EOSC Support Office Austria: Visions, needs and requirements for research data and practices

Marcus Huber (TU Wien), Bernd Saurugger (TU Wien)

This interview is also available for download: https://doi.org/10.5281/zenodo.8398358

In 2015 the vision of a federated system of infrastructures supporting research by providing an open multi-disciplinary environment to publish, find and re-use data, tools and services led to the launch of the European Open Science Cloud (EOSC). Against this background, bodies such as the EOSC Association on the European level and the EOSC Support Office Austria on the national one have been established.

Within this framework and since research has always been at the heart of EOSC, we are eliciting visions, needs and requirements for research data and practices from researchers who are located at public universities in Austria. Let's see what quantum physicist Marcus Huber has to say!

"Efficient Open Data Preservation and Grant Application Processes"

BS: What is your work currently focused on?

MH: Our group is largely based on theory work predominately, in collaboration with various experimental groups in the area of quantum science and technology. In terms of theory we do spend a lot of time on foundational questions, like the underlying essence of quantum theory and what it tells us about information processing. What can we know about the world? Also, in conjunction with thermodynamics: What are the fundamental thermodynamic limitations? One popular example you might have heard of is that every CPU produces some heat, and this is not due to some technical design. It is an intrinsic connection between information processing and physics, that we are trying to uncover at a quantum level.

BS: What kind of data are you working with?

MH: Of course, in the pure theory part of proving theorems and conceptualizing things, there's no data to speak of. All the things we do can be written up in a self-contained manner, in a manuscript. Sometimes there is a speculation whether the process should be documented. In the end, there is a beautiful mathematical proof that is half a page long, but to actually get there it needed 500 pages of notes with full of mistakes, which are usually not published. Of course, we also work with experimentalists on different platforms. They often generate data in the sense of measurement outcomes that were recorded in certain devices, which are then statistically analysed and interpreted by us.

BS: Do you face some issues in terms of trust?

MH: No, in theory the nice thing is, that if a mathematical proof is complete and correct, everybody can check it in a very simple and quick manner. So, the fact that it took you forever to get there is meaningless. I mean, the only thing that impacts you is when you need some other theorists work and you get like: "Oh my God! This is so short and beautiful. How did you ever get there?" You don't see the work that went into it. It is more a psychological issue than a trust issue.

However, with experimental data we work with, we face other problems. The problems we face are mostly the data handling problem. Say there's a Quantum experiment where we distribute entangled photons. It is nothing world moving. It has been done 100 times in 100 labs around the world and yet you record every click with a timestamp. In the end, after a day or two, you suddenly have terabytes full of random click data, out of which you distil or post-process the properties you want to analyse the data for. Then in the end, it's a single plot. But then the big question is: To what extent does this data need to be preserved? If every inane Quantum experiment, that is basically not challenging any of the foundations, had to keep all the terabytes of data, that went into creating every single plot, we would also have a serious issue about data storage. I mean, we would have billions of terabytes of data already and how long do you want to keep them, if almost nobody's ever going to look at them again? When, however, there is an experiment that really promises a new effect or challenges established theory, you better well make sure that all of it is preserved and that people can double-check.

> "Preserving all data is causing a serious data storage issue. A decision is needed on what data is worth to be preserved and opened to the public"

Often, there's a random code you or one of your PhD students wrote at some point. It's a bit incomprehensible. The data is in a completely unreadable format, without knowing exactly which column and row means what. A data matrix, especially huge terabytes of ones and zeros is worthless, without the right context to read it. It is a lot of extra work to make it openly accessible. I could give you the data of all our experiments, but it would be useless, even to experts in the field, without a long explanation of how to analyse the data. Then the problem is that, so far, this has not been rewarded in the scientific publishing system. Optimistically, in a very short experiment, you spend 200 hours on the experiment and writing the paper. It would take another 200 hours to make the data openly accessible. If this is not rewarded in the slightest, why would anyone do that? So, I think if we want to move forward in open data, we need to have some mandates to some extent. But you shouldn't overshoot to the extent that we're now forced to keep billions of terabytes of unusable, unnecessary and uninteresting data, nobody's ever going to look at again.

BS: What burdens do you face as a researcher?

MH: I think there's a large overhead in the way peer review is organised. The same is true for science funding. Just from an economical point of view: I wrote many ERC applications before I got this grant. Not every proposal is successful, but every application takes up a lot of time. Here, it's not that the peer review rejects your application as it finds some fatal flaws in it, and you need to improve to be scientifically sound. It's mostly that there is just not enough funding for these kinds of grant competitions. So, 90% of fundable grants are rejected without funding, because there's not enough money in them, and the feedback you get is: "Well, that's a nice grant. Would be nice if we had enough money to fund it – thank you!".

> "A certain level of external evaluation is important, but the distribution of scientific work and invested resources in the review process needs to be balanced."

Given the amount of resources invested, if a grant has a success rate of 10%, which is higher than ERCs, that means the other 90% of grants are also written by top researchers, where the majority of them are technically sound and fundable. All of them were seen by a host of referees and researchers, which take time from their actual jobs to evaluate them. Then, panels get together, including senior researchers, who have many other duties, and discuss all the reports. Then, you're invited for interviews. Then, there's a whole department managing this. Then, there's a department at our university, trying to communicate with their department and so on. I think here there's maybe a lesson to be learned, that of course sanity checking research and forcing every once in a while an external evaluation on your research ideas is a good exercise for the researcher, to think of the big picture and explain to people why the research is interesting. If you never do this, you might just fall into a rabbit hole of something weird. Also, you make sure that most of the science funding, which is public money, goes towards projects which make sense. I think a certain level of external evaluation is a good thing. I'm just not sure that the distribution is currently in good balance.

BS: How do you see the future for your young researchers?

MH: I hope that the university, recognizes the value of permanently employed staff, not at the full professor level. A full professor nowadays is expected to have 50 roles at once. You're supposed to be a good mentor, a good public speaker, a good manager, a good organizer, a good admin, a good scientist, a good writer, a good teacher or lecturer and all of these things. Sometimes there are absolutely brilliant scientists who might not be brilliant at all of these tasks. Universities then sometimes promote people to professors who are really terrible at one of these tasks. If you're a terrible manager, then that basically means that the people in your group suffer. Or if you're a terrible teacher, that means the students at the university really suffer, while you might do brilliant science. So, I do hope that there's also a greater recognition of a path, what is called in German "Academischer Mittelbau", something that you permanently employ people at universities that are not full professors. I see a move toward this, and I do hope this continues.

Marcus Huber studied physics at the University of Vienna, where he also completed his doctorate in 2010. He then worked at the University of Bristol, the University of Barcelona and the University of Geneva. In 2016, he returned to Vienna as a group leader at IQOQI (the Institute for Quantum Optics and Quantum Information of the Austrian Academy of Sciences). Since 2020, he has been a professor (assoc. Prof. Full professor since 2023) at TU Wien, his diverse and international research groups is based at both IQOQI and TU Wien. Marcus Huber has already been able to attract many highly endowed research grants and win prestigious science awards – including a Marie Curie Fellowship from the EU (in 2011), the START Prize from the FWF (2015) and an ERC consolidator. He is also the founder of the Vienna-based journal Quantum, which is now one of the top journals in quantum research.