- Accès ouvert après l'embargo
- Licence CC-BY

4. Exploitation des Données
- Service d'analyse des données à distance

5. Communication

Science ouverte à SOLEIL: en appui sur des collaborations, projets, services mutualisés, …

Mise en œuvre de la **politique de données** à **SOLEIL**

1. La Qualification des Données

➔ DMP OPIDoR pour la création de DMP à partir des modèles disponibles

➔ ExPaNDS (European Open Science Cloud Photon and Neutron Data Services) métadonnées standardisées, ontologies PaN

modèle PaN Active DMP en préparation

2. L'Ouverture des Données

➔ Adoption du format NeXus / HDF5

➔ Catalogue de données permettant l'attribution de DOI

➔ API de recherche fédérée pour les catalogues de données du PaN

3. La Préservation des Données

➔ SOLEIL: 1er cas d'usage de FITS

Infrastructure répartie fédérant les services et savoir-faire des 2 centres de calcul CNRS, IDRIS et CC-IN2P3

Projet démarré mi-2021

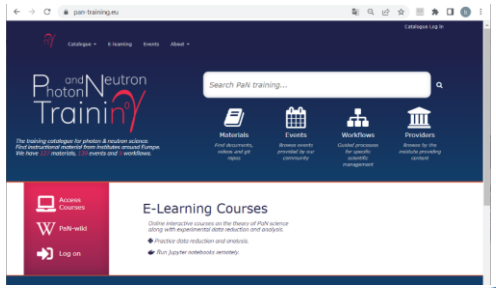
4. L'Exploitation des Données

➔ Plateforme VISA d'analyse à distance

https://bit.ly/VISA-video

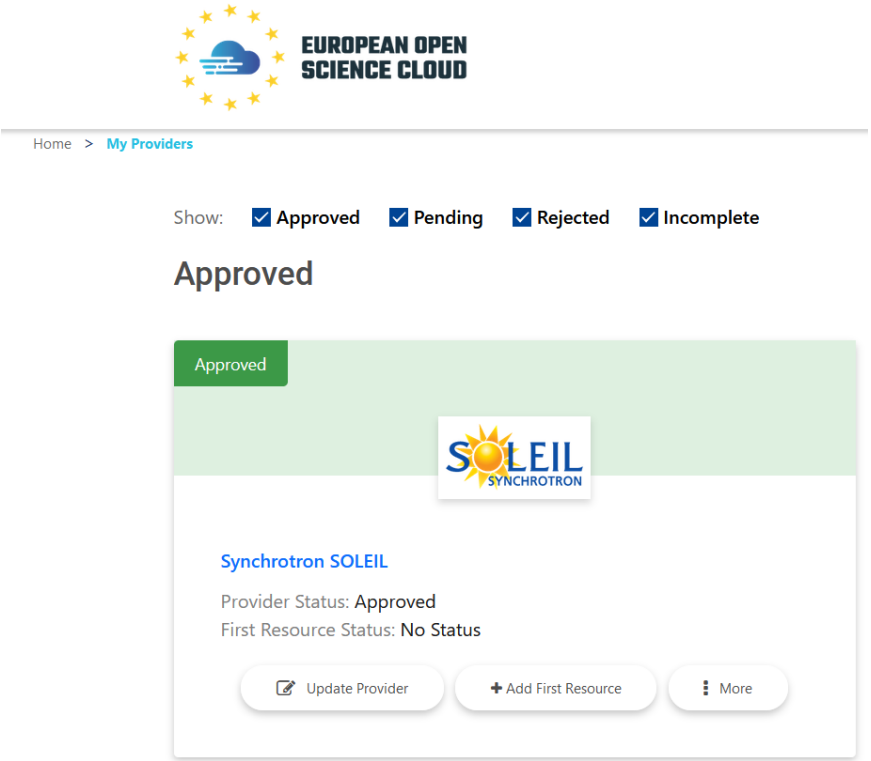
5. Communication

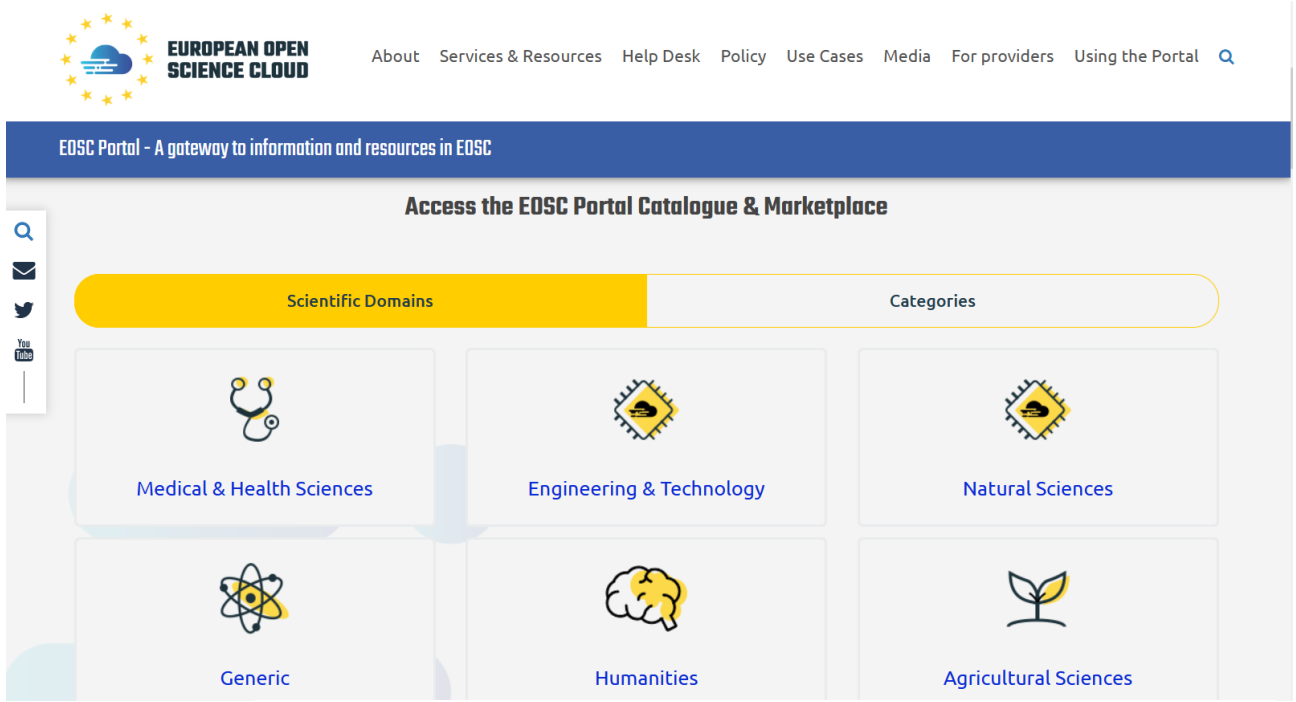
➔ Plateforme PaN-training.eu

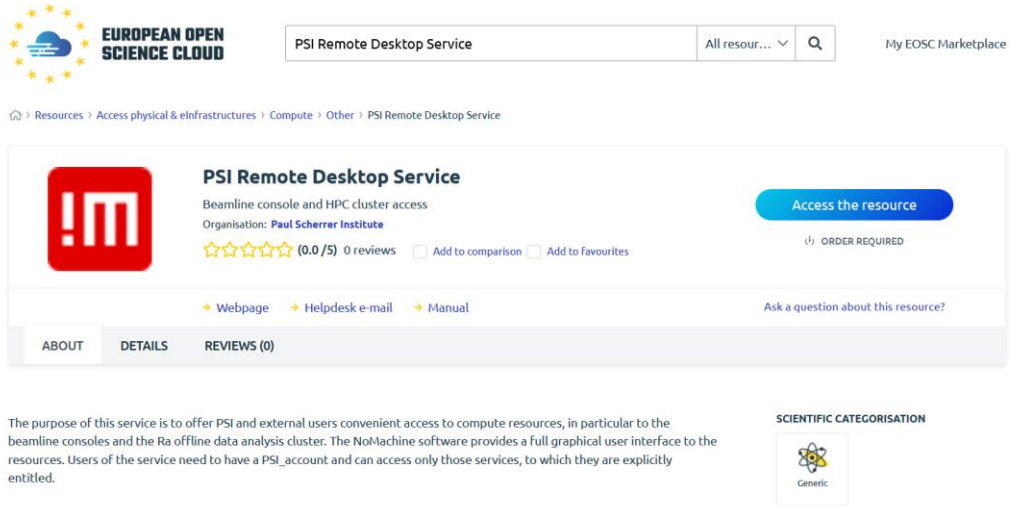


Et en continuant à s'appuyer sur l'expertise CNRS en Science Ouverte (https://www.science-ouverte.cnrs.fr/)

Science ouverte à SOLEIL: adhésion à EOSC (European Open Science Cloud) …







Portail EOSC (https://eosc-portal.eu/) :
SOLEIL enregistré comme potentiel fournisseur de services à l'exemple de PSI ci-contre

EOSC Association (https://eosc.eu/) :
SOLEIL bientôt observateur (AG de fin mai)