

Webinaire RDA France

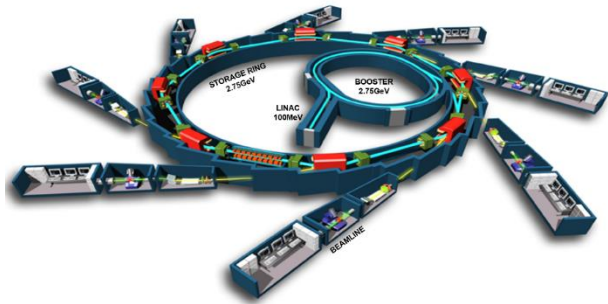
Synchrotron SOLEIL : Politique des Données

B.Gagey – Stratégie des Systèmes d'Information SOLEIL

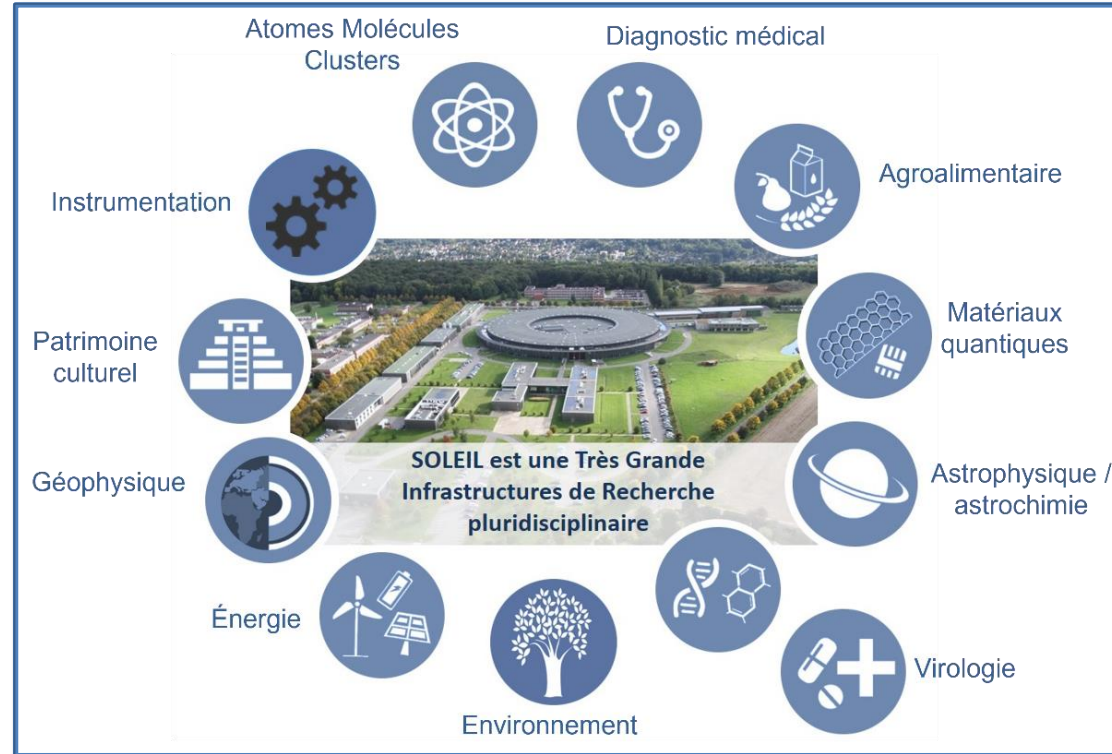
26 mai 2023



Un instrument de haute technologie au service de la recherche, et des grands enjeux nationaux et de l'industrie ouvert aux utilisateurs en 2008



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

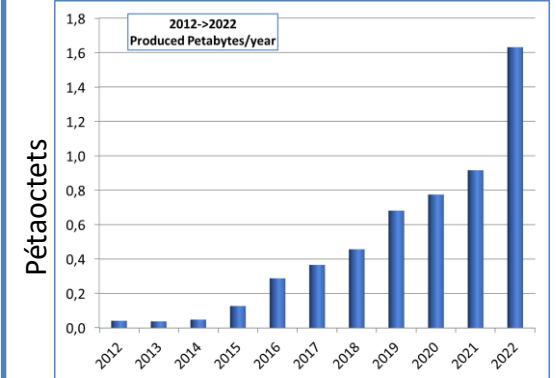


72%



28%

Données produites par les Expériences



Vers + de 6 Pétaoctets/an

Par an	~ 5.000 heures de temps faisceau Utilisateurs ~ 1.200 propositions d'expériences, < 50 % acceptées
2022	4128 visites utilisateurs sur site + 668 à distance, 704 publications

SOLEIL → SOLEIL II : en cours d'avant-projet

- Des expériences jusqu'à 10 000 fois plus rapides
- Des expériences jusqu'à 1 000 fois plus sensibles
- Résolution à l'échelle nanométrique
- Etudes en conditions réelles de fonctionnement



Février 2011 : un [modèle de politique de données scientifiques](#) posant des principes communs, élaboré par les installations photons et neutrons (PaN) d'Europe

2013 – 2015 : groupe de travail sur les données avec représentants des Utilisateurs, des scientifiques et de l'informatique SOLEIL, responsable du Bureau des Utilisateurs, juriste ; animé par DSI SOLEIL pour la Direction SOLEIL → V0

2016 – 2017 : refonte en appui sur deux experts du CNRS (F.André, MC.Jacquemot) → meilleure identification des rôles, responsabilités et droits de chacun (Utilisateur / SOLEIL) ; meilleure intégration des [principes FAIR](#)

2018 : ajustements après présentation au Comité ORGUES des Utilisateurs de SOLEIL → [Politique SOLEIL](#) validée Direction SOLEIL → présentée au [SOLEIL Users' Meeting 2019](#)

- **Données des projets d'expériences de la recherche publique conduits sur SOLEIL.** Possibilité accords spécifiques
- Responsable du Projet d'Expériences : établit PGD, autorise accès pdt embargo, ..., peut déléguer/transférer ses droits
- Equipe expérimentale : s'assure de la complétude des métadonnées

- Stockage long terme des données brutes et métadonnées
- Format HDF5/NeXus pour la préservation des données
- DOI données ; Embargo 3+1+1 ans; Licence [CC-BY](#)
- Catalogue en ligne des métadonnées : accès restreint pendant embargo (équipe expérimentale et personnes autorisées par Responsable du Projet d'Expériences) ; accès ouvert après

2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
<i>En //, adoption progressive du modèle PaN-Data par :</i>		ILL, ISIS	ELETTRA		ESRF, MAXIV	HZDR, HZB, PSI	ALBA, DESY, EuXFEL	SOLEIL	DIAMOND, FELIX

Le texte adopté reflète concrètement l'adhésion de SOLEIL à la Science Ouverte et sa volonté d'encourager la production de données FAIR tout en tenant compte des préoccupations de ses Utilisateurs

SOLEIL :
80% des publications
(Utilisateurs et staff SOLEIL)
en accès ouvert


- ✓ Démarche progressive et menée en cohérence avec les autres installations PaN → **communautés partagées d'Utilisateurs**
- ✓ Démarche qui s'est poursuivie en 2019-2023 avec les projets communs [PaNOSC](#) et [ExPaNDS](#)

→ **des outils coconstruits et prêts à l'emploi** pour mettre en œuvre et faire évoluer les politiques de données dont :

- Modèle de PGD pour les expériences sur les installations PaN,
- Format de données NeXus/HDF5 : outils de visualisation, ontologie [PaNET](#) des techniques PaN
- Services de catalogue de métadonnées, d'analyse de données à distance, ..., fédérés via EOSC
- Plateforme PaN-training : accès centralisé aux documents, événements, formations en ligne, ...
- [Modèle évolué de politique de données](#)



- ✓ Démarche qui nous prépare bien aux étapes qui suivront **une fois finalisé le déploiement de ces outils**
→ efforts à venir : PGD semi-automatique, entrepôt de données certifié, etc.
- ✓ Des échanges renforcés avec les autres installations PaN → **des MoU en préparation**
- ✓ Nous a permis d'intégrer **EOSC**, d'élargir nos contacts et relais sur la thématique FAIR dans la communauté scientifique
- ✓ Le tout dans l'objectif de servir au mieux nos Utilisateurs dans l'esprit de la Science Ouverte.

 PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

✓ *Inquiétudes des Utilisateurs et des Scientifiques SOLEIL :*

Responsabilités (et charge additionnelle) du Responsable du Projet d'Expériences : référent pour SOLEIL tout au long du cycle des données	<ul style="list-style-type: none"> → possibilité de délégation ou transfert de droits → Support SOLEIL, mise à disposition d'outils
Expériences dont les données doivent rester confidentielles	→ possibilité d'accords spécifiques
Ouverture des données avant que l'équipe expérimentale ait eu le temps de les exploiter et de publier	→ alertes à l'approche de la fin de l'embargo, possibilité de faire prolonger l'embargo, stratégie d'association DOI aux jeux de données
Réutilisation non maîtrisée	→ recommandations bonnes pratiques (citation, ...), préservation des données originales, traçabilité de l'accès au catalogue, ...

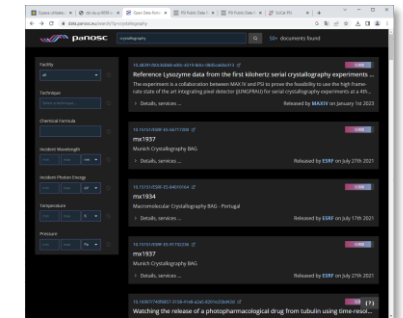
✓ *Délais et ressources :*

Construire une politique qui reflète le plus de consensus	→ Impliquer toutes les parties prenantes
Mettre en place l'environnement informatique	<ul style="list-style-type: none"> → Outils : en appui sur les projets PaNOSC et ExPaNDS → Infrastructures : SOLEIL cas d'usage FITS (CNRS/CC-IN2P3, IDRIS, GENCI)
Préciser les modalités d'application pendant et après embargo : génération DOI, curation et sécurité, conseil d'entrepôt après rétention sur stockage long terme SOLEIL, ...	<ul style="list-style-type: none"> → Comité de suivi SOLEIL avec la Direction → Tests sur lignes pilotes
Accompagner Scientifiques et Utilisateurs	→ Informer, montrer, tester avec, s'appuyer sur des convaincus

- ✓ Support de la Direction
- ✓ Efforts partagés avec d'autres entités, Retours d'expériences, Conseils experts, ...
- ✓ Promotion auprès des Utilisateurs et Scientifiques par des « convaincus »
- ✓ Démonstration de bénéfices immédiats / rapides
 - Grâce à des outils utilisables du laboratoire de l'Utilisateur ne lui nécessitant pas d'importantes ressources locales
 - ➔ Catalogue de données qui lui permet de retrouver ses jeux de données et les métadonnées associées
 - ➔ Plateforme d'analyse à distance
 - Grâce à des outils connectés à l'EOSC :
 - ➔ portail de recherche qui fédère les catalogues PaN : pe. rechercher / accéder à d'autres jeux de données pour faciliter l'analyse de ses propres données, ...



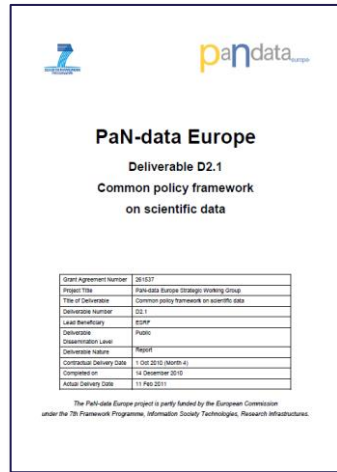
Plateforme VISA d'analyse à distance
<https://bit.ly/VISA-video>



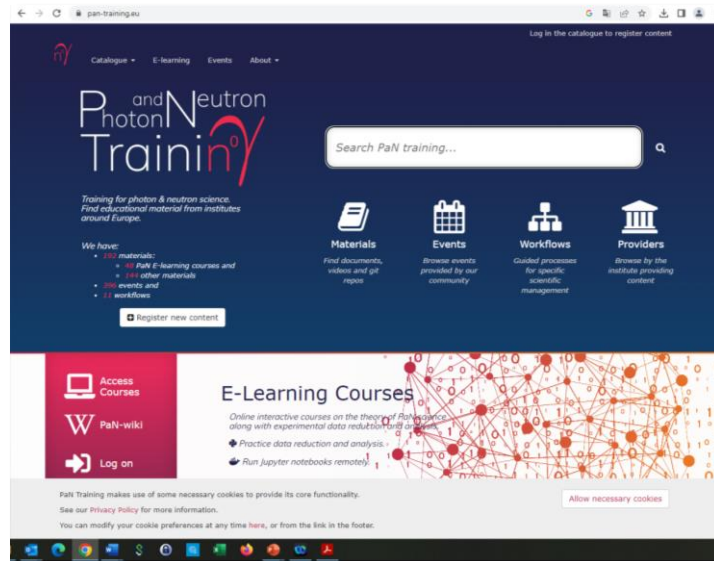
Portail fédéré de recherche
<https://data.panosc.eu/search>

✓ et mettre en place un processus d'évaluation et de revue de la politique : à maintenir dans la durée

Modèle PaN-Data Europe



Politique de Gestion des données de SOLEIL



Plateforme PaN-training

Livrables
dont :

- PaNOSC D2.2 : [DMP Template for facility users](#)
- ExPaNDS D2.3 : [Final data policy framework for Photon and Neutron Ris](#)
- ExPaNDS D2.7 : [Final recommendations for FAIR Photon and Neutron Data Management](#)

Merci pour votre attention !