

Ouverture des données : l'exemple du synchrotron SOLEIL

La Science Ouverte
« une révolution nécessaire »

J. Daillant

Auditorium Marie Curie,
CNRS, 08/10/2019





72 %



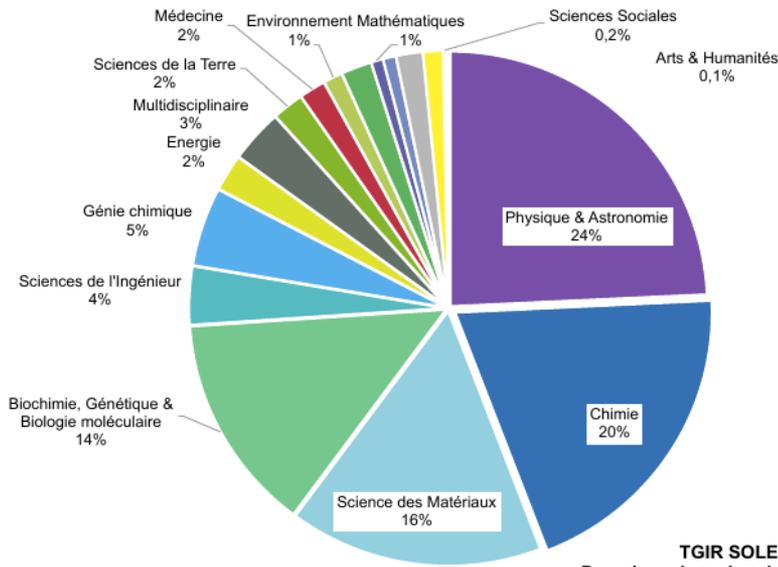
28 %

29 expériences

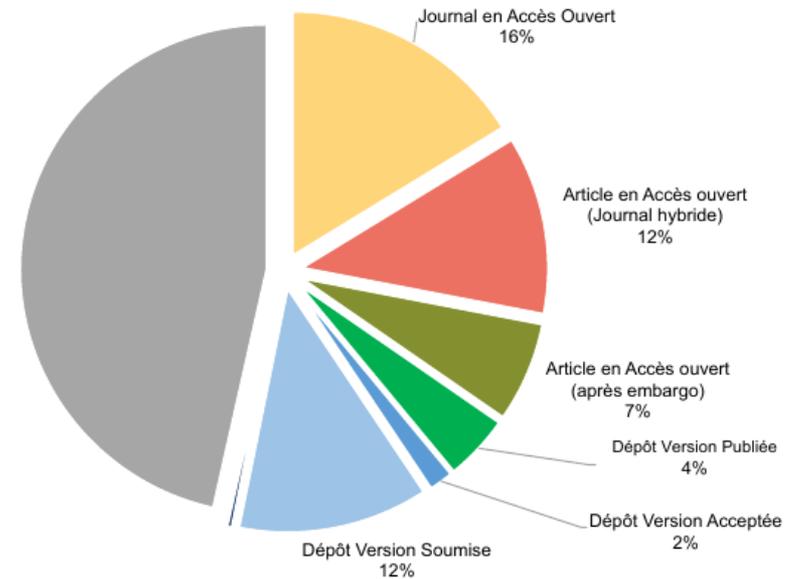
~ 700 projets acceptés par an

~ 2500 utilisateurs par an

~ 900 laboratoires



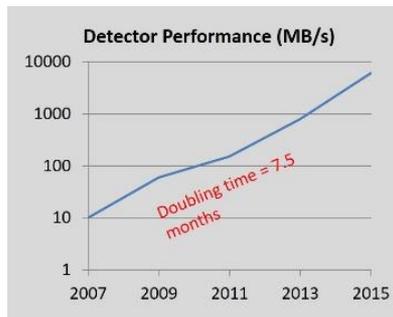
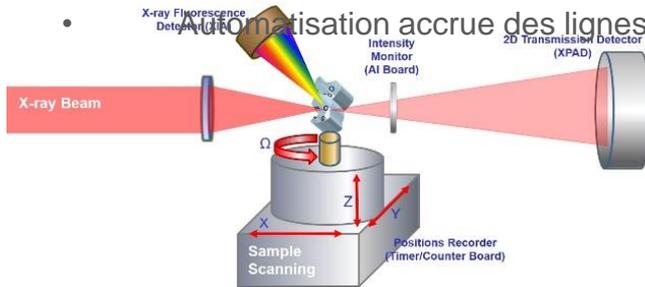
TGIR SOLEIL
Domaines de recherche (2013-2018)
 Source : Publications TGIR - SciVal - ASJC - 3 400 articles



→ Ouverture des données, plan national pour la science ouverte

Vers une explosion de la production de données issues des expériences :

- Acquisition Flyscan : multi-technique / multi-détecteurs, rapide, synchrone
- Détecteurs à haut débit et haut volume
- Automatisation accrue des lignes



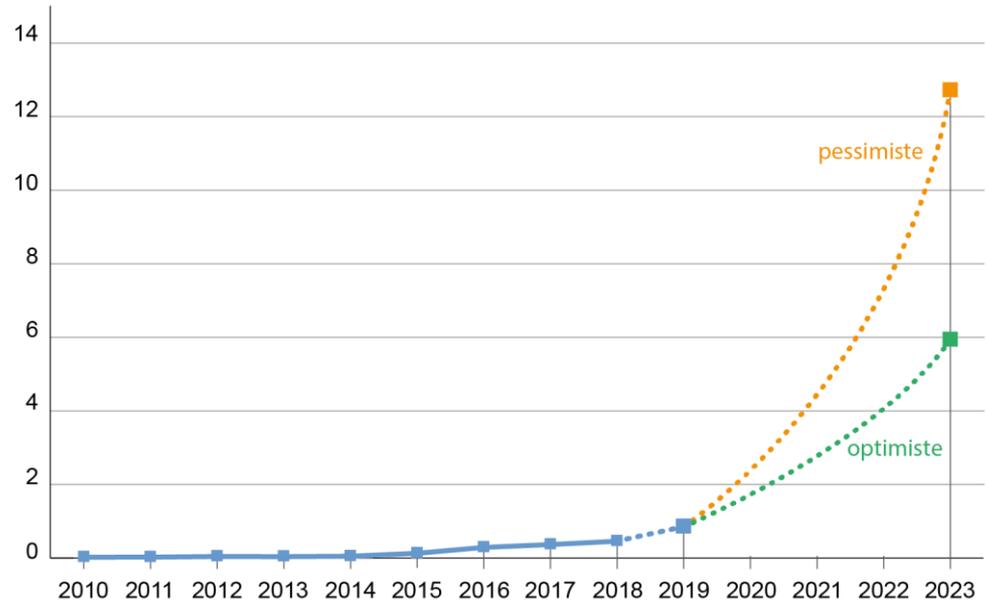
Données brutes : production annuelle passée et à venir

2010-2018 Données stockées.

2019 Projection basée sur les données stockées lors des premiers essais en 2019.

2022-2023 Projection basée sur les réponses des scientifiques des lignes de lumière à une enquête.

Pétaoctets



Politique des données expérimentales à SOLEIL (Octobre 2018) :
Encourager et aider Utilisateurs et Chercheurs à produire des données
Faciles à trouver, Accessibles, Interopérables, Réutilisables (principes FAIR)
→ Nombreux échanges avec le CNRS (en particulier F. André et nos utilisateurs)

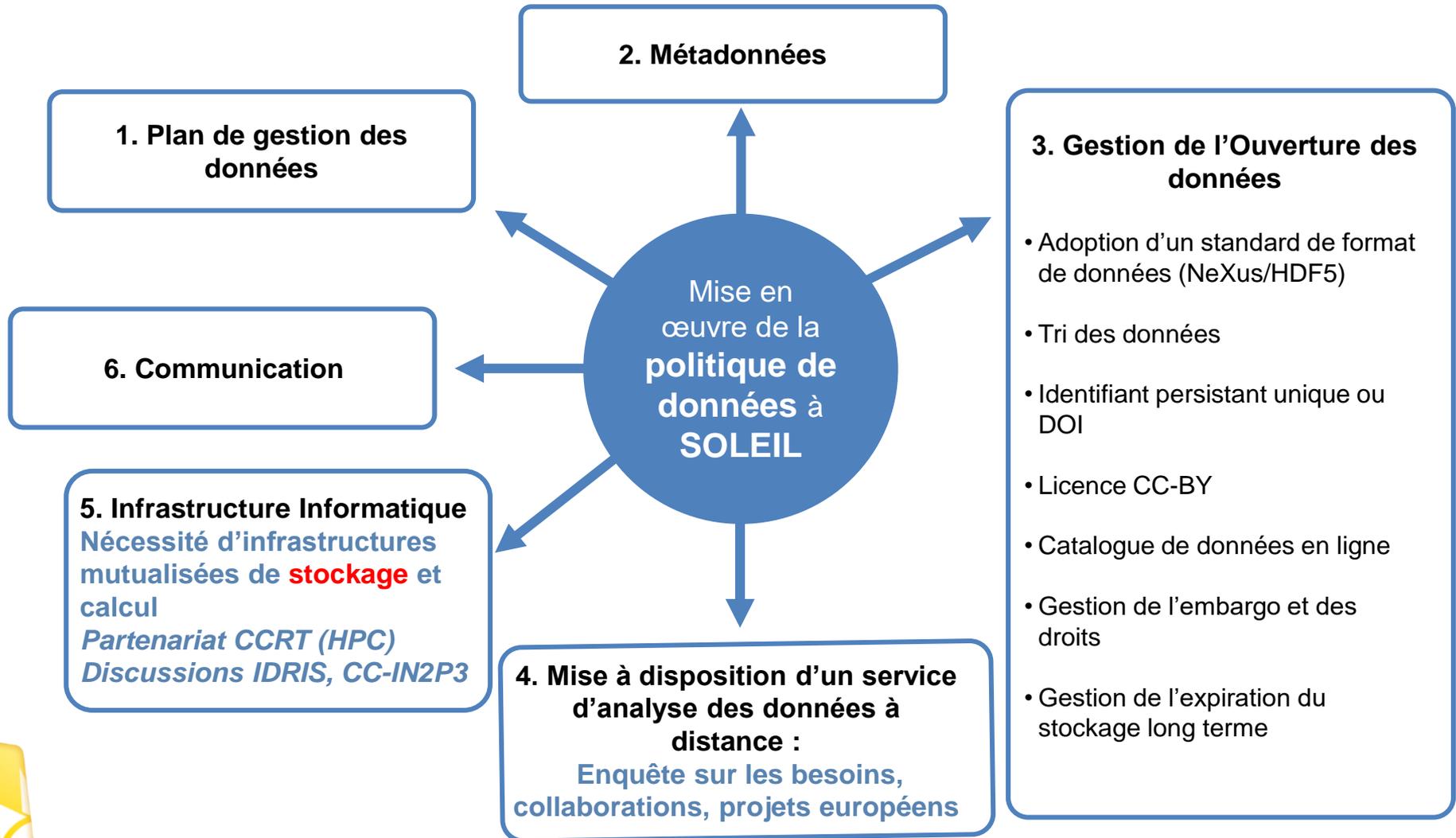
- Applicable aux données de la recherche publique
- Préservation des données brutes et métadonnées jusqu'à 5 ans (SOLEIL visant jusqu'à 10 ans à terme)
- dans des formats standards (HDF5/NeXus)
- avec un identifiant unique pérenne, DOI
- Accès via un catalogue en ligne
- Embargo de 3+1+1 ans; accès ouvert après l'embargo
- Licence [CC-BY](#) pour les données préservées



The screenshot shows a news article on the SOLEIL website. The article title is "SOLEIL establishes a Data Management Policy". Below the title, there are social media sharing icons for Facebook, Twitter, LinkedIn, and Email. The article text states: "In October 2018, SOLEIL Management approved the implementation of a SOLEIL data management policy for experimental data, and associated metadata, which are collected, reduced, processed and / or stored at SOLEIL (read the document on the right). Through this data management policy, SOLEIL is encouraging its Users and Scientists to produce FAIR (Findable, Accessible, Interoperable, Reusable) data according to principles of Open Science approach." To the right of the text, there is a section titled "ADDITIONAL RESOURCES" with a link to "SOLEIL Data Management Policy (603.25 KB - pdf)". Below the article text, there are icons representing the FAIR principles: Findable (magnifying glass), Accessible (hand cursor), Interoperable (gears), and Reusable (recycling symbol). To the right, there is a "READ ALSO" section with a link to "Synchrotron SOLEIL and the CEAS synchrotron shed light (by X-ray diffraction) on a nuclear fuel".

L'équipe expérimentale doit établir un plan de gestion des données (DMP) et s'assurer de la complétude des métadonnées.

SOLEIL s'engage à fournir des moyens pour la capture automatique des métadonnées, pour la réduction et le traitement des données brutes.



Collaboration envisagée avec l'Institut Laue Langevin (Grenoble) et les infrastructures analytiques de l'Institut de Chimie (IR-RMN, Renard, FT-ICR – S. Lecocq) :

- Pratiques communes de gestion des identifiants permanents (DOIs)
- Modèles harmonisés de plans de gestion de données
- Définition de bonnes pratiques pour la définition des métadonnées.

Techniques considérées : SAXS, SANS, RMN, RPE, spectro de masse, tomographie.



Cristallographie macromoléculaire

> 150.000 structures dans la « Protein Data Bank »

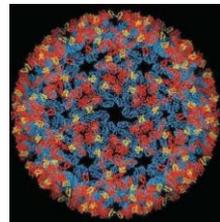
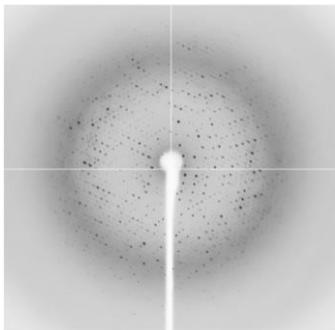
Communauté avancée dans l'ouverture des données

Collecte de données automatisée (robotisation) en quelques secondes.

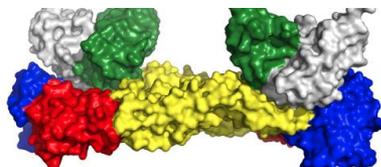
Jusqu'à 1.5To de données par 24h pour une ligne de lumière.

10 minutes de collecte des meilleurs jeux de données / 24h.

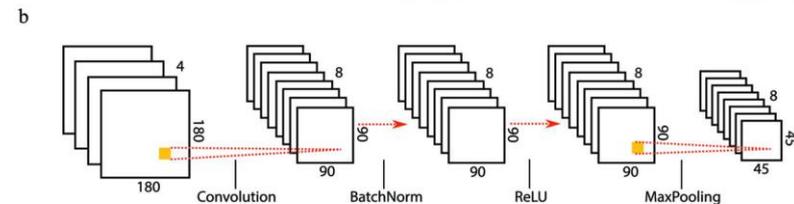
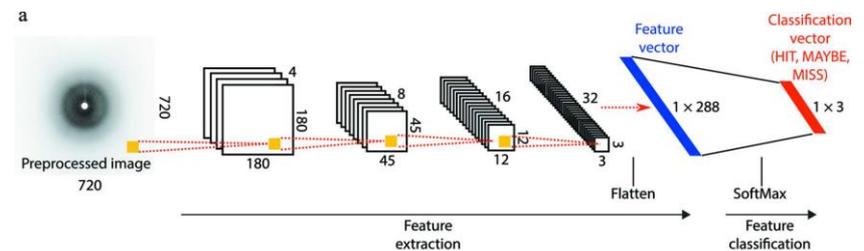
→ Tri des données à l'aide de réseaux de neurones convolutifs en apprentissage profond (coll. G. Evans, DLS and M. Sauter, Stanford).



J.-E. Voss et al.,
Nature 468, **709**
(2010)



G. Barba Spaeth et al., Nature **536** 48 (2016)



Ke et al., *J.S.R.*, **25**, 655–670 (2018)

Favoriser la diffusion des principes Science Ouverte et l'utilisation des services EOSC dans les Infrastructures Photon et Neutron

PANOSC : Photon and Neutron Open Science Cloud
ESFRI infrastructures ; SOLEIL observateur.

EXPANDS : EOSC Photon and Neutron Data Services
Infrastructures nationales

Services de données compatibles FAIR cohérents entre les infrastructures



JRA2 - Démonstrateur pour un service DaaS d'analyse des données Photon Science

- **Data Citation Working Group :**

Recommandations pour la publication des données, spécifications pour l'implémentations des DOIs.

- **Active Data Management Plan Interest Group:**

Interfaces logicielles pour l'intégration digitalisée des DMPs avec les systèmes d'information des IRs

- **DMP common standards Working Group :**

Définir un modèle de DMP constitué d'un ensemble d'éléments de base et une architecture modulaire permettant des extensions spécifiques selon la discipline scientifique.

- **Physical Samples and Collections in the Research Data Ecosystem Interest Group:**

Associer des PIDs uniques aux échantillons, lien publications, échantillons, métadonnées standards



Merci de votre
attention

