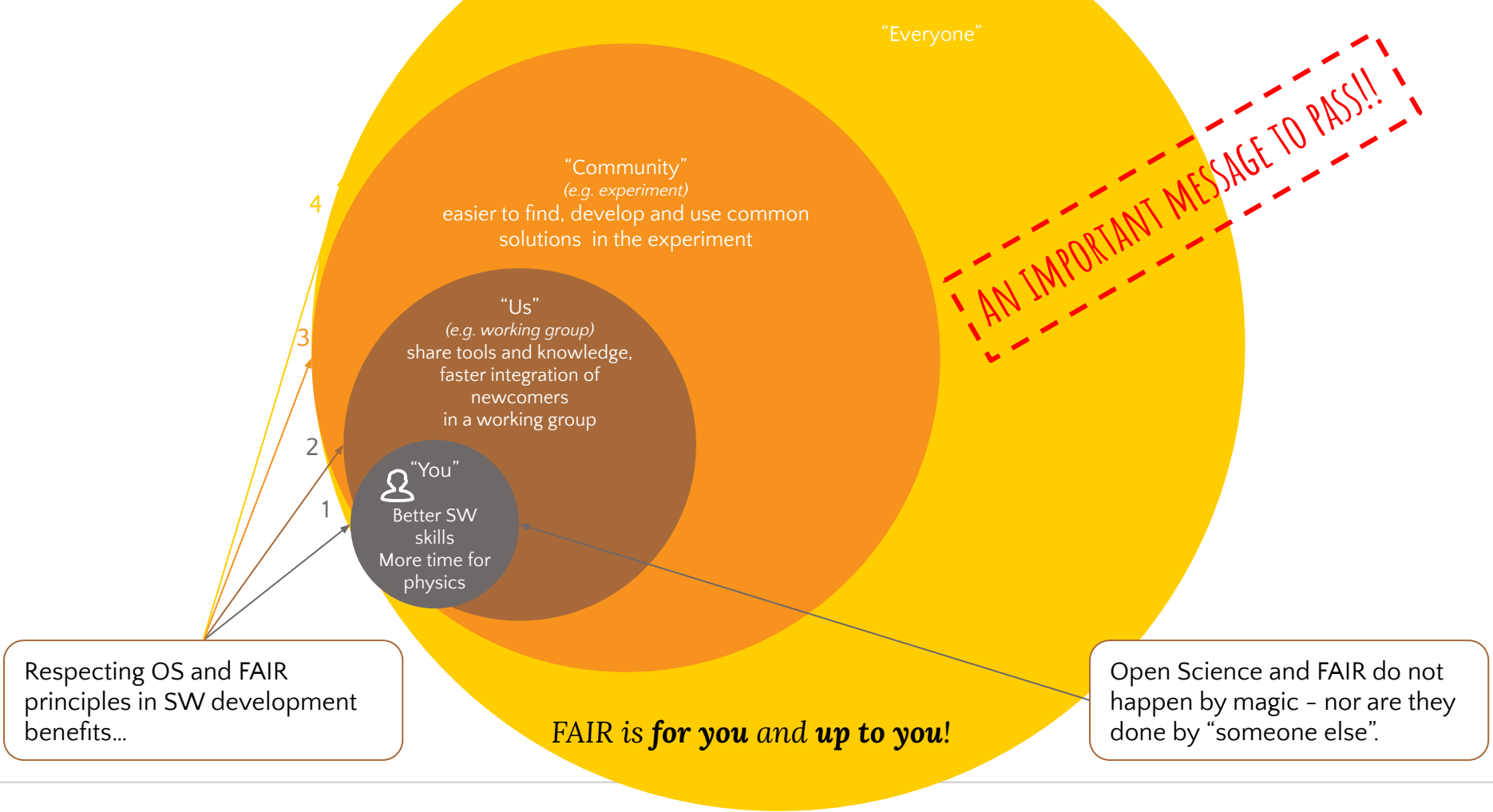
ICFA Data Lifecycle Panel: Recommendations - audiences and themes

ICFA Data Lifecycle Panel meeting - November 5, 2024

Kati Lassila-Perini
Helsinki Institute of Physics - Finland

Recommendations

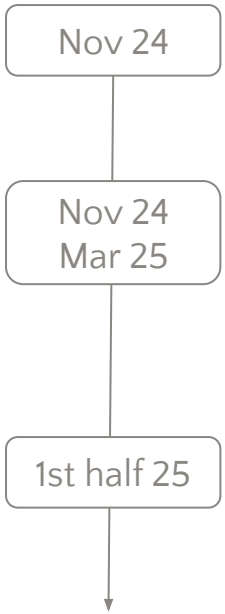
- Collect and compile best practices to achieve Open Science and FAIR into recommendations
  - concrete, specific and relevant to our domain.
  - understandable to all stakeholders: from students and analysts to the experiment management → direct the reader to the relevant content
- Input from *enablers* in our domain
  - From this panel
  - DPHEP workshop
  - HSF training workshop
- Follow the ongoing work for KPIs (Key Performance Indicators) for Open Science at CERN (and elsewhere?)
  - Recommendations and KPIs should match.


"Everyone"
"Community"
(e.g. experiment)
easier to find, develop and use common solutions in the experiment
"Us"
(e.g. working group)
share tools and knowledge, faster integration of newcomers in a working group
"You"
Better SW skills
More time for physics
4
3
2
1
AN IMPORTANT MESSAGE TO PASS!!
Respecting OS and FAIR principles in SW development benefits...
Open Science and FAIR do not happen by magic – nor are they done by "someone else".
FAIR is for you and up to you!

What to avoid?

- Repeating FAIR principles is not very useful
- Provide concrete recommendations at the level in which the reader can take action.

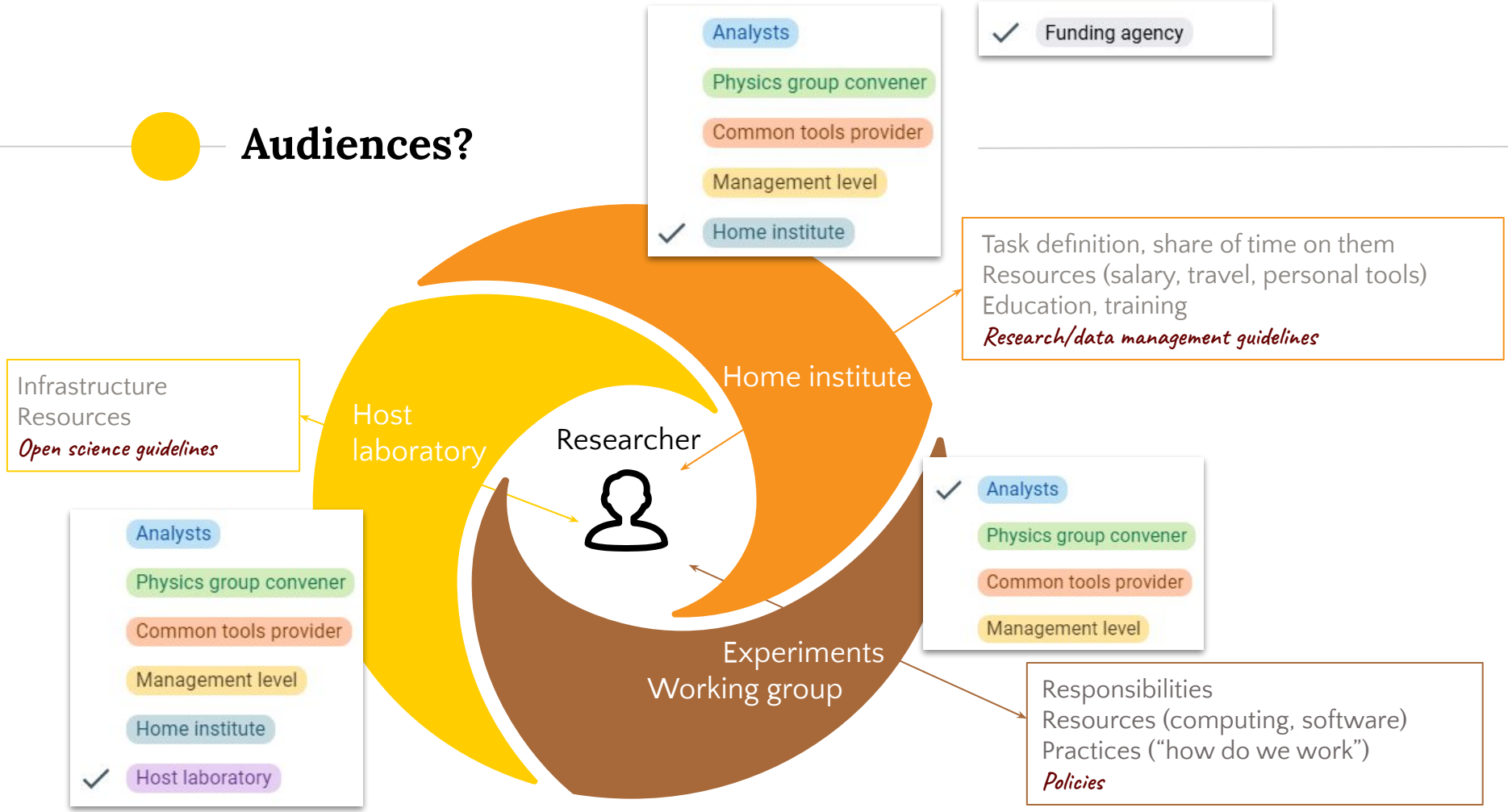
Proposal for a program of work:

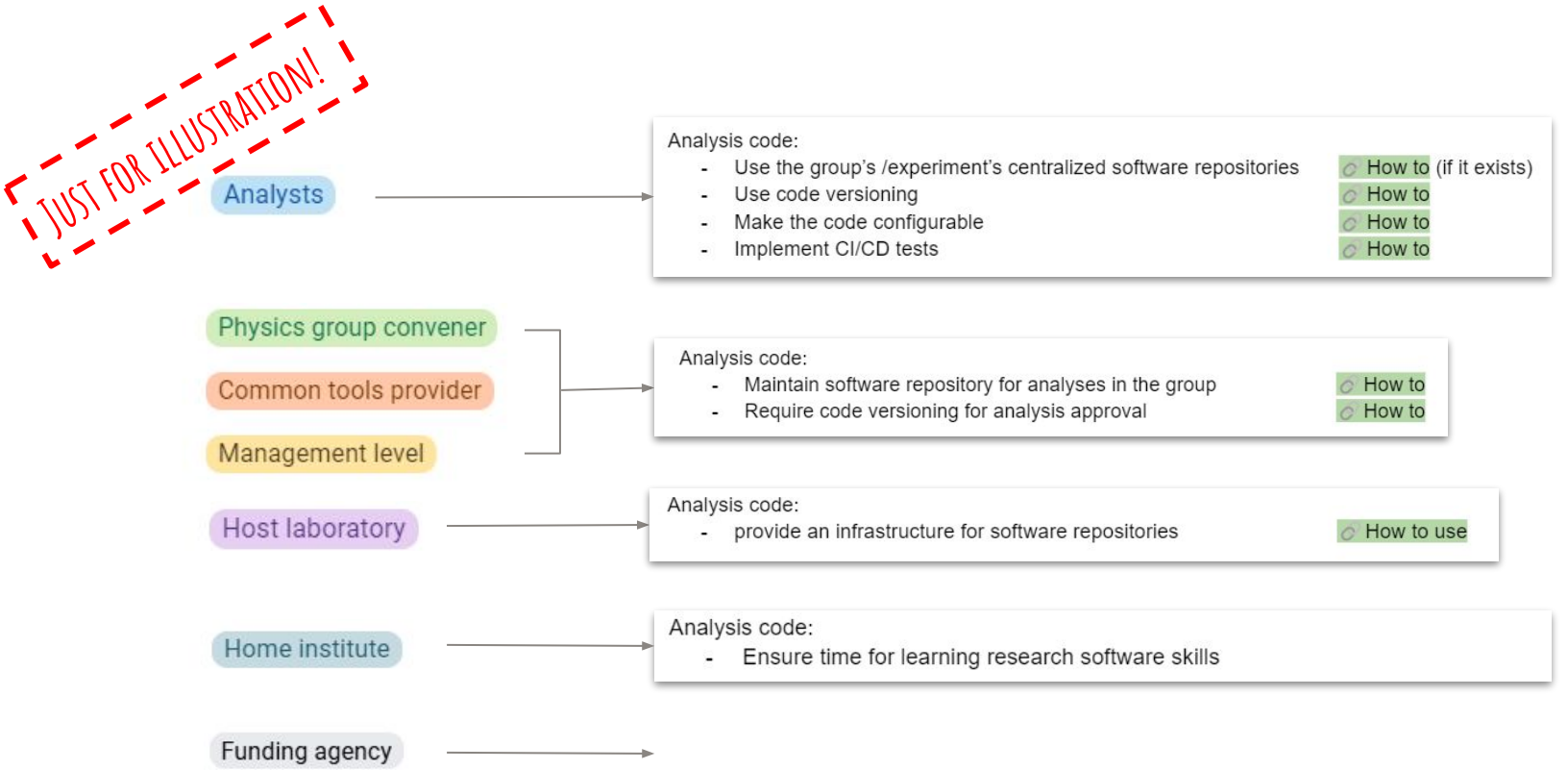


- Define initial themes and audiences - Discussion today
  - Fill in the initial draft using input from surveys
    - Match to the mandate of the DLC panel
- Elaborate the draft by a working group with volunteers from
  - this panel
  - people involved in DP/AP/docs in the past and present experiments
  - people involved in the OS/FAIR initiatives (HSF / EVERSE / FAIROS-HEP)
  - people involved in the OS KPI (Key Performance Indicator) definition
- Organize a workshop / a retreat to finalize the details
  - FAIROS-HEP can assist with funding.
- Circulate for a wider feedback.

Audiences?

Audiences?


Analysts
Physics group convener
Common tools provider
Management level
Home institute
Funding agency
Task definition, share of time on them
Resources (salary, travel, personal tools)
Education, training
Research/data management guidelines
Home institute
Infrastructure
Resources
Open science guidelines
Host laboratory
Researcher
Analysts
Physics group convener
Common tools provider
Management level
Home institute
Host laboratory
Experiments
Working group
Analysts
Physics group convener
Common tools provider
Management level
Responsibilities
Resources (computing, software)
Practices ("how do we work")
Policies


JUST FOR ILLUSTRATION!
Analysts
Analysis code:
- Use the group's /experiment's centralized software repositories
- Use code versioning
- Make the code configurable
- Implement CI/CD tests
How to (if it exists)
How to
How to
How to
Physics group convener
Common tools provider
Management level
Analysis code:
- Maintain software repository for analyses in the group
- Require code versioning for analysis approval
How to
How to
Host laboratory
Analysis code:
- provide an infrastructure for software repositories
How to use
Home institute
Analysis code:
- Ensure time for learning research software skills
Funding agency

Themes?

Initial suggestion - to be modified as needed

Themes v0.0

- Policy and management
- Infrastructure and resources
- Software skills development
- Software and workflow management
- Analysis preservation tools and practices
- Data management tools and practices
- Documentation and knowledge preservation
- Long-term sustainability

- Some examples from the survey input in the following slides

Themes v0.0

- Policy and management
  - Policies at the experiment, host lab and home institute level
    - e.g. Exp. management: Early approval of the Open Data policy
  - Dedicated groups to implement the policies
  - Dedicated SW & Computing positions
  - Recognition
  - Career evolution
  - ...

Themes v0.0

- Infrastructure and resources
  - Code repositories
    - e.g. Host lab: Provide infrastructure for software repositories
  - Open data repositories
    - Open data, HEPData
  - Analysis-specific data resources
    - Code, intermediate data, etc
  - Dedicated personpower
  - ...

Themes v0.0

- Software skills development
  - Training opportunities
    - E.g. Home inst.: Allow time for SW skill development
    - E.g. Experiment: Organize regular training on...
  - Common training efforts
    - E.g. Experiment: Contribute to and promote common training initiatives
    - E.g.: Common training curriculum
  - Recognition
    - Eg. Experiment / home institutes: Create proper and stable reward/evaluation system
  - ...

Themes v0.0

- Software and workflow management
  - Open-source SW and tools only
  - Basics SW best practices
    - Central (findable & accessible) analysis code repos
    - Git and code versioning
    - Configurable code
    - Automated tests
    - Well-defined workflows
    - Well-defined environments
  - Recognition
  - Career evolution
  - ...

Themes v0.0

- Analysis preservation tools and practices
  - Apply SW best practices to analysis code
  - Record and store analysis input in human and machine-readable form
  - Allow time for setting up AP machinery
  - Provide easy-to-use tools
  - Promote best practices
    - e.g. Exp. management: Senior leadership as the advocate and the driver of the analysis preservation practices
  - Enforce best practices from the early stage of the analysis
  - Recognition
  - Career evolution
  - ...

Themes v0.0

- Data management tools and practices
  - Open-source solutions only
  - Backward compatibility of the major data management tools
  - Data provenance information for legacy data
  - Long-term availability of key information over disruptive changes
    - E.g. Availability of file catalogs and condition and luminosity data beyond the lifetime of specific tool
  - Recognition
  - Career evolution
  - ...

Themes v0.0

- Documentation and knowledge preservation
  - Preserved analyses
  - Actionable examples
  - Preserve “silent“ knowledge
  - Recognition
  - Career evolution
  - ...

Themes v0.0

- Long-term sustainability
  - Affects most (all) of the previous themes
  - ...

Themes v0.0

- Others?
  - ...
  - ...

Next steps?


Q4
Q1
Q2
Q3
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
LOR
IPS
DOL
Collect input
Define basics
Call for volunteers
Work on details
Finalize
Circulate

1. Define basics

- Next weeks
  - 5 Nov – 22 Nov
- Contributors from this panel
  - You expertise in needed!
- Fill in the initial input
  - An introduction with a motivation text
  - Group recommendations by Themes and Audience
  - Input from the surveys
- Working mode:
  - Google doc – discuss through comments
  - Asynchronous work – Working meeting if needed

2. Call for volunteers

- Starting in the next days
  - 8 Nov - 22 Nov
- Bottom-up approach, people with hands-on experience
  - From DPHEP
    - Contact presenters from the experiments
    - Make sure not to be CERN-LHC only!
    - Make sure to have physics group convenership level presented
  - From HSF training panel
    - Contact volunteers
- Others:
  - EVERSE - RECFA panels - FAIROS-HEP - theory - CERN OS
  - mostly present in this panel

Discussion

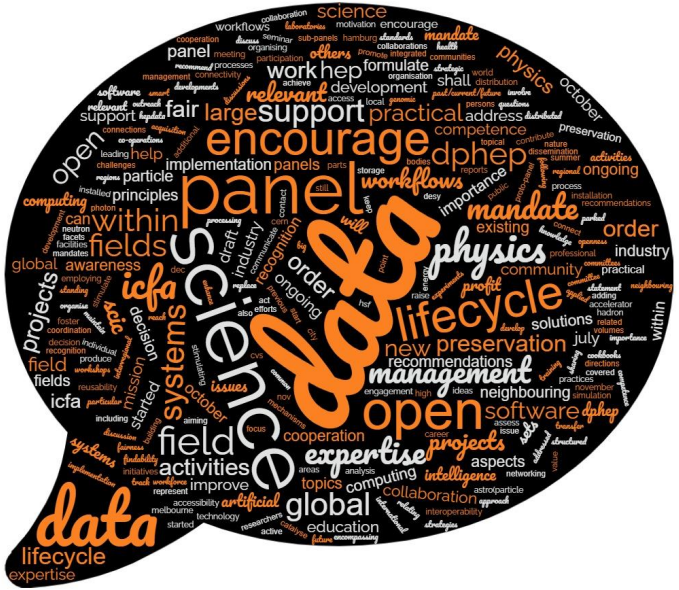
Questions?

Suggestions?

Outlook

Collect and compile best practices to achieve Open Science and FAIR into recommendations.

Program of work defined, looking forward to getting started!



ICFA statement on the Data Lifecycle Panel
Mandate of the Data Lifecycle Panel

“