Visions, needs and requirements for (future) research environments: An exploration with ERC Grantee Toma Susi

Katharina Flicker (TU Wien), Barbara Sanchéz (TU Wien), Toma Susi (Universität Wien)

Researchers are at the very heart of the EOSC: So what do researchers really need to do cutting-edge research? How do they think the EOSC could support them in their endeavours? Let's see what ERC grantee and physicist Toma Susi has to say.

"Data Sharing in the interest of reproducibility and re-usability"

TU Wien: What is your work currently focused on?

TS: My work focuses on material science and more specifically electro-microscopy. What we do is we look at the atomic structure of different materials using electro-microscopy and their defects and dynamics and things like that. In addition to the experimental work, we do quite a bit of modelling. We use methods based on quantum mechanics to model the properties of these materials and then try to understand what is going on by combining the experiments with the theory and then correlating the two.

TU Wien: Do you have data challenges that EOSC could help you with?

TS: Yes! We generate two types of data. We generate electro-microscopic images, which are in principle normal image files except that they are 32 bit image files. If you open such a file on your normal computer without using a specific program, you won't see much because the contrast range in such images is much wider. And if there was a way to have an open data repository where we could put these images we could just upload a bunch of images and the website would make preview images that you could just browse to in your normal web browser. Then you could also download the original data files. That would be very useful. The other challenge for making

data available is the simulation data. You have to run a simulation to model the properties of materials. That generates files that are pretty big and are typically not shared at all between researchers. We generate such 1 to 2 TB per year. We keep them for the time until the publication is accepted. Maybe a little bit beyond that. Then we have to delete it. We normally cannot store long-term such big files.

TU Wien: What would you need from the EOSC in order to resolve the challenges you face? Which services would help you most?

> "Having one place for all European research data that is openly available on the web. I think that is the valuable and worthwhile goal"

TS: I do not currently see us doing a lot of this work we do in the cloud. We can do the analysis locally and it probably is easier. For us it is about open science. That is where a cloud could be very useful. And of course, not just have static files sitting on a server, but something that you and other people could access in various ways and re-use in various ways.

TU Wien: What is the main added value that you see in the EOSC?

TS: Having one place for all European research data that is openly available on the web. I think that is the valuable and worthwhile goal.

TU Wien: Do you believe that the EOSC can boost interdisciplinary research in Europe?

> "If more of the data was more openly available and easily accessible, it would improve our ability of verifying results"

TS: I don't think that interdisciplinary research really is a data problem. That is more of a knowledge and communication and networking problem and of people actually finding these collaborations. You cannot just take a dataset that you don't understand and do research with that. You have to understand the dataset and to do that you have to cross boundaries and need somebody from the other side telling you and instructing you. Even good metadata probably isn't enough. You can describe the data. But unless you understand the discipline specific meanings of those descriptions, it doesn't get you very far. In terms of helping within the discipline of nanoscience and physics – if more of the data that we generate or that people in the field generate was more openly available and easily accessible and perhaps even some analysis that we could run in the cloud, we would use that. I think it would improve at least our ability of verifying some results. But the question is then at which point of the research cycle this is going to be incorporated in. Is it post-publication, or even part of the peer-review process? Or Is the data already part of the submission? Or even before the submission like in a pure open science approach?

TU Wien: Do you have preferences when it comes to *when* to share the datasets?

TS: That also depends on how easy it is to share the data. Currently you have to do most of the work manually. If you'd had to just drag one zipped file to a cloud and everything else is handled somehow automatically, it would be much easier to do share the data in the beginning of the peer-review process. Sharing something before that is another story. There you do get into – and I think legitimate – fears of being scooped, if you share your latest findings months before you have a manuscript ready or before you submitted something. I do think that we would be hesitant with that. Because we can collect very specific types of data. And just finding the right samples, making the right measurements, getting the good dataset – that is a lot of work and if you just share it before you are ready to publish, perhaps then somebody will take the data and will do the analysis that you wanted to do. That accelerates science, but it doesn't accelerate our careers in any way. I think that at submission we would be comfortable with sharing data.

TU Wien: In your domain, which challenges can you or will you address with the EOSC, or with more open data available which would be impossible to tackle alone?

> "Anybody should be able to reproduce the analysis and whatever steps are necessary to reach the conclusions that are in the paper"

TS: It would be valuable to have a way of fulfilling open data mandates in a user-easy and painless way of putting our data out there, with our papers and having it potentially cited and re-used. That for us is the biggest motivation for doing open data. Currently we do some open data, but it feels

very often that you are just uploading stuff on the cloud or on the server and that is it. Nobody ever uses it. Nobody ever cites it. It is nice to be open, but if there is no benefit of it to anybody, then it feels a little bit like a waste of effort.

TU Wien: What is the meaning of open science and how does it increase the value of your work as a researcher?

TS: For me it has to do with reliability, verifiability, and the re-use of the research we do. Currently at least it is tied to publication. Every single part of the publication needs to be open. The paper itself needs to be open. The data needs to be open. Anybody should be able to reproduce the analysis and whatever steps are necessary to reach the conclusions that are in the paper. So they should have the data available, and all the analysis available and they should be able to reproduce this whole chain and reach the same conclusions. That will help to improve the reproducibility of research.

> "It is nice to be open, but if there is no benefit of it to anybody, then it feels a little bit like a waste of effort"

TU Wien: How does the EOSC help to drive open science and facilitate open access?

TS: Making it very easy to share data and to share various types of different data in a way that is helpful and useful.

TU Wien: What are you missing in the discussion? What do I need to tell everyone?

TS: There needs to be a very clear explanation of what the EOSC is. Why are we doing this? How is it going to work in practice? There needs to be a single webpage. Not tens of pages or a couple of

pages. One webpage where I can go to understand what is going on and why I should care. There needs to be a clear message about what it is and why we are doing this. In terms of adoption, it needs to be simple and painless. If we have to jump to hoops or if fill in pages of metadata manually it is not going to be widely adopted.

Born in 1983 and educated in Helsinki, Finland, Toma Susi received his award-winning doctorate in nanomaterials from Aalto University in 2011. After moving to Austria for a two-year Austrian Science Fund (FWF) fellowship in 2013, he stayed to lead an FWF Stand-alone project. In 2017, Toma received the prestigious ERC Starting Grant of the European Research Council that funds pioneering basic research to create a new way to manipulate materials at the atomic level. The five-year grant of 1.5 million euros allowed him to establish a research group at the Faculty of Physics of the University of Vienna, where he is now an Assistant Professor on tenure track. Toma has worked on materials synthesis, spectroscopy, electron microscopy and modelling, authoring over 50 peer-reviewed articles and reviews mostly on heteroatom-doped graphene and carbon nanotubes. Toma has long been an advocate for open access publishing and was responsible for the open science policy work of the Young Academy of Europe as its Vice-Chair in 2018-2020.