

ICFA Data Lifecycle Panel: Medium-term goals Plan of action

ICFA Data Lifecycle panel meeting - May 21, 2024



Kati Lassila-Perini
Helsinki Institute of Physics - Finland



Your input

The mandate is broad, even overwhelming...

My expertise is CMS/CERN/Open data/FAIR-centric

Diverse range of expertise within the panel - your input counts



Survey questions

Goals:

“Suggest medium-terms goals for the panel, i.e. what would you like to see achieved in 1-2 years.”

Actions:

“Suggest actions through which we could reach those goals.”



Stefan

Goals

A lot of training initiatives starting around now, e.g. in EVERSE, in Next Gen Triggers and others.

ECFA is trying to setup a training working group.

Help to raise the awareness for training and the recognition of people who engage in training. M1.3 M2 M5

Come up with a set of training materials which are needed in software and computing for accelerator physics. M2

Actions

Use synergies among the activities that have started or are running.

ICFA could be body to endorse a set of training materials useful for the community or even suggest a curriculum which can be proposed to schools and universities.



Takanori

Goals

[...] in order to indicate the future direction of the data lifecycle in HEP, I feel it is necessary to **list the problems** we are currently facing and come up with a rough plan for how to deal with them before setting the medium-term goals.

[...] to establish a panel on computing broadly under ICFA will have great benefits for the HEP community, such as **disseminating the importance of computing** to all community members. As written in the "Mission" and "Mandate" in the PDF "ICFA Panel on the Data Lifecycle", this panel must cover a wide range of topics.

Actions

it may be better to start the following items...

- 1) list what problems we are currently facing in the data lifecycle
- 2) Prioritize which of these issues is most pressing
- 3) install sub-panels as necessary



Mihoko

Goals

I think we should first **invite talks related to each mandate** rather than focusing on everyone's current interests.

I felt that in the previous meeting, the chair focused on open data and resources for that, but it is not clear to me if we have enough attention from experimental groups to push the direction.

We probably have to listen to the opinion of the experiment groups to make our report useful for the community.

Actions



Seema

Goals

- (1) Consolidate current status of various computing resources, tools (e.g. file systems) and limitations (mostly from user perspective). This should be used to project challenges expected for data preservation in coming years e.g. can the data used for publications (both from experimental and theory communities) now would be readable in a few years even these exist on some storage ? M1.4 M2 M4
- (2) Facilitating more awareness about open data e.g. CMS & ATLAS. Unless we have more and more people using the data available, we do not really know what are the bottlenecks. M2
- (3) There is a surge of AI/ML tools but without much awareness on what are actually beneficial vs computing resources heavy. Can something be done about it ? M3

Actions

- (1) Might contact various collaborations and major groups to get a broad idea of the situation.
- (2) The efforts in terms of workshops are already going on, may be increase the ongoing ones and add more.
- (3) Not sure if this goes beyond the scope of discussion here. May be some working groups ?



Harvey

Goals

- (1) Beginning to deploy/integrate SENSE and related advanced network+Site services with the mainstream LHC data management tools. M1,M3
- (2) Build a paragon network + data management system to show current capabilities (an order of magnitude beyond DC24 for example) – both scale and already existing functionalities and tools. This, and projections of technologies for 2026 and beyond are essential for setting real requirements for the HL LHC era; this will also enable more effective production and analysis workflows and their management. M1 M3

Actions

Engage with the GNA-G and its working groups.

Teach HEP about current server, network and interface technologies, and their projected evolution over the next 1-5 years

Learn from the SENSE network + site management teams

Learn about programmable network capabilities and methods also in the Global P4 Lab

Oversee and enable stronger ties and begin closer joint work between the above efforts and the LHC data management teams as well as the at-large HEP communities.



Liz

Goals

1. An important goal would be to make sure that the importance of the Data Lifecycle is highlighted in the EU strategy for particle physics '25. **M1.1 M2**
2. Be an advocate for common software solutions, since the HSF is losing support from CERN. **M1.1 M2**

Actions

For 1. make sure that someone from the panel can represent us and go to the meetings.



Cristinel

Goals

Enhance careers in HEP Data Lifecycle, including aspects of training, ethics, gender equity etc. **M5 M2**

Actions

Community workshop (possible attached to one CHEP/ICHEP/EPS conference): issue a panel report with proposed measures.



Caterina

Goals

provide information and pointers to tools that make it easier to achieve FAIR data management for current and future experiments [M1](#) [M2](#)

Actions

a meta-repository of data sources (based on what exists, without duplicating)

lobby for rewards for those who work on this topic



Kati

Goals

1. A well-defined, HEP-specific training curriculum on research software best practices based on the existing training resources (eventually in agreement with universities: ECTS credits) M1.1 M2
2. **Practical** guide for different stakeholders (researchers, group leaders, experiment management) on actions to achieve FAIRness of research software and workflows. M1.1 M2
3. Promote the work that goes in organizing training and learning, and value the time that is invested in following best practices M5

Actions

For 1. work together with the current training projects on their plans and challenges.

For 2,

Discussion

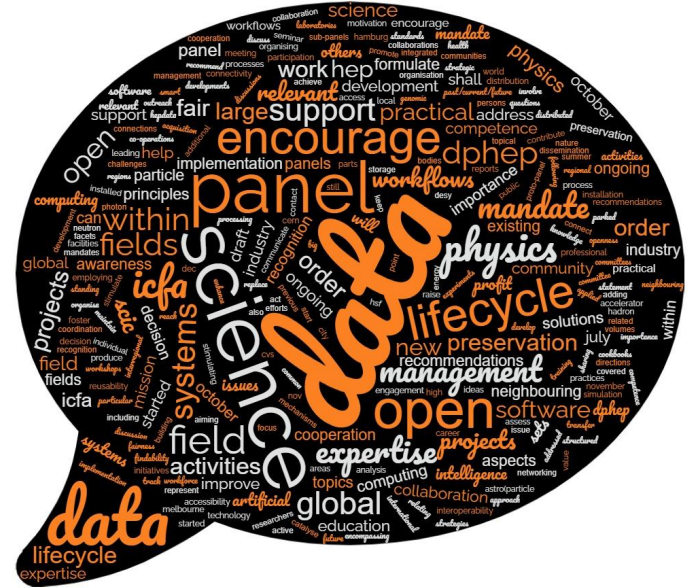




Outlook

Next meetings on June 18th and July 16th

- Complete the input to the goals & actions survey
- Collect a list of problems w.r.t panel's mandate topics
- Hear from DPHEP (Cristi) and SCIC (Harvey)
- Define preliminary goals





Thank you!

Questions?

And thanks to [SlidesCarnival](#) for this free presentation template



Mission

Mission

The mission of the panel is to enhance global coordination on all aspects of the data lifecycle including acquisition, processing, distribution, storage, access, analysis, simulation, preservation, management, software, workflows, computing and networking in particle physics, with a focus on open science and FAIR practices.

In order to achieve this, the panel will

- A. address all aspects of the data lifecycle, encompassing the efforts and expertise from previous panels, and relating to and building on activities of other relevant bodies and committees;
- B. encourage global cooperation on the above topics in particle physics and with neighbouring fields;
- C. discuss strategic questions and recommend to the community future directions;
- D. encourage engagement with and profit from industry expertise in data management solutions, in artificial intelligence, and in systems competence;
- E. develop ideas and strategies for the workforce development and for professional recognition mechanisms within the topical areas of the panel.



Mandate 1

Mandate

1. Address the data lifecycle within a structured and integrated systems approach in HEP
 - 1.1. Formulate recommendations on organisation, technology, standards, outreach, education for past/current/future experiments.
 - 1.2. Connect regional and local activities in the field and encourage international cooperation, aiming at stimulating active participation from the global HEP community.
 - 1.3. Raise awareness of open science and the FAIR principles applied to data, software and workflows, and stimulate relevant developments.
 - 1.4. Assess the openness and FAIRness of the field.
 - 1.5. Encourage transfer of knowledge
 - 1.6. Support the ongoing projects and collaborations started within the “Data Preservation in High Energy Physics” collaboration (DPHEP) and the “Standing Committee on Interregional Connectivity” (SCIC).



Mandate 2

Mandate (cont)

2. Improve the awareness for the importance of the data lifecycle in HEP
 - 2.1. Work out and communicate the motivation of FAIR (findability, accessibility, interoperability, and reusability) principles and open science and encourage its dissemination.
 - 2.2. Organise workshops, formulate recommendations and cookbooks, issue global reports
 - 2.3. Contribute to the training and education on open science issues in all world regions, employing in particular the facilities of the large laboratories in the field.
 - 2.4. Help in sharing expertise and existing solutions; catalyse new common projects; promote collaboration.



Mandate 3

Mandate (cont)

3. Encourage and foster connections to other fields of science, to industry and to open science initiatives in order to profit from their expertise and competence in the following fields:
 - 3.1. Big and distributed data management.
 - 3.2. Data management systems.
 - 3.3. Artificial intelligence.
 - 3.4. Open science processes.
 - 3.5. Data preservation systems.
 - 3.6. Reach out to neighbouring fields such as astro(particle) physics, hadron physics, and accelerator science, but also to the communities of photon and neutron science and others with large data volumes and related data challenges (genomic, public health, smart city, ...)



Mandate 4 & 5

Mandate (cont)

4. Help in organising practical support and act as point of contact for practical issues in the field of data, software, workflows and computing
 - 4.1. Support the ongoing projects and co-operations started within DPHEP in order to maintain data sets that (can) still produce science, keep track on parked data sets
5. Improve recognition of the nature and value of work on the data lifecycle in researchers' CVs and support their career development.