Analysis Facilities

User interface perspective

Pierre Schnizer, Tim Ruhe et al.
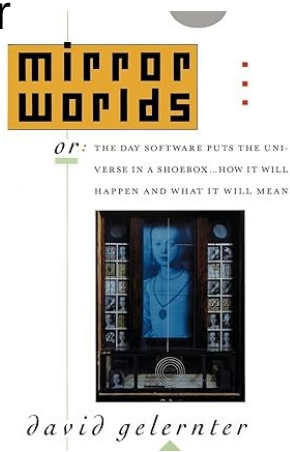für DIGUM Topic Group User Interface

DIG-UM

User Interface: Motivation

- Large scale data → exponential groth rate
- HEP phyics data rates → other ERUM Communities

- Challenges for users (typical response)
  - Get their data
  - Archive their data
  - Reduce their data
  - Get measurement result

- Users: fairly good idea of their job
- Need
  - Standard task → off load
  - Clean stable interface
  - „Hang their job on the trellis"

We are drowing in information
we are starving for knowledge
Rutherford D. Roger

Mankind was always better in acquiring data than interpreting it.
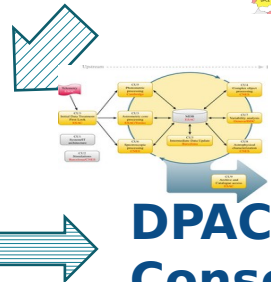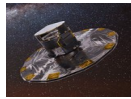
-- Mirror Worlds, David Gelernter

mirror worlds

or: THE DAY SOFTWARE PUTS THE UNI-
VERSE IN A SHOEBOX...HOW IT WILL
HAPPEN AND WHAT IT WILL MEAN

david gelernter

P. Schnizer T. Ruhe *et al.*

DIG-UM

Users, large scale facilities

Data taking	Data processing Data curation	Data publication	Data exploitation

ESA:
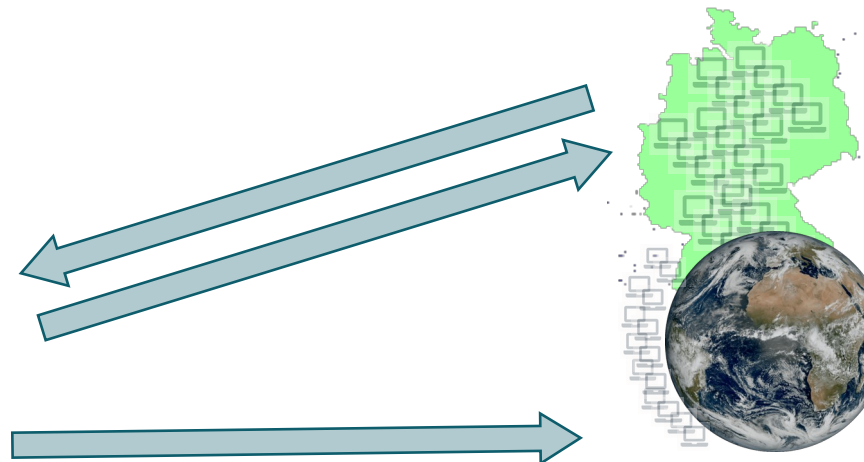Satellite
Operation

**DPAC+
(Partner)
DataCenters:**

**DPAC
Consortium**

Images: ESA + DPAC

Good solutions appied here

Example GAIA

UI ↔ Federated infrastructure: Server ↔ client access
UI ↔ Research data management:
    * Data metadata Layout
    * Workflow
UI ↔ Big data analytics: Algorithm interface

Outreach: spread the info

3

P. Schnizer T. Ruhe *et al.*

DIG-UM

Users, large scale facilities

Data taking	Data processing Data curation	Data publication	Data exploitation

UI ↔ Federated infrastructure: Server ↔ client access
UI ↔ Research data management:
    * Data metadata Layout
    * Workflow
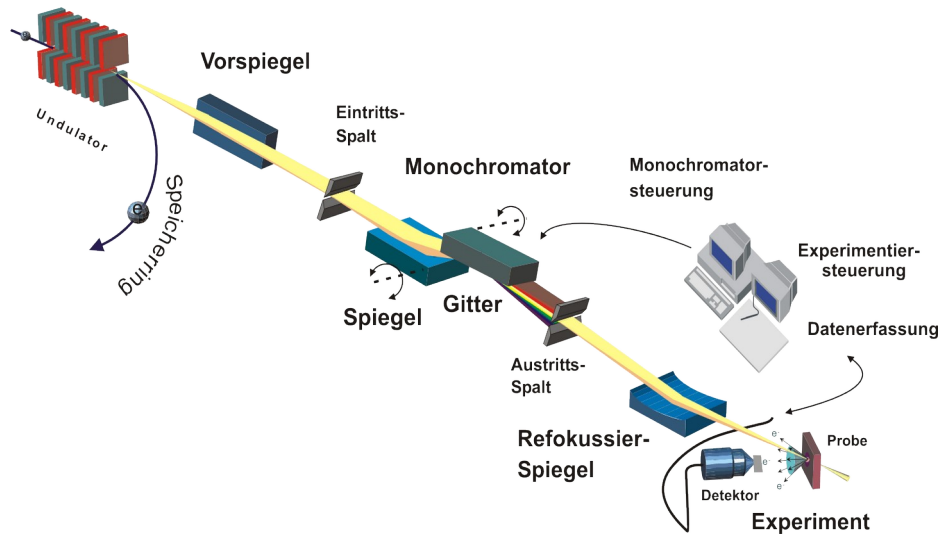UI ↔ Big data analytics: Algorithm interface

Outreach: spread the info

Good solutions appied here

P. Schnizer T. Ruhe *et al.*

DIG-UM
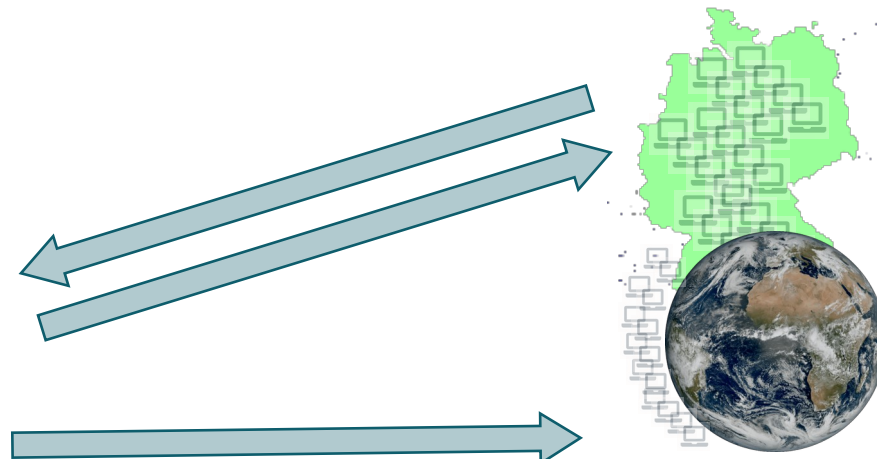
User Interface: expectation

Target: Ideal User Interface (Requirement specification)

- Data
  - search: as simple as using Scooogle
  - Load: as straight forward as a „pip install"
- Analysis
  - Scientific analysis as simple as an Afternoon Stroll
  - Intuitive, validated Workflows
  - Scientific work bench recurrent tasks made easy
- → Let's scientists refocus on sciences (again)

Dieter Rams: Principles of good design
- innovate
- Maked product useful & understandable
- Unobstructive / honset
- Long lasting
- Thought trough to the end

P. Schnizer T. Ruhe *et al.*

DIG-UM

User Interface: user centric

Handling user demands:

- Put first: what user **really needs**

- Implicit demand (seldom mentioned)
  available soon
  experience to build up

- Does user consider?
  Work part of a large(r) work flow?
  Need to be rerun in 10 years?

Don't forget service recovery paradoxon

P. Schnizer T. Ruhe *et al.*

DIG-UM

2023 Workshop @ BERLIN

Foto: Angela Warkentin, ERUM Data Hub

Experten Workshop
„Next Generation Environment for Interoperable Data Analysis"

3-4 Mai 2023, HZB Berlin

Sprecher:

- Nicolas Eich, VISPA, RWTH
- Verena Kain, CERN
- Mohammad Al-Turany, GSI / FAIR
- Kai Polsterer, HITS, Heidelberg

P. Schnizer T. Ruhe *et al.*

DIG-UM

Workshop Participants: Data products

„Higher Order Data Products"

- Similar to data catalogs
- Data and work flows
- Versioning:
  - Original data
  - Analysis pipeline
- Motivation:
  - Flat basis curve
  - Based on established design patterns

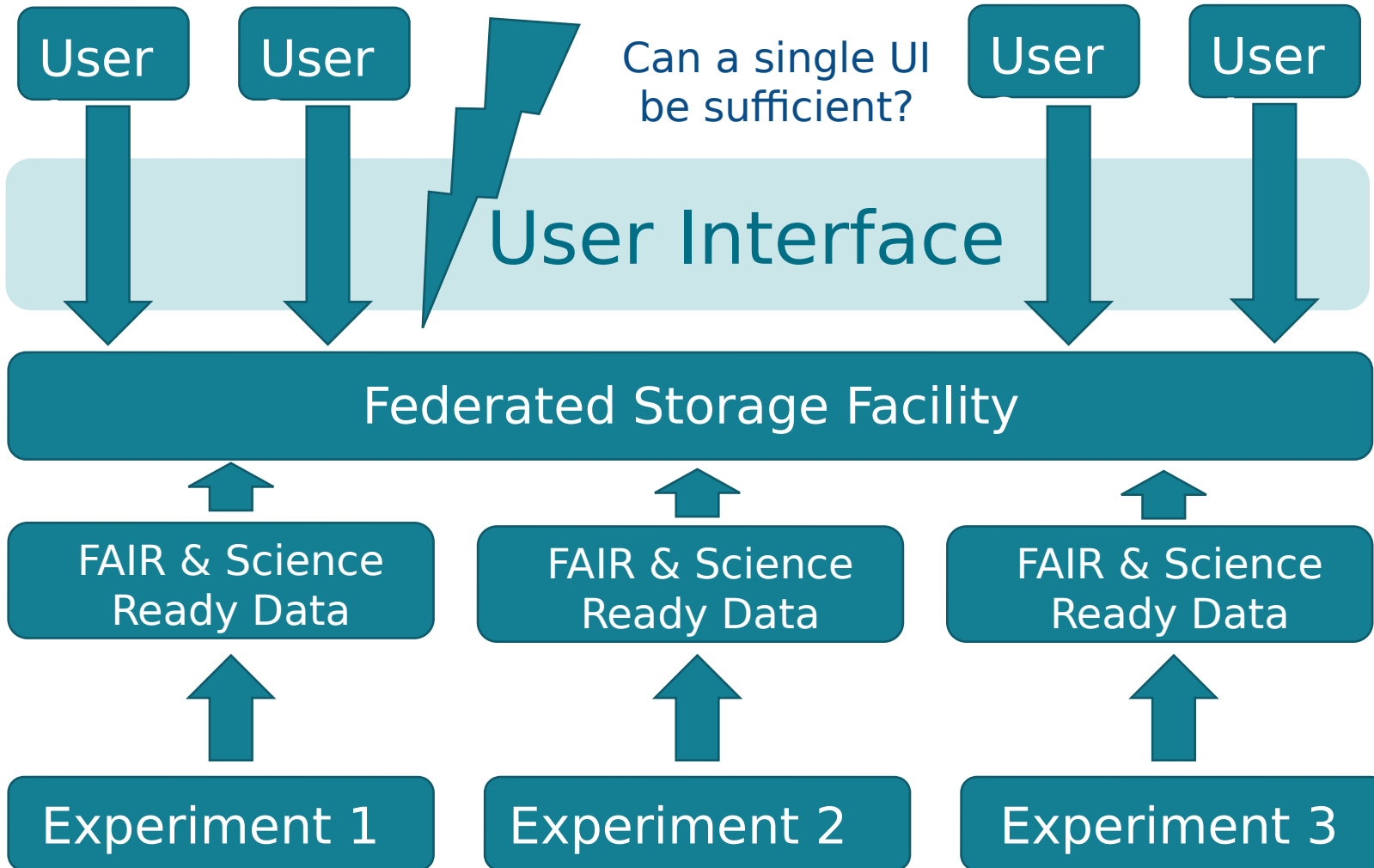- Avoid: "not invented here antipattern"

Workflow: Support
- Validated
- Dokumented
- Archived

Many communities:

Will require support for data product design
→ example / best practice repositories
→ not one fits all
→ design decision guidelines

P. Schnizer T. Ruhe *et al.*

DIG-UM

Enabling Interdisciplinary Discoveries

| User | User | | User | User |

Can a single UI be sufficient?

User Interface

Federated Storage Facility

| FAIR & Science Ready Data | FAIR & Science Ready Data | FAIR & Science Ready Data |

| Experiment 1 | Experiment 2 | Experiment 3 |

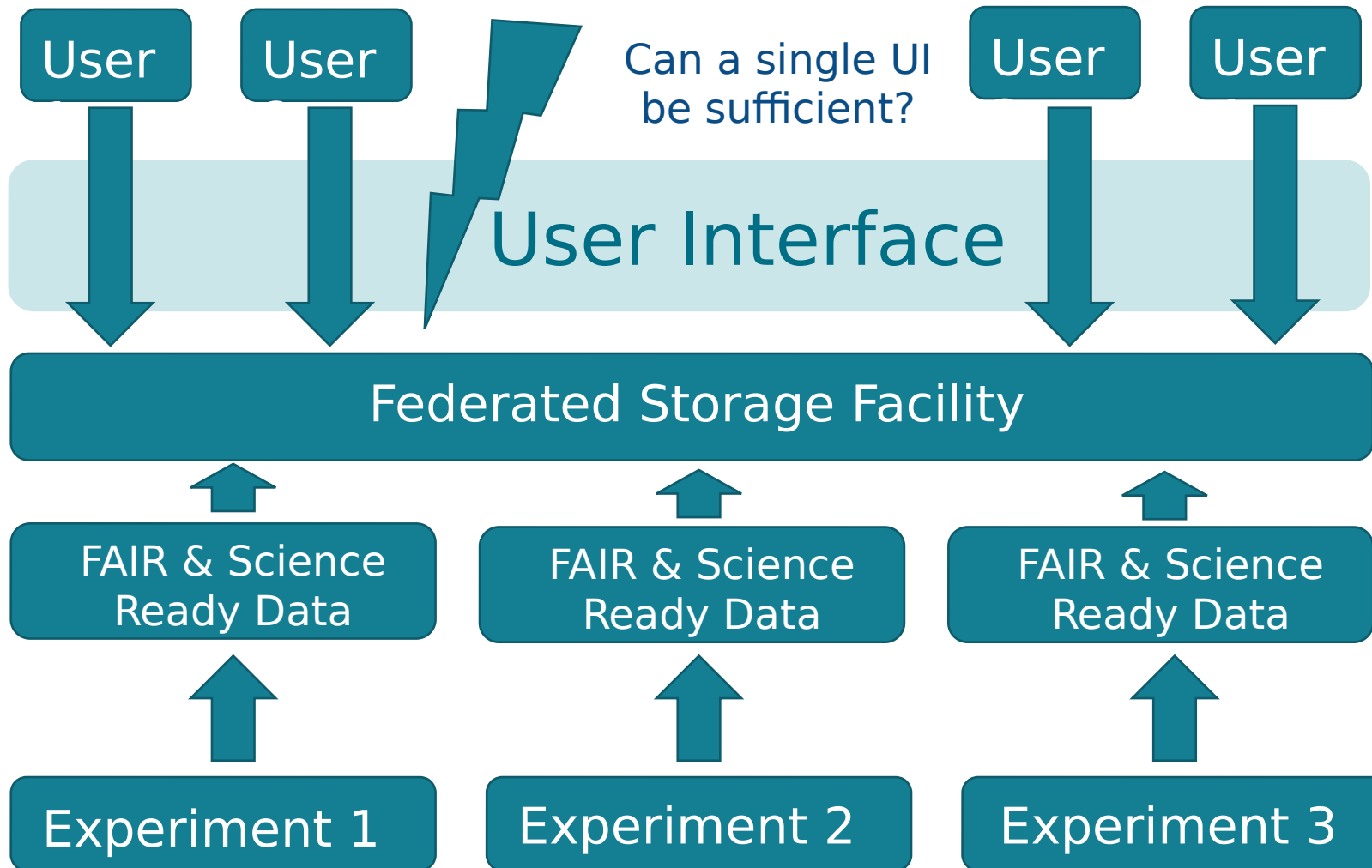The best user interface is probably the one you don't notice.

Comfortable look and feel.

Data processing and data curation remains with the experiments.

Experiments should be encouraged to publish their data in a timely manner.

Interdisciplinary analyses require cross-community standards.

DIG-UM

Enabling Interdisciplinary Discoveries

User

User

User

User

Can a single UI be sufficient?

User Interface

Federated Storage Facility

FAIR & Science Ready Data

FAIR & Science Ready Data

FAIR & Science Ready Data

Experiment 1
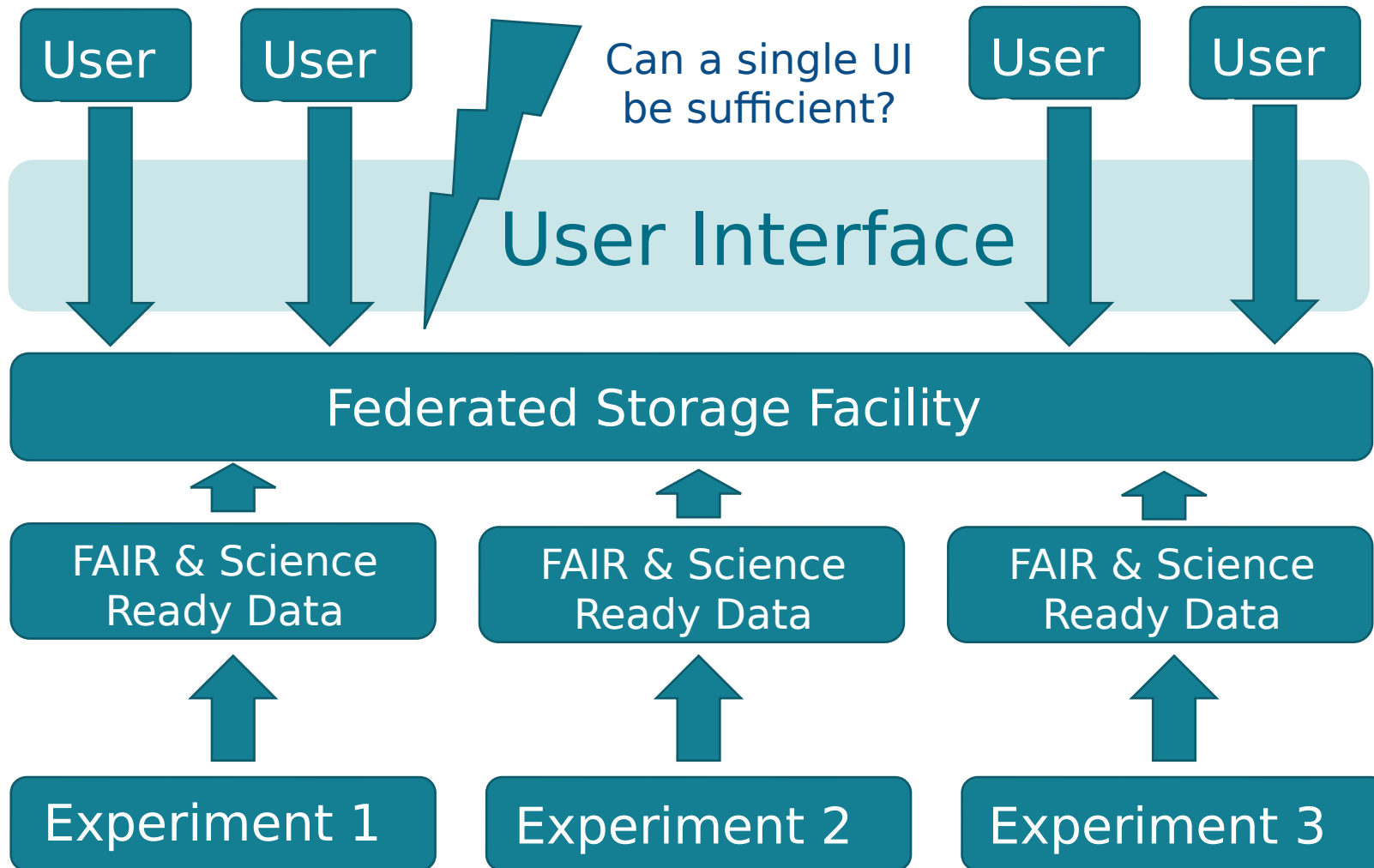
Experiment 2

Experiment 3

The UI is the duct-tape connecting the user to the data…

… and the WD-40 that helps to run everything smoothly.

WD-40

DIG-UM

Enabling Interdisciplinary Discoveries

User | User | Can a single UI be sufficient? | User | User

User Interface

Federated Storage Facility

FAIR & Science Ready Data | FAIR & Science Ready Data | FAIR & Science Ready Data

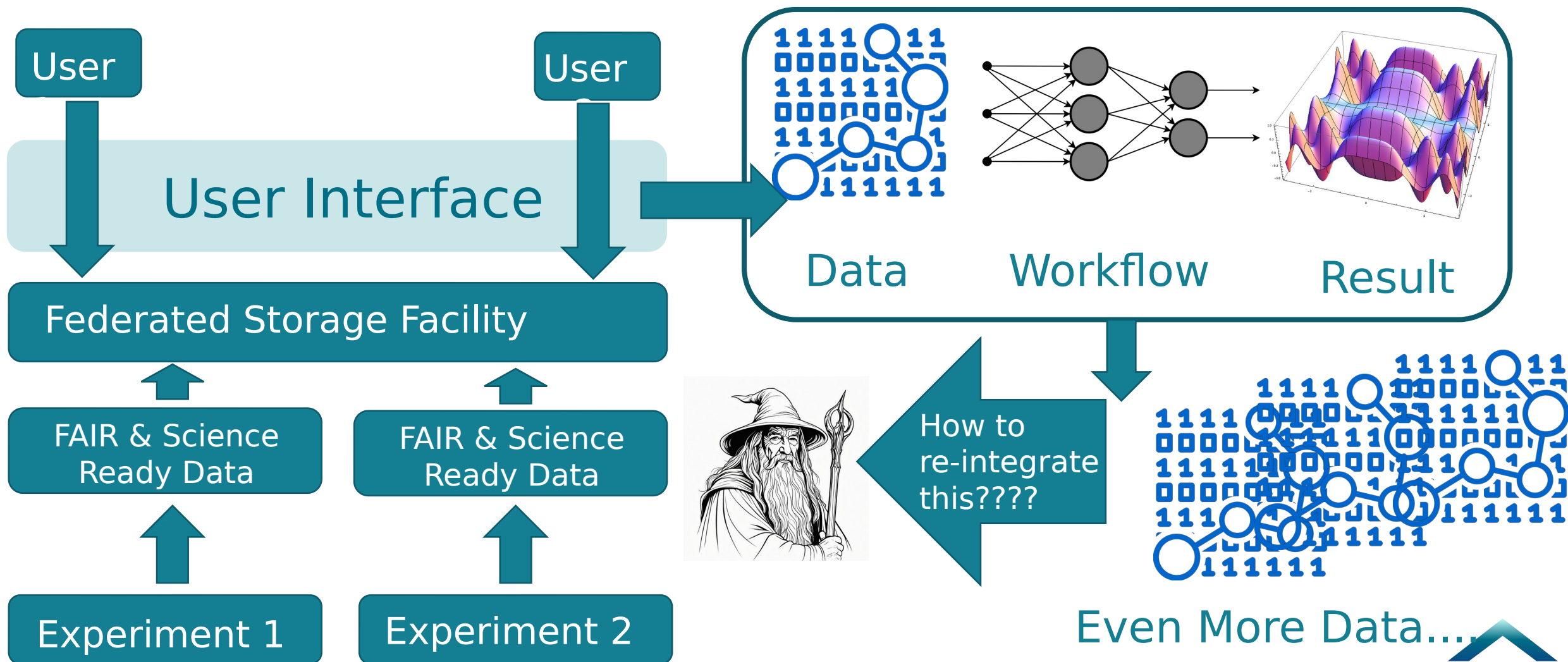Experiment 1 | Experiment 2 | Experiment 3

Data processing and data curation remains with the experiments.

Easy for large international collaboration, but a challenging task for smaller groups of researchers!

Requires awareness and training!

P. Schnizer T. Ruhe *et al.*

DIG-UM

Additional Aspects: reintegartion of results

User

User

User Interface

Federated Storage Facility

FAIR & Science Ready Data

FAIR & Science Ready Data

Experiment 1

Experiment 2

Data   Workflow   Result

How to re-integrate this????

Even More Data....

DIG-UM

User Interface: European history lesson

(Western) European history:

- Roman empire heritage
  - Common language → Lingua franca
  - On common values → understood by many
  - Legal systems → required abstractions

- Lesson
  - Language understood by many
    Python? (3rd millenium's BASIC)
    Equivalent workflow language ?
  - Abstraction
    - Workflow organisation
    - Common language (need for dialects?)
    - Split into managable task

Trellis Johann Jaritz Wikipedia

Work flow: Analysis split up

- Separate computing steps:
  User defined (small) programs
- Combining them to full analysis Workflow

P. Schnizer T. Ruhe *et al.*

DIG-UM

User Interface: analysis facilities requirements

- Work flow orchestration
  - Common language: variations if required
  - Can be validated without being executed?
  - Support user in writing
    beyond e.g. gitlab's pipeline editor
    IDE like → suggest proper usage
    Notebook like → evaluate while you write

- User programs
  → Single-responsibiltiy task
  → Only report errors

Analysis split up

learn from Erlang

- Separate computing steps:
  User defined (small) programs
- Combining them to full analysis Workflow
- Analysis facility
  - Orchestration
  - Nodes: report errors and fail
    → simplifies node implementatoin

- Need to satisfy requests
  - Interactive
  - Data scientist work flow

P. Schnizer T. Ruhe *et al.*

DIG-UM

User: why does it take so long

- Work flow run inspection
  - Data access / cache
  - Working nodes
  - Wait for scheduling
    - Processor / nodes / GPU available
    - Tasks running
    - Recovery
  - Forecast to end: finished after
    - Coffee ?
    - Lunch ?
    - Meeting

  - In this case:
    consistency is more important than
    correctness

Inspiration from

- Parallel programming profiler tools
- OS & applications profiler

  Visualise:
- Jobs running
- Waiting for
  - Data transfer?
  - Calculation completed

P. Schnizer T. Ruhe *et al.*

DIG-UM

Analysis facilities: interoperability

- Current technology driver: HEP
- These days: centred on CERN
  - Smallest common denominator → large

- Other ErUM Communities
  - Small, university size
  - Embedded in community
    - loosely connected
    - Changing partners → different fields

- → to be successful
  - Combining different data sources
  - Different fields

User questions

- Entry leve / learning curve ?
- Access → administrative barrier ?
- Availabe facilities: given
  - Dataset A
  - Code B
  - Access C

    Which facilities can the analysis be executed on ?

- Can data from different providers being combined?

P. Schnizer T. Ruhe *et al.*

DIG-UM

Beyond Workflows and Data

Who knows what?

? ?

? ?

? ?

Version 0.1

Image Source: By MCruz (WMF) - Own work, CC BY-SA 4.0,
https://commons.wikimedia.org/w/index.php?curid=34370213

Knowledge needs to be decoupled from individual persons!

 Peer Programming/Co-Pilot
 Utilize existing code and existing workflows
 Organize existing code
 Suggestions for re-usable code
 Pattern recognition
 Semi-automatic generation of code

~~Chat~~GPT

Physics

DIG-UM

How Can I Support the Effort?

Next Generation Environment for
Interoperable Data Analsyis –
Second Annual Expert Workshop

Dortmund: September 17th to 18th
https://indico.desy.de/event/45148

Bi-weekly Zoom-
Meetings!
Join the Email List:

https://lists.rwth-aachen.de/postorius/lists/erum-data-user-interface.lists.rwth-aachen.de/

Conclusion

- User interface: fit to what user needs
  - Today
  - Tomorrow
  - Student joining lab / collab in 10 years
- Intuitive design → long lasting
- Use same (similar) interface across facilities → inter disciplinary work
- Support
  - Non expert user
  - Working within small lab
- Split in program ↔ workflow
  - Can parts be validated without being executed ?

- Analysis facilites ← source of data
- (Re)integration of analysis results Linked to
  - analysis workflow
  - Used data

- Be prepared for users beyond your home community

P. Schnizer T. Ruhe *et al.*

DIG-UM

The 10 year from now challenge

Source code available ?
Language version still supported ?
Can run original binary ?

Helpful ?
- High level abstraction: eg. Work flow in work flow language
- Reuse common packages
-

10.06.2024

T. Ruhe

DIG-UM