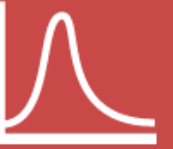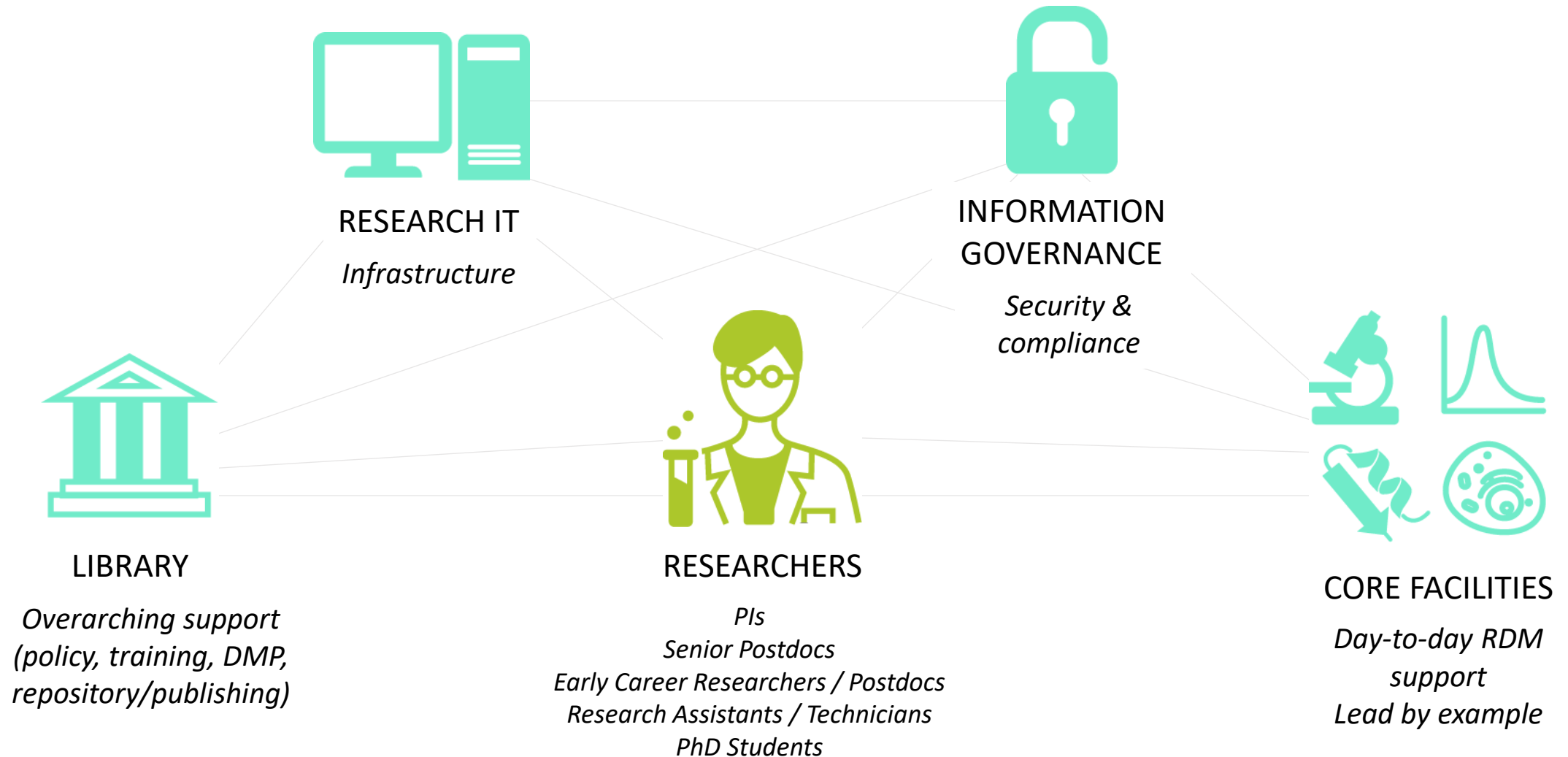# Building RDM best practice into Core Facility workflows

# FBMH Core Research Facilities

- 9 Core Facilities:

  Bioimaging, Sequencing, Mass Spectrometry, Electron Microscopy, Bioinformatics, Bimolecular Analysis, Genome Editing, Fly Facility, Flow Cytometry.

- Major data-generators in the Faculty.

- Core Facility data central to many research projects.

- Central managerial structure.

# Where do Core Facilities fit in the RDM landscape?



**RESEARCH IT**

*Infrastructure*

**INFORMATION GOVERNANCE**

*Security & compliance*

**LIBRARY**

*Overarching support (policy, training, DMP, repository/publishing)*

**RESEARCHERS**

*PIs*
*Senior Postdocs*
*Early Career Researchers / Postdocs*
*Research Assistants / Technicians*
*PhD Students*

**CORE FACILITIES**

*Day-to-day RDM support*
*Lead by example*

# How can Core Facilities help?

- Survey / discussions with PIs and Postdocs

STORAGE & TRANSFER

- Accessibility
- Capacity
- GDPR/security
- Moving TBs of data

ANNOTATION

- Metadata collection & sharing (meaningful data)
- Controlled vocabularies

SHARING

- Internally
- External collaborators
- Publicly
- +/-

TOOLS

- eLab books
- Data/project management tools

# RDM strategy for FBMH Core Research Facilities

**Mission Statement:**
Efficient, streamlined data management solution to support researchers and maximise the use of Core Facility-generated data - **with minimal administrative hurdles!**

**Goals:**

**1) Single storage location**

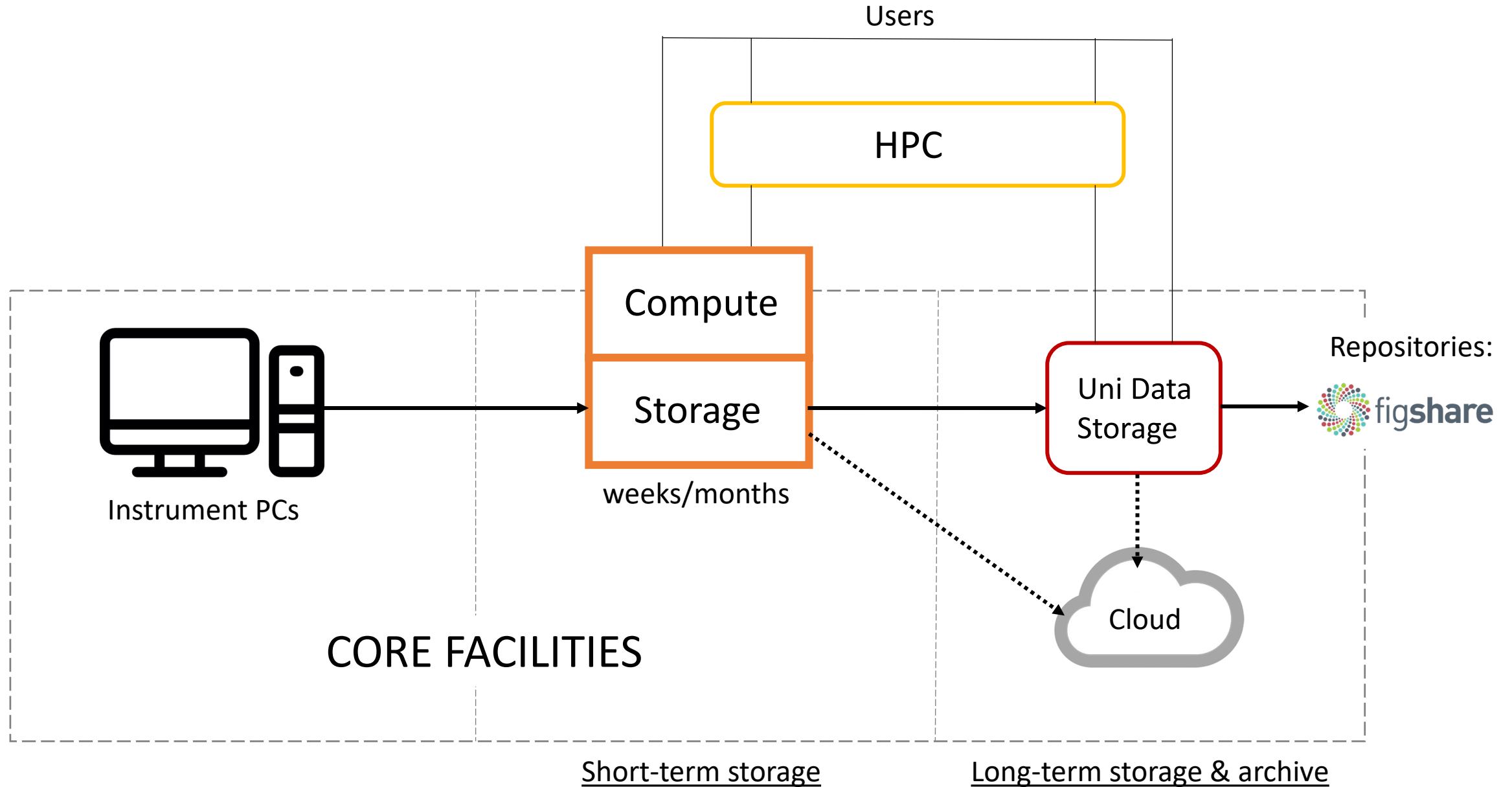- **Known and accessible to the PI**
- Avoid duplication

**2) Streamline data transfer**

- Moving data between short/long-term storage
- Minimise data transfer steps…
- Bringing compute to the data

**3) Metadata collection**

- Providing meaning, maximise data use/reuse
- Important for reproducibility & data integrity

# Our Approach: Local Storage and Compute Platform Model



Users

HPC

Compute

Storage

weeks/months

Instrument PCs

CORE FACILITIES

Uni Data Storage

Repositories:

figshare

Cloud

Short-term storage

Long-term storage & archive

# How we can embed best practice into our workflows?

1) Metadata collection

→ high-level via PPMS (Findable)
→ detailed through Core Facility metadata templates (Reusable)
→ controlled vocabularies (Interoperable/Reusable)
→ possibility of porting metadata to the institutional repository (private/internal/Findable/Accessible)

2) File naming/formatting (Interoperable)

3) Training – do I really need that data?
   → Quality control step
   → Experimental design
   → Wider context
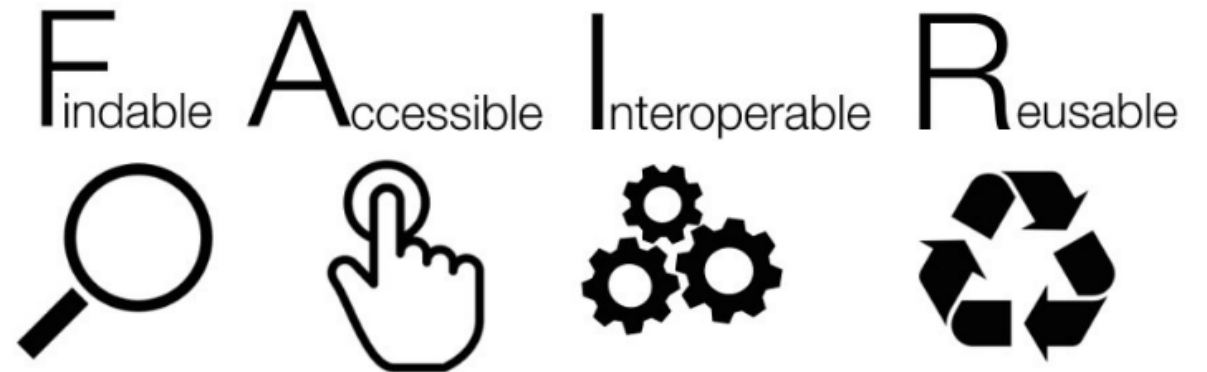
# Challenges we face…

- Volume and complexity of the data

- Technical

- Design for scalability & adaptability

- Transfer of ownership