

***The Scholarly Kitchen Podcast:  
“101 Innovations” and Scientific Workflow***  
**Transcript of 22 July 2015 show**

**Stewart Wills (Host):** Welcome to the Scholarly Kitchen Podcast for July 22<sup>nd</sup>, 2015. I'm Stewart Wills, and today the topic is scientific workflow. It's a term that publishers and others concerned with scholarly communication throw around a lot these days, as they seek ways to understand and serve the needs of their target audience. But it's a moving target—what does workflow really mean in an era in which scholars are continually adopting new digital tools to solve some aspect of getting the job done?

Well, fortunately we have two people on the line today who've thought a lot about that very question. Jeroen Bosman is the geoscience librarian at Utrecht University Library, and Bianca Kramer is the life science and medicine librarian at the same institution. Among other things, they've started a project called “101 Innovations” that's seeking to amass data and insight on how digital tools are reshaping how scientists and other scholars do their work. Jeroen, Bianca—thanks very much for being with us today.

**Jeroen Bosman:** It's my pleasure.

**Bianca Kramer:** Thanks very much for inviting us.

**SW:** Well, first of bit of background. What is the 101 Innovations project, and how did you come to get this started? Bianca, could you start us off about that?

**BK:** Sure. It's a project, basically, that's tracking changes in the scholarly communications landscape, and doing that by looking at the tools people are using, and the changes in the tools that they're using. And we started about one and a half years ago, partly out of our own curiosity, and also based on the idea that we want to broaden our services to researchers beyond just supporting discovery and publication, but also reaching out into other areas of research support.

**SW:** So why “101 Innovations”?

**JB:** Well, let me take that question. First of all, just, the title catches the attention, and also it makes clear that “101” is a lot, but it's not exactly one hundred—it can be any number. Actually it is by now a lot more; we have charted some five hundred of these tools. What we also like about this name is that it's also a sort of reference to all the names that people give to their introductory courses. So it's also a kind of introduction into thinking about tools and especially thinking about creation of new tools. And finally this is also a tactical choice, of course, because any title beginning with the number always sorts nicely; comes up on top of all kinds of lists. So that's tactical as well.

**SW:** Okay; well, let's talk a little bit about some of the details here—starting with, I suppose, what scientific workflow actually means. As I as I alluded it's kind of, kind of a slippery term. You have identified, I think, six phases of scientific workflow. Can you really just quickly take us through those phases?

**BK:** Yeah, sure. First of all, like you say, it's a slippery phrase. And we made a very simple model, but like all, models reality is more complex, and we're now seeing rather than a really cyclic model, like we describe, it's becoming multicyclic, and it's also becoming multiordered. It's not—these activities that we describe are not gone through subsequently, as you see in many interfaces now.

But the phases we distinguished were discovery, so looking for information, looking at what's known about the subject that you're researching; analysis, gathering data, analyzing data; then writing that up for publication; sharing your findings with others; doing outreach and making your findings known to other scientists and also to the general public; and finally assessment, so, assessing your own impact, and the impact of others. And you might say that there's even seventh phase—or maybe a first phase—and that's preparation: grant application; getting ideas; finding collaborations and so on.

**SW:** Well, I'm curious. Because, you know, you mentioned this notion that it's not necessarily cyclic, but multicyclic. I do remember you had diagrammed this as a kind of a circle, where, where the workflow kind of kept feeding—phases of the work flow kind of kept feeding into the next iteration. What do you mean by “multicyclic”?

**BK:** That within each of these phases, people cycle within these phases. Like, when you do writing, you don't write in one go—you do a draft, you do comments on a draft, you rewrite it. And the same for publication; you might submit, go through peer review; you might have to resubmit, revise, or submit to another journal. Those kind of internal cycles within each of the phases.

**SW:** I see. And it's interesting, because that suggests that each of those phases kind of has a different set of drivers for, you know, what the researcher is looking to do, and for sort of user behavior. What are the kinds of things that, in dealing with these various phases, that you think drive researchers in the toolset that they choose to kind of master those things?

**JB:** Well I think there are indeed many different considerations there. For instance, it might be what people were brought up with—not in their early youth, but as an early-stage scholar, or what they get kind of advised from colleagues, from their P.I.'s. Also just, plainly, what is available; if there's a cost attached. But of course in the end it comes to the functionality that the tool offers, and its friendliness. But then there is this problem that people have also invested in tools, and the more they have invested, the more that there will be a kind of burden for them to switch tools, because they have to learn new tools, and the things they have invested in might be thrown away, sort of. And also of course differences arise over academic fields, and researcher roles, and even, perhaps, between characters. Some people just don't like switching; they'd like to stick to what they know.

To really get a grip on this, what we did is, we introduced a very simple model. We termed that the “GEO” model—so it's “G,” “E,” “O.” And that looks at what tools contribute to making science either more efficient, so that “G”; more open, that's the “O” of course, or more reproducible, that's the “G”; we call it *gut*, efficient open, but *gut*, you might interpret that as being more reproducible or fair.

So for instance, the DOI—you could call that a tool that that is just for efficiency, to make some scholarly communication more efficient. But a thing like Archiv, much used in physics

and astronomy and maths, is a tool that makes, of course, a lot of research more open. Whereas a very recent tool, like Publons—I don't know if the listeners know that tool, where people can claim their peer reviews—might be called a tool that makes science more reproducible, by getting more attention to doing your peer review and getting the credit for that.

**SW:** Well you know, it's interesting that you mentioned that there's a certain, sort of, activation energy in getting people to switch tools. One aspect I found particularly interesting is that you have identified in some of your work, for each of the workflow phases that you talk about, one development that—on the horizon or ongoing, that's potentially most significant and one that's potentially most disruptive. And I was wondering if you could talk a little bit, just give some examples of those, of those observations.

**BK:** Yeah, sure. These took an enormous amount of discussion among us, to get to this point. And they're mainly aimed at triggering discussion, and framing developments, and getting people to talk about these kinds of developments. We base them on our own experience, our own ideas; also the literature, and looking at all the different tools that we saw in doing our inventorization, our inventory. And some examples—when you look at publishing, one thing that will be interesting is to look at the effect of journal and publisher status; whether that will remain the driving force in how people decide where to publish, or whether that will diminish and other publishing models will gain some traction, or have the opportunity to gain traction, like PeerJ, or Science Open, or The Winnower. So that's one issue, for publication.

And when you look at assessments, an interesting thing there is to see whether we'll keep the focus on the really quantitative measures, like citation and impact factor, all these internal quantitative measures, or whether it will become more important how scientists can explain to the public the importance of their research—to prove their contribution to society; whether that will become a more important factor in assessing impact.

**SW:** So, you know, as you mentioned, a lot of this was kind of based on discussion and, and on, in essence, kind of your modeling of the process. But as you know, earlier this year you actually started up a survey to try to get a handle on the tools that people are actually using. Maybe you could talk a little bit about that survey—how it's set up and what sort of audience it's trying to tap.

**JB:** Yeah, sure, I'm glad to do that. This survey was intended, was always intended to be just a very broad assessment of what people are really using. And that information was not widely available, is not still available yet. We do have, for instance, the three-yearly faculty surveys, by Ithaca, which are very nice, but they focus more on general academic behavior and not on the tools and sites themselves. And what we do with this survey is really dive into the usage of those tools. And we do that with a Web-based survey, and that has all the pros and cons associated with that, of course. It's very broad, but it's a largely unknown sample, of course, that you get, with all kinds of biases in there that we will have to remove afterwards, try to remove. It will be a quite long running survey; we'll keep it open until February 10 next year, so that people from all kinds of countries and all areas will have a lot of time to really respond to our survey, and also for us, that we have time to distribute it widely.

In the survey, we ask for tool usage in seventeen activities, over those six research phases. And what we also do, of course, we have some basic demographics of all respondents. So

their fields, the country that they are in, their position in academia, their career length—we measure that by their first peer reviewed journal article that was published; in which year was that. So we have those demographics, so we can break down afterwards over those four categories. And what we also purposely, on purpose did is try to make it very user friendly. So it's short; it's graphical; people really tell us that it's easy to do, fun to do. I mean even that people learn a lot from it—they get into contact with all kinds of new tools that they try afterwards.

**SW:** Okay; so you as you as you point out it's still an ongoing survey; it's going to be up through February. But you did recently release some very preliminary results based on I think the first thousand responses to the survey. What—can you say anything about what those early results are telling you, and whether there are some, you know, some of these many available tools that are that are either starting to gain scale, or that are finding—that, you know, as you suggested, maybe old tools that are that people are still really stubbornly sticking to?

**JB:** Well, I don't want to disappoint you there, but actually it's really too early to tell with any confidence, although we did publish that, but more as a kind of teaser to let people know what can come out of this if they participate and, really, if a lot of people start participating. So really the sample size, and the bias in those first one thousand is just—the sample size is too small and the bias is too strong. For instance, we started distributing it in our own circle, so that's mainly librarians. So we have a lot of librarians there right now. That's diminishing, and we've got a lot more researches, and we see people from other countries coming in.

So it's difficult to tell. And also, what we didn't analyze yet in those preliminary results is all those tools that people mentioned that we didn't preselect for them. For each activity they can select one of seven tools that we preselected that we find important or stimulating. But of course if they use different tools, they can add them themselves—and they massively did! So right now, with these one thousand first results, we already had some eighteen hundred tools that we didn't list before, that people mentioned. So we have to analyze them, and have to see what is all that they are using, and the reason . . .

**SW:** So it created a lot of new work for you!

**JB:** Yeah but, but we like it! It's very interesting to see what comes out of that.

**SW:** Well, notwithstanding the fact that it's too early to tell, as you say, there were a couple of things that really jumped out at me, at least. You looked at authoring tools, for example, which I think a lot of organizations are looking at as a potential growth area, but it really looks, from the early survey at least, like a lot of people are still using Word.

**BK:** Yes, definitely, and we can, of course, hypothesize a bit about why that is. I think especially in the area of writing and authoring tools, developments are very slow, because these are tools that people use every day, and that might explain why they're slower to change, why they're not so easy to give up on. For this new generation of authoring tools, the benefit for switching may lie in either the ease of collaboration, or the integration of different activities, like reference management, even journal submission; if these tools make that easier to do that in the right format. And it interesting that, a couple years ago, Peter Kreger did also a kind of survey. And at that time he found that researchers do use Google Docs for instance

for informal collaboration, but not for more formal writing. The reason for that is the lack of the more sophisticated aspects like reference management and those kind of things. So maybe these new kinds of tools will gain traction, but it will probably be slow and we'll have to see.

**SW:** So what about the way that researchers are archiving and sharing information—you know, where they go to set up research profiles and the like. Any preliminary thoughts about what you're seeing there?

**JB:** Yeah, well, that's an interesting set of tools. What we see there is that there's actually quite fierce competition, and what's interesting is that many of these tools have to potential to be used for various activities in various phases. They can be used as reference manager; they can be used for sharing publications, uploading your own publications, writing to your profile; of course, also getting out with your profile, with all the things you are doing—not only your publications, but also your presentations, etc., and having contact about all of the things that you're doing, about the projects that you're doing with your peers worldwide. And what you see is that many publishers also invest, and I think strategically, into these kinds of tools. For instance, a tool like Mendeley for Elsevier, and also Papers for Springer; they can be used in a lot of these phases, and I think they are probably the tools to watch out for. And it might be that users really start using them for more than just storing your papers.

**SW:** So almost, kind of like “one-stop shopping.”

**JB:** Yeah, perhaps. But also, on the other end, this is also the area where there is really a lot of new contenders. Very broad things, like—technical things like ORCID, but also think of Loop, or Incent, which is a Dutch initiative, or ScholarBridge, or ResearchConnection. Or also think of Google Scholar citations, which is a kind of passive profile right now that a lot of people have. But Google could build on that quite drastically, and perhaps even blow away all the others.

**SW:** Well, I guess what you're saying kind of leads into another thought that I had, looking at the work that you've done in the initial results of the survey. It occurred to me that we're really talking about here potentially thousands of different workflow chains—maybe even tens of thousands or hundreds of thousands—because we have all of these different tools kind of chained together in different ways. You know, as you, as you are looking at pulling these results together and trying to interpret them, you know, over the rest of the year and into early 2016, do you have any thoughts about how you can roll up what you find into kind of a general, a larger set of conclusions? How are you going to manage that?

**BK:** Yes, you're definitely right. It's going to be a lot of number crunching; we're dealing with a lot of results, and like Jeroen already said, we need to correct for bias in various ways. But it's important to realize that simply using a different tool doesn't mean that you're having a different kind of workflow. It doesn't necessarily take the character of your workflow. And we regularly use look at the bigger picture, and identify some dominant workflows that are emerging, and looking at the different types of tools that people are using together.

And also, for instance, looking at whether people are going to use sets of tools that are supplied by one provider. We see publishers strategically buying tools to provide this one-stop shop, maybe, this ecosystem, in which the whole research workflow can be done within their ecosystem. It will be very interesting to see if people are actually going to work that

way. But these are some of the bigger patterns they we hope to identify.

**SW:** So beyond completing the survey, what do you view as the next steps for 101 Innovations?

**JB:** Well, actually, we are right now in the midst of this survey, and we want to get it out to more people. We want to cooperate with people distributing the survey; so quite a lot of institutions already are doing that. And work also toward a more global response, by having the survey translated in Chinese, in Russian, but also in Spanish and French—looking out for partners in those language areas to work with. But really, after the survey and after we have done the analysis of the results, we have another quite nice option, which is a follow-up with a selection of people that have indicated that we can contact them to ask additional questions. So that offers us the opportunity to look into the why and the how—why do people use tools in some combinations but not in others, and how do they work together? Why do people switch from one tool to another? Is that out of frustration, or because a colleague says that it's a nice option to use, or whatever. So really right now we are trying to collect the basic numbers and data, and afterwards we do more qualitative follow up into the how and why.

**SW:** So where can our listeners go on the Web to learn more about the project, keep track of it, and take the survey?

**BK:** We've got the Wordpress site; it's called [101innovations.wordpress.com](http://101innovations.wordpress.com). And that has all the information about the survey; also about the option for institutions and societies to get a custom URL, so they can spread it among their members and get the data for their own people. And the preliminary results are there, and also some information about the other around this project. So it's [101innovations.wordpress.com](http://101innovations.wordpress.com).

**SW:** Well Bianca Kramer, Jeroen Bosman—thanks very much for being with us.

**BK:** Thank you very much.

**JB:** Thank you very much; it was my pleasure.

**SW:** And thank you for dropping in to The Scholarly Kitchen Podcast for July 22<sup>nd</sup>, 2015. Be sure to visit [scholarlykitchen.sspnet.org](http://scholarlykitchen.sspnet.org), where every day, the Kitchen's team of pundit-chefs serves up a fresh helping of what's hot and cooking in the scholarly publishing world. You can also comment on this podcast episode on its blog page, and we'd love to hear from you. Thanks to the Society for Scholarly Publishing for its support of this project, and for hosting our audio files, and to the American Association for the Advancement of Science for use of its studio and production facilities. This is Stewart Wills from The Optical Society; until next time, on behalf of SSP and all of the chefs in The Scholarly Kitchen—*bon appétit!*