

# VARSHA KHODIYAR, SPRINGER NATURE

DATA CURATION EDITOR, *SCIENTIFIC DATA*

---

PLEASE TELL US ABOUT THE ROLE OF RESEARCH DATA IN SPRINGER NATURE. COULD YOU GIVE EXAMPLES OF THE PRODUCTS YOU WORK WITH?

Springer Nature provides several research data products: For example, *Scientific Data* is a peer-reviewed, open-access journal for descriptions of scientifically valuable datasets, and research that advances the sharing and reuse of scientific data. Springer Nature has made the research data policy texts, unless otherwise stated, available for reuse by the research data community under a Creative Commons attribution license. We also provide a date-stamped archive of our recommended repository list, which we will update every quarter. The archived list is available for use under the CC-BY license.

Moreover, the Research Data Support helpdesk is a service that provides free advice on research data policies of funders, institutions and journals and on finding research data repositories. It is independent of journal, book and conference proceedings editorial offices and does not advise on specific manuscripts. And, to help Springer Nature authors and journals follow good practice in sharing and archiving of research data, we're providing optional data deposition and curation services.

WHAT ARE THE MOST DIFFICULT DATA MANAGEMENT REQUIREMENTS FOR YOUR COMPANY? WHAT ARE THE RECURRING BOTTLENECKS?

I guess getting journal editors to thinking about research data, and how to do citation. A lot of help desk requirements come from our internal editors. We are kind of informally training Springer Nature people. Skills is something that people do need to work on.

Also, editors who have been editing scientific journals for a long time without necessarily working with data, see that, at the moment, they need to manage data in addition to everything else they need to do. *I guess it is a cultural change to consider data as being integral to what we are doing in science publications.*

A lot of people are now thinking about the types of tools you need if you want to get your research data properly available... It's quite scary for people who have not actually thought about those tools or haven't heard about the associated terminology. It's how to get those parts of the equation to the people that need to use them in a way that doesn't scare them. That is why we need to think about the user interface. There are people who are very qualified in their fields but not necessarily know about these aspects of data. *We need to figure out about how to present those existing tools and resources in a way that it's not scary but is also easy to use.*

#### HOW DO YOU DECIDE ON IMPLEMENTING SPECIFIC TOOLS AND RESOURCES?

In order to implement technical recommendations, we need to have the space and resources to trial different technical solutions and find something that works. The space to try different things out: that's something that we don't have.

It quite depends on who is your leadership team: what's their background, their knowledge and understanding of what this is and how much of these resources could be dedicated to it. In Springer Nature we are thinking about these things. That's in part why we have a research data team within Springer Nature. We are still in the phase where we are trying to make that change happen. We come from being in a position where the changes happened and now we are actually getting on and trying different solutions to see what works.

My feeling is that publishers have been pretty proactive in this. There are certainly things that we publishers can't do, things that we need to get going. If the community decides "this is standard for this discipline, for this type of assignment, we expect you publishers to enforce it". I don't think you'd get any journal that is going to say "no, we are not going to enforce that standard".

#### REGARDING THESE DIFFERENT SOLUTIONS, COULD YOU MENTION THE INITIATIVES IN WHICH SPRINGER NATURE IS CURRENTLY INVOLVED?

Springer Nature staff have taken active roles in many of the external initiatives currently underway to promote research data sharing. These include multiple RDA groups, Force11, Credit for Data Sharing and ELIXIR. It is easier to implement policies and guidelines such as Force11 data citation principles, and the FAIR principles, and less easy to implement technical outputs such as Scholix. **For outputs to be easily implemented by publishers, these must be funded and maintained for the longer term.** To date, the majority of outputs have been proofs of concept, rather than ready to use outputs.

#### WHY ARE TECHNICAL OUTPUTS SUCH AS SCHOLIX MORE DIFFICULT TO IMPLEMENT?

At the moment Scholix works because it is looking at citations. And citations, though not completely standardized, are much more standardized than most metadata around and other kind of research objects. Thinking about how you put metadata in a standardized way, you need quite simple tools that can do quite sophisticated things, but the actual using of the tool has to be quite simple. It's technically not difficult to do. You want to start by the foundation and then everything will follow.

#### COULD YOU IDENTIFY OPPORTUNITIES FOR ACADEMIA AND THE PUBLISHING INDUSTRY TO COLLABORATE TO THEIR MUTUAL BENEFIT?

The development of discipline specific academic data repositories, (especially in non-life sciences disciplines) would be welcome. Research data repositories must be able to provide long term persistency and accessibility of hosted research data, so as to be usable by both researchers and publishers.

Standardization of metadata for all aspects of data sharing and interoperability (e.g. citation, discoverability, indexing) is a key area that will require joint publisher and academic collaborative efforts.

#### IS THERE ANYTHING ELSE THAT PUBLISHERS COULD DO TO SHARE OR REUSE THE RESEARCH DATA?

Making every journal editor to think about what data they need to focus on... Things like that, over time, will definitely make research data easier to access. **We are still in the phase where we need to make the research data findable. Moving along, we need to think about what happens when we get to the point where the data that is being shared is actually reusable.** That's quite difficult; it's technically difficult. Different experiments from different disciplines will have different standards that need to be implemented in order for that data to actually be reusable. The community is coming out with their standards and we, as publishers, convey with those standards. As publishers, we are not in the right position to come up with their standards ourselves.