"Open access to the data will, in the long term, allow the maximum realization of their scientific potential." --CMS Collaboration

Brookhaven National Laboratory

U.S. DEPARTMENT OF ENERGY

Overview of US Data Preservation
(Open Data in Low Energy Nuclear Physics)

Jin Wu

National Nuclear Data Center

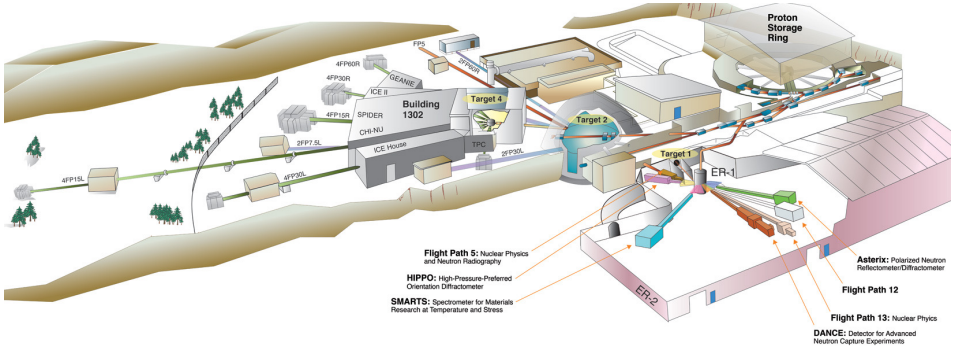
Workshop for Applied Nuclear Data Activities (WANDA) – 3/1/23

@BrookhavenLab

Where are we?

Los Alamos National Laboratory

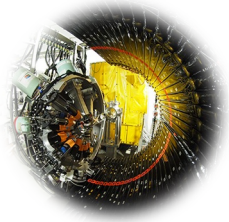
LANSCE



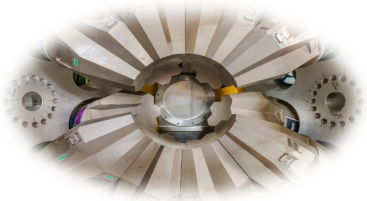
FRIB

GEANIE

DANCE



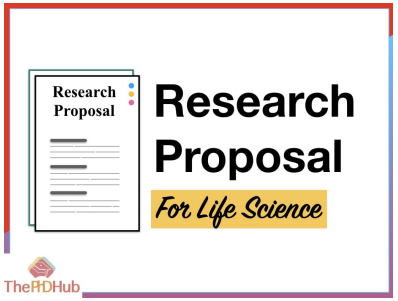
FDS



GRETA

- Large amounts of experimental data will be produced attributing to the powerful accelerators and complex detection systems.
- This “self-curation” by individual research groups lacks uniformity and results in a situation where data discovery and reuse are often difficult or impossible.

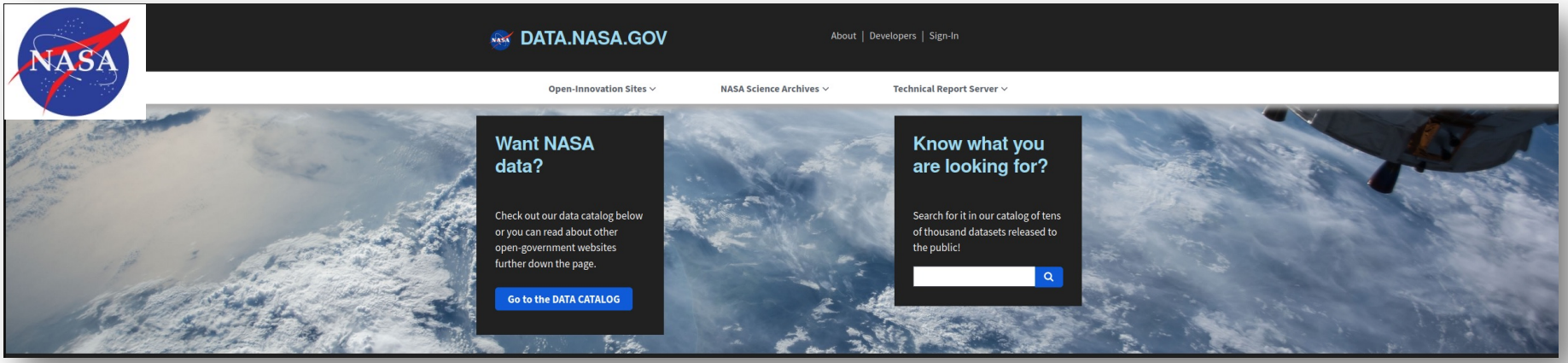
Many scientists could do ...



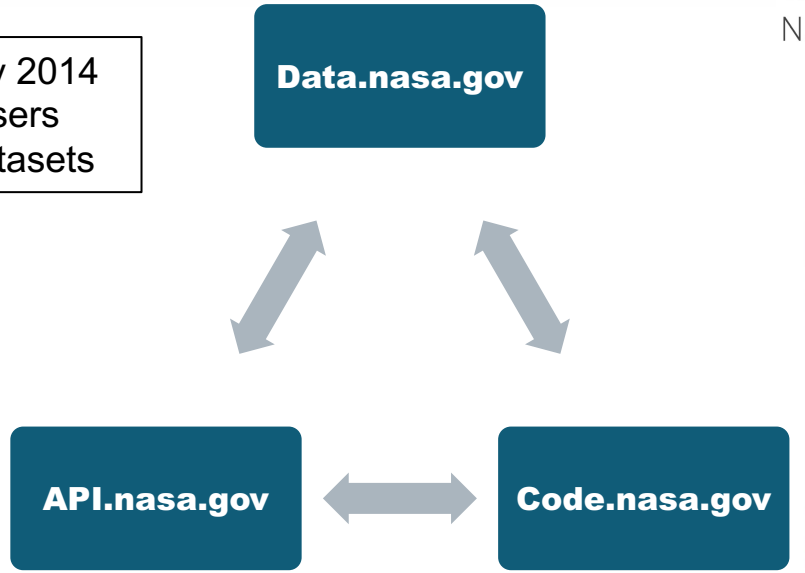
MEMORANDUM from White House (issued on 8/25/2022)

See presentation “Public Access Policy and Data Management Plans” by Dr. Michael Cooke

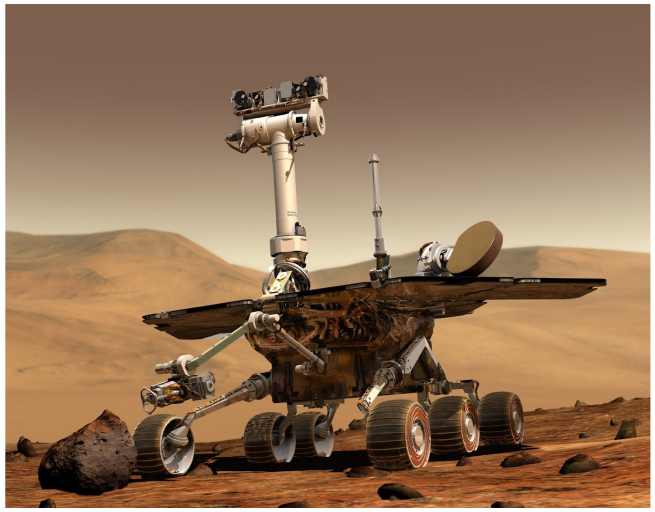
NASA Open Data



- Started in July 2014
- 20K unique users
- 40K active datasets



Texting robots on Mars using Python, Flask, NASA APIs and Twilio MMS



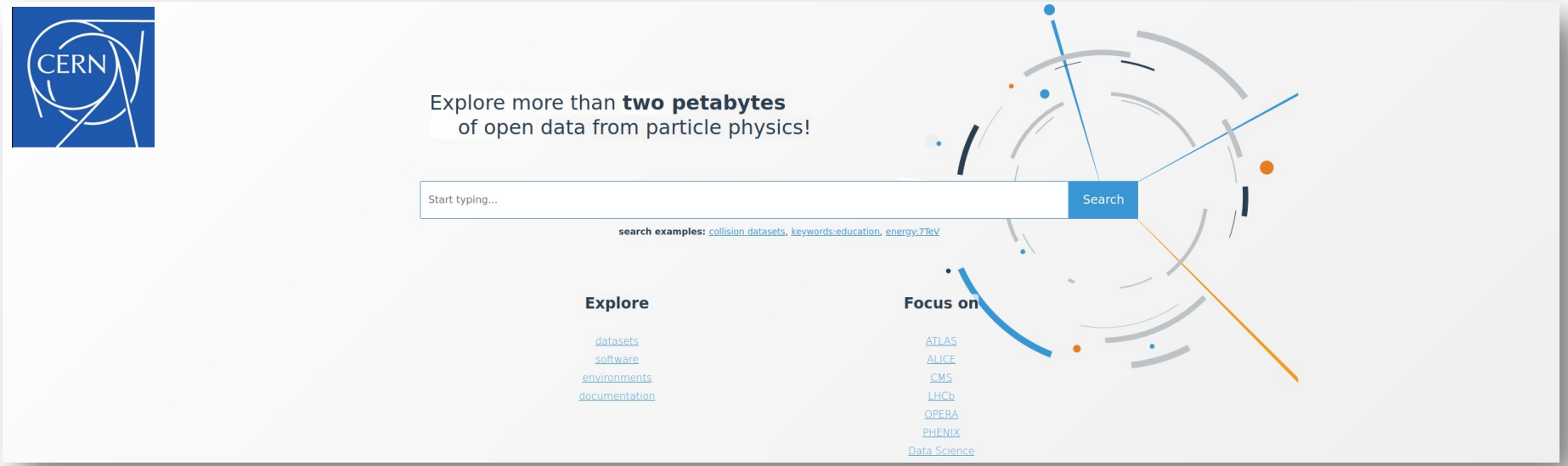
CERN Open Data

01 Sep 2020 | Network Updates

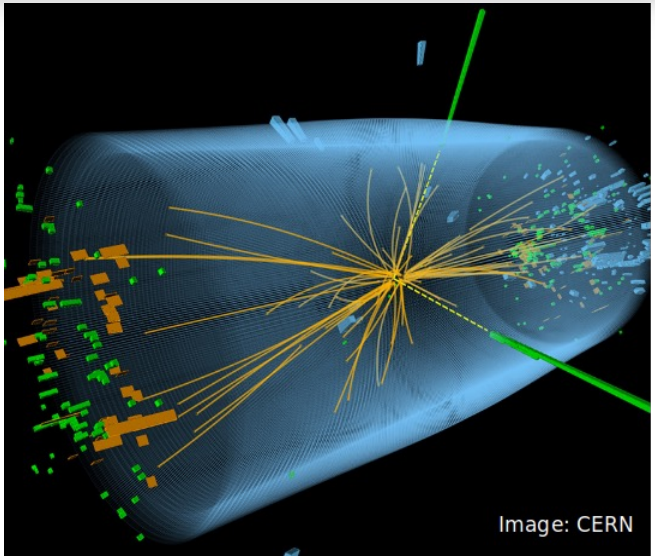
CERN releases fifth batch of open data recorded from Large Hadron Collider experiment

By Communication from CERN

All research-quality data recorded by CMS during the first two years of LHC operation are now publicly available.



- >2 petabytes of data available to the public LHC, ATLAS, CMS, ALICE, etc.
- Preprocessed data
- Example source code



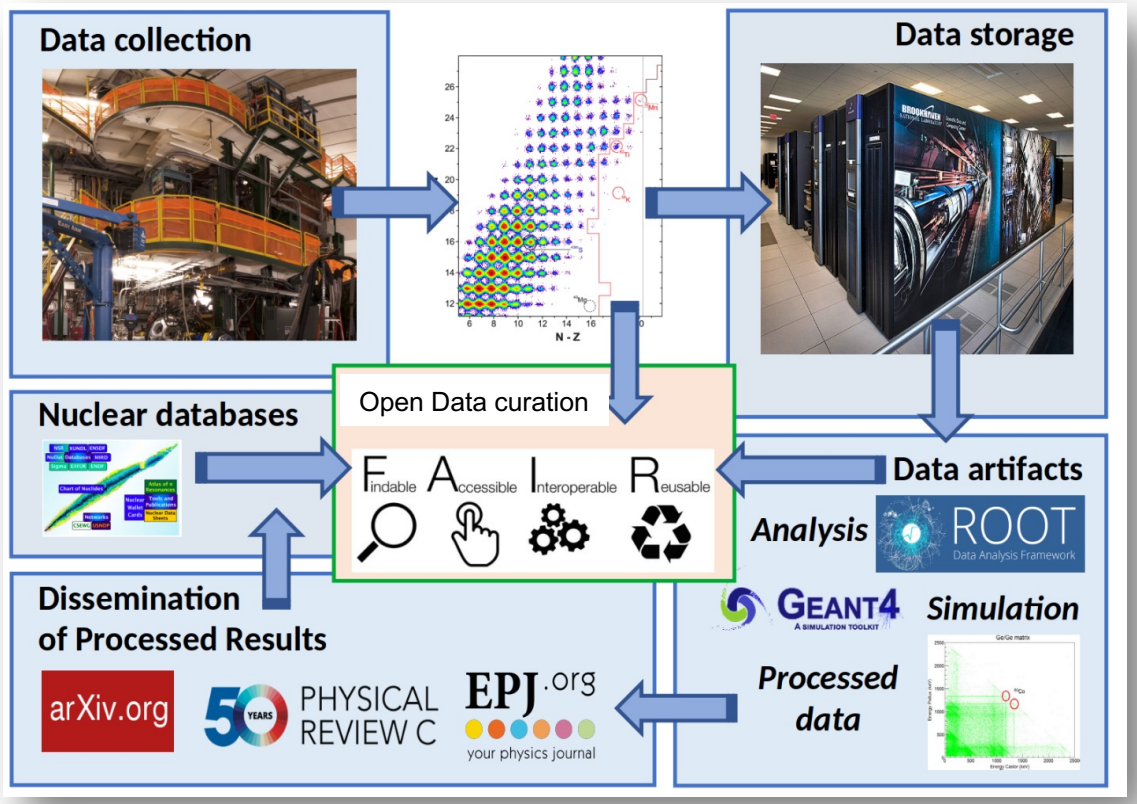
European Open Data in Nuclear Physics

See Presentation "openNP"
by Dr Adrien Matta (Ganil)

European Open Science Cloud (EOSC) – ESCAPE Program

- ✓ Brings together the astronomy, astro-particle and particle physics communities, aiming to produce versatile solutions for the open data management, cross-border and multi-disciplinary open environment.
- ✓ Composed by 31 European organizations with wealth expertise and experience on various of research fields, which are committed to contribute to the construction of the data research open environment

Modern database technology: curating open data



Credit: Adam Hayes

Acceptance of Proposal → **During Experiment** → **Acceptance of publication** → **The end of embargo period**

The "FAIR" data principle

1. **Findable** with a globally-unique identifier and rich metadata
2. **Accessible** through a free and standard protocol
3. **Interoperable** with as much standardization as possible
4. **Reusable** with accurate and rich provenance metadata

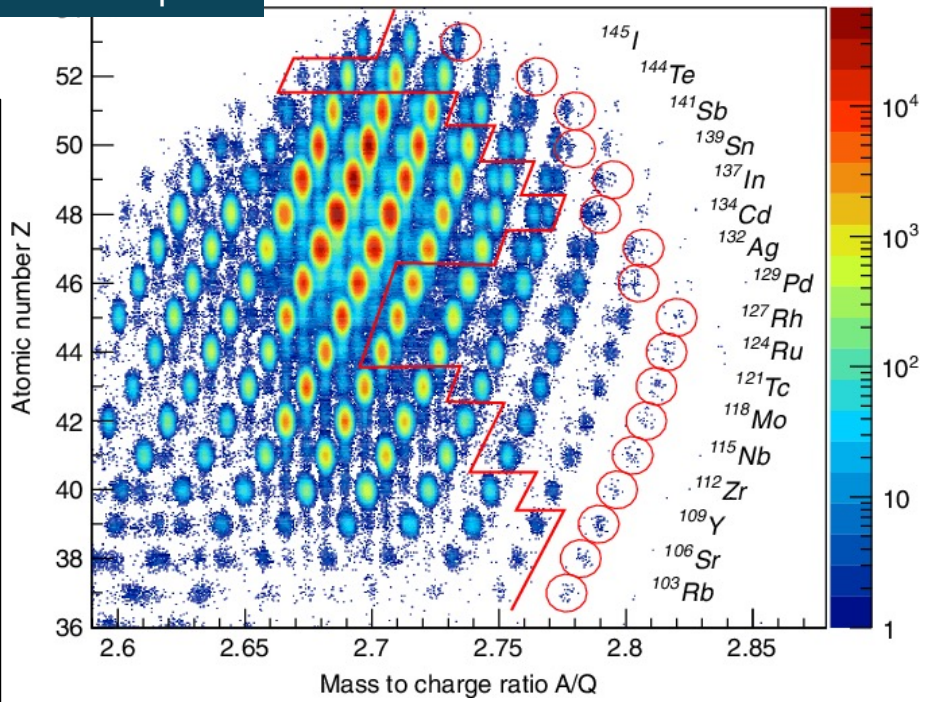
Findable	Accessible	Interoperable	Expansionable and Reusable
• Metadata • Persistent Identifiers (e.g. DOI) • Data Management Plan	• Web portal • APIs • Data embargo	• Workflow starting from acceptance of proposal	• Policy • Expansionable • Infrastructure

What we get from Open Data

- Validation of results
- Rescue / improve / repurpose existing data
- Byproducts – time and grants saving (e.g. fission, fragmentation, deep-inelastic, etc)
- Education
- Dataset citation
- Future experiment guidance
- Implemented ML/AL: Experiment optimization, and data analysis, data management.

110 Isotopes

G. Lorusso et al, PRL 114, 192501 (2015)



Measured: β-$T_{1/2}$, P_n, β-γ, Isomers, ICe

Published:
β-$T_{1/2}$ (110 Isotopes)
^{126}Pd, ^{128}Pd Isomers
^{126}Pd, β-γ
^{123}Ag, ^{125}Ag: β-γ
^{125}Pd, ^{127}Pd: isomers

Large amounts of byproducts are unanalyzed!

Future Perspective

1. Developing & maintaining Open Data website
2. Organizing Community workshops
3. Developing Data/Metadata, Persistent Identifiers system
4. Recruiting future contributors

Jin Wu
jwu2@bnl.gov

Elizabeth McCutchan
mccutchan@bnl.gov

David Brown
dbrown@bnl.gov

Data Preservation Survey

https://docs.google.com/forms/d/e/1FAIpQLSeyhMTibWwfqbcXa4r8nYE9NFS_dLgTNVpx2ZTUEjiwky_3fg/viewform?usp=sf_link

Thank you!