

0 0

0 0

0 0

PP&B - Computer Support Group



THE FUTURE ROLE OF PP&B AT THE FHI

TACKLING CURRENT AND FUTURE CHALLENGES OF SCIENTIFIC IT

Science

Simeon D. Beinlich^{1,*} and Heinz Junkes¹

PP&B - Computer Support Group Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin, Germany

*beinlich@fhi.mpg.de



(SOME) CHALLENGES IN SCIENTIFIC IT

Doing this analysis the 100th time...

What did he/she do?

Where did I store the data?

My Nextcloud is full...

I can't get this to work again...

(----------

FAIR principles?

Experiment and simulations?

What happened to this sample before?

My data is gone...

My USB stick broke...

Funding!

Recreating this plot for the 10th time...

This only runs on Windows XP...

We could do this smarter...

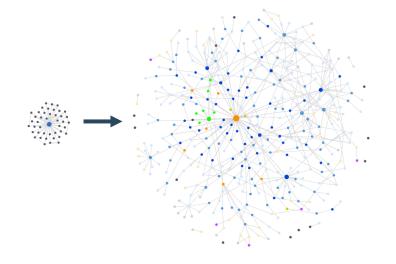
Someone changed the settings...



CHALLENGES IN SCIENTIFIC IT

More data

More complex data



More complex analyses

More complex setups

More requirements



CHALLENGES IN SCIENTIFIC IT

More data

More complex data



More complex analyses

More complex setups

More requirements



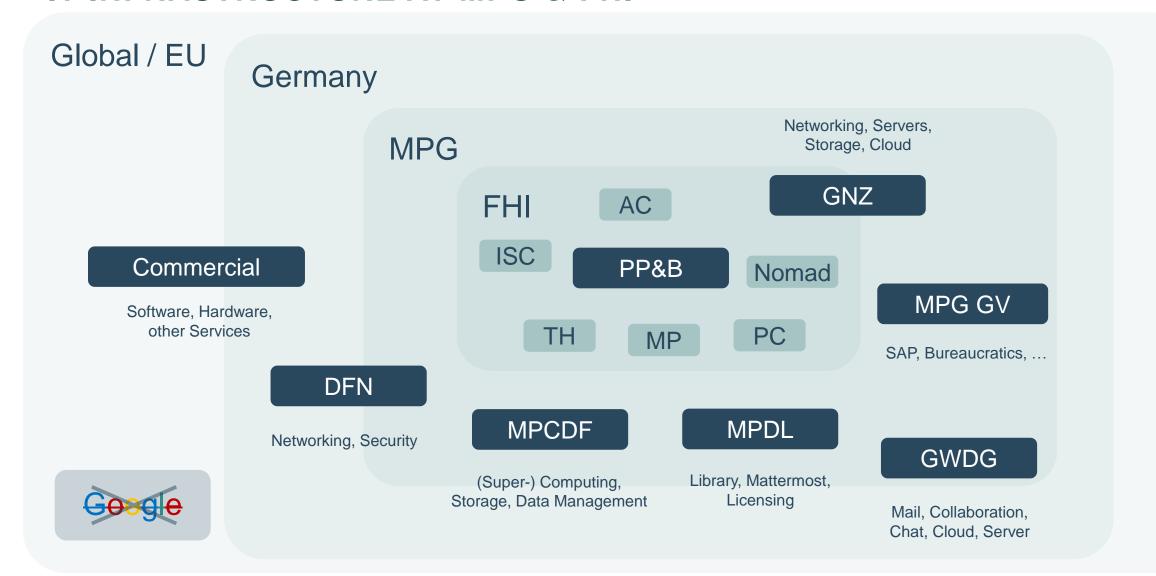
IT INFRASTRUCTURE AT MPG & FHI

&

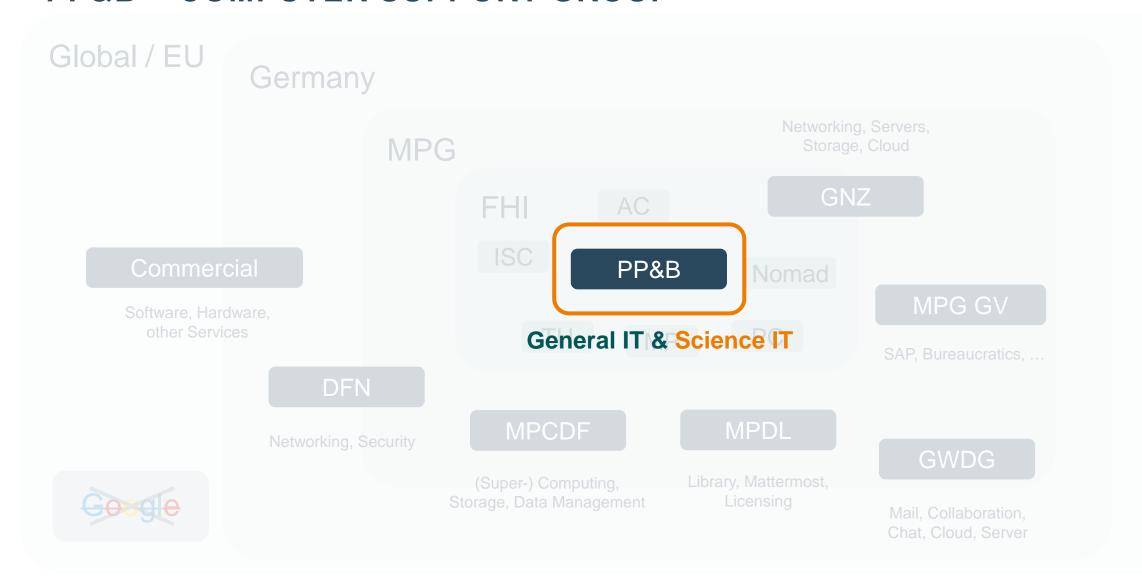
THE ROLE OF PP&B



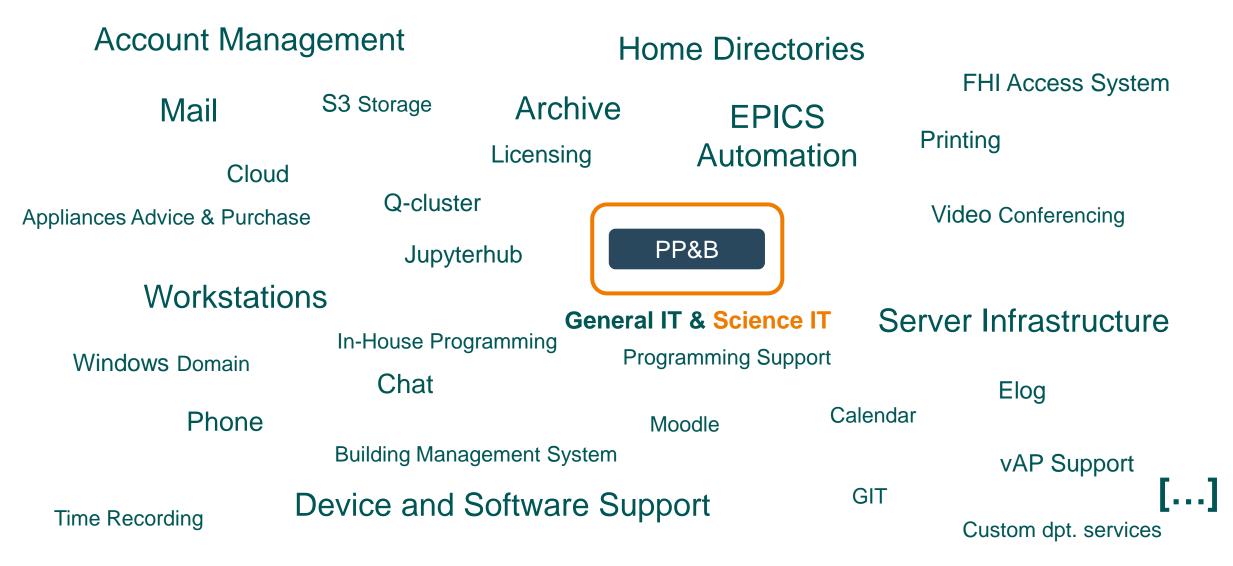
IT INFRASTRUCTURE AT MPG & FHI



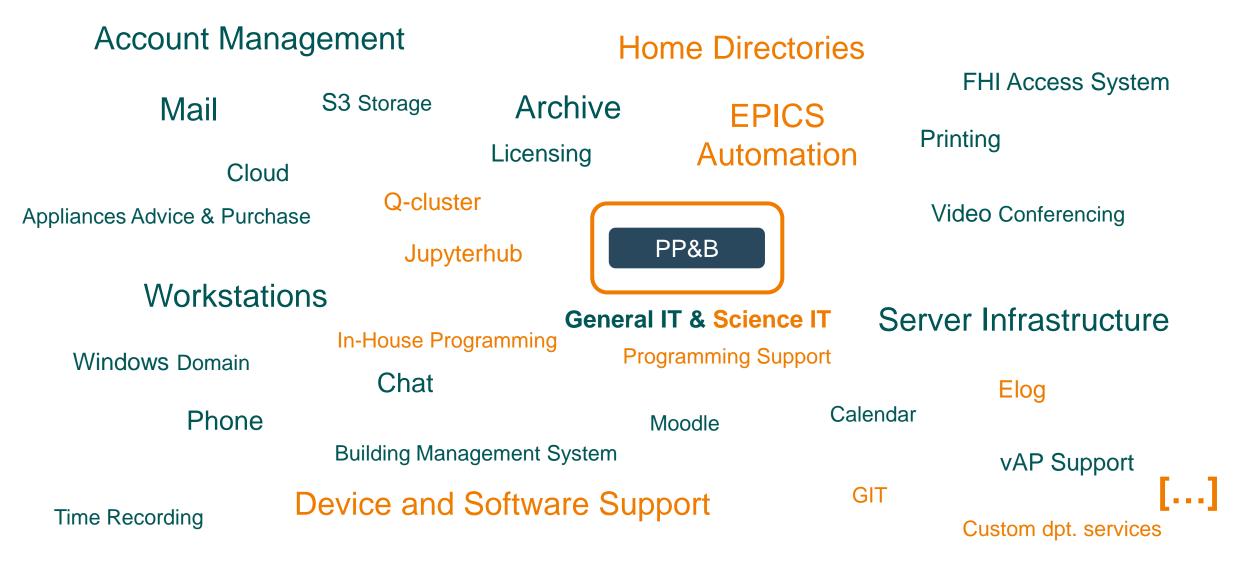














Where is it going? <u>uto</u>mation **Stronger Focus** on Scientific IT



Move Services

(In-house)

E.g. Access System, *TimeRec*

→ Facility management

Stronger Focus on Scientific IT

Unify Services

(or drop)

Single, not multiple solutions for the same problem

Move Services

(External)

- E.g. Mail, Cloud \rightarrow GWDG
- E.g. Cluster Hardware, JupyterHub, Storage, Virtual Servers → MPCDF

FRITZ-HABER-INSTITUT | SIMEON BEINLICH

THE FUTURE ROLE OF PP&B AT THE FHI

BERLIN, GERMANY | 11TH JULY 2024



Automation Support

In-house programming

Printing

Move Services

Automation

(In-house)

• E.g. Access System,
TimeRec
→ Facility management

Stronger Focus on Scientific IT

Unify Services

(or drop)

• **Single, not multiple** solutions for the same problem

Windows Domain

Phone

Programming Support

Time Recording

Move Services

(External)

- E.g. Mail, Cloud

 → GWDG
- E.g. Cluster Hardware, JupyterHub, Storage, Virtual Servers
 → MPCDF

Elog

Calendar

Computing Support

Custom dpt. services





Sven Pokrzykowski

FHI accounts, mail, phones server & Windows administration, licensing, IT support ...



Heinz Junkes

Head of PP&B, management, EPICS automation ...



Mike Wesemann

Storage, backup, archive system, server administration, software development



Marco Hollemann

Workstation infrastructure
Windows administration & support,
client monitoring



Ina Richter

FHI accounts, phones, door access system, IT support ...



4 new Apprentices

Mathematical technical software development, System Integration, Application development (→ 08.2024)



Who we are ...

William Kirstaedter

Software development, system integration, EPICS automation



Falk Rosenhahn Elab-IT / PP&B

Simeon Beinlich

HPC systems, scientific computing, scientific programming



CHALLENGES IN SCIENTIFIC IT

&

POSSIBLE WAYS TO TACKLE THESE



CHALLENGES IN SCIENTIFIC IT - MORE COMPLEXITY

More data

More complex data



More complex analyses

More complex setups

More requirements



CHALLENGES IN SCIENTIFIC IT - MORE REQUIREMENTS

Reproducibility & Documentation

Data Management

What did I store and where?

Data Security

- Hardware failures
- Cyberattacks (HZB, BHT, ...)
 - Human errors ...

Data Provenance

What happened to the sample until it was measured?

How to tackle this?

Setup Complexity & Stability

Accessibility & Findability

Internally & externally

Reusability & Uniformity

Data & procedures!

Independence from vendors

AI?!



CHALLENGES IN SCIENTIFIC IT - MORE REQUIREMENTS

Reproducibility & Documentation

Data Provenance

What happened to the sample until it was measured?

Accessibility & Findability

Internally & externally

Data Management

What did I store and where?

How to tackle this?

INFRASTRUCTURE AUTOMATION COLLABORATION

Reusability & Uniformity

Data & procedures!

Data Security

- Hardware failures
- Cyberattacks (HZB, BHT, ...)
 - Human errors ...

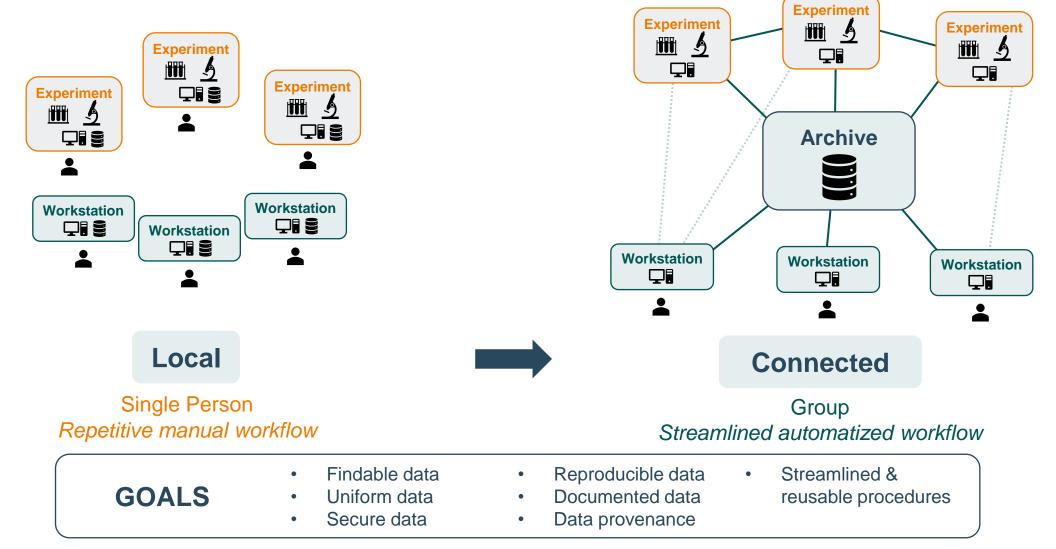
Setup Complexity & Stability

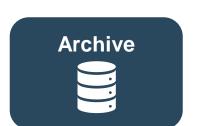
Independence from vendors

A1?!

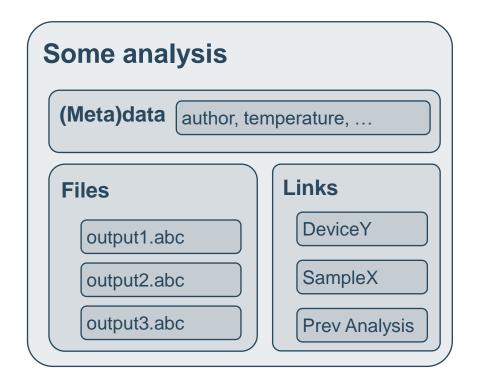


INFRASTRUCTURE, AUTOMATION, & COLLABORATION





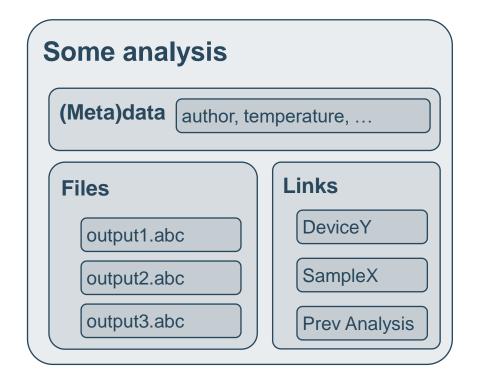




Reproducibility **Data Management** & Documentation **Data Security Data Provenance Accessibility & Setup Complexity Findability** & Stability Independence Reusability & from vendors Uniformity A1?!



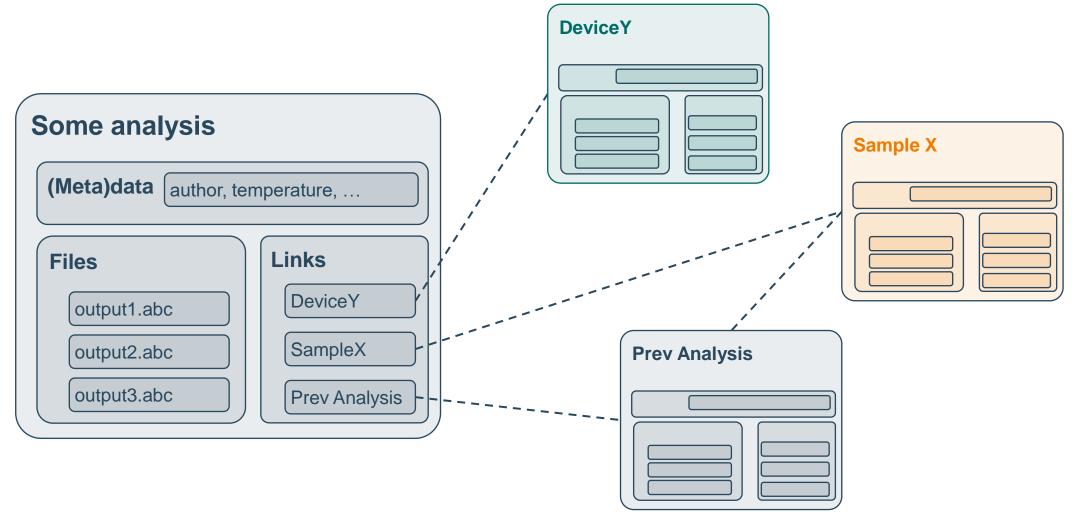




Reproducibility **Data Management** & Documentation **Data Security Data Provenance Accessibility & Setup Complexity Findability** & Stability Independence Reusability & from vendors Uniformity A1?!

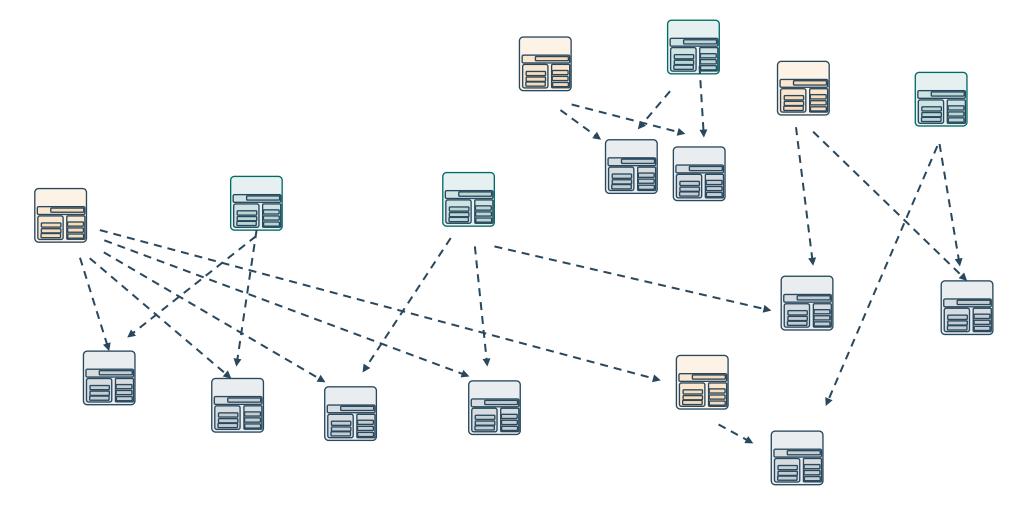






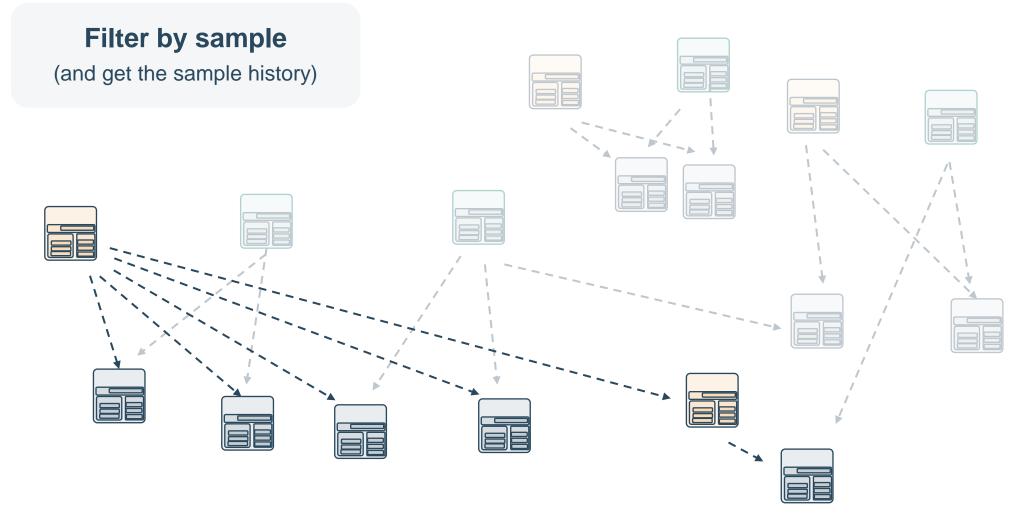






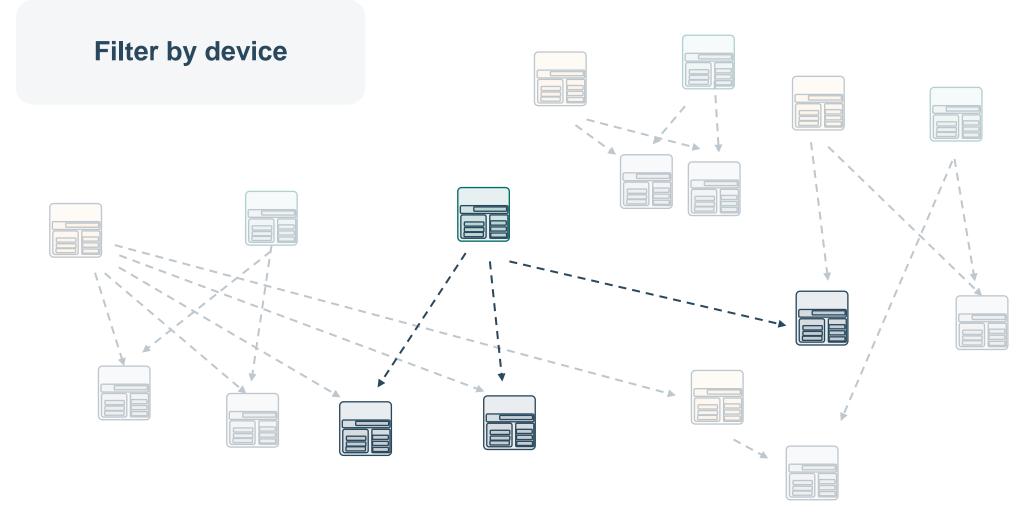






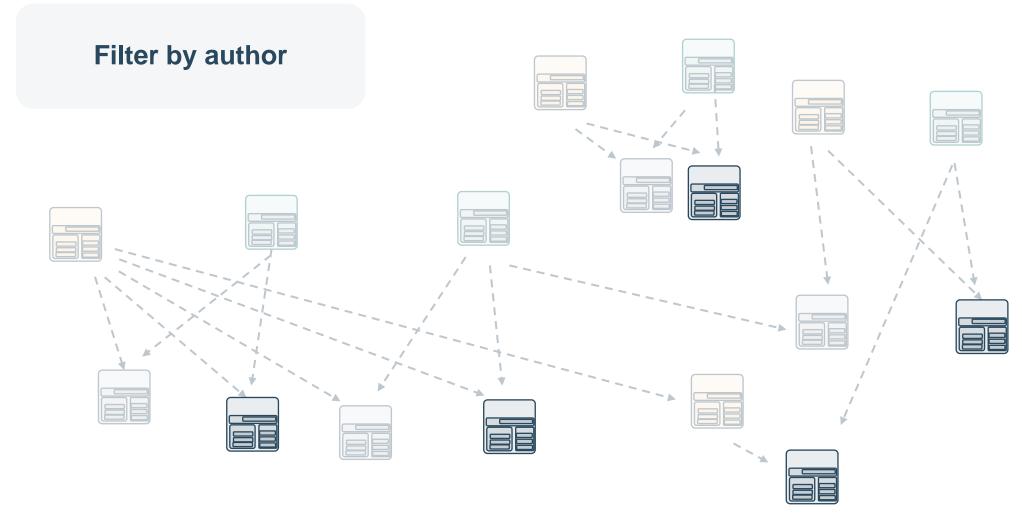






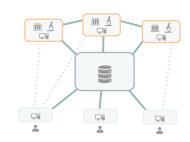




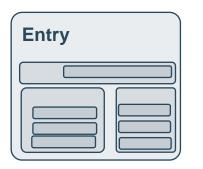


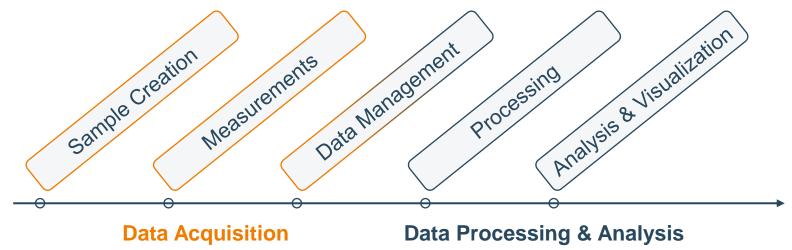
CHALLENGES IN SCIENTIFIC IT











EPICS Automation

Scripted Analyses

(Python, Archive, ...)

Classical Setups

(Local Software and Data)

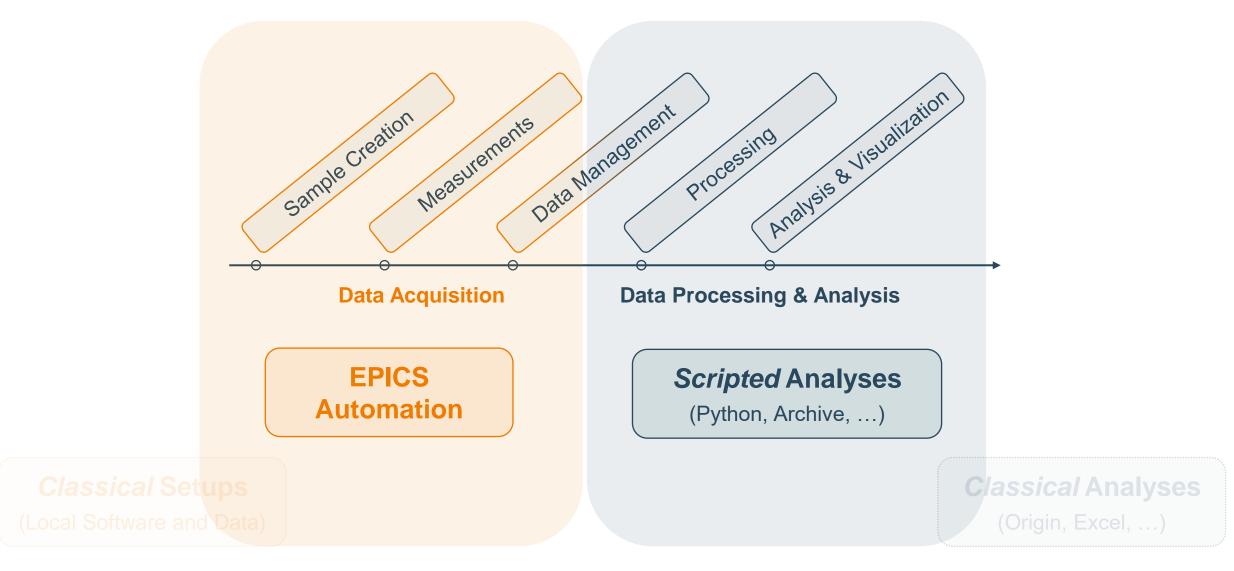
Automatized
--- Manual -----

Classical Analyses

(Origin, Excel, ...)



A FULLY AUTOMATIZED EXAMPLE - HABER REACTOR (AC)

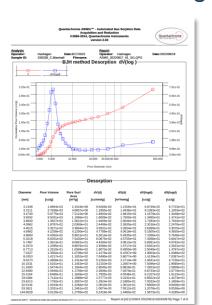




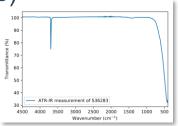
A FULLY AUTOMATIZED EXAMPLE - HABER REACTOR (AC)

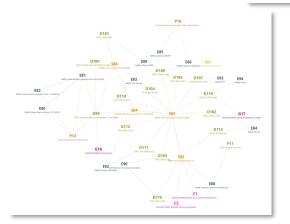
Automatized:

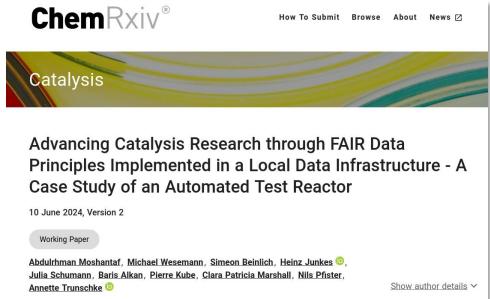
- Experimental control (EPICS)
- Data collection & linking (Archive)
- Data processing (Python)
- Plotting / Reporting (Python)











A. Trunschke et al. ChemRxiv (2024)

Demo Archive (Open Access) haber.archive.fhi.mpg.de





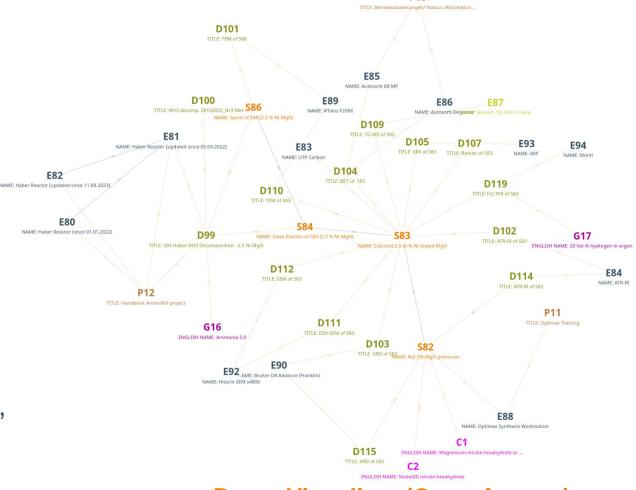
A FULLY AUTOMATIZED EXAMPLE - HABER REACTOR (AC)

Automatized:

- Experimental control (EPICS)
- Data collection & linking (Archive)
- Data processing (Python)
- Plotting / Reporting (Python)

Provenance from start to finish:

- Sample creation, modification & characterization
- Measurement parameters, raw data, processing, plotting & reporting.
- Access to other parameters (room temperature, humidity, ...)

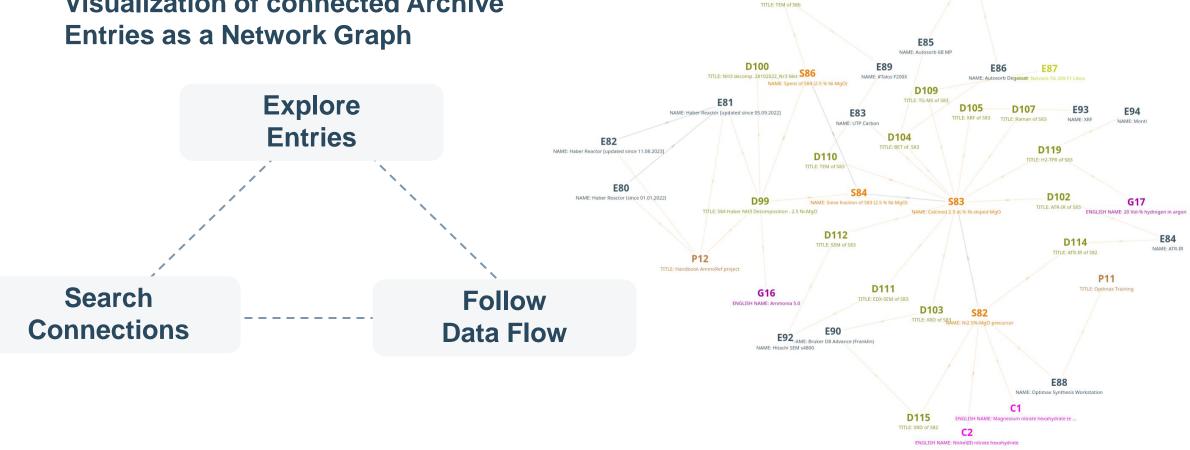


Demo Visualizer (Open Access) visualizer.fhi.mpg.de/haber



EXAMPLE FOR TAILORED SERVICES: ARCHIVE VISUALIZER

Visualization of connected Archive Entries as a Network Graph

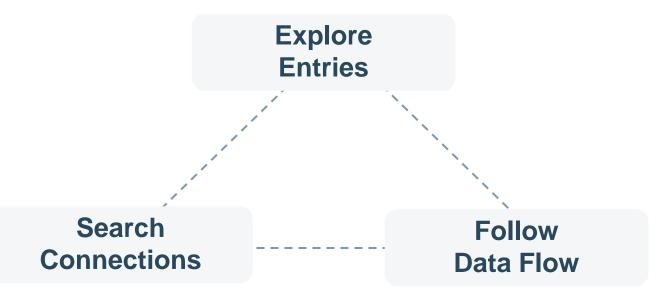


Demo Visualizer (Open Access) visualizer.fhi.mpg.de/haber



EXAMPLE FOR TAILORED SERVICES: ARCHIVE VISUALIZER

Visualization of connected Archive Entries as a Network Graph

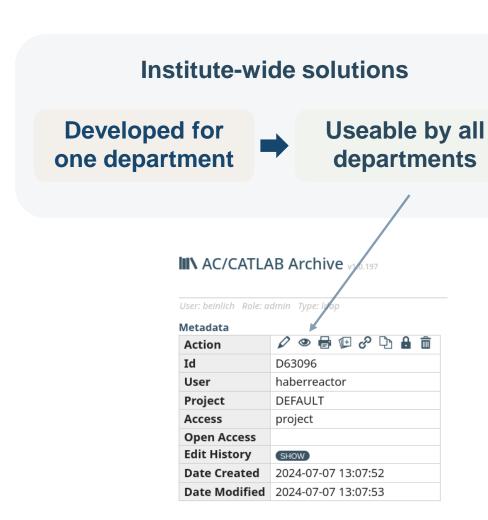


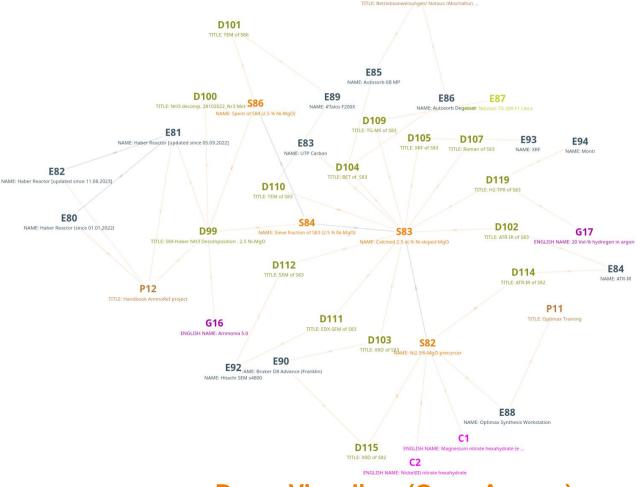
Demo Visualizer (Open Access) visualizer.fhi.mpg.de/haber





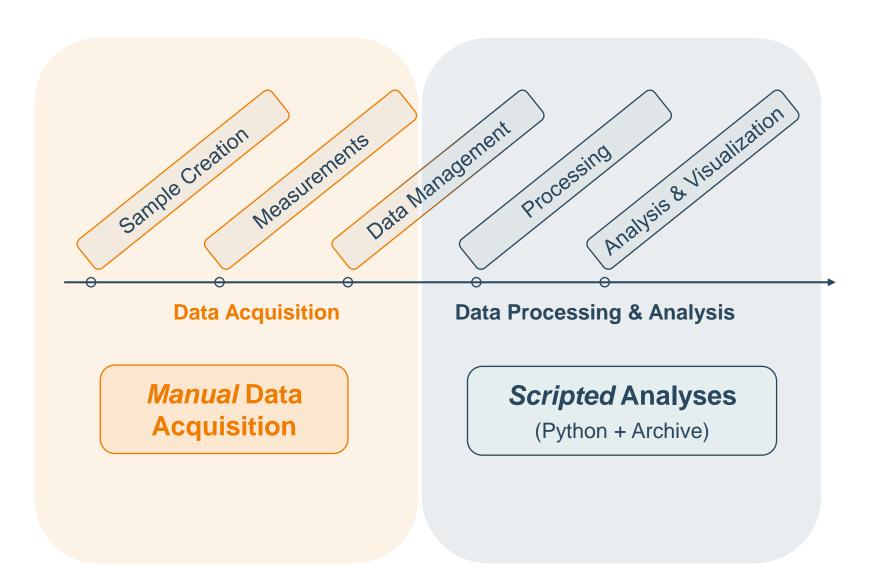
EXAMPLE FOR TAILORED SERVICES: ARCHIVE VISUALIZER





Demo Visualizer (Open Access) visualizer.fhi.mpg.de/haber





THE FUTURE ROLE OF PP&B AT THE FHI



Structured Raw Data (local)



Raw Data
+ Meta Data
(Archive)

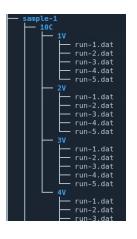


Processed
Data
(Archive)



Analyzed
Data
(local/Archive)







Metadata from:

- folder structure
- data header
- manually entered'



```
"temperature": 30,
"scan": 1,
"sampleID": "S307",
"scan_direction": "positive",
"step_number": "1"
```



```
"VSTEP1": 0.1,

"VSTEP2": 0.15,

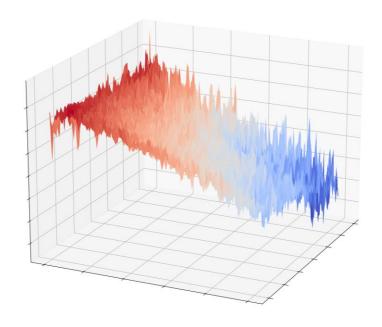
"TSTEP1": 0.15,

"TSTEP2": 0.15,

"SAMPLETIME": 0.0000033334,

"start_index": 0,

"end_index": 90006,
```





Structured Raw Data (local)



Raw Data
+ Meta Data
(Archive)



Processed
Data
(Archive)



Analyzed
Data
(local/Archive)

Automatize repetitive tasks

- Data collection
- Data processing
- Analysis
- Plotting

Python: simple & powerful

- Local workstation
- JupyterHub
- HPC

PYTHON

IPY-Notebooks

- raw_data.ipynb
- (pre-)processing.ipynb
- analysis.ipynb

Here: provided ready to run by PP&B

Modules

- Archive Rest API
- Archive Python API
- Archive Python DAO
 (Database Access Object)
- Archive Dataframe Extension (Pandas)



Structured Raw Data (local)



Raw Data+ Meta Data(Archive)



Processed
Data
(Archive)



Analyzed
Data
(local/Archive)

But I don't know Programing?

- Use pre-made scripts/notebooks from:
 Colleagues, PP&B, Theory?, ...
- Modify existing notebooks
- Then maybe: Write your own!

PYTHON

IPY-Notebooks

- raw_data.ipynb
- (pre-)processing.ipynb
- analysis.ipynb

Here: provided ready to run by PP&B

Modules

- Archive Rest API
- Archive Python API
- Archive Python DAO
 (Database Access Object)
- Archive Dataframe
 Extension (Pandas)



A HALF-AUTOMATIZED EXAMPLE

Structured Raw Data (local)



Raw Data+ Meta Data(Archive)



Processed
Data
(Archive)

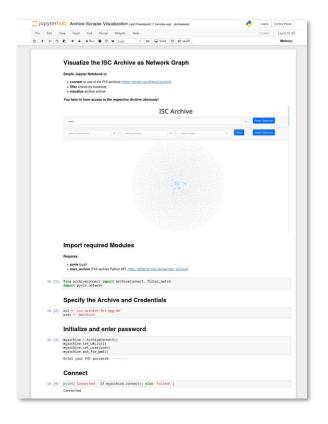


Analyzed
Data
(local/Archive)

But I don't know Programing?

- Use pre-made scripts/notebooks from:
 Colleagues, PP&B, Theory?, ...
- Modify existing notebooks
- Then maybe: Write your own!

Ask for ready-made Notebooks on JupyterHub as a start!







Advice on how to tackle IT problems Science Science-IT

Help with programming

Python

Other

What the PP&B can do for you

Help with / provide small services

- Jupyter Notebooks
 Python modules
- Python scripts
- Small services

Help with existing services

- Archive
- Cluster
- Storage
- JupyterHub
- EPICS

Help with script sharing

• GIT





Advice on how to tackle IT problems

Science Science-IT

Help with programming

Python

Other

What the PP&B can do for you

Help with / provide small services

- Jupyter Notebooks
 Python modules
- Python scripts
- Small services

Help with existing services

- Archive
- Cluster
- Storage
- JupyterHub
- EPICS

Help with script sharing

• GIT



SHARING & COLLECTING ANALYSIS PROCEDURES - GIT

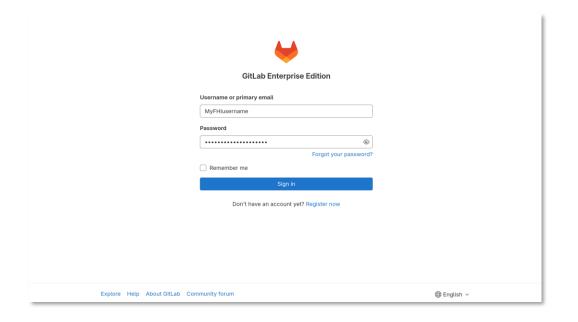
GitLab

Collaboration & Versioning Tool

Knowledge is useless if you don't share it.

Use Git for:

- collecting analysis procedures...
- sharing...
- improving...
- documenting (Git Wikis!)



FHI GitLab
https://gitlab.fhi.mpg.de
(register first)



SHARING & COLLECTING ANALYSIS PROCEDURES - GIT

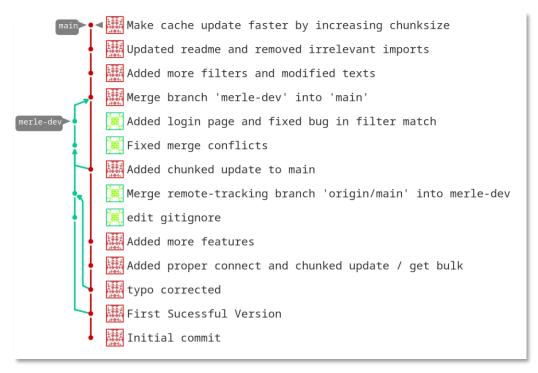
GitLab

Collaboration & Versioning Tool

Starting to use Git can be troublesome..

- Need a kickstart?
- Got lost in branches, forks, commits, merges?
- Push, pull, fetch, stash, or rebase?

→ Get in touch!



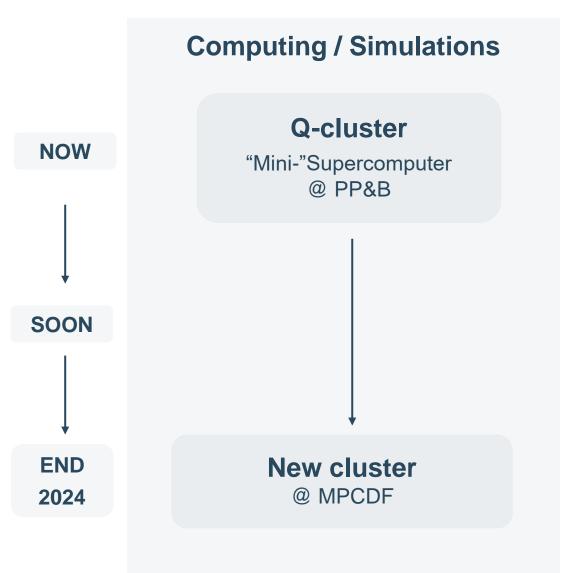
FHI GitLab

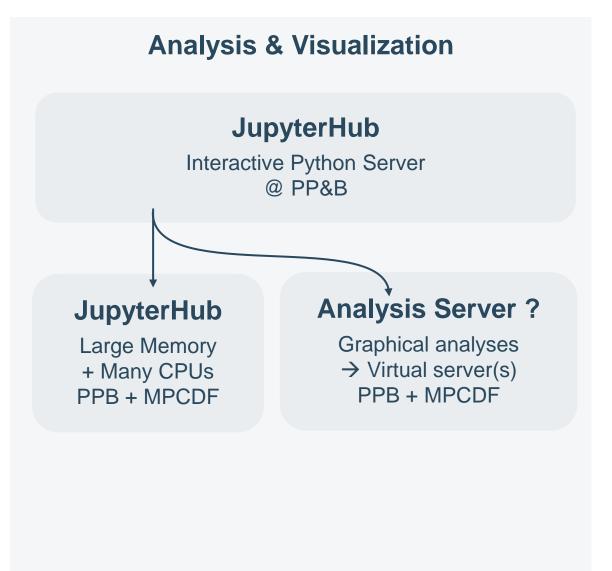
https://gitlab.fhi.mpg.de (register first)



RUNNING LARGE ANALYSES & & COMPLEMENTING EXPERIMENTS WITH SIMULATIONS









Simulations

Density Functional Theory

- Energetics
- Thermodynamics & Kinetics
- Structure, Geometry, Stability
- STM image simulation
- Vibrational Spectra
- Electron spectroscopy ...

Kinetic Modeling

- Micro Kinetic Modeling
- Kinetic Monte Carlo ...

Large Analyses

Resource intensive scripts and programs

AI Training

Anything that requires more than a desktop PC



https://doku.lrz.de/supermuc-ng-10745965.html



Simulations

Density Functional Theory

- Energetics
- Thermodynamics & Kinetics
- Structure, Geometry, Stability
- STM image simulation
- Vibrational Spectra
- Electron spectroscopy ...

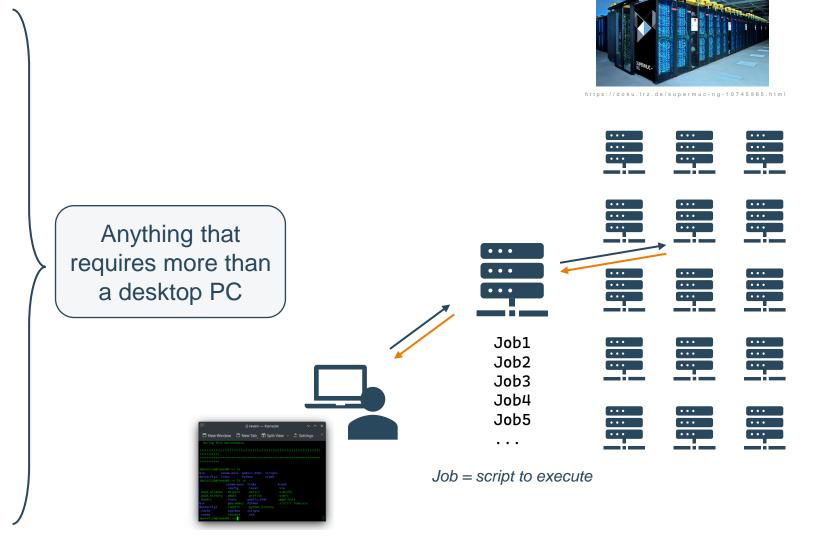
Kinetic Modeling

- Micro Kinetic Modeling
- Kinetic Monte Carlo ...

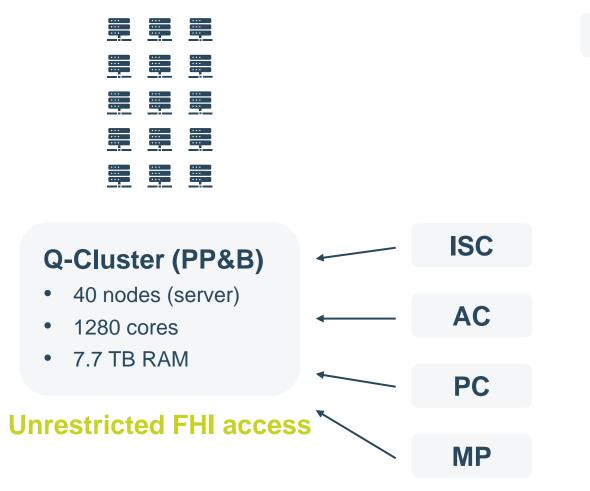
Large Analyses

Resource intensive scripts and programs

AI Training







Raven (MPCDF)

- 1592 nodes (server)
- 114624 cores
- 447 TB RAM

Department based access

(and single users)

Supermuc (LRZ)

- 6480 nodes (server)
- 311040 cores
- 719 TB RAM

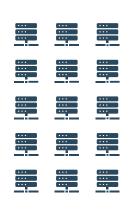
Juwels (JSC)

- >2600 nodes (server)
- >130000 cores

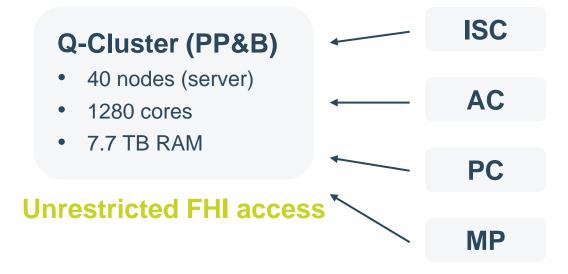
Grant based access

TH

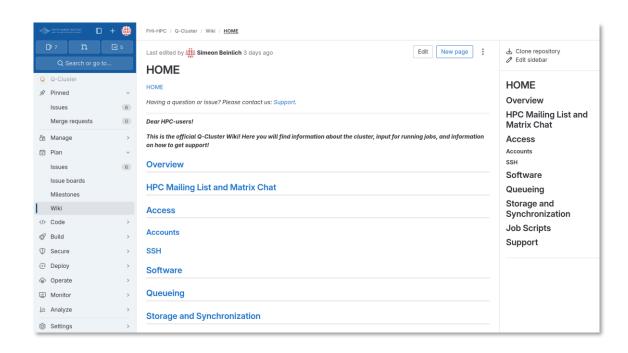




- → Medium-sized simulations & analyses
- → Simulations independent from theory groups







Q-cluster wiki

https://gitlab.fhi.mpg.de/fhi-hpc/q-cluster/~/wikis (register first)

PPB Wiki > HPC

Q-cluster wiki

Mailing List

Matrix Channel

- → Medium-sized simulations & analyses
- → Simulations independent from theory groups

Technical Issues?

Plan to run simulations?

Compiling new code?

Get in touch!

Ideas?

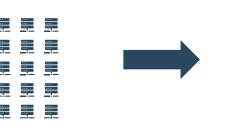
Advice?

Support?



Q-Cluster (PP&B)

- System & software updates
- Hardware repairs



New Cluster

- Currently in planning
- Hosted at MPCDF

Hardware and Software reaching their End of Life

+ Less maintenance and administration work

→ More support

+ State-of-the-art configuration

Software & Hardware

+ Ideally: Software stack identical to Raven

Compatibility and Transferability

Sun Grid Engine	Scheduler	SLURM
CentOS (Linux)	Operating System	SUSE Enterprise (Linux)
RocksCluster	Cluster Setup	-



py-3.12 O

INTERACTIVE PYTHON SERVER - JUPYTERHUB

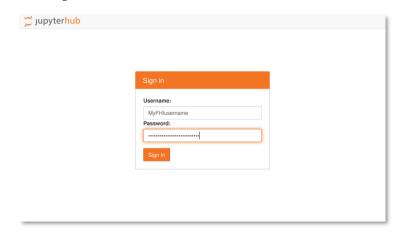
Jupyter Notebook = *Interactive Script*

Combining the best of

- Python scripts
- Python terminal
- Markdown documentation

JupyterHub:

- Requires no local installation
- Directly use ready-made notebooks!
- Soon: very powerful hardware!



Tiny Little JupyterHub tljh.fhi.mpg.de

Simple

No installation

Uniform Setup

Visualize the ISC Archive as Network Graph

ISC Archive

· connect to one of the FHI archives (http:



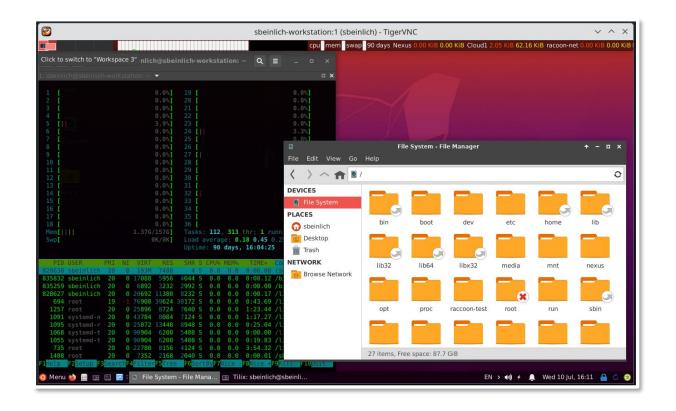
GRAPHICAL ANALYSES – VIRTUAL SERVER(S)

Not Python and not runnable on a Cluster?

→ Graphical Virtual Analysis Server!

- Graphical Desktop sessions (VNC)
- High resource (Memory, CPU, ...)
- Direct support
- Ready to use scripts etc. possible.
- Own Software installable on request.

→ ~ Theory department workstations





SUMMARY

Challenges and Pathways in Scientific IT

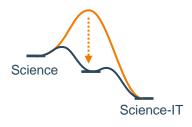
- More (complex) data & setups + More requirements
 - → Suitable Infrastructure and Automation
- From Repetitive & Local to Automatized & Connected
- Joint approach to tackle challenges
- From manual to scripted analyses (where possible)

Role of PP&B

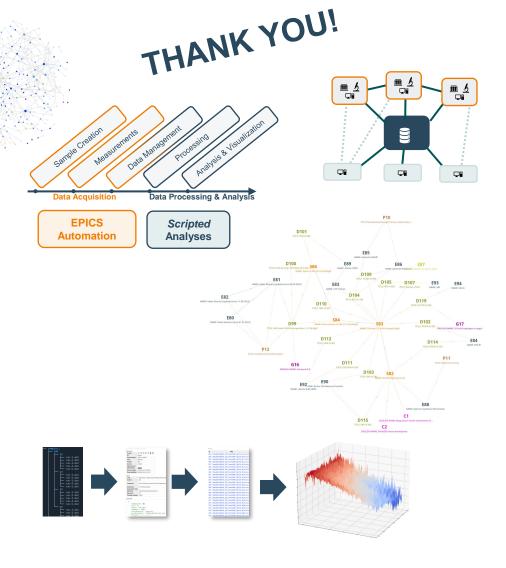
- Focus on Science IT
- More support and scientific services

HPC at FHI

- New infrastructure, more support
- Less maintenance and administration







Need something similar? Get in touch!



PC DEPARTMENT - TOPICS, QUESTIONS, IDEAS 1/2

STORAGE?

Microscopy raw data?

- ~10TB / a
- ✓ Both locally and remote (MPCDF) directly feasible.

Raw-Data-Processing?

- > 10h processing time, Parallelization?
- ✓ Real or trivial Parallelization on cluster or Analysis Server (Software-dependent?!)

Automation!

LabView preferred: fast and simple set up.

- Some PP&B Ideas:
 - Archive integration + Scripted Analyses?
 - LabView-EPICS Interface (Server/Client)?

Data-Analysis?

SVD / ICA?

(Single Value Decomp. / Indep. Comp. Analysis)

- > 100 000 entry Arrays
- ✓ Where? → Cluster / Jupyterhub / Analysis server
- ✓ Support?
 → Me!
- ✓ Data Format? → Any common/ open/ readable/ non-proprietary (.csv, .pkl, .txt, .hdf5, ...)
- ✓ Software?
 → Python / Scikit-Learn?
- ✓ Waiting Time?
 - Analysis Server (new): → 0
 - JupyterHub (new): → 0
 - Cluster (old/new): $\rightarrow \sim 1 4h$.
- ✓ MPCDF Raven/Viper? → TH Budget?



PC DEPARTMENT - TOPICS, QUESTIONS, IDEAS 2/2

EPICS AUTOMATION?

Large activation barrier!

- ✓ Introductory Courses? → Yes, please get in touch!
- Maybe: Joint Demo Project?
- ✓ Fully set-up systems from PP&B
- ✓ General:
 - What is EPICS?
 - How large is the activation barrier?
 - What is an IOC?
 - Platform dependence?
 - Drivers for Instruments?
 - Large Data Rates (Microscopy-) pictures?

COMSOL?

HPC - runnable?

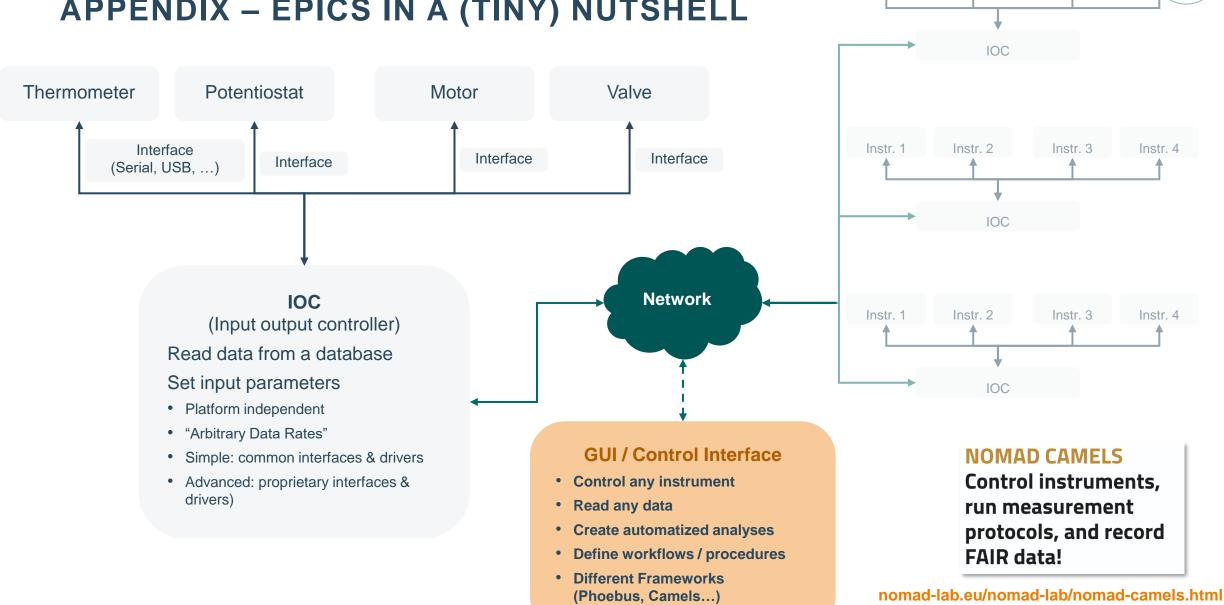
- Cluster: → Yes, if job is scriptable
 Does a CLI exist for your application type?
- Analysis Server → Yes
 (GUI + full software & network control)

Licensing?

Floating Network Licenses @ HPC, e.g. Comsol?

- Q-cluster: → yes (same network)
- New cluster / Analysis Server:
- → soon (network integration is currently implemented

APPENDIX - EPICS IN A (TINY) NUTSHELL



Instr. 2

Instr. 3