

TECH[NOCULTURE

The Open Science movement

Episode 13

Full transcript

Guest: Sabina Leonelli [Sabina]

Host: Federica Bressan [Federica]

[Federica]: Welcome to a new episode of Technoculture. I'm your host, Federica Bressan, and today my guest is Sabina Leonelli, who is Professor of Philosophy and History of Science at the University of Exeter in the UK, where she co-directs the Centre for the Study of the Life Sciences. Her research focuses on the methods and assumptions involved in the use of big data for discovery, the challenges involved in the extraction of knowledge from digital infrastructure, and the role of the Open Science movement within current landscapes of knowledge production. Today we will try to zoom in on the topic of Open Science, which is very actual in all fields of research, but it's also important for every citizen and society at large. So, first of all, welcome to Technoculture, Sabina.

[Sabina]: Thank you.

[Federica]: I first got in touch with your work precisely because of Open Science. You do many different things, some of which we just mentioned, but when did you develop this interest in Open Science?

[Sabina]: Well, my main work is in philosophy of science. So, I have a strong interest in the ways in which science is governed, and what does it mean to do research, what are the components of research, what is the epistemology of research, how actually we get to know what we know, how do we prove it, how people use models, data, etc. And so, as part of those interests, I focus particularly on the study of data, and in my interest in data I encounter the increasing discussions around open data, in what it means to disseminate and publish data in the first place. And that led me to think of it more broadly, about Open Science. So, I would

say I've been involved in discussions around openness and Open Science since maybe 2012-2013.

[Federica]: We hear a lot about Open Science today, like with capital letters O-A, it's a thing, it's been promoted very much: in the scientific community sometimes, I would also say, it's been "pushed" on researchers, like this is the direction where research practices need to go now. And there's much good in this, but it seems like, it's a new thing, it's been talked about now. So I would like to ask you if Open Science is really something new, or maybe it was already there in the past but it wasn't called Open Science.

[Sabina]: Well, I think Open Science is really a subset of a much broader movement which involves things like open government, open knowledge, around how do we disseminate information. So, part of the reason why it has become so popular and so visible these days is the dissemination of technology and the availability of digital technology, ICTs, the Internet, and the fact that people are feeling more and more that there are alternative ways of disseminating information than the usual official channels, you know, the typical traditional publishing industry or, you know, formal policy documents, things like that. So, partly there is that, and also partly there is a strong feeling, and in particular among governments, that there are problems in communication, in particular the communication of expertise, at this point in time, that there is a lot of mistrust, particularly among, you know, developed democracies, in the developed world, of government, of expertise, of the credibility, really, or research itself. And so, what is needed as an antidote to this increasing mistrust, is to make knowledge as open as possible, so that people have the feeling that things are transparent and they're not being hidden from them in any way. So, you know the idea of Open Science per se, so the idea that you can actually freely disseminate and make available the results of research is very old, I mean it wasn't called Open Science, but many communities in the sciences have been operating like this since hundreds of years, I mean, in astronomy...

[Federica]: It's in the spirit of science.

[Sabina]: Exactly. I mean, it's a very very old idea. I mean, one could say, you know, the idea that there is scrutiny over each other's results, and all the results of science are laid bare to the world so that people can criticize them if they want, I mean, that's really what makes science science, right? So, that's really not a new idea at all, but I think what makes it new or at least attractive as a movement now are these components. On one hand trying to address the mistrust of the population towards expertise, and on the other this idea that we need to anyhow **rethink communication** because of the technological opportunities we have.

[Federica]: I think that we've already touched on a very crucial point, and that is trust. Open Science is about opening up your data, and like you just said, the transparency, but there is definitely an element of wanting to see your data also to be sure that you didn't tamper with

it, didn't massage it in favor of the hypothesis you wanted to prove. I guess that Open Science is not about policing your peers' and colleagues' job, but there is an element of that built into it, built in the narrative of how it's presented. The will of science to ascertain things, to verify things, to know for sure, is not motivated by a mistrust in the beginning. So, I'm not saying it's not a good thing to want to check things, but not because there's mistrust. And in the case of opening up the community's data, that shouldn't be, I think, the main motivation, but indeed I am aware also of the mistrust that the generic public has for science in general. It's done with public money, sometimes, you know, like in every field, scandals make it to the mainstream media, and there has been a process of eroding indeed the respect besides the trust that people have in science. So, I think that mistrust here can be intended in two ways: it's the mistrust that peers and colleagues have for each other's work, and also the mistrust that the generic public has for science as one social collective endeavor that is being funded with public money. Can you tell me when this feeling of mistrust started? How did it come to grow?

[Sabina]: Well, I mean, so there's two different things here, I think, that are going on. First of all, there is the fact that science, certainly over the last thirty years, has changed enormously, it has become more and more specialized, more and more people are coming into the field of scientific research around the world, in many nations there are kind of flourishing research programs, in a way that wasn't the case, say, 50 years ago. Now, that means that because research keeps being more and more and more specialized, it also becomes more and more tribalized. So you get people that work in fields which are more and more narrow, that are losing the capability to talk to each other, because people are adopting very specific terminology, very specific instrumentation to look at very particular parts of the world. In a situation like this, it's quite easy to understand why the general public would feel that actually all this work that's happening in academia and in research is incomprehensible to anybody who doesn't have many many years of training to understand what's going on. And, I think, partly it is really a result of how science is being done right now, it's almost unavoidable. On the other hand, it's also because there is little incentive really within the scientific world to devote some time towards engaging, like, other types of people in the work that you do, explaining to the general public what is going on in your research, and how that may affect what they do. So, you know, I think there are actually good reasons in contemporary science to think that there is a bit of a gap between how science is represented and publicized and understood in general terms in the media, and what actually is going on in scientific laboratories, in academic environments. So, in that sense, I think, it's justified, the fact that there is a growing mistrust. I mean, there's not necessarily a ground for this exactly, the thought that people are doing things which are not relevant or credible, but it's just that it's becoming more and more difficult for people who don't have certain kinds of training to assess or in any way even just interpret what people are doing in these [?] specialized jobs.

[Federica]: I struggle with accepting this feeling of mistrust, because I personally don't

share it - although I might have a special perspective on this, because I am a researcher, so I see the processes of science also happening from the inside. Although, I would say I have a fundamental trust for colleagues in fields of which I know nothing about. And I will admit this is something, of course, that cannot be denied, that there is a lot of pressure on researchers today to publish as much as possible: that means that you have to show that you have a lot of results, and not just results but positive results at that. That means, I have a hypothesis, and I test it, and I need to be proven right. Now, that's a shame, because sometimes you're just curious to find something out, and if you're proven wrong, as a scientist I think you should be kind of equally happy, because you have learned something new. But we all know that to say "I had this fancy hypothesis and it was proven right", that's considered more a success than of course a negative result. So, I think that with digital technology used to generate, extract, process, store, transfer, share data, which makes it fairly easy to tamper with it, and in a few clicks you can change parameter in large datasets and still have a clean result to present, that, well, if we don't want to live in a naive world, then, yes, there will be groups of people who tamper with their data, precisely because, I think, there is a tendency to want to steer data in your favor. Now, that's always been happening, probably, but now maybe more so because not only digital technology makes it easier, but because there's more people involved in the community, like you were saying before. So, with these two elements combined, the pressure on researchers to produce a lot, and the ease to massage your data, then, yes, I can imagine that the will to police peers' and colleagues' job becomes a necessity. What I would like you to tell us is something about how much this aspect is relevant within the Open Science movement. How important is this aspect of policing data in the overall narrative of how transparency and access to data is presented?

[Sabina]: Well, the idea is that in order to be able to assess what is going on in science - even for an expert - ideally, you should have access to as many components of the research process as possible. And that hasn't really been the trend in scientific publishing over the last 30 or 40 years. So, people have gotten more and more used to value, you know, hypotheses more than anything else. So, so you have a hypothesis [?] in your work, you then go off, you set up a whole research and experiment, to try and prove your hypothesis, you maybe develop a particular software, or particular instruments that allow you to produce data and analyze data. You then do these analyses, you kind of tweak it in lots of different ways. But ultimately, what people are encouraged to publish is only the claims that they get out of this research, that they feel may be true, and the little bit of data that proves, kind of, in a sense, incontrovertibly, that these claims are true. And for a long time this has become really what we think about when we think about knowledge in general. For a lot of academics knowledge is written a textbook, whatever is written or published in a paper, not all the activities, the process, the protocols, the techniques, the instruments that are actually used to perform those activities. So, I think, part of the motivation behind Open Science is a great recognition that actually a lot of different types of work goes into doing research, and only small parts of this work

is now recognized and rewarded as what really is knowledge, right? So, people who become famous and get Nobel Prizes are typically people who write very theoretical papers, that prove a hypothesis. People who set up very complex instruments, people who spend a lot of time curating data, so that other people can pick them up and reuse them, these people are not typically rewarded. Sometimes [they are] not even seen as scientists. But that work, arguably, is actually as important for the production of knowledge as is the work of people who work on hypotheses and theoretical knowledge. So, part of the idea behind Open Science is to make sure that people are valuing and also evaluating and rewarding every part of the whole process that is the scientific process of knowledge production. And that means, actually, make it more evident and managing to document things like how you set up a project, how you design it, how you set up instruments, what protocols you're using, what software you're using, however many data you've produced, even if in the end you're only using a certain subset of data to prove whatever hypothesis you want to support. So, the idea being that, you know, if you manage to document all of this, and to make it public, first of all you give people a much better sense of how you've done your research. You make it easier to check its validity, you make it easier, possibly, if you are able to reproduce it, and at the same time you also make it possible to pick up some of these other elements of research and reuse them, even for people that maybe are not that interested in the hypothesis, but maybe they're interested in your data, or maybe they're interested in your software. I mean, software is a easy example, I think, because it's the typical case in which you develop an algorithm or a program to analyze a certain kind of data, maybe in cell biology, because you want to be able to check whether a certain set of genetic data is associated to a certain phenotype, a certain trait of the cell, but in fact it turns out that that particular algorithm you've developed could be applicable to compute different types of data, maybe in musicology, maybe in archeology, maybe in text mining. And [experts] who come from those fields could look at the software and think about a very different use, just as productive, but in a different field. If that software hasn't been made public, because people who developed it thought "well, it's not really relevant, what's really relevant is just our publication at the end, it's not really all this stuff that we've developed to get there," well, then you're never gonna get that kind of reuse. And so, this is really a core motivation for Open Science.

[Federica]: Can you talk a little bit about the problems and limitations that we may encounter when we actually want to implement this very good concept for which you've made a great case, that is opening up data. I imagine copyright issues or, to begin with, the fact that it's very time consuming to curate the data before you share it?

[Sabina]: Yeah. So, it is indeed extremely difficult to open data, I mean, anybody who's ever tried to do that knows it, partly because what you don't want is to simply dump your data somewhere. You know, this difficult problem of data dumps, that you create these archives which are not organized. It's very difficult to find data of relevance in there, or to retrieve them

in any way. I mean, to make data public in that way is almost equivalent to not making them public, because if there's no structured way to look for your data and organize them, they really are not usable to anybody. So, it is true that curating the data is absolutely essential. Now, I will be actually the first to say that I don't think opening up all the data is really necessarily what we really aim for here. I mean, precisely because it's so time-consuming, so laborious to put data online, what people really need to think about is okay, which data are really precious here, how do we choose which data we prioritize in this exercise? So, you know, if you look in a simulation data, for instance, well, maybe these are not a big priority, because at that point it's better to make your software available online, rather than the data themselves, they're less valuable. If you're thinking about very specialized and very laboriously produced experimental data, or a unique sample, well, those may be exactly data you want to be able to put online, because they're very very difficult in fact to reproduce, and that makes them precious as a resource that somebody else may want to access and reevaluate. So, that is just a beginning consideration. But, I think, the biggest battle really at this point in the whole field of open data and more generally Open Science is actually the battle of rewards and incentives. So, in that sense, a lot of the work also done with the European Commission on Open Science is about the rewards that are attached to working openly in research. Because at this point in time, really, there is basically no reward. So, anybody who works in academia has to get promoted and hired, and basically make a living out of that job, well, the criteria on which people are evaluated have very often nothing to do with Open Science behavior. All that many governments do, or many universities, is to look at your impact factor, look at in which prestigious journals you're published, and how many papers you've published, and that's how they decide who to hire. Obviously this is completely contrary to trying and incentive a behavior where you actually do spend time taking care of your data, you do spend time putting them online, because these activities don't really lead to any publication necessarily. Not in this traditional sense. So, the first step towards trying and implement open data in any meaningful sense is basically to get rid of impact factor as a way to evaluate researchers. Because as long as that is in place, there's not a really full of Open Science to go.

[Federica]: I have tried to engage with Open Science myself, especially with the Marie Curie project that I'm doing now. I wanted to be transparent, that means of course that I try to publish Open Access, but also that I thought how I could open my data up. And you know what was interesting in the beginning? That asking myself the very question "what is my data?" was not obvious, was not easy to answer. I think that it's a very good thing if new generations of scientists are trained from the beginning, you know, in the mindset of Open Science, because there are some questions that come with it that I find positively stimulating. For example, what are my data? How should I curate them? Provided that I have the time and resources to curate them, how should I do that to make them the most readable and useful to other people who might access them? And I blamed this hard time I had in the beginning on the research field, because there are other research fields that are not my own, so I might be wrong here,

and that's my question for you, because I know you have a background in biology and I would imagine that biology and the medical field in general have it easier, because practices are more standardized, whereas, you know, in the Digital Humanities, where I work, it's really not clear, nobody does exactly what I do, and so it's hard to structure my data and decide what data to share, etc. Would just say that biology is one of those fields where we observe more engagement with Open Access, because it's easier to? And I don't mean easy in absolute terms, I mean easier than other fields.

[Sabina]: Well, I'm personally attracted to fields where that's absolutely not the case, and I would take biology to be a very very good example of a field where there's no agreement whatsoever on standards, on methods, on techniques, even on what counts as data very often changes, even within the same subfield. And what interests me is the fact that there's a very good reason for this. And the reason for this is that the work that people do in different parts of biology, as in environmental science, as in biomedicine, is very specifically tailored to the particular objects of the research. So, somebody who studies salamanders would not necessarily be using the same methods, the same tools, or even the same types of data, as somebody who studies worms or somebody who studies gorillas. This will have to vary, because your methods will be more and more tailored and adequate to what you have to be looking at in the world. So, in biology, the fact that these methods have been pursued over hundreds of years has created a situation where almost every tiny little group uses slightly different measures, slightly different criteria, and of course that makes it very difficult to share the data, and to interpret them, because everybody has different points of view.

[Federica]: So, even to people studying gorillas, but independently, will produce different data sets?

[Sabina]: It is very likely, because they probably will study in different locations, and would have slightly different theoretical perspectives, they would use different presuppositions. I mean, of course it depends on the fields, but there tends to be a very very little standardization in the ways in which biologists work.

[Federica]: There is something that I have learned in my experience with Open Science so far, not just opening up my data but also trying to access somebody else's data for my own research, and that is irrespective of how well-meaning the individual researcher can be: in order to achieve long-term structural change, there needs to be a clear decision at policy level, which needs to translate to platforms and even a reward system like you are saying before. I know that you have conducted a study where you have actually interviewed many scientists across the world, and asked them how they feel about Open Science, how far they are in complying with this new guidelines. What have you learned from that study?

[Sabina]: I've spent a lot of time arguing back against people who work in universities management or even politicians, who think that researchers are very resistant to Open Science, they hate it, we have to impose it on researchers. I really don't think that's the case, partly because the history of many many fields in research for centuries as being one of Open Science. So, I think it's kind of ridiculous to think that Open Science is a political invention that is now being imposed on research. What I, however, also think that is very problematic is the fact that because of the incentives we just talked about, researchers will have a tendency of closing down more and more, and in fact people who behave in Open Science are punished in the current way of doing research. And so, that is what creates the confusion and, you know, the tensions that a lot of researchers feel when they think about Open Science. So, we did indeed a lot of in-depth interviews with researchers both in the UK, and in Europe, in the States, and in parts of Africa, around the perception of openness, what they associate with that word, what they think about Open Science, if they have heard of it, and these kinds of things. And generally, the message was that everybody was very attracted to Open Science as an idea, they all wish that they actually could behave in such a way and that there were incentives towards that. But they are all pretty much recognized that there were big obstacles in their way. And that it was really not easy for them, given their own institutions and their own jobs and the responsibilities, also towards other people in their group, to implement these behaviors. For instance, a lot of PIs, a lot of heads of labs, told us that they would love to be able to only publish in Open Access journals, but these are often not necessarily the ones that give you the highest impact factors. They are senior people that don't necessarily care too much about this. But they're worried that the postdocs and the PhD students that published with them would lose out if they decide to only publish in Open Access journals, because these journals are not as well recognized as journals which are proprietary, which are closed to the public, but actually have a highly impact factor. So, things like this are really what's on the mind of researchers. One other thing that causes a lot of confusion is **the intersection between Open Science and intellectual property rights**. So, this is partly because researchers typically are parts of lots and lots and lots of different networks all at the same time. So, on one hand of course they work in one or two institutions, and so on, you would think that they would just have to follow the policy of their own institution, if there is a clear policy, which is not always the case. So, you know, for instance: my institution thinks that whatever data are produced in the institution are at least to some extent property of the institution. So you can follow this guideline. However, very often research is funded by external bodies. So, what happens here: these external bodies and external sponsors also make claims of property over the data, for instance. It's also the case that researchers collaborate a lot. So, they are in very complex networks of collaborators. So, what are their responsibilities vis--vis the institutions of their collaborators? What happens if a funder that funds one of your collaborators and not you requires certain things of your project? How do you deal with that? And then of course there are national governments. Which are even more complicated, because typically now, especially in Europe, there is a lot of legislation that is coming down from the governmental side

about how to deal with Open Science, and is also very often in contrast or at least in tension with some of the other guidelines that you get from other institutions or networks you're part of.

[Federica]: Can you give an example of that? Something that is in contrast?

[Sabina]: Yeah. So, for instance, France has very recently decided to introduce a new national legislation that favors Open Access, and in fact that legalized the publication of preprints. Basically, the legislation states that if research is - well, broadly, of course, I don't want to be liable for this, but my prediction of it is that if you publish work, if you produce papers with public funding, then by law you are basically obliged to make at least a preprint of these papers available online, free of charge. The interesting thing about this legislation is that it goes directly against some of the embargo policies of publishers, or at least some publishers, which actually state that if you publish with them, you're not allowed to put your work online, and certainly not for the first, say, six or twelve months after publication in their journals. So, this is an example of a situation where things are really unclear, and this actually was very deliberate on the side of the French state, they're just trying to push the idea and push publishers into admitting the fact that it is important to have work published Open Access. But at the moment that means that a lot of people are caught in between, it is very not clear to them what to do. For data, very similarly, I mean, one case I also didn't mention before but it's also problematic, is what happens when you do work which is at least partially sponsored by private money. So, you do work which is a public-private partnership, in collaboration with an industry. Very often, those kinds of collaborations dictate terms of data closure to researchers. So, you know, if you produce data in the context of project with Shell, or with a pharmaceutical company, or Monsanto, very often there are requirements from these companies that you don't disclose your data publicly for at least a few months after you have actually produced them. And this is because that is for them a way to preserve competitive advantage. So they give you the funding and you know as a, kind of, in return they get the opportunity to use the data before anybody else. Now, these kinds of policies are in direct conflict with the policies of the European Commission now, which recommends that data are released as soon as possible after they're generated in the lab. So, if somebody, in their own research group, is funded both by a body like the European Research Council and by a private company, and there's many many people who are exactly in that position, what are they to do? I mean, there's a lot of confusion around this and rightly so. I mean, it is really not clear what to do in these cases yet.

[Sabina]: In your surveys, did you observe differences in how these policies are currently being perceived, received by different scientific communities across the world? For example, Europe versus North America: are they unanimous or are there maybe cultural differences, also, at play?

[Sabina]: It is of course difficult to generalize, because these are big nations and there are

big disciplinary differences, too. But I would say that a lot of the response to the idea of Open Science depends on how much access to infrastructures, and how much expertise in these matters, and how much exposure researchers have.

[Federica]: What do you mean by infrastructure? Just the platform for sharing?

[Sabina]: So, in the case of, you know, different kinds of research in Europe, for instance, in different European nations, even within that, there are huge disparities in the kind of environment that researchers are in when they're conducting the research. I mean, there are institutions which are more peripheral, don't have that much money to invest in a very fast broadband, maybe they don't have the latest version of software for their computers, you know, their apparatus maybe is a little bit more old fashioned because they don't have the money to renew it every year. There are other institutions, particularly Centers of Excellence, like in Oxford or in Cambridge or Sourbonne, where researchers tend to have much better access to the latest technology, the latest software, etc. Now, the problem sometimes, and that's something discussed a lot in my research, is that the institutions that end up leading the way in Open Science, and particularly in certain of the infrastructures for Open Science like databases and things like that, are actually the most prominent and best funded institutions. Partly, because of this, probably, incentives, that they are the ones who have some surplus, some spare time and some spare capability, that they can devote towards Open Science activities. Now, the problem, however, is that [this] means, that a lot of Open Science infrastructures like databases, are set up with the needs of this very well served researchers in mind. They're not necessarily well adapted to situations where people don't have maybe access to a very fast broadband, don't have access to the latest version of software, don't have access to the latest experimental tools. And this actually is creating a lot of problems and disparities in the uptake of Open Science, because all of a sudden we get a situation where actually the richer you are and the better equipped you are, the more you can be open. While people who actually have less opportunities in that sense, feel that they cannot be quite as open, because they don't have the resources, they don't have the technology and they do not have the time. So, it's a difficult situation because it's creating a vicious circle where, you know, something that Open Science is supposed to be, is that it's supposed to help you to level out the disparities in equipment and opportunities across researchers around the world, and if that's not what we've seen now, we've seen quite the opposite: a situation where it is actually making some of these disparities even more noticeable.

[Federica]: There is something that I like about your approach very much, and that is that you keep the picture very real. You don't just promote Open Science like a concept and like a good thing - which it is, - but you make it real in that you talk about the problems, the obstacles, that research groups and individual researchers encounter when they want to open up their data, and make the research more transparent, and reach a wider audience. Because it's not about the data, it's about the people, it's about the people who produce the knowledge,

and who use the knowledge, and then benefits from it. And I like this approach very much. You have conducted another study that talks about the importance of turning a research project into a community, or to build a community around a research project. And that has to do also with data sharing. Can you talk a little bit about that?

[Sabina]: Well, I mean, I think one of the interesting questions was for me philosophically that comes out of looking at Open Science, particularly in relation to data and this whole idea about Big Data, is the fact that one of the reasons why it's so difficult to implement Open Science is that the science now is very very highly distributed. So, you have so many different communities, so many different groups, so many different types of work in science, and that means that when you release results in this kind of open way, it's very difficult to predict where they're gonna end up. I mean, there can be lots of different uptakes from different people. And that, on one hand, makes a lot of people, certainly in philosophy, very nervous, because the question becomes, well, I mean, how do you check that people understood it appropriately, the quality of the data is assessed, you know, are we building a house of cards here? I mean, what is actually going on. But at the same time, I think, it's also very interesting this idea of trying to get your head around this unavoidable distributed nature of scientific work. I mean, we are in a moment where science becomes so complex, I mean, there's so many different things to think about when you're doing a certain type of project, there is just impossible for one person to control them all. So, we are anyhow shifted more and more towards recognizing that thinking about science as a social activity, an activity where even your own understanding is distributed across lots of different minds, it can never be all in your head, but it's just distributed across the team, is fundamental. And so, a work, as in philosophy science and other fields, sociology of science, which tries to understand how scientific communities are composed, what brings them together, how the people communicate with each other, and between groups, I think is very important, particularly now, because otherwise we don't really understand what it is that we're calling science at this point, and how can we even think about disseminating all these different things across huge fields where different individuals may just do very very different things with them.

[Federica]: I like the idea of science as a social activity very much, and it doesn't come without ethical implications. Are there ethical implications in opening up your data, and Open Science in general? Is it just about personal details, like gender, religion, and those things, or there are others.

[Sabina]: Well, I mean, that's certainly the type of data that is talked about a lot these days as having lots of ethical connotations. I actually believe the problem is much larger, and I'm particularly concerned by data which have in fact no obvious association to particular individuals, like climate data, for instance, or data about temperature in a particular location. I mean, [data] that affect communities in particular locations, for instance, in studies which are involved

in now, which are studies that are trying to bring together data about the temperature, the climate, the vegetation of a particular area, with medical data about, you know, the symptoms that people feel in the area, the epidemics that they go through, these kinds of things. And I mean, when you bring in all these different data together in this way, one of the things that people are trying to produce are predictive models that tell you, ok, so, for instance, people in this little village, you know, in Belgium, they have a likelihood of having a heart attack which is lower than the national average, or their likelihood of being exposed to asthma is higher than a national average. Once you start to make these predictions, what you're really doing is, like, potentially affecting very severely the livelihoods of people who live in the particular village. Because, for instance, the government can come and say okay, so, if people here are more likely to have a heart attack, or less likely to have a heart attack and more likely to have asthma, we're going to go to the local hospital, and we're gonna have much less people there who can take care of cardiology, and many more who can take care of respiratory diseases, right? So, I think there are implications around how one deals with data, which data is being linked with each other, and what kind of conclusions we draw after the analysis of this data, which go well beyond just worrying about privacy. They're really about what kind of knowledge do we think is worth producing, but also what kind of knowledge do we think is best suited to human flourishing, you know, to the well-being of people. And that means that ethical considerations around, you know, **social implications of knowledge production**, really permeate every aspect of knowledge production, in my view. I mean, they're not something that should be assessed by an expert in ethics at the beginning of a project and then never talked about again. It's something that researchers probably really need to think about as they're developing their projects, particularly when they're dealing with these Big Data conglomerates, which can, you know, give you lots of materials to reach all sorts of conclusions about society.

[Federica]: A discussion around Open Science is definitely interesting, worth having, and I'm sure we will keep discussing it. To try to bring this episode of Technoculture home, I would like to ask you: having said all we have said, do you have an advice for young researchers who actually want to make a difference, want to open up their their data, they really want to do something about it but they're not sure how to act, because they have different sorts of constraints, work in different situations and research groups. What is it that all of us can easily do? Is there something, an action we can take, to start contributing to Open Science?

[Sabina]: Yeah. So, thanks. There's lots of things that one can do, I think, that are not too time-consuming, but are definitely going in the right direction. So, one sort of in-between measures that many people take these days is to actually publish their data, separately from publishing papers out of the data.

[Federica]: Like in an annex?

[Sabina]: Actually in things like journals. So, there are data journals that only publish data sets. But also there are repositories, where you can go and you get a DOI, so a Digital Object Identifier for your dataset. And I think publishing in those fora actually adds a lot of visibility to your work, particularly when you're a young researcher. So, a lot of people are very afraid that if they make their data public at that stage of their work, they may be scooped and other people may come in and steal their ideas, but actually there's very little evidence that that's the case. There's much more evidence that people who do open up at least some parts of their research components early on, actually acquire more visibility in the field. Their peers become more aware of their work, the work gets reused more. So, I think, you know, people have to be very aware of that, of the fact that behaving in Open Science way, can be extremely beneficial to your career later on, even at a moment when people don't immediately recognize the importance of these behaviors. I mean, it's also been actually increasingly recognized anyhow. But even without that, there are [positive effects] on your career, which are really good. So, you know, a place, for instance, a very good place now to go and put your data is now Zenodo, which is this big kind of data repository, which is run by the CERN. And there you can deposit your data for free, you get a DOI, so people can share your data, but at the same time they're yours, they need to acknowledge you when they're using your data set. If they decide to reuse it, you may get very useful feedback on your research, when you put out your results like that. And all of this, may help you both to publish eventually, you know, actual papers on your work, but also to make your name and to add to your reputation in the field. So, there are all sorts of behaviors like this. And of course, Open Access publications, I think, by now, is a no-brainer. There start to be in every discipline options for very good journals which are well recognized for publishing good work which are Open Access. And I would really recommend to any young researcher: do everything you can to publish in an Open Access journal, for the simple reason that these papers are much more widely read, and more widely cited than papers which are published in proprietary journals. And that really has been demonstrated by lots of empirical studies. So, there are serious immediate advantages to behaving like this.

[Federica]: Thank you so much, Sabina, for your time, for sharing your perspective and expertise on Open Science, and also for the advice you just gave on how we can take action. I hope you've had a good time in Ghent, and that you have a nice trip back home to the UK.

Thank you for listening to Technoculture! Check out more episodes at technoculture-podcast.com or visit our Facebook page at [technoculturepodcast](https://www.facebook.com/technoculturepodcast), and our Twitter account, hashtag [technoculturepodcast](https://twitter.com/technoculturepodcast).

Index

Big Data, 11, 12

community, 2–4, 11

data, 1–13

knowledge, 1, 2, 4, 5, 10–13

Open Access, 6–9, 13

Open Science, 1–13

philosophy, 1, 11

research, 1–13

science, 1–4, 7, 11

society, 1, 12

technology, 2, 4, 10