Données Ouvertes de la Physique des 2 Infinis (DOP2I)

Antoine Lemasson
GANIL

XV^{ème} Journées Informatiques IN2P3/IRFU

Données == Objets Numériques

Objets numériques

- Ensembles de données expérimentales (données brutes, auxiliaires, traitées, données des publications, ...)
- Simulations
- Résultats des calculs
- Bases de Données
- Logiciels (code source, Workflows, ...)
- Rapports, publications, diaporamas, sites web,
- Photos...

Données == Objets Numériques

Objets numériques

- Ensembles de données expérimentales (données brutes, auxiliaires, traitées, données des publications, ...)
- Simulations
- Résultats des calculs
- Bases de Données
- Logiciels (code source, Workflows, ...)
- Rapports, publications, diaporamas, sites web,
- Photos...

Qui suivent un cycle de vie ...

Données == Objets Numériques

Objets numériques

- Ensembles de données expérimentales (données brutes, auxiliaires, traitées, données des publications, ...)
- Simulations
- Résultats des calculs
- Bases de Données
- Logiciels (code source, Workflows, ...)
- Rapports, publications, diaporamas, sites web,
- Photos...

Les données sont le fondement du processus de recherche scientifique

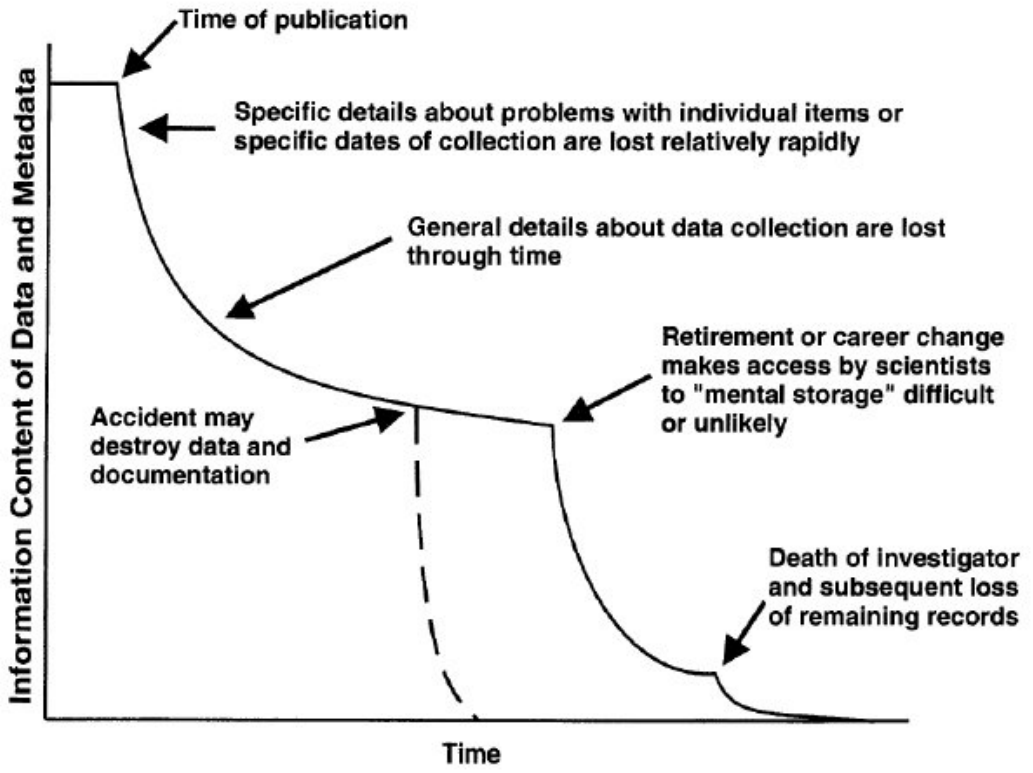
Qui suivent un cycle de vie ...

Pourquoi se soucier de la gestion des données (ouvertes ou non !) ?

Aucun ensemble de données n'est parfait et explicite

- S'appuye trop souvent sur le « stockage » humain/mental
- Crucial pour interpréter avec précision les résultats et leur origine (à partir du traitement, de l'analyse et de la modélisation)
- Accessibilité et reproductibilité des résultats de recherche
- Améliorer la visibilité de la recherche à l'intérieur et à l'extérieur du domaine de recherche

Data and Metadata Entropy



W. K. Michener et al., Eco. App. 7 (1997) 330-342

Pourquoi se soucier de la gestion des données (ouvertes ou non !) ?

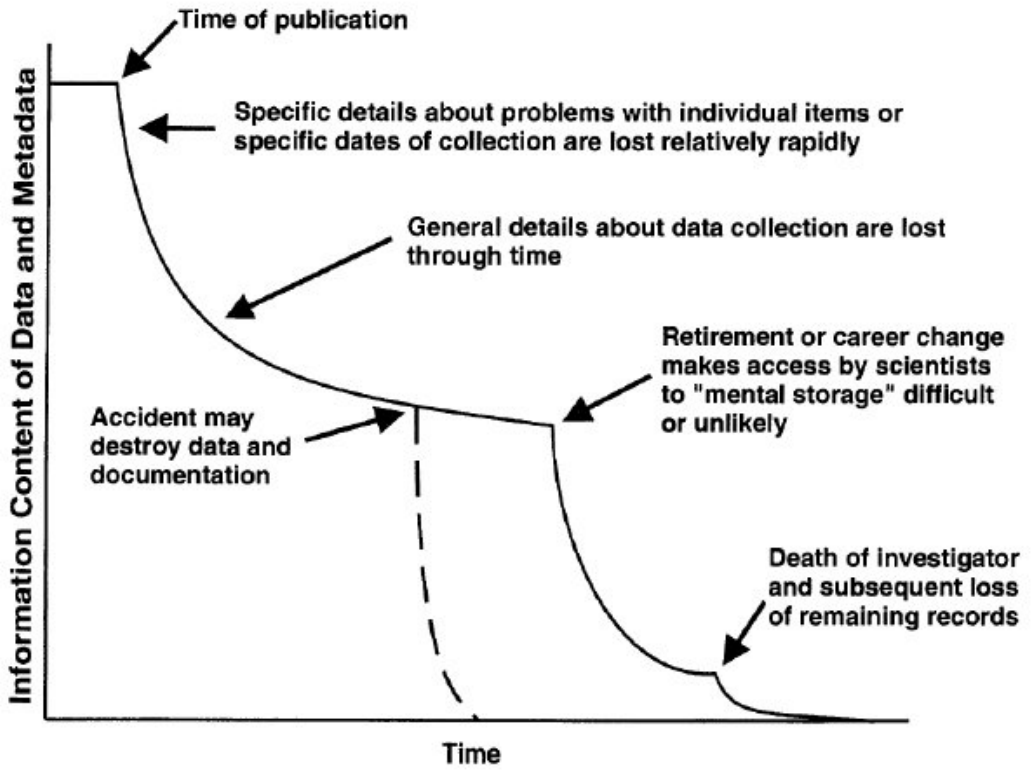
Aucun ensemble de données n'est parfait et explicite

- S'appuye trop souvent sur le « stockage » humain/mental
- Crucial pour interpréter avec précision les résultats et leur origine (à partir du traitement, de l'analyse et de la modélisation)
- Accessibilité et reproductibilité des résultats de recherche
- Améliorer la visibilité de la recherche à l'intérieur et à l'extérieur du domaine de recherche

Préservation et gestion à long terme

- Définition de politiques de données (accès, partage, conservation, réutilisation, …)
- opportunités de réutilisation, faciliter la recherche inter-domaines
- Comment choisir si un ensemble de données doit être conservé (Le stockage illimité est derrière nous !)

Data and Metadata Entropy



W. K. Michener et al., Eco. App. 7 (1997) 330-342

Les données dans la Science Ouverte

Les ambitions de la science ouverte

- Changer la perception des citoyens sur la recherche et l'investissement public dans la recherche
- Accélérer les découvertes et augmenter la valeur scientifique par le partage et le transfert de connaissances avec les communautés scientifiques
- Contribuer à la formation de la relève scientifique
- Saisir les opportunités offertes par la révolution numérique pour permettre à chacun de contribuer au processus scientifique.

Les données sont le produit précieux de l'activité scientifique.
Dans le processus scientifique, bien en amont des résultats et des publications.

Données Ouvertes, Données "FAIR"

- Doivent suivre les principes "FAIR"
  (Findable, Accessible, Interoperable, Reusable)
- Données Ouvertes, mais pas gratuites !
- **Aussi ouvertes que possible, aussi fermées que nécessaire**

Ouvrir les données, facile ?

quelques serveurs et c'est parti ...

- Mettre les données à disposition
  c'est un est excellent point départ …

  mais c'est loin d'être suffisant !
- Si seul le code et/ou les données sont mises à disposition, l' utilité pour les utilisateurs en dehors des auteurs est très faible (voire nulle ?).

  -> des millions de lignes de code, des peta-octects de données
- Les experiences sont complexes et les codes et procedures d'analyses sont compliquées.
- Pour que ces données soient utiles, elles doivent être « accessibles » et « explicites » pour être comprises et utilisées par la communauté scientifique

Perspective | Open Access | Published: 15 November 2018

Open is not enough

Xiaoli Chen, Sünje Dallmeier-Tiessen ✉, Robin Dasler, Sebastian Feger, Pamfilos Fokianos, Jose Benito Gonzalez, Harri Hirvonsalo, Dinos Kousidis, Artemis Lavasa, Salvatore Mele, Diego Rodriguez Rodriguez, Tibor Šimko ✉, Tim Smith, Ana Trisovic ✉, Anna Trzcinska, Ioannis Tsanaktsidis, Markus Zimmermann, Kyle Cranmer, Lukas Heinrich, Gordon Watts, Michael Hildreth, Lara Lloret Iglesias, Kati Lassila-Perini & Sebastian Neubert

Nature Physics **15**, 113–119(2019) | Cite this article

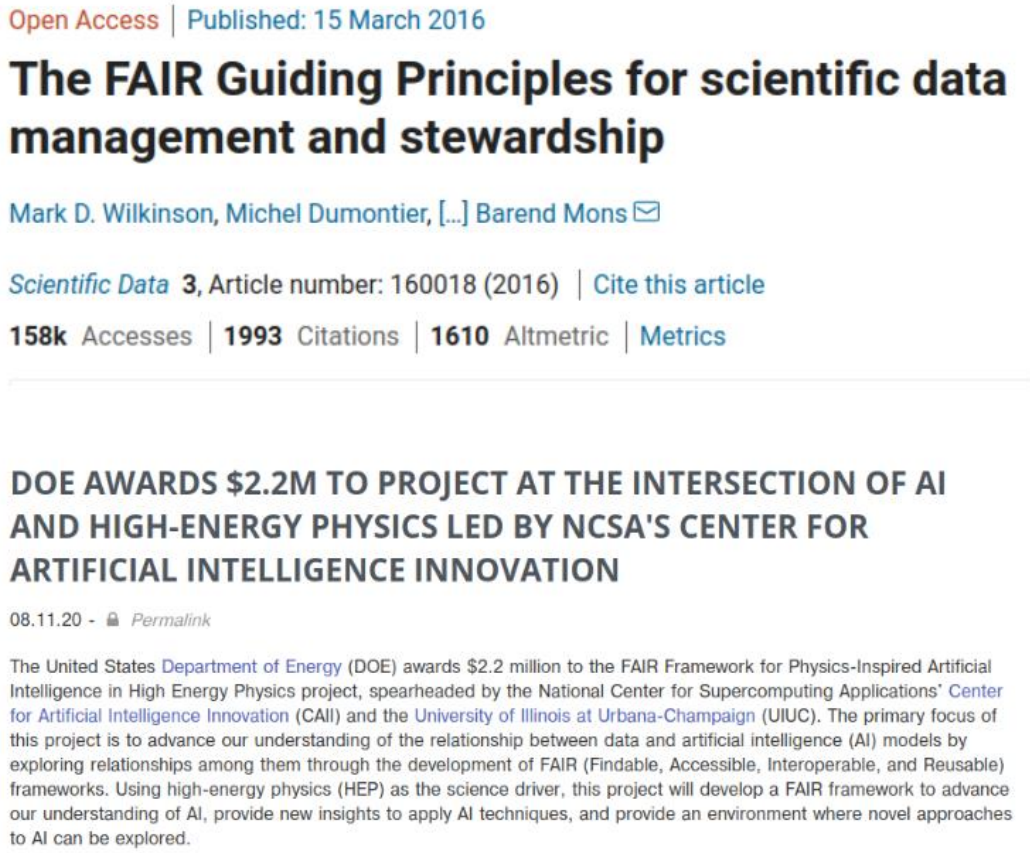
Challenging, but possible

In this paper we have discussed how open sharing enables certain types of data and software reuse, arguing that simple compliance with openness is not sufficient to foster reuse and reproducibility in particle physics. Sharing data is not enough; it is also essential to capture the structured information about the research data analysis workflows and processes to ensure the usability and longevity of results.

Rendre les données « FAIR »

Pour que les données soient utiles, elle doivent être “FAIR”

- **Findable**: Metadata and data should be easy to find for both humans and computers.
  - Unique persistent identifier (digital object identifier (DOI)) and rich metadata
- **Accessible:** Retrievable by identifier using a standardized communications protocol
  - HTTPS, public APIs
- **Interoperable:** Interoperate with applications or workflows for analysis, storage, and processing
  - Schemas and serialization
  - Formal, shared, broadly applicable
- **Reuseable:** Well-described so can be replicated and/or combined in different settings
  - Annotated metadata (Codemeta JSON-LD)



Open Access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, [...] Barend Mons

Scientific Data **3**, Article number: 160018 (2016) | Cite this article

158k Accesses | **1993** Citations | **1610** Altmetric | Metrics

DOE AWARDS $2.2M TO PROJECT AT THE INTERSECTION OF AI AND HIGH-ENERGY PHYSICS LED BY NCSA'S CENTER FOR ARTIFICIAL INTELLIGENCE INNOVATION

08.11.20 - *Permalink*

The United States Department of Energy (DOE) awards $2.2 million to the FAIR Framework for Physics-Inspired Artificial Intelligence in High Energy Physics project, spearheaded by the National Center for Supercomputing Applications' Center for Artificial Intelligence Innovation (CAII) and the University of Illinois at Urbana-Champaign (UIUC). The primary focus of this project is to advance our understanding of the relationship between data and artificial intelligence (AI) models by exploring relationships among them through the development of FAIR (Findable, Accessible, Interoperable, and Reusable) frameworks. Using high-energy physics (HEP) as the science driver, this project will develop a FAIR framework to advance our understanding of AI, provide new insights to apply AI techniques, and provide an environment where novel approaches to AI can be explored.

Large interest in FAIR data across all levels

Se soucier de la longévité des données et de la science

Changement de mentalité (également dans notre travail quotidien) :

- passer plus de temps à familiariser les nouveaux membres aux analyses (passées et en cours)
- se demander si notre prochain doctorant/collègue peut s'appuyer sur le travail actuel
- s'assurer la vérification et la validation de nos codes avant d'effectuer de nouvelles analyses.
- discuter avec les collaborateurs si nos résultats et les leurs sont conservés et réutilisables à long terme.
- Penser à publier le code, les données et des notes sur les analyses au moins dans des dépôts internes.

Changement d'organisation :

- passer du temps au début d'un projet à planifier les données produites par l'expérience, comment elles peuvent être organisées non seulement pour les collaborateurs, mais aussi pour le partage (pendant le fonctionnement et à la fin du projet)
- besoin d'un nouveau type de physicien qui ne soit ni un « expérimentateur » ou un «théoricien », mais un « data physicist » qui a une formation en physique, mais aussi une formation en statistiques/ Science des données /Machine Learning et en informatique scientifique.
- la nécessité d'envisager de trajectoires de carrière pour les physiciens et ingénieurs des données.

Faire progresser la gestion des données de la recherche vers la science ouverte

Gestion et préservation des données

- Principes de données « FAIR »
  - Plan de Gestion des Données
    => G. Debaecker (Mardi)
  - Catalogues et Entrepôts de données
    => M. Dubois (Mardi)
- Environnements logiciels standardisés
- Gestion des accès aux données
  - Infrastructure d'autorisation d'authentification
    => C. L'orphelin (Mercredi)
  - Plateformes de données (DataLakes, …)
    => **F. Gillardo (Mardi)**
- Mise en place de collaborations au delà des thématiques scientifiques (ESCAPE, …) pour le développement de solution communes aux problématiques techniques.

Qualité des données / Reproductibilité

- Collecte améliorée et automatisée de métadonnées (physique, détection, accélérateur)
  => AMI (F. Lambert/J. Odier) Mercredi
- Gestion des Logiciels (suivi de version, déploiement, catalogues, …)
  => T. Vuillaume (Mardi)
- Environnements logiciels standardisés (Containeur, Workflows, …)
  => P. Davis (in 20 min !) + C. Santos

Plateformes d'analyse

- Pour les scientifiques (analyse collaborative, organisations autour d'un sujet scientifique…)
  => D. Savchenko (mardi)
- Pour la formation de la prochaine génération de scientifiques

Des problématiques aux interfaces entre la physique, l'informatique, les infrastructures et l'institut

Les données et l'IN2P3

TEXTES GÉNÉRAUX

MINISTÈRE DE L'ÉDUCATION NATIONALE, DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE

Arrêté du 29 avril 2016 relatif à l'Institut national de physique nucléaire et de physique des particules du Centre national de la recherche scientifique

NOR : *MENR1611333A*

Arrête :

Art. 1er. – L'Institut national de physique nucléaire et de physique des particules du Centre national de la recherche scientifique exerce les missions nationales d'animation et de coordination dans les domaines de la physique nucléaire, de la physique des particules et des astroparticules, des développements technologiques et des applications associées, notamment dans le champ de la santé et de l'énergie, en ce compris la radiochimie.

Pour la réalisation de ces missions, l'Institut national de physique nucléaire et de physique des particules :

– conçoit, coordonne et anime des programmes de recherche nationaux et internationaux dans ses domaines de compétence ;
– organise et conduit, en y associant les organismes et acteurs concernés, des exercices de prospective nationale permettant de définir la stratégie scientifique de long terme et d'identifier les équipements nationaux et internationaux nécessaires à sa mise en œuvre. Il veille à la plus large diffusion des résultats de ces travaux et favorise leur prise en compte dans l'élaboration des programmes de recherche et d'équipement à l'échelle nationale et internationale ;
– favorise et coordonne la participation des opérateurs de recherche aux structures d'intérêt national ainsi qu'aux très grandes infrastructures de recherche et aux programmes scientifiques qu'elles permettent de réaliser ;
– coordonne la mise en place de systèmes d'information permettant le stockage, la mise à disposition auprès de la communauté scientifique, le traitement et la valorisation de l'ensemble des données scientifiques concernées, ainsi que leur archivage.

Données (ouvertes) à l'IN2P3

Thématiques scientifiques

- Physiques des Hautes Energies
- Physique Nucléaire
- Astroparticules/Cosmologie
- Physique Médicale
- Physique des Accélérateurs

...

Infrastructures : IR*/IR

- CERN / LHC
- DUNE
- EGO/VIRGO
- GANIL
- AGATA
- KM3NET
- LSST
- ...

Plateformes
Collaborations
Centre de Calcul et Données
CC-IN2P3

Grande variété de pratiques de gestion des données
Infrastructures <-> Collaborations <-> Equipes

Physique des Hautes Energies

S’appuye sur de larges collaborations structurées, principalement autour du CERN.

Très tôt, la question de la préservation des données et des environnements d’analyse a été discuté (2008)

DPHEP : Collaboration for Data Preservation and Long Term Analysis in High Energy Physics

➔ Roadmap mise à jour plusieurs fois depuis 2008; la dernière en date 2023

Eur. Phys. J. C (2023) 83:795
https://doi.org/10.1140/epjc/s10052-023-11885-1

THE EUROPEAN PHYSICAL JOURNAL C

Review

Data preservation in high energy physics

DPHEP Collaboration

T. Basaglia[1], M. Bellis[2,b], J. Blomer[1], J. Boyd[1], C. Bozzi[3], D. Britzger[4], S. Campana[1], C. Cartaro[5], G. Chen[6], B. Couturier[1], G. David[7,c], C. Diaconu[8,a], A. Dobrin[9], D. Duellmann[1], M. Ebert[10], P. Elmer[11], J. Fernandes[1], L. Fields[21], P. Fokianos[1], G. Ganis[1], A. Geiser[12], M. Gheata[9], J. B. Gonzalez Lopez[1], T. Hara[13], L. Heinrich[1], M. Hildreth[21], K. Herner[14], B. Jayatilaka[14], M. Kado[1], O. Keeble[1], A. Kohls[1], K. Naim[1], C. Lange[20], K. Lassila-Perini[15], S. Levonian[12], M. Maggi[22], Z. Marshall[18], P. Mato Vila[1], A. Mečionis[1], A. Morris[17], S. Piano[16], M. Potekhin[7], M. Schröder[1], U. Schwickerath[1], E. Sexton-Kennedy[14], T. Šimko[1], T. Smith[1], D. South[12], A. Verbytskyi[4], M. Vidal[1], A. Vivace[1], L. Wang[6], G. Watt[19], T. Wenaus[7]

[1] CERN, Geneva, Switzerland
[2] Cornell University, Ithaca, USA
[3] INFN Ferrara, Ferrara, Italy
[4] Max-Planck-Institut für Physik, Munich, Germany
[5] SLAC National Accelerator Laboratory, Menlo Park, USA
[6] Institute of High Energy Physics, IHEP, CAS, Beijing, China
[7] Brookhaven National Laboratory, BNL, Upton, USA
[8] Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France
[9] Institute of Space Science, ISS, Bucharest, Măgurele, Romania
[10] HEP Research Computing, University of Victoria, Victoria, BC, Canada
[11] Princeton University, Princeton, USA
[12] Deutsches Elektronen Synchrotron, DESY, Hamburg, Germany
[13] High Energy Accelerator Research Organization, KEK, Tsukuba, Japan
[14] Fermi National Accelerator Laboratory, Batavia, USA
[15] Helsinki Institute of Physics, Helsinki, Finland
[16] INFN Trieste, Trieste, Italy
[17] University of Bonn, Bonn, Germany
[18] Lawrence Berkeley National Laboratory, Berkeley, USA
[19] IPPP, Durham University, Durham, UK
[20] Paul Scherrer Institut, Villigen, Switzerland
[21] University of Notre Dame, Notre Dame, USA
[22] INFN Bari, Bari, Italy

Received: 18 April 2023 / Accepted: 20 July 2023 / Published online: 8 September 2023

Physique des Hautes Energies

S'appuye sur de larges collaborations structurées, principalement autour du CERN.

Très tôt, la question de la préservation des données et des environnements d'analyse a été discuté (2008)

DPHEP : Collaboration for Data Preservation and Long Term Analysis in High Energy Physics

- ➔ Roadmap mise à jour plusieurs fois depuis 2008; la dernière en date 2023
- ➔ Différents modèles de préservation (Level1-Level 4) pour la gestion et de la préservation des données et des analyses
- ➔ Préservation des procédures d'analyse (REANA, RECAST)

Depuis 1975 : HEPData https://www.hepdata.net/
(Principalement Level 1 : les données des publications)

Data Preservation Models identified by DPHEP

Preservation Model	Use case
1. Provide additional documentation	Publication-related information search
2. Preserve the data in a simplified format	Outreach, simple training analyses
3. Preserve the analysis level software and data format	Full scientific analysis based on existing reconstruction
4. Preserve the reconstruction and simulation software and basic level data	Full potential of the experimental data

Physique des Hautes Energies

S’appuye sur de larges collaborations structurées, principalement autour du CERN.

Très tôt, la question de la préservation des données et des environnements d’analyse a été discuté (2008)

DPHEP : Collaboration for Data Preservation and Long Term Analysis in High Energy Physics

➔ Roadmap mise à jour plusieurs fois depuis 2008; la dernière en date 2023

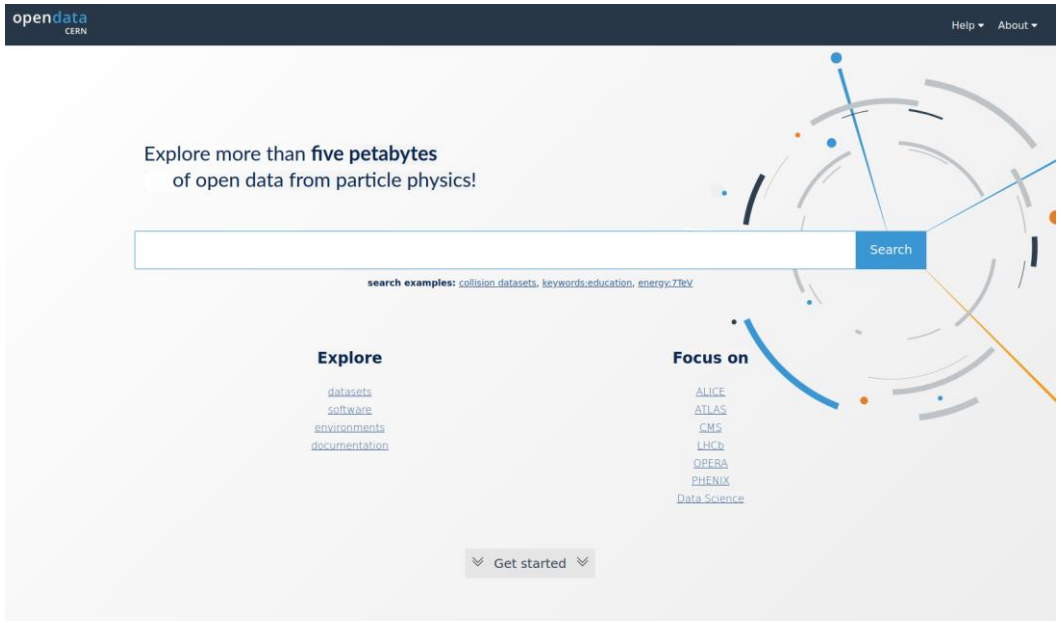
➔ Différents modèles de préservation (Level1-Level 4) pour la gestion et de la préservation des données et des analyses

➔ Préservation des procédures d’analyse (REANA, RECAST)

Depuis 1975 : HEPData https://www.hepdata.net/
(Principalement Level 1 : les données des publications)

Depuis 2014 : CERN Open Data portal - http://opendata.cern.ch

- Ouverture de jeux de données (Level 4 to 2)
- Data Release régulières par les collaborations (ATLAS, CMS, LHCb, ALICE, …)
- communauté scientifique et diffusion de la culture scientifique



Astroparticules / Cosmologie

A community in-between astronomy and particle physics

Astronomy

Astronomy pioneered Open Access to publications and data.

Data from large sky surveys are archived by public institutions (e.g., NASA, ESA), and made publicly available (e.g., http://archive.stsci.edu/, http://ned.ipac.caltech.edu/, http://skyview.gsfc.nasa.gov/, http://simbad.u-strasbg.fr/ simbad/).

Also, astronomical data have been integrated globally through the Virtual Observatory (VO), set up in 2002. The VO enable access to a multitude of on-line resources through a framework of interoperable tools.

Particle physics

opendata.cern.ch

Public access to (some, and growing) data produced at CERN (ALICE, ATLAS, CMS, LHCb, OPERA).

Not only data: software and documentation needed to understand and analyse them.

Aimed at outreach (all experiments) and research (only CMS at the moment)

The products are shared under open licenses; issued with a digital object identifier (DOI) to make them citable objects.

Courtesy P. Ghia

Astroparticle experiments: how are they doing with open data?

Present vs future

20 years ago (i.e., old experiments):
open data were a vague idea

Internal-only data and software management
(Data and codes access for the collaborations only)

Open data conceived "a posteriori"
(during the experiment):
Additional work to render data and codes accessible by anyone (even by companion collaborations)

Open data realised by "volunteers" collaborators:
a variety of format, styles and tools

Available open data "independent" ("un-coordinated"):
No single place where one can find all of them

New experiments (i.e,. CTA, Km3Net…):
open data are designed since the project start

Data and software management for external users too
(Data and codes external-access already in the projects)

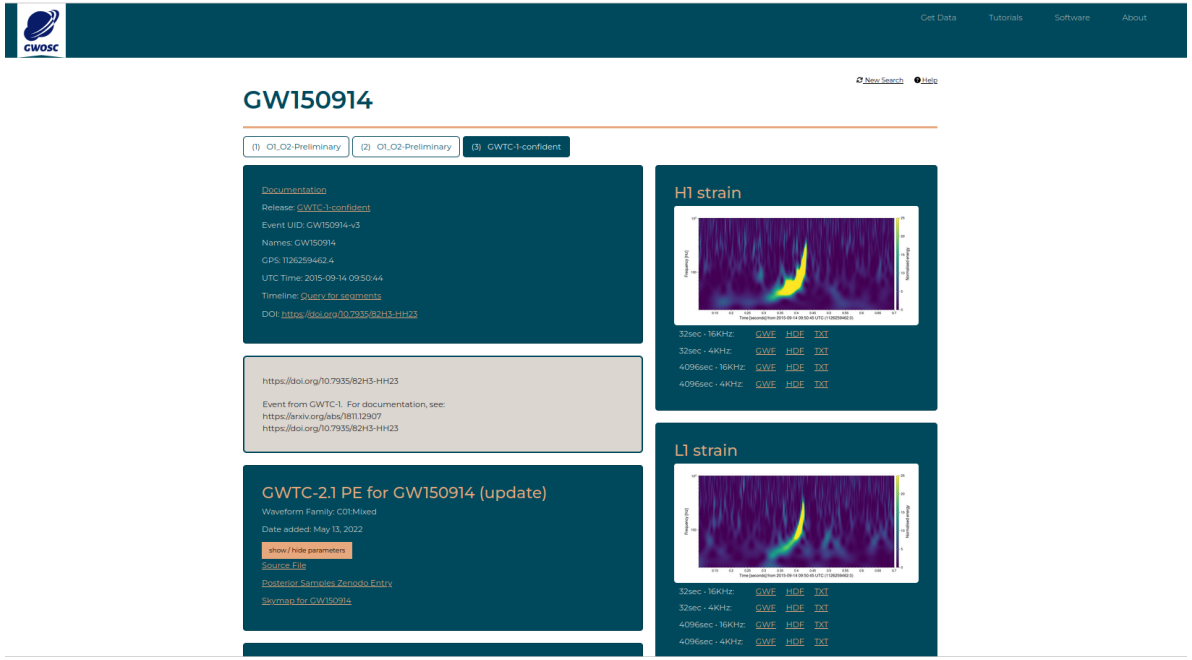
Open data conceived "a priori"
(planned before the experiment):
Work on data format and tools as open-source as possible

Open data realised by "specialised" collaborators:
effort towards common format and tools

Coordinated effort to make open data interfaced in a Virtual Observatory (e.g., ESCAPE infrastructure)

Courtesy P. Ghia

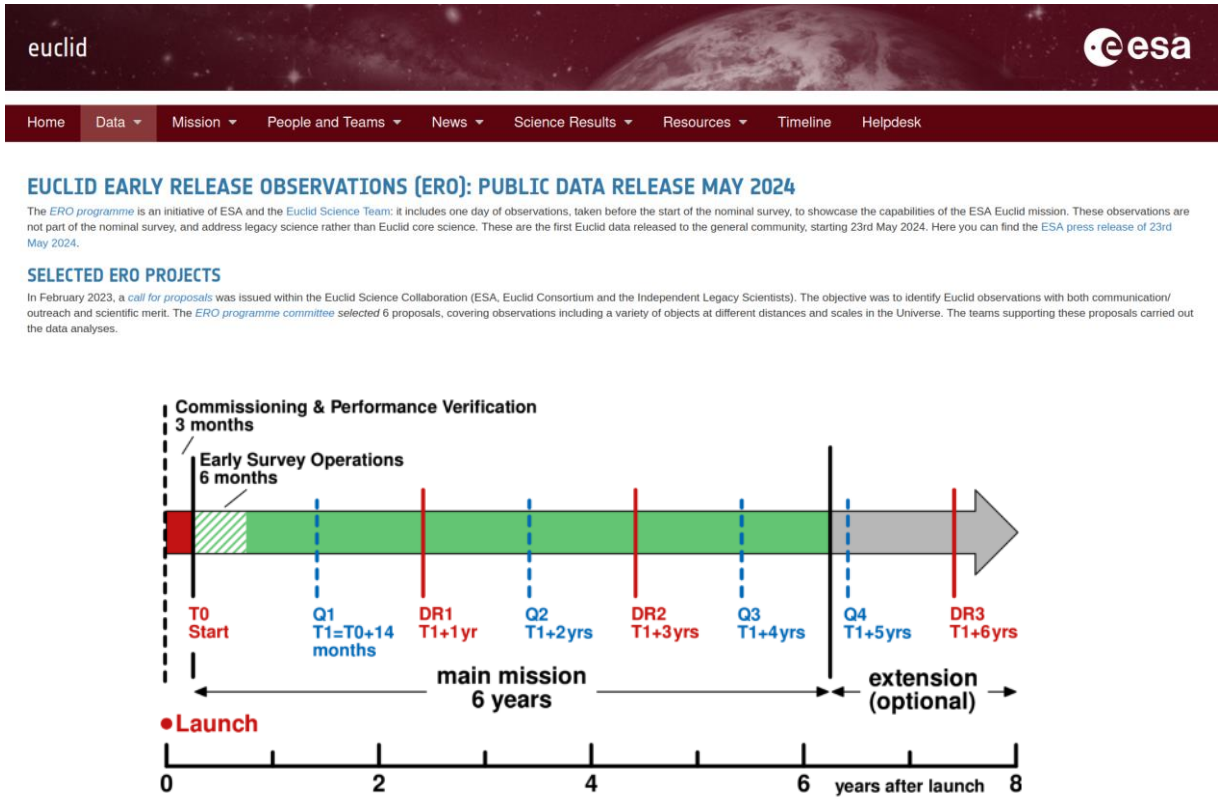
Gravitational Wave Open Science Center LIGO/VIRGO/KAGRA



https://gwosc.org/

-> Voir la présentation de demain de M. Dubois

Euclid Early Release Observation (ERO) Mai 2024



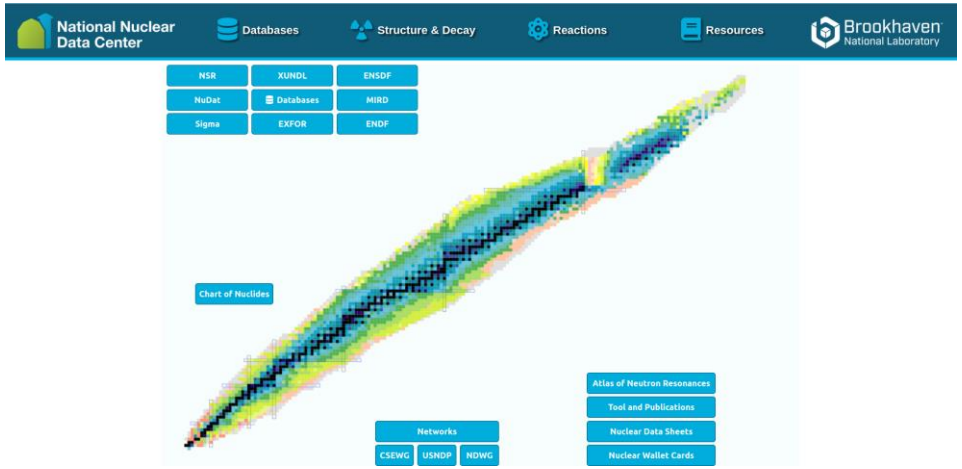
Physique Nucléaire

- Une longue histoire de données évaluées ouvertes (60 ans) (les bases de données nucléaires IAEA (EU) / NNDC (USA)) : post publications (équivalent au Level 1 de HEP)

- Effort d'ouverture des données expérimentales / logiciels
  - DMP Institut Laue Langevin (2010) : ouverture depuis 2010
  - DMP-GANIL (2019) : ouverture des données brutes des experiences
  - Framework d'analyse (AGATA, GRIT, FAZIA, VAMOS, ...)
  - Tkn (access au bases de données nucléaires) ...
- 2024 : Long Range Plan NUPPEC pour une coordination européenne sur les données en physique nucléaire :
  - Plans de Gestion de données
  - Standardisation des Méta données
  - Mise en place de catalogues
  - Entrepots de données des infrastructures

- Un domaine assez hétérogène dans ses pratiques, avec peu de ressources dédiées à la gestion des données (pour le moment !)

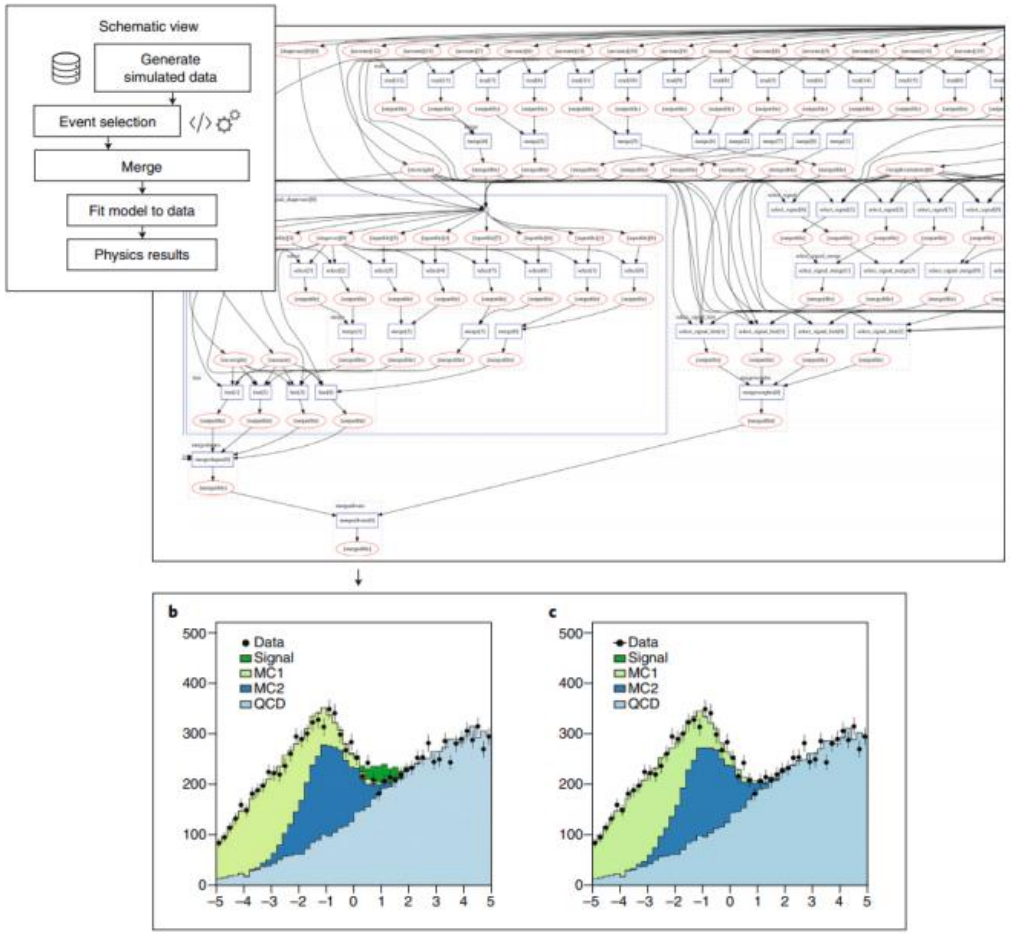
IAEA Nuclear Data Section (NDS) depuis 1964 (Photo: IAEA)



Simulations / Modélisation / Théorie

- Nombreuses discussions et réalisations en cours sur la ré-interpréation des résultats
- Bénéfices de publier les données issues des modèles (Modèles statistiques HEP, …)
- Cout de calculs vs gestion et stockage des résultats
- Rôle primordial d'une bonne gestion de ces données pour la reproductibilité et la pérennité des travaux.
- Importance de la gestion logicielle et des workflows

⇒ Voir la présentation de P. Davis



REANA computational workflow for a beyond the Standard Model (BSM) full analysis

https://www.nature.com/articles/s41567-018-0342-2

Et tous les autres cas identifiés (ou à identifier)

- Physique des accélérateurs
- Instrumentation
- Physique médicale
- …

DOP2I : Données Ouvertes de la Physique des 2 Infinis

- Groupe de travail dont les objectifs sont :
  - de faire un état des lieux des pratiques de gestion et d'ouverture des données à l'IN2P3
  - de proposer des recommandations à la direction de l'institut pour le développement et l'amélioration des pratiques d'ouverture des données
    - Les infrastructures à développer
    - Les ressources nécessaires pour prendre en charges ces nouveaux besoins/nouvelles obligations
    - Reconnaissance des activités de gestion des données
- Groupe de travail constitué de personnels issus des différents laboratoires et thématiques de l'institut

APC : Eric Chassande-Mottin, Bruno Khelifi
CC-IN2P3: Frederic Azevedo , Yonni Cardenas, Sébastien Gadrat (RI3), Gino MARCHETTI , Jean-Yves Nief
CPPM : Cristinel Diaconu
GANIL : Antoine Lemasson
IJCLAB : Piera Ghia, Olivier Brand-Foissac
IP2I : Olivier Stézowski, Sara Marcatili
IPHC : Jerôme Pansanel
IN2P3 : GRIVES Mathieu, Sabine Crépé-Renaudin
LAPP: Thomas Vuillaume
LPC Caen : Daniel Cussol, Phil Davies, Adrien Matta
LPC Clermont : Louie Corpe
LPNHE : Olivier Dadoun
LPSC : Sabine Kraml (* Theorie)

Quels sont les besoins ?

=> Développer et transmettre les bonnes pratiques :

- « Référents données » au sein des projets / Laboratoires
- Information sur les ressources disponibles
- Sensibiliser à l'ouverture des données / logiciels
- Former à la gestion des données / logiciels

=> Développer les infrastructures et les outils :

- Cycle de vie des données
  - Outils de génération de Plan de Gestion des Données
- Stockage et accessibilité des données
  - Entrepôts Thématiques (IN2P3) + Collaborations
  - Catalogues
  - Système d'authentification et d'autorisation
- Logiciels :
  - Plateforme de gestion des versions, de déploiements (CI/CD)
  - Catalogue de logiciels
  - Publications des logiciels
- Plateforme d'analyses :
  - Accès / Authentification / Autorisation
  - Analyse / Computing
  - Mise à disposition / (Ré) Utilisation
- Publications / Data Papers

Quels sont les défis ?

Des besoins différents selon les cas (collaborations vs équipes locales)
et les pratiques dans les différentes communautés

des défis :

- techniques (quelles solutions adaptées aux différentes besoins communautés, volumétries, ...)
  => /-\ Adaptation de solutions existantes (mais pas seulement ?)
- organisationnels (accords au sein des collaborations internationales, anticiper la gestion de la données, Ressources Humaines, qui assure la curation des données dans les entrepôts, reconnaissance du travail effectué sur les données)
  => /!\ effort important
- de mentalité (formation initiale et formation des différents acteurs)
  => /-\ dépends aussi de la qualité des solutions proposées
- éthiques (comment maîtriser la ré-utilisation des données hors des collaborations instrumentales, le cas des données anonymisées de santé, ...)

Quelques pistes de recommandations (IN2P3)

- **sur la définition d'une politique d'institut sur les données ouvertes**
  - au sein de collaborations
  - au sein de l'institut et du CNRS (politique de citation des jeux de données, des softwares, …)
  - provisionner dans le budget et les ressources des expériences une enveloppe dédiée à la Science Ouverte (dépôts données /plateforme d'analyse/curation …)
- **sur les besoins d'infrastructures**
  - centre de référencement thématique,
  - entrepôt de données thématique,
  - catalogue de logiciels
- **sur la valorisation des activités de gestion des données**
  - composante Data Scientists dans les activités de recherche, de support
  - sur la valorisation des activités de Science Ouverte dans l'évaluation
- **Sur la promotion des pratiques d'ouverture et de gestion des données et la formation**
  - nomination de "référents Science Ouverte" dans chaque expérience ou projet (à l'instar des czars pour le calcul)
  - Les "référents SO" pourrait interagir et partager des bonnes pratiques lors de journées Science Ouverte (comme celles qui sont régulièrement organisées pour le calcul)
  - Portail de ressources : Bien qu'il existe de multiples ressources (formations, écoles, documents), les chercheurs ne savent pas toujours quoi faire.

Quelques pistes de recommandations (IN2P3)

- **sur la définition d'une politique d'institut sur les données ouvertes**
  - au sein de collaborations
  - au sein de l'institut et du CNRS (politique de citation des jeux de données, des softwares, …)
  - provisionner dans le budget et les ressources des expériences une enveloppe dédiée à la Science Ouverte (dépôts données /plateforme d'analyse/curation …)
- **sur les besoins d'infrastructures**
  - centre de référencement thématique,
  - entrepôt de données thématique,
  - catalogue de logiciels
- **sur la valorisation des activités de gestion des données**
  - composante Data Scientists dans les activités de recherche, de support
  - sur la valorisation des activités de Science Ouverte dans l'évaluation
- **Sur la promotion des pratiques d'ouverture et de gestion des données et la formation**
  - nomination de "référents Science Ouverte" dans chaque expérience ou projet (à l'instar des czars pour le calcul)
  - Les "référents SO" pourrait interagir et partager des bonnes pratiques lors de journées Science Ouverte (comme celles qui sont régulièrement organisées pour le calcul)
  - Portail de ressources : Bien qu'il existe de multiples ressources (formations, écoles, documents), les chercheurs ne savent pas toujours quoi faire.

Réflexion IRFU/IN2P3 au sein du RI3 ?

Journées des données ouvertes IN2P3

16 et 17 décembre 2024 au CC-IN2P3

Inscription et programme

https://indico.in2p3.fr/event/31614/

Venez participer à ces journées pour faire remonter vos besoins et échanger sur vos réalisations.

En résumé

- Des décennies d'expertise scientifique et technologique dans la gestion et le traitement des données de recherche
- L'ouverture des données ouvre de nouvelles opportunités scientifiques, mais aussi des nouvelles problématiques sur la gestion de ces données
- A ce jour, les résultats scientifiques (publications) dans nos domaines sont majoritairement ouvertes (arxiv, HAL, SCOAP3, ...).
- Les pratiques d'ouvertures de codes et des données sont en nette progression, mais de façon hétérogène ...
- Nécessité d'accélérer la mise en œuvre des ressources indispensables à la gestion et l'ouverture large des données pour la réalisation concrète de la Science Ouverte dans les différents domaines.