# Panel discussion on Open Science and Reproducibility

LATTICE 2024, LIVERPOOL, 2024-08-02

# Plan

- Definitions and motivations
- Panelist introductions and lightning talks
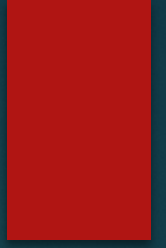- Q&A via Slido

Join at
slido.com
#9989 867

# Who am I?
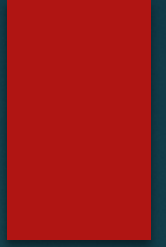
- STFC Research Software Engineer Fellow

- Background in BSM lattice quantum field theory

- Strong interest in reproducibility and openness of data analysis workflows

# Definitions

# Reproducibility

# Reproducibility

Same data + same analysis → Same results

# Reproducibility

Same data + same analysis → Same results

(from The Turing Way project)

# Reproducibility

Same data + same analysis → Same results

(from The Turing Way project)

- Related concepts:

# Reproducibility

Same data + same analysis → Same results

(from The Turing Way project)

- Related concepts:
  - Replicability:
    New data + same analysis → same results

# Reproducibility

Same data + same analysis → Same results

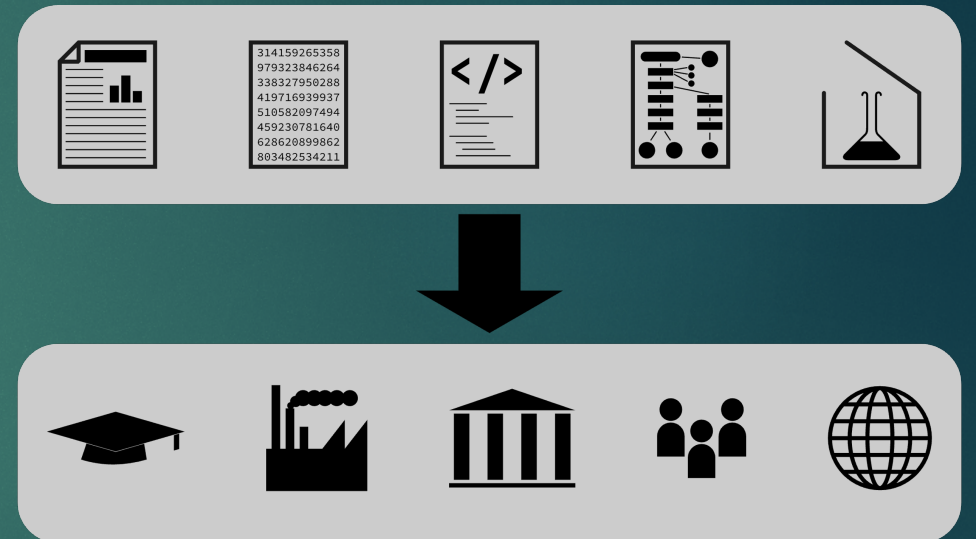(from The Turing Way project)

▶ Related concepts:

▶ Replicability:
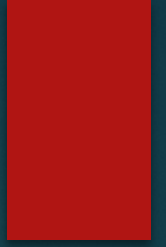New data + same analysis → same results

▶ Robustness:
Same data + new analysis → same results

# Open science

▶ The movement to make all research accessible to all levels of society.

▶ Including, but not limited to:

    ▶ Publications

    ▶ Physical samples

    ▶ Data

    ▶ Software

▶ A direction to be strived in

# FAIR

# FAIR

Research data (and software) should be:

# FAIR

Research data (and software) should be:

▶ Findable

# FAIR

Research data (and software) should be:

- Findable
- Accessible

# FAIR

Research data (and software) should be:

- Findable
- Accessible
- Interoperable

# FAIR

Research data (and software) should be:

- Findable
- Accessible
- Interoperable
- Reusable

# FAIR

Research data (and software) should be:

- Findable
- Accessible
- Interoperable
- Reusable

*Independently of whether the data are open*

# Motivations

# Nullius in verba



THE ROYAL SOCIETY'S HOUSE IN CRANE COURT (*see page* 104).

# Public money ⇒ Public results

# Our funders are requiring it!

**Science and Technology Facilities Council**

Policy

**STFC scientific data policy**

Policy providing guidance on data management to the staff and communities of the Science and Technology Facilities Council (STFC)

From: STFC
Published: 1 April 2016

*Data resulting from publicly funded research should be made publicly available... unless there are specific reasons (e.g. legislation, ethical, privacy and security) why this should not happen*

# Introductions

# Who am I?

- Associate Research Scientist at The Cyprus Institute

- Background in Lattice Gauge Theories

- Strong interest and involvement in European Commission projects on Open Science and the European Open Science Cloud (EOSC)

# More to be done for Open Science!

- ▶ LGT community has launched sharing data
  - ○ ILDG (2002)
- ▶ However, OS has evolved little since then
- ▶ Funding Agencies are imposing OS
  - ○ in Horizon Europe, OS is "modus operandi"
- ▶ LGT community is falling behind
- ▶ Need to catch up…
- ▶ Harmonize with bigger initiatives (EOSC)

# Do we need Initiatives?

- ▶ Open Science initiative in LGT?
  - ○ Identifying the needs
  - ○ Creation of tools for RDM
  - ○ Semantic advancements
    - ▪ Metadata schemas?
    - ▪ Ontological schemas?
  - ○ Use Cases
  - ○ Training agenda
- ▶ Begin with something of a smaller scale
  - ○ Support capacity building
  - ○ Support ILDG?

European Commission

**INFRA-EOSC calls?**

OSCARS
Open Science Clusters' Action
for Research & Society

www.oscars-project.eu

# Who am I?

▶ Professor for theoretical physics at University of Bonn

▶ Background in Lattice QCD and algorithms

▶ ILDG data sharer for 15 years

▶ Practical interest in research data management

# ILDG is our Asset

- ▶ Lattice community was way ahead of its time
- ▶ First of all, ILDG enables collaborations
  - ○ Within the community
  - ○ Within collaborations
- ▶ In addition,
  - ○ Supports good scientific practice (reproducibility, etc...)
  - ○ Helps with funding agencies

# ILDG is our Asset

- ▶ We should try to maintain and develop it further!
  - ○ And we made big progress recently!
- ▶ While keeping in mind
  - ○ Our needs and requirements
  - ○ Open science shouldn't end in itself
- ▶ But it requires a community wide effort!
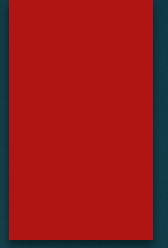- ▶ Should build on new national funding schemes for RDM, like PUNCH4NFDI in Germany

# Anna Hasenfratz

▶ Professor of physics at University of Colorado Boulder, USA

▶ Background in lattice QFT: QCD, BSM, new phases (SMG); improved actions, RG on lattice

▶ Interested in research data management and sharing

# Goal of data sharing:

- ❖ Validate numerical predictions
  - ○ Investigate unconventional systems:
    - ▪ Conformal or exotic phases – UV and IR fixed points
    - ▪ Emerging UVFP ? Symmetric Mass Generation
  - ○ Calculate "unconventional" quantities:
    - ▪ Order parameter of exotic phases
    - ▪ Scaling dimensions, RG beta function, hyperscaling, etc
- ❖ When controversies arise, subject all data from different groups to the same analysis
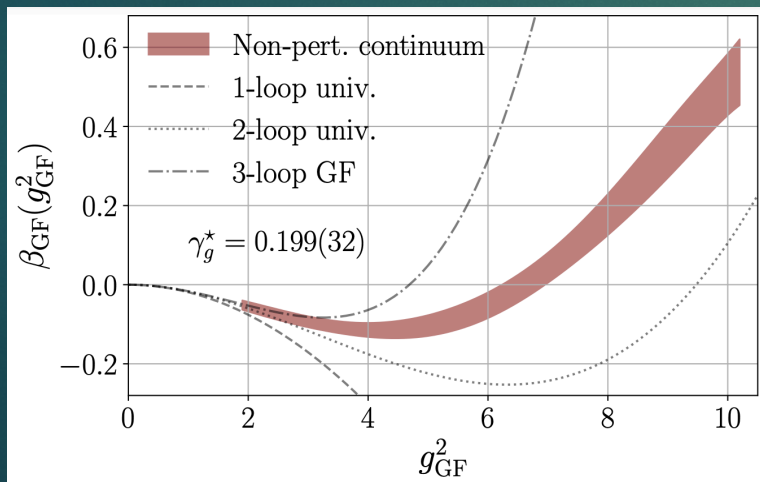- ❖ Share data, share analysis code ?

# How much do we share?

How much effort is required from both parties?
Example:

AH, Peterson, PRD 109 (2024) 11, 114507
12-flavor beta function
Data uploaded to Zenodo

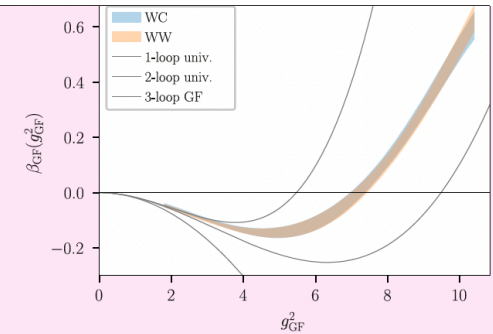Re-analyzed and validated by Bennett et al
(poster Lattice 2024)

# Louise Chisholm

- Joint Director UK Square Kilometer Array (SKA) Regional Centre

- Background: Professional Research & Strategy Manager, launched and led UCL-wide eResearch Community, PhD in Epigentics

- Interested in developing digital research infrastructure ecosystems, FAIR Data sharing communities
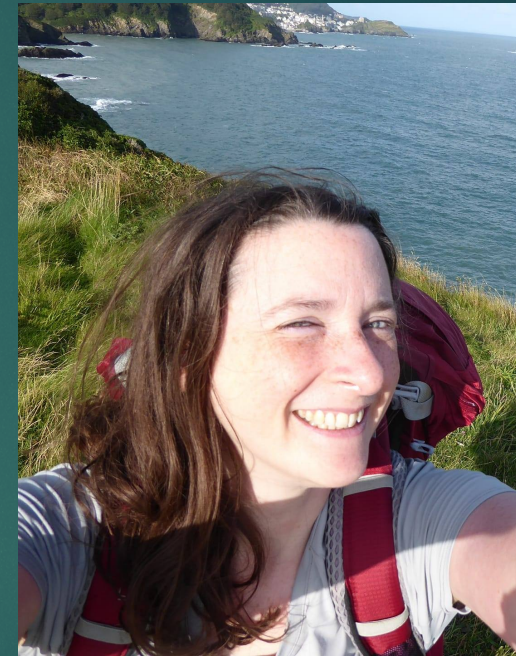
# Square Kilometre Array: next-generation radio astronomy facility

- Developing an end-to-end system from telescope time proposal to science delivery
- Enabling FAIR data sharing: metadata standards, data sharing policies, proprietary periods

**UNDER CONSTRUCTION**

SKA-Mid

SKA-Low

17.01.2024

## SRC | Net
### SKAO Regional Centre Network

**Science Enabling Applications**
Analysis Tools, Notebooks, Workflows execution Machine Learning, etc

**Distributed Data Processing**
Computing capabilities provided by the SRCNet to allow data processing

**Data Discovery**
Discovery of SKA data from the SRCNet, local or remote, transparently to the user

**Visualization**
Advanced visualizers for SKA data and data from other observatories

**Support to Science Community**
Support community on SKA data use, SRC services use, Training, Project Impact Dissemination

**Interoperability**
Heterogeneous SKA data from different SRCs and other observatories

**Data Management**
Dissemination of Data to SRCs and Distributed Data Storage

# Challenges & opportunities

**FAIR Data Accelerator**
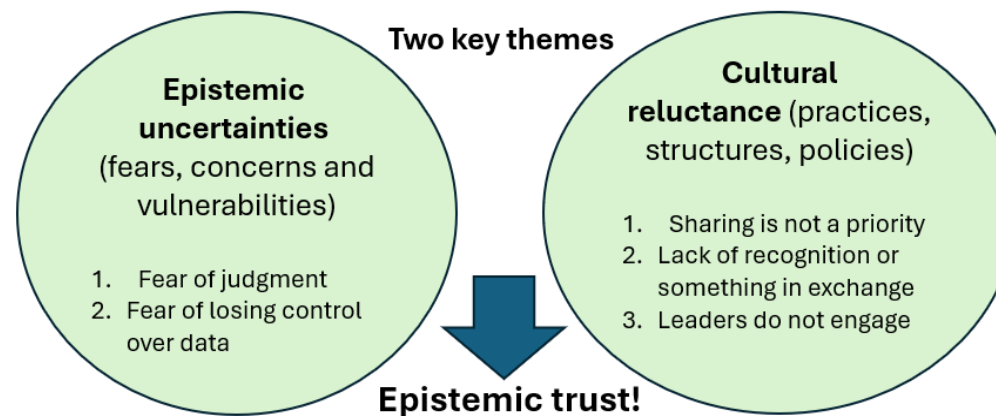
- Learning across disciplines
- Approach to proposing and undertaking research
- Team Science
  - New professions
- Funding mechanisms
  - Harness data resources
  - Infrastructure
  - People



**Ethnography Study identified social and cultural barriers to implementing the FAIR principles**

Two key themes

**Epistemic uncertainties** (fears, concerns and vulnerabilities)

1. Fear of judgment
2. Fear of losing control over data

**Cultural reluctance** (practices, structures, policies)

1. Sharing is not a priority
2. Lack of recognition or something in exchange
3. Leaders do not engage

**Epistemic trust!**

http://www.uksrc.org/fair-data-pilot/

Identity Negotiation and Professional Development in the Context of the Digital Transformation of Research Communities
Duran del Fierro et al.   https://doi.org/10.5281/zenodo.12806507

Q&A