FAIROS-HEP “Kick Off” Workshop @ CERN

Kyle Cranmer (Wisconsin), Peter Elmer (Princeton)

Mike Hildreth (Notre Dame)

FAIROS-RCN: way too many acronyms?

We are supported by the US National Science Foundation (NSF) under a program called:

Findable Accessible Interoperable Reusable (FAIR)

Open Science (OS)

Research Coordination Networks (RCN)

"The goal of the FAIROS RCN program is to advance FAIR and Open Science capabilities, both human and technical, by supporting groups of investigators to communicate, innovate, coordinate, and standardize research practices, training, and educational activities across disciplinary, organizational, geographic and international boundaries."

We are in the "Disciplinary Track", i.e., focused on HEP data and the HEP community

FAIROS-RCN Landscape

We are one of 10 funded projects:

- DBER+ Commons
  - FAIR in Education research
- MaRCN
  - Open science in materials
- FAIR facilities and instruments
  - PIDs for research instruments
- FARR
  - Best practices for ML/AI
- Geospatial Big Data Infrastructure
  - Environmental research
- Paleobio/zooarchaeology database project
  - Community-coordinated resources
- SEEKCommons project
  - Bridge social and environ. sciences
- NoCTURN
  - Non-clinical tomography
- REPETO
  - Reproducibility in CS Education

More info here: https://www.arl.org/news/arl-applauds-nsf-open-science-investment/

FAIROS-HEP: Motivations

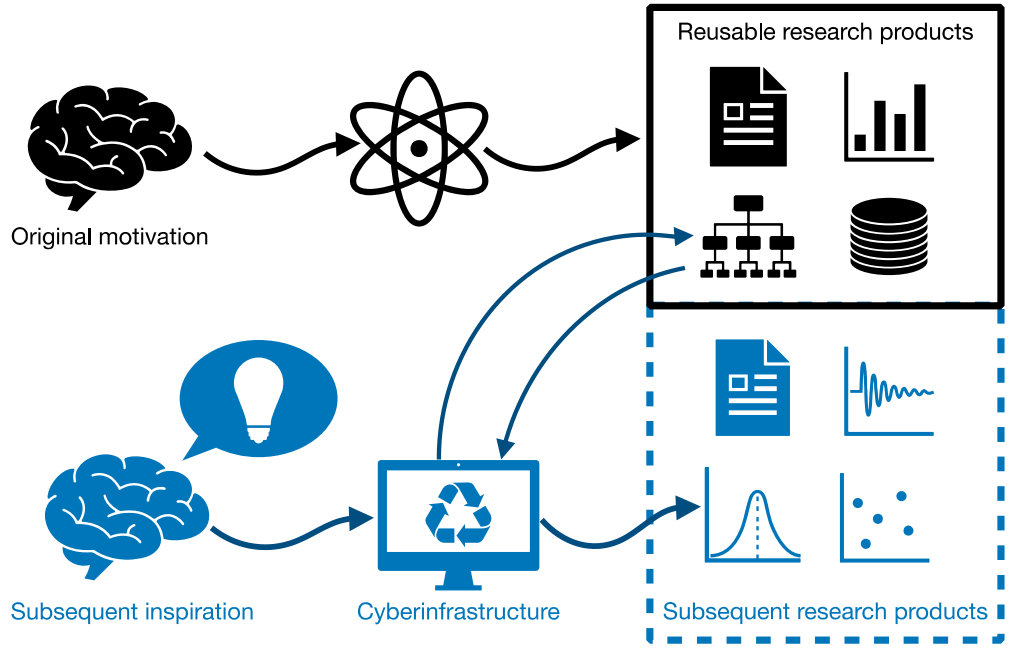
- The HEP global cyberinfrastructure encompassing publications, data products, related software and analysis workflow preservation is maturing.
- Possible to develop new analyses of archived data
- Possible to add scientific value to the archived data, analysis pipelines, and other research artifacts through patterns of reuse such as reinterpretation or integration into a statistical combination or meta-analysis.
- Underlying infrastructure will support the preservation and reuse of practically any data analysis based on computational workflows.

FAIROS-HEP: Goals

- **Network Building**
  - **Bring together open data practitioners from past, existing and future experiments, existing cyberinfrastructure projects (INSPIRE, Zenodo, etc.), funded projects in this space (IRIS-HEP, Punch4NDFI, FAIRHEP, etc.), DPHEP, and Coalition for Networked Information (CNI)**
- **Ecosystem Coordination**
  - **engage the U.S. and global particle physics community to translate a shared vision of data reuse and combination into actionable recommendations grounded in the technical language of the physics use-case and cyberinfrastructure components**
- **Technical Contributions**
  - We will make specific technical contributions to the cyberinfrastructure ecosystem, with a focus on interoperability and integration. These are areas that slip through the cracks in the status quo because they are often seen as out of scope for each individual component

FAIROS-HEP: Goals

- Larger goal: create a "living HEP publication" to facilitate distributed, asynchronous reuse of HEP research products
  - This will require the completion and integration of the combined infrastructure supporting publications, data, software, analysis and workflow preservation.



We'll come back to what this might represent in reality during the discussions tomorrow and Fri.

FAIROS-HEP: Structure

- Duration of project: three years
- Three US institutions:
  - Notre Dame
  - Princeton
  - Wisconsin
- Each has funding support for hosting or standing up two workshops
  - Participant support (travel, lodging, etc.)
  - Approximately 6 workshops total are planned
- Each has funding for 0.5 FTE of Software Engineering to actually build some infrastructure
  - What that is depends on needs and opinions of the community

FAIROS-HEP: ”The Network”

Try to connect people working on open data issues and infrastructure scattered across the science landscape. Each group probably has a few people focused on these problems – find them and include

- Past, Present, and Future Particle Physics Experiments
  - LHC experiments
  - DUNE, Minerva, Nova, …
  - Belle-II
  - IceCube
  - Xenon, …
- DPHEP
- CNI: Consortium for Networked Information. More than 200 institutions representing higher education, publishing, IT, scholarly societies, libraries

FAIROS-HEP: ”The Network”

- GO FAIR US: US office of GO FAIR
- DANCE Network: “Dark matter and Neutrino Computation Explored”
  - Dark matter experiments (XENONnT, DarkSide, SuperCDMS, LZ, DEAP, ARGO), neutrino-oriented experiments (incl. µBooNe, NOvA, COHERENT, HyperK), and double- beta-decay oriented experiments (incl. EXO, nEXO)
- DOE Labs
- Electron-Ion collider
- Zenodo
- HEPData
- InspireHEP
- CERN Open Data Portal

FAIROS-HEP: ”The Network”

- CERN Analysis Preservation Portal
- REANA
- RECAST
- arXiv
- FAIR4HEP
- DASPOS
- PUNCH4NFDI

FAIROS-HEP: Workshop Plans (1)

Workshop 2: Broader Community Engagement and Theory Reinterpretation

Attendees: Particle Physics Experimentalists and Theorists

Location: American Physical Society Division of Particles and Fields (DPF) 2023 meeting

Purpose:

(i) Establish use cases for data access, interoperability, and reuse, especially from the theory community,

(ii) define what data and associated information supports these use cases, and

(iii) broad community engagement.

Technical Recommendations: Initial specifications of interfaces needed for data access aligned with physics use cases.

Workshop 3: Specific Reuse Case: Deriving EFT Results from Future LHC Data

Attendees: Participants from the LHC experiments and theorists working on EFT interpretations

Location: US University

Purpose:

(i) Establish theoretical frameworks, data formats, and the means of reporting measurement results that will enable reuse and meta-analyses,

(ii) (ii) define access methods for finding, assembling and re-analyzing data.

Technical Recommendations: Specification on FAIR properties of EFT analyses, including data discover, access, and interoperability details.

FAIROS-HEP: Workshop Plans (2)

Workshop 4: Broader Engagement: CNI and External Science Partners

Attendees: Core RCN members, CNI membership, interested External Science Partners

Location: Semi-Annual CNI Membership Meeting

Purpose:

(i) Disseminate concepts, results and future prospects for RCN work to broader com- munity,

(ii) (ii) Engage with CNI community and broader science participants to obtain feedback on directions and understand their research needs.

Technical Recommendations: Discussions may provide ideas for more generic implementations of common tools.

Workshop 5: Specific Reuse Case: Kinematic RECASTing for New Physics Discovery

Attendees: Participants from the LHC experiments and theorists working on new physics searches

Location: US University

Purpose:

(i) Establish theoretical frameworks, data formats, and the means of reporting mea- surement results that will enable reuse and meta-analyses,

(ii) (ii) define access methods for finding, assembling and re-analyzing data.

Technical Recommendations: Specification on FAIR properties of search analyses, including data discover, access, and interoperability details.

CERN Workshop Goals

1. Assess progress by each experiment in producing reusable data,
2. Establish updated ideas regarding the use cases for data access, interoperability, and reuse across the different experiments and experimental fields,
3. Define what data and associated information supports the use cases, and
4. Identify a preliminary set of access methods and infrastructure that would support these use cases.

Technical Recommendations: Provide initial direction for which elements of the cyber ecosystem will be most relevant for the first round of technical improvements. Initiate investigations.

Workshop Logistics

- Today: Plenary Overview of DPHEP, CERN, and LHC efforts
- Thursday morning (Bat. 40, Salle Bohr): more experimental presentations, start of discussion
- Thursday afternoon (Bat. 40, Salle Bohr): overview of FAIR themes, publication ecosystem, more discussion
- Thursday evening: Workshop Dinner, Café du Soleil
- Friday morning (Bat. 40, Salle Bohr): discussion of next steps, wrap-up
- Friday afternoon (13/2-005): Panel Discussion - How can academic publishing support FAIR and Open Science?