

Interview Transcript: Melanie Henry interviews Magnus Palmblad on Curating Research Software in Bioinformatics and Omics

Transcribed by Melanie from interview 11/7/2025

Melanie: Magnus, thanks for joining me. To start, could you briefly describe your work?

Magnus: Thanks, Melanie. My work focuses on collecting, organizing, and curating research software used in bioinformatics, mass spectrometry, proteomics, and metabolomics. That includes identifying useful tools, describing what they do, tracking how they are maintained, and helping make them easier for researchers to discover through community resources such as *ms-utils.org* and broader platforms like *bio.tools*.

Melanie: That sounds incredibly useful, but also very broad. Why is software curation so important in these fields?

Magnus: These fields move quickly, and there are thousands of tools available. New methods are published all the time, but many researchers do not know which software exists, which tools are actively maintained, or which ones are appropriate for a specific task. Good curation reduces that friction. It helps researchers find reliable tools faster, compare options more fairly, and avoid wasting time on software that is outdated, poorly documented, or not suitable for their data.

Melanie: When you say “curating” software, what does that actually involve day to day?

Magnus: It is a mix of technical review and structured organization. We look at what problem a tool solves, what type of input and output it uses, whether it is open source, whether it has documentation, whether it is actively maintained, and how it fits into a broader workflow. I also try to standardize metadata so tools can be searched and compared in meaningful ways.

Melanie: What kinds of software are you most often dealing with?

Magnus: A wide range. In proteomics and metabolomics, that includes tools for raw mass spectrometry data conversion, peak picking, spectral library searching, peptide and metabolite identification, quantification, statistical analysis, pathway interpretation, visualization, and workflow automation. In bioinformatics more broadly, it can include sequence analysis, annotation tools, machine learning packages, and data integration platforms. And many tools to visualize your data of course.

Melanie: What makes software in mass spectrometry and omics particularly difficult to organize?

Magnus: One challenge is that the software landscape is very fragmented. Tools may be highly specialized, designed for a narrow instrument type, data format, or analysis problem. Another issue is terminology: two tools may do similar things but describe themselves very differently. And then there is the question of reproducibility—software versions, dependencies, containers, and workflows all matter. So organizing software is not just listing names; it requires context.

Melanie: How do resources like *ms-utils.org* help the community?

Magnus: They provide a more structured entry point into a very complex ecosystem. Ideally, such a resource lets users search by application, data type, algorithmic approach, or domain. It can also highlight links to source code, publications, documentation, benchmarks, and training materials. For new researchers especially, that can make the field much more accessible.

Melanie: And what role does a broader community-driven platform like *bio.tools* play?

Magnus: Community-supported platforms such as ELIXIR's *bio.tools* registry are especially valuable because they allow knowledge to be distributed rather than centralized in one person or one lab. The community can help identify missing tools, update metadata, flag broken links, and improve descriptions. That makes the resource more scalable and more representative of the field as a whole.

Melanie: What criteria do you use to decide whether a tool should be included?

Magnus: First, relevance: does it address a meaningful research need in the domain? Second, discoverability and accessibility: is the tool available to the community, and is there enough documentation to understand how to use it? Third, credibility: is there a publication, source repository, or evidence of use? Inclusion does not necessarily mean endorsement, but there should be enough information for users to evaluate it.

Melanie: Do you also assess software quality?

Magnus: To some degree, yes, though carefully. I try not to reduce quality to a single score because different tools serve different purposes. But I do look at indicators such as maintenance activity, documentation quality, licensing clarity, installability, test coverage when available, interoperability, and whether the tool has been used and cited by others. Those are all useful signals.

Melanie: How do you handle tools that are no longer maintained but still widely cited?

Magnus: That is a very important issue. Older tools often remain scientifically influential, even if they are no longer actively developed. I think they should still be documented, but clearly labeled as being "legacy" software, which is exactly what we do in *bio.tools*. Researchers need to know both that a tool was historically important and that it may no longer be the best choice for a new analysis pipeline.

Melanie: How much of this work is technical, and how much is community management?

Magnus: It is both. The technical side involves understanding the methods, formats, and workflows. The community side involves communication, consensus, and trust. Curation works best when researchers, developers, and users all feel that the resource is fair, useful, and open to improvement.

Melanie: Do researchers usually know what they are looking for when they come to these resources?

Magnus: Sometimes, but not always. Experienced users may search for a very specific type of tool. Others may only know the problem they want to solve—for example, peptide identification from DIA data, metabolite annotation, or differential expression analysis. Good software curation should support both kinds of users: those searching by name, and those searching by need.

Melanie: What are the biggest problems researchers face when choosing software?

Magnus: Information overload or overchoice are major ones. Another is incomplete or inconsistent metadata. Many tools are described in publications, but not in a form that is easy to compare. And of course there is survivorship bias: the most visible tools are not always the most appropriate. A curated resource can help bring lesser-known but excellent tools into view.

Melanie: How important is standardization in this work?

Magnus: Extremely important. Without some shared structure—consistent categories, metadata fields, and terminology—it becomes very hard to search or compare tools. Standardization also helps connect software resources with publications, data repositories, training materials, and workflow systems. This is why we develop and use the EDAM ontology.

Melanie: Are there differences in how software is curated in bioinformatics versus proteomics or metabolomics?

Magnus: Yes. Bioinformatics is broader and often more mature in terms of ecosystem size and infrastructure. Proteomics and metabolomics can be more instrument- and workflow-dependent, and data formats can be especially important. In mass spectrometry, for example, the relationship between raw data, spectral processing, identification, and quantification tools adds extra complexity.

Melanie: Have you seen the culture around research software change over time?

Magnus: Definitely. There is much more emphasis now on open source development, reproducibility, containers, workflow languages, and FAIR principles for software and data. There is also greater recognition that research software is scholarly infrastructure, not just a byproduct of a paper. That cultural shift is very encouraging.

Melanie: What do you think makes a research software resource genuinely useful rather than just a directory?

Magnus: Context. A simple list of tools is only the beginning. What makes a resource useful is when it helps users understand relationships: which tools solve similar problems, which fit into the same pipeline, which are interoperable, and which are actively maintained. The best resources support decision-making, not just discovery.

Melanie: What motivates you personally to do this kind of work?

Magnus: I like the idea of making scientific knowledge more navigable. A lot of excellent software is created by talented researchers, but it can remain hard to find or hard to evaluate. Helping connect people with the right tools feels like a practical way to strengthen the research ecosystem and support better science.

Melanie: What would success look like for this kind of curation effort in the next few years?

Magnus: I would like to see more community participation, better metadata standards, tighter integration with publications and workflows, and more sustainable maintenance models. Success would mean that researchers can move from a scientific question to an informed software choice much more quickly and with more confidence.

Melanie: Finally, Magnus, one personal question: what is something not many people know about you?

Magnus: Not many people know that I am actually left-handed. Some people notice because of the way I hold my cutlery, or eat soup holding the spoon in my left hand.

Melanie: Finally, Magnus, one personal question: what is something not many people know about you?

Magnus: Not many people know that I am actually left-handed. Some people notice because of the way I hold my cutlery, or eat soup holding the spoon in my left hand.

Melanie: That is interesting, has that influenced the work we talked about in some ways?

Magnus: Possibly, yes. Left-handed people may experience and interact with software through graphical user interfaces differently than right-handed folks. Perhaps even preferring menus or information to be organized or displayed differently. There are probably some studies of this I should look up, now that we got talking about it.

Melanie: Magnus, thank you again for speaking with me. It has been fascinating to hear about the work you do to make research software more visible, organized, and useful for the community.

Magnus: Thank you, Melanie. I am grateful for the chance to share a bit about it.

Melanie: And thank you to our audience for listening. We look forward to seeing you next time.

Magnus: Thanks everyone.