

Future Research Infrastructures

Andreas Rauber

Vienna University of Technology
Favoritenstr. 9-11/188
1040 Vienna, Austria
rauber@ifs.tuwien.ac.at
<http://www.ifs.tuwien.ac.at/~andi>

Science is evolving...

- Globalization of scientific cooperation and competition
- Digitization and digital skills
- Increasing interdisciplinarity
- Data-driven research
- Increasing processing power
- Ubiquitous connectivity
- Miniaturization, nano-technology, genetics,
- Smart-*, machine-actionable *, autonomous systems, ...
- "Artificial Intelligence"
- Increased pace of change

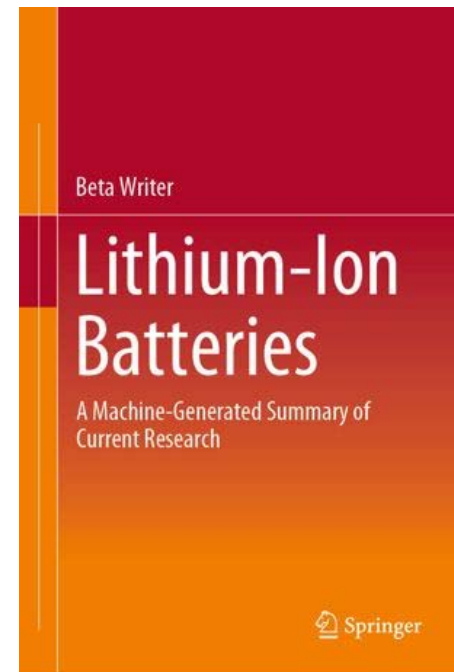


Requirements for Future Research Infrastructures

- Understanding what we might need 10-15 years from now
 - Think back 20-30 years (increased pace of change)
 - Consider impact of observable trends
 - A bit of Science Fiction (but with a focus on science)
 - Remaining just shy of the singularity predicted for 2050
- How will we / our PhD students do research?
- Who will we / they be working with?
- In what kind of environment will we / they work?
- What should we do to start preparing for it now?
- Focus on technology, infrastructure, but also legal aspects, societal, ...
- Everything beyond "more of the same"

Just as an example...

- <https://link.springer.com/book/10.1007/978-3-030-16800-1>
- doi.org/10.1007/978-3-030-16800-1
- **Authors and affiliations:**
Beta Writer



EOSC Workshop Series

- EOSC as one example of future research environments
- By-invitation-only workshops 25-35 participants
- Programme / Setup
 - What will we need?
 - 2 to 3 breakout sessions with 3 groups of around 5-10 people
 - Reporting-back sessions after each breakout session
 - Plenary: consolidation and feedback processes
 - Final wrap up
- Follow-Up Actions
 - Reports
 - Takeaway messages



Results

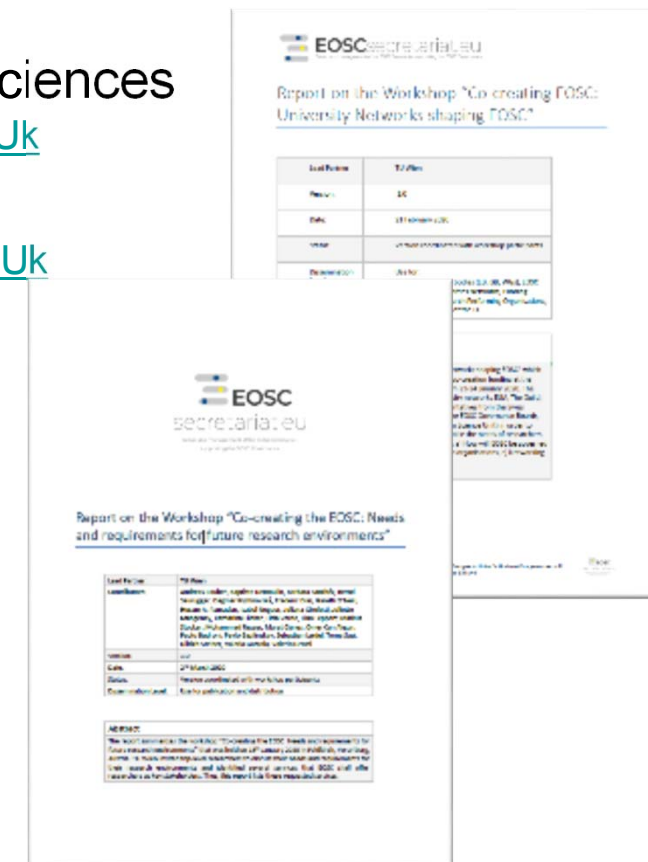
Reports on the workshop with:

- Researchers: natural and technical sciences
https://zenodo.org/record/3701194#.XmY_1KhKiUk
- Members of university networks
<https://zenodo.org/record/3693914#.XmZMYahKiUk>

Key take-away messages:

<https://zenodo.org/record/3701269#.XmZKrqhKiUk>

- Members of university networks
- Researchers: natural and technical sciences
- Science Europe



We need more...

- Only have starting points so far
 - Strongly influenced by current problems
 - Strongly influenced by what seems feasible
 - Focus on immediate next steps
 - Assuming a rather standard scientific process
- Need to think further ahead to reveal what we might need
- In spite of increasing speeds: infrastructure set-up takes time
- Need to start preparing now
 - How do we get there?
 - What services (beyond the obvious) do we need?
 - How to balance technical feasibility and future needs?

Observations from today

- Matthias:
 - *Keep the waste basket*: “publishing” raw data
 - *Smart knowledge extraction*: Tools exist, but need tuning, parameters
 - *Bring code to data (data volume, privacy)*: Federated ML, Migrating code
 - *Complete characterization*: Provenance extraction/monitoring
 - Practical AI specialists: hmm...
- Andre:
 - *Don't share the data: bring research to data*: Federated ML, Data shielding, Cryptographic Databases
 - *Data governance*: social contracts, consent, trust, ...?
 - *Reproducibility, distributed validation*: structure + semantics!?

Observations from today

■ Arturo:

- *Scientific data collections of specimens (objects):* what can / cannot be digitally represented?
- *Importance of old legacy data:* limits of automation
- *Bias in data:* how to represent, allow machines to understand?
- *Complexity of data:* how to represent?
- *Extraction, Representation, Infrastructure:* what do we need in each of these areas?
- 7 big questions

THE K-D MATRIX

Under-developed developing established

NOW

	K-EXTRACTION	K-REPRESENTATION	K-INFRASTRUCTURES
D-FLOODING	CS, crowdsourcing	maps, summary plots	CS platforms
D-QTF	Mostly ad-hoc	developing: ignorance maps	annotations
D-DESILOING	Test-level	underdeveloped	Aggregators, crawlers e.g. Eol, openUP
D-REPLICABILITY	n/a or problematic	doi for datasets, standards, metadata	doi, persistent repositories
D-STEWARDSHIP	IPT	Virtual notebooks	GBIF
D-AUTOMATION	opportunistic	plugins, dashboards	DOIP, GHR
D-EVOLUTION	CS: iNaturalist	EOL	Post-GBIF?

Logos: Universidad de Navarra, ERM, GED

THE K-D MATRIX

sooner later

THEN

	K-EXTRACTION	K-REPRESENTATION	K-INFRASTRUCTURES
D-FLOODING	-OMICS, RS	Dynamic, 4-5D	Exascale HPC, EGI
D-QTF	Big Data Validation	QTF Metrics	Blockchain
D-DESILOING	DO crawlers	Imported visualizations	DOIP, GHR
D-REPLICABILITY	Automated proxying	Automated provenance plotting	Open Science Peerage
D-STEWARDSHIP	AI-enabled semantics	Embedded analytics	EOSC/EUDAT, DISSCo
D-AUTOMATION	IoT, RS, Intelligent Assistants	LW-ERIC and the like	PAAS
D-EVOLUTION	AAAS?		Autonomous AAAS?

Logos: Universidad de Navarra, ERM, GED