



Open Science @ IN2P3

Open science : context

International context

- UNESCO : recommandations for open science, OECD: recommandations on research data
- Europe
 - the commission requests open publications and open data for research funded by Europe
 - since 2021, EC defines open science as a criteria of scientific excellence
 - EOSC – European open science cloud
- International working groups : RDA, GO FAIR...

CNRS

- [CNRS roadmap for Open Science](#) (2019)
- [CNRS research data plan](#) (2020)
- DDOR creation : Direction des Données Ouvertes de la Recherche (2020)
 - open science from publication to data, participation of all CNRS institutes

National initiatives

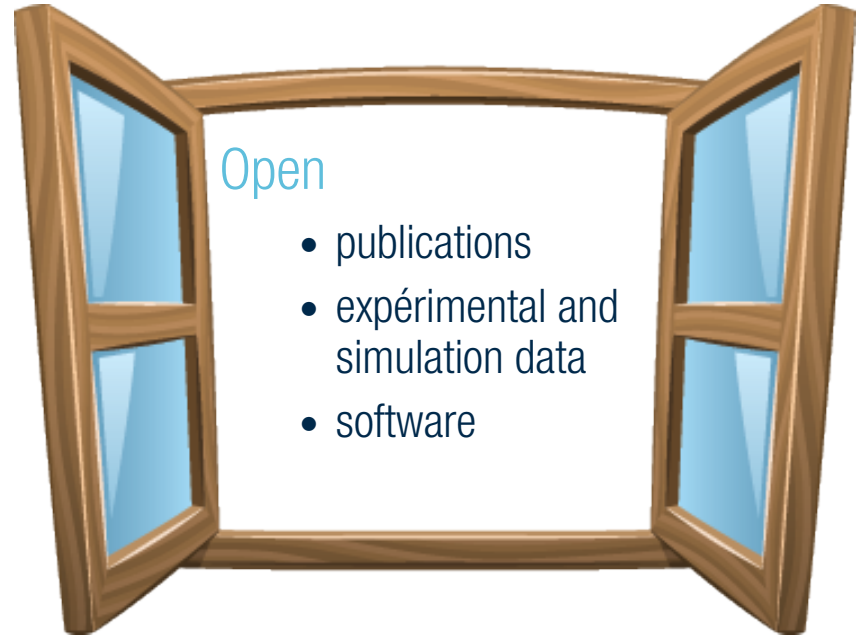
- Loi pour une République numérique (2016)
 - data as open as possible as close as necessary
- [National plan for Open Science](#) (2018)
- [2nd plan for Open Science](#) (2021)
 - open access publication for calls
 - open data platform data.gouv.fr
- [Roadmap 2021-2024](#) (2021)



Open Science

Organisation at IN2P3

- Scientific direction for data and computing along with in particular scientific direction for the 3 main research domains (HEP, astroparticle and cosmology and nuclear physics)
 - publications, computing and data
 - strong link with CNRS DDOR
- IST (Information Scientifique et Technique) department for publication managed centrally at IN2P3 with people distributed in the IN2P3 labs
- organisation specific for open data in discussion
- International collaborations with their own open data policies
- Excellent computing and data infrastructures
 - CC-IN2P3 national computing infrastructure
 - regional and local platforms and computing centers



Publications

Publication treatment

- Publications from arXiv and editors fed into INSPIRE
 - curation, metadata enrichment by the IN2P3 team
 - not only IN2P3 : all publications with French affiliation in our fields
- Automatic exportation from INSPIRE → HAL (CNRS open archive)

Key Numbers

- ~3200 publications each year in INSPIRE, of which ~1800 with authors from IN2P3
- HAL-IN2P3 portal : 68 000 records, 2 000-2 500 records per year
- 90 % in open access
 - [SCOAP3 partnership](#) : Sponsoring Consortium for Open Access Publishing in Particle Physics
 - publication pipeline

Target

- 100% of publication in open access

INSPIRE HEP

- interlinked databases on literature, conferences, institutions, journals, researchers, experiments, jobs and data.
- Run in collaboration by CERN, DESY, Fermilab, IHEP, IN2P3, and SLAC, it has been serving the scientific community for almost 50 years

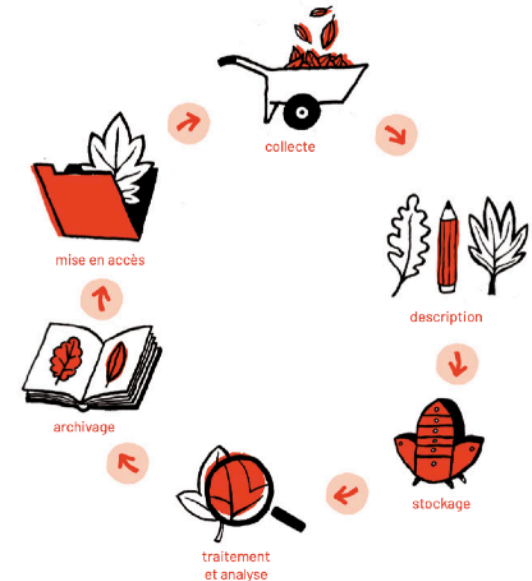


Open/FAIR data

- doesn't only mean provide access to your data to the outside world
- but means good practice to manage the data all over their lifecycle
 - collect good quality data → describe (with rich metadata) and identify (DOI) → store on adapted and reliable storage → process and analyze → delete, clean up or archive with inventory → open... or not if sensitive data
- software are part of the process and also a type of data that follow the same rules : should be referenced, versioned, stored and open
 - data and linked software should be associated

Outcome

- good and properly referenced data with complete set of metadata and software means better science, easier analysis, and possible reuse in the communities that produced the data and outside
 - avoid duplication of experiments and allows new analyses using data from different experiments



Data management at IN2P3

IN2P3 rules

- Computing needs and data policy should be thought of and assessed before the beginning of any experiment at the designed level
 - the experiment needs are discussed and the level of IN2P3 implication is decided
- Data should be properly described and stored in a reliable computing center, preferentially at CC-IN2P3 or at international sites chosen by the collaborations
 - a DMP (Data Management Plan) should be defined and reviewed each year : it defines the data produced by the experiment and their metadata, the computing and storage needs, the person responsible for the data, the future of the data
 - CC-IN2P3 DMP template is recorded in [DMP-Opidor](#)

Expertise available at IN2P3

- description, management, and processing of large and multiple datasets
 - also in distributed environment
- long term data storage at CC-IN2P3
- feasibility study done on data archive at CC-IN2P3
- experience on open data in international collaborations
- but no large repository hosted by IN2P3



Open data status

Open data policy is defined by the collaborations

- Long tradition of open access in astroparticle physics and cosmology
 - image catalogs in open access, data released after rather short period of embargo
 - a non negligible amount of publications are made outside the collaboration
 - next generation of experiments discussing on how to share and understand their data : just open is not enough, need upstream work of experts to allow multi messenger exploration of our universe
- In HEP
 - most publications now associated with their data in [HEPData](#) openly
 - discussions with theoreticians to improve the informations provided in publication by the experimental collaborations
 - small datasets with software released publicly since years for outreach and education purposes
 - most software open source
 - CERN open data policy for LHC experiment released in 2020 → data from run 1 will be soon available in [CERN Open data portal](#)
- In nuclear physics
 - Results (cross-sections. . .) opened in the international databases (EXFOR. . .) since decades
 - Increasing set of open source and shared softwares (Kaliveda, Fairoot, nptool, Agapro. . .)
 - Need to consolidate data and their referencing
 - complexity due to use of multiple accelerator complex by one experiment, multiplicity of physics programs and users using different infrastructures
 - need cooperation between accelerators and experiment communities and between different accelerator labs
 - OpenNP in European project Euro-Labs will help to build open science for nuclear physics in collaborative and coherent way

Conclusion

Open science means

- publications in open access
- open source softwares
- data of good quality, well described, referenced, stored on reliable storage, archived or deleted, opened and shared

→ allows efficient use of data, avoid duplication, allows new research projects and results based on the interplay of datasets from different experiments

IN2P3

- supports open publications thanks to the IST team and the participation to Scoap³ initiative, INSPIRE and HAL
- encourage open source software
- ensure data produced and stored at IN2P3 have a DMP, are well referenced and stored on reliable infrastructure
- support open data in collaborations

GANIL initiatives to reinforce good practice for data and Eurolabs OpenNP project are important and will bring large improvement on data management and open science in nuclear physics