



La science ouverte: une perspective INP

Laurent Lellouch (CMI Calcul, données et science ouverte)

16 novembre 2020

CNRS – Institut de Physique

16.11E20

Science ouverte ?

Une perspective INP

Science ouverte: dans la mesure du possible, rendre accessible publiquement tous les produits du cycle de la recherche financée par des fonds publics

Objectifs:

- Rendre le processus scientifique plus efficace en facilitant la réutilisation de résultats et d'outils
- Permettre l'émergence de nouvelles formes de « méta-sciences » interdisciplinaires ou non
- Assurer la fiabilité des résultats publiés en rendant accessible toutes les étapes du processus
- Rendre différents résultats intermédiaires ou outils directement utilisables et citables, notamment pour une évaluation plus complète des chercheurs
- Réduire les coûts de publication et de lecture d'articles
- Ouvrir la science aux citoyens et la rendre plus accessible aux étudiants
- Mieux valoriser la diversité des produits de la recherche

Science ouverte ?

Une perspective INP

Qu'est-ce que cela concerne?

- Les données primaires d'expériences et/ou de simulations
- Les protocoles utilisés pour obtenir les données expérimentales (descriptions de l'échantillon, de l'instrument de mesure, des conditions et protocoles des mesures)
→ cahier de laboratoire électronique
- Et/ou les logiciels de simulation
- Les « workflows », les méthodes et les outils d'analyse
- Possiblement des données intermédiaires
- Les pré-tirages
- Les versions éditées des publications

Les données de la recherche

Une perspective INP

- De plus en plus de données: résolution plus fine, temps réel, 3 dimensions, etc.
- Relative homogénéité du type de données: données de mesure ou images numériques \neq SHS
- Des pratiques très hétérogènes:
du disque dur dans un tiroir \rightarrow hébergement cloud de données FAIRisée
(Findable, Accessible, Interoperable, Reusable)

\rightarrow demande au CSI de nous aider à répertorier ces pratiques

« Nos » bons élèves: les TGIRs PaN (Photon and Neutron), i.e. les synchrotrons SOLEIL, ESRF, XFEL et les sources à neutrons ILL, ESS

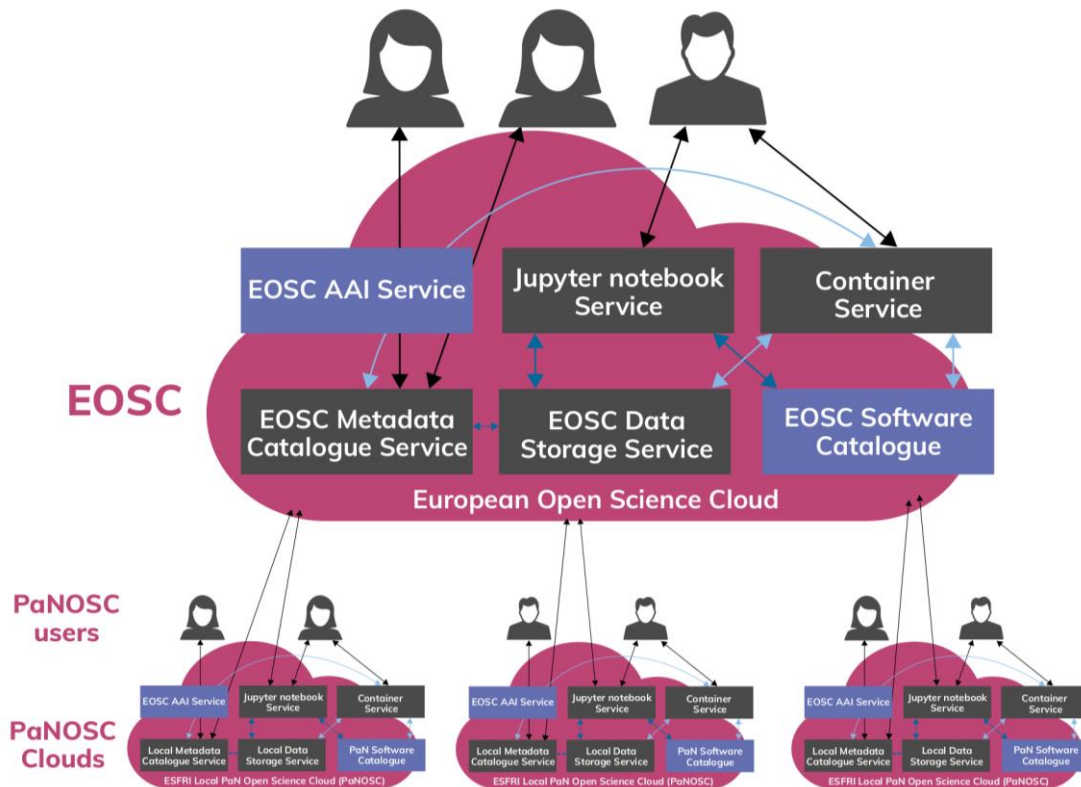
PaNOSC

PaNOSC (ESRF, ILL, ...
INFRA-EOSC-04-2018):

- FAIRiser les données de 6 IRs européen
- Développer/déployer des services pour les analyser et les partager dans un cloud PaN
- Interfacer le tout avec EOSC

ExPaNDS (Soleil, ... INFRA-EOSC-05-2019):

- Déclinaison de PaNOSC pour 11 IRs nationaux



Les données de la recherche

Stratégie INP v1

TGIRs PaN: expertise très utile car ils ont l'habitude de gérer des données de communautés très différentes

Stratégie: (avec E. Lacaze et soutien de D. Veynante) utiliser expertise et outils développés (sans pénaliser ceux concernés) pour disséminer les bonnes pratiques dans différents domaines de la physique



- Projet RD4E1 (Research Data for Everyone) (coord. S. Lecocq, INC)
- Mise en contact acteurs Equipex TEMPOS (TEM et croissance cristalline en temps réel) ↔ ILL (avec N. Keller)
- A faire: aider réseau METSA (TEM et sondes atomiques) à mettre en place une politique et infrastructure des données (avec E. Lacaze, B. Devindre)

Les données de la recherche

Stratégie INP v1

Projet RD4E1: Research data for everyone (coord. S. Lecocq, INC)

- Partenaires: INC, comité TGIR-CNRS, Soleil, ILL, IRs de l'INC, INP
- Attribution systématique et automatique de DOIs (échantillons, données, etc.)
- Création de DMPs modulaires pour répondre aux spécificités d'une discipline
- Métadonnées d'échantillons spécifiques à une discipline, mais indépendantes de l'instrumentation
- Très bonne dynamique avec transfert de connaissances TGIRs ↔ IRs INC

Discussions sur les données avec lignes de lumières ESRF françaises (CRG) et IDRIS, CC-IN2P3 autour des projets EquipEx+ MAGNIFIX et IDRIS/CC-IN2P3

Participation
active
à
CatOpiDoR



Accueil
À propos
Modifications récentes
Naviguer par
Type de service
Stade du cycle de vie
Domaine scientifique
Service
Structure d'appartenance
Contribuer
Ajouter un service

Catégorie Discussion

Catégorie:Annuaire CNRS-INP

Page en construction

Cette page liste les dépôts de données et les services dont le CNRS-INP est responsable et auxquels il participe.

Pages dans la catégorie « Annuaire CNRS-INP »

Cette catégorie comprend 10 pages, dont les 10 ci-dessous.

D

- Data.esrf.fr

E

- EELS Data Base
- ESRF

• ESS

I

- ILL Data Portal
- Institut Laue Langevin

S

- SOLEIL
- Synchrotron ESRF
- Synchrotron SOLEIL

X

- XFEL

Les données de la recherche

Stratégie INP v2

Acquérir une meilleure compréhension de la diversité des pratiques à l'INP, avec le CSI, la DDOR, l'INIST

→ puis voir comment la stratégie décrite plus haut peut être adaptée pour aider à développer les bonnes pratiques dans l'ensemble des domaines de l'INP

Encourager une participation française aux grandes initiatives Européennes
e.g. FAIR-DI e.V.: FAIR Data Infrastructure for Physics, Chemistry, Material Science and Astronomy



Computational
Materials Science -
NOMAD Laboratory



Experimental
Materials Science



Soft-matter &
Biomolecular
Simulations



Heterogeneous
Catalysis



Astronomy &
Space-situational
Awareness



Artificial-
intelligence Tools



Digital Research
Infrastructures

Participations principalement allemande, néerlandaise, belge et espagnole

FAIRmat Workshop
18-20 Nov, 2020 - register now!

Les logiciels

Une perspective INP

Recensement des codes produits par les laboratoires INP: (avec A. Lambrecht, B. Devincré, N. Dimarcq)

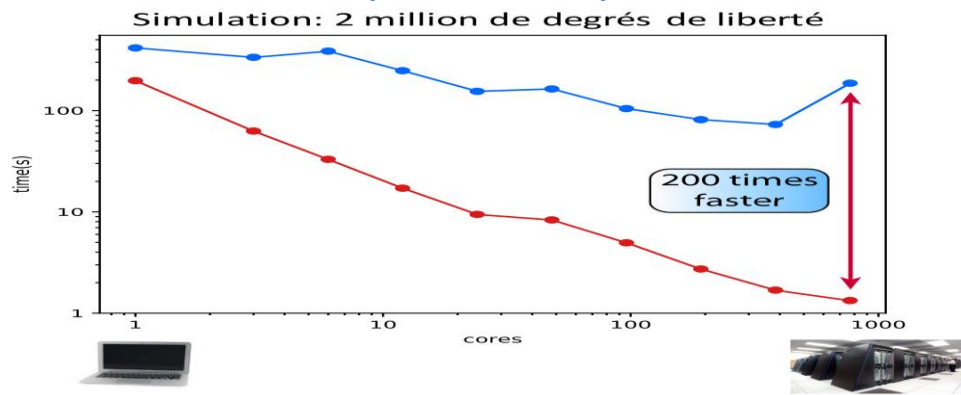
- Codes importants pour les activités de recherche des laboratoires
- 84 codes recensés dans 21 laboratoires
- Hétérogénéité des licences (sans, GNU, GPL, protégés, ...)
- Parfois rendus disponibles au sein de communautés, ou non partagés
- Modes de distribution hétérogènes (sur demande, sur sites web personnels, ad hoc ou communautaires, jamais, ...)
- Modes de citation très divers: article publiés qui présente le code, nom du code, ...
- Comment: Assurer l'évolution du code? Leur préservation et celle du savoir faire? Les professionnaliser si nécessaire? Les rendre plus largement disponibles? ...
- Valorisation? (cf. ci-dessous)
- Il serait intéressant de travailler avec d'autres instituts sur ces sujets (me contacter)

Les logiciels

Une perspective INP

Groupe de travail sur la valorisation des logiciels: (dir. J.-M. Schmittbiel, CNRS Innovation)

- Participation des 10 instituts
- Echanges sur les expériences de valorisation, les différents types de licence, les modèles économiques, les besoins en personnel, ...
- Etudes de cas concrets
- INP: pour l'instant, un logiciel de simulation d'ordinateurs quantiques en prématuration: réduire significativement le temps de conception de ces ordinateurs
→ recherche de solution d'hébergement, de sécurisation technique et juridique, de modèle économique



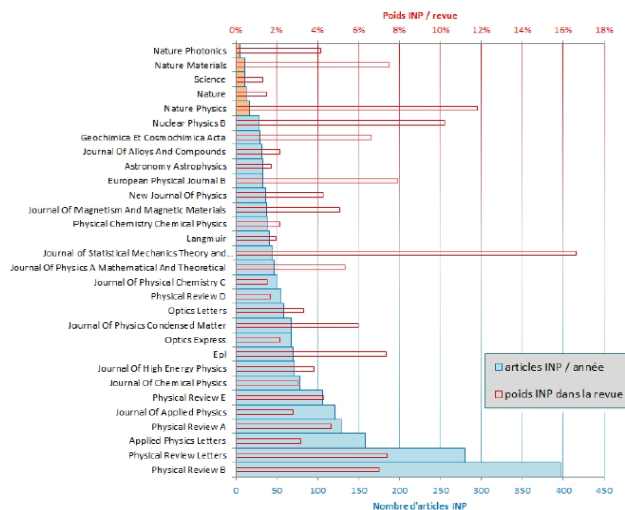
Les publications

Une perspective INP

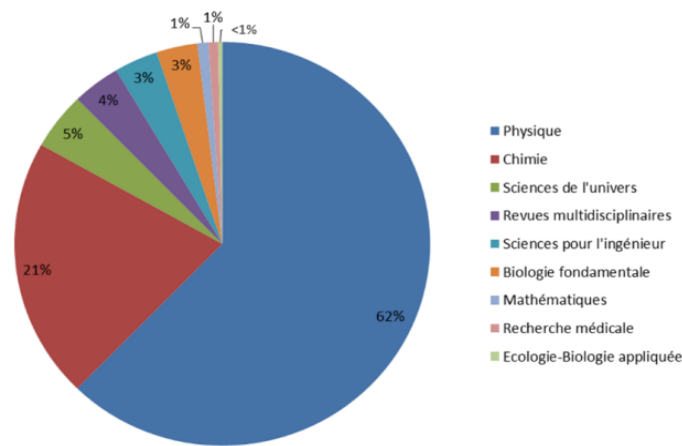
Une tradition du OA qui remonte à la nuit des temps (en physique des hautes énergies): années 60, prétirages; version électronique ~1991 avec arXiv (hep-th, hep-lat, hep-ph); SPIRES → INSPIRE (outil de bibliographie HEP) ~1970; www @ CERN, 1991; etc.

Gros travail de M. Knoop et B. van Tiggelen

- Etude des pratiques (avec INIST): hétérogénéité



Domaines de publication de l'INP



Les publications

Une perspective INP

- Guide « Open Access pour les nuls » avec la SFP (<https://www.sfpnet.fr/l-open-access-pour-les-nuls>)
- Présentation annuelle sur l'OA aux Journées des nouveaux entrants en physique (JePhy)
- M. Knoop fait partie du collège « Publication » du Comité pour la science ouverte (CoSO) du MESRI
- Mise à jour en cours avec le CSI

Participation de l'INP au travers de la DIST à



SCOAP³ – Sponsoring Consortium for
Open Access Publishing in Particle Physics

SCOAP3

- Porté par le CERN
- CNRS représente la France au CA (S. Hohenegger, IN2P3)
- Concerne les publications arXiv hep-ph, hep-lat, hep-th et hep-ex
- Participation d'instituts de recherche de plus de 44 pays
- S'occupe de collecter les fonds pour financer les publication du domaine en OA au prorata des publications
- Négocie avec les éditeurs pour obtenir les meilleurs conditions possibles (Elsevier, Springer, APS, IOP, Hindawi, ...)
- Tous les articles en OA
- Complètement transparent pour les chercheurs
- > 35.000 articles disponibles et ~ 360 €/article payé directement par les instituts
- Projet pilote: SCOAP3 pour les livres
- Reconduit jusqu'en 2022

Dans quelle mesure cet exemple vertueux peut-il être répliqué dans d'autre domaines?

Conclusion

- Le chantier de la SO est immense
- La SO requiert une très grande diversité de compétences allant d'une connaissance pointue du monde de l'édition à l'utilisation de l'IA pour la fouille de données
- Un institut ne peut pas surmonter ces défis seul: il faut travailler ensemble
 - l'INP est preneur de toutes les bonnes idées et est prêt à collaborer activement sur ces sujets
 - la nouvelle Direction des données ouvertes de la recherche (DDoR) a un rôle très important à jouer
 - encourager les communautés à s'engager, à prendre des initiatives et trouver des moyens de les aider à le faire
 - encourager les initiatives personnelles, notamment celles des nouvelles recrues (PRIMEs pour projets de science ouverte?)
 - la SO doit être présentée aux nouveaux entrants comme faisant partie intégrante de ce que l'on attend d'eux
- Contactez-moi si vous avez des suggestions, commentaires, questions