The Open Science Interviews by Kaja Scheliga

Interview with Jon Crowcroft, Brussels, April 2013

KS: Could you please introduce yourself?

JC: I am Jon Crowcroft, I am a professor of communication systems at the University of Cambridge in the faculty of computer science.

KS: Great thank you. What is your understanding of open science?

JC: That is a good question, I am not sure I have come across the phrase before you mentioned it yesterday. So what we do in computer science involves the thing called open source which is, if you are a computer scientist/ so not everyone does open source obviously but, almost all our projects are, when we look for funding from UK government agencies and European government agencies and US government agencies which we get money from all of those, we usually say that the output of our work, if that is a piece of software, that that will be open source. And then the other open, the sort of recent trendy thing, is open access to publications, and obviously publication could be data as well. We are pretty strongly in favour of that, in a lot of the computer science community that is not a big deal for us because many of the publishers for the work we do have a reasonable policy of open access. For example the ACM allow you to keep a copy of any submitted paper, any published paper in their conference or journals on your own site or an institutional repository and so that is open access. That is one way, there are other ways, but that is easy for us, of course other scientists they may have other barriers to that. So, open access and open access to publications, open source for another kind of publication which is software are things we do. I am not quite sure what open science means though.

KS: So do you see that as rather two separate things that kind of go along each other or do you see one being part of the other?

JC: Oh, well, okay. There are links between these. There is a recent book by a guy called Ben Goldacre, who is a medical teacher in the UK, a practicing medical guy who teaches.

KS: Bad Pharma?

JC: Yes. But he has recently been part of a group that has succeeded in getting at least one pharmaceutical company, to publish results of all drug trials including negative results, so that you have a kind of level playing field. In what we do in computing, particularly in systems work, where we often do performance measurements, we have some conferences that insist that you publish the data in the same way. So that when you submit a paper to the Internet Measurement Conference they require you to make the data accessible. That may put a lot of pressure on you to have to anonymise data if it involves any personal identifying information. And that is fine, there are techniques that are reasonable for that, you might control access to

the data if you cannot fully anonymise it. But the reviewer is going to look at it and other people on request that would probably be acceptable. But the point that goes with open source is that a programme is a method for doing something, an algorithm for doing something, and when you publish data and you publish a paper, the paper as analysis of the data, so actually you should put all three together. You should have the publication that says here is this result and the data which is what the result is about and the way you got from that data to your result. You prove your hypothesis, or you do not disprove your hypothesis, using some algorithm or whatever method, or statistical approach, whatever you want to call it. So actually these fit together. And the fact that I am talking about computer science, it does not have to be that. I mentioned the pharmaceutical [sector] because there is not much difference. In a pharmaceutical drug trial, you do some hopefully properly randomised trials with some control groups and the usual blinding and everything and then you get your anonymised data back and you do your statistical analysis, which is just algorithms running over that data, and then you publish your result. The result could be, X percent of people appear to do better on this but not with much confidence why, here is the t-test we used or whatever we used to do that, then everyone can validate if what we have done is okay, or even better, add it in. They might invalidate it, but they can add it in to further trials, so actually everyone just gets these large numbers of small trials but they aggregate together into meta-studies.

KS: Okay. So do you think openness, how it is happening now, is something new?

CJ: I think it must be because / if you go back to Newton and Leibniz and people, there was not any technical mechanism to share anything apart from writing short papers with results and mathematics, and your reasoning I suppose could be in the paper. But the work that Newton did looking at lots of glass for doing optics or doing mechanical experiments to figure out laws of motion, then inferring things about gravity, the experimental work was, well you can go and find his notebooks if you want to go and look at them, I suppose they are actually open access but that was not the intention and it is not a very good scale, whereas now there is almost no cost to copying data and sharing data so it is trivial to do. The only sensible objection for, at least in the long term, is if there are privacy implications. In the short term, for commercial organisations, there may be commercial imperatives about protecting something until they got their intellectual protection for the thing they base on the data. But that does not take that long, so you might get ten years patterns on a drug, but it does not take ten years to get the patent. Once you got your drug trial results and they say yes this is really good and some other people get / reproduce those results, you can then publish the results. There is no reason not to, in fact you should be definitely required to because that is the basis on which we are selling this drug to a health practitioner. This is the same when you do performance measurements, of a computing system or network that / it seems reasonable to do that. There are other places in systems engineering, it does not have to be computing systems, where you are looking at. A performance result might be safety and reliability and so obviously a place you might look at for that is aerospace, were every single time there is an accident in a plane there is a very public analysis of all the data. All the data is shared by every one of these agencies, so they do really detailed breakdown and it is absolutely / would be against anyone's interest to start keeping that secret. They have an awesome safety record.

If you look at it, it works. They have got an evidentially approach to agree to share the material and investigation outcomes have been a very positive impact on the airline industry. Of course it has costs, one thing goes wrong with the Dreamliner recently and they have to pull all the planes until they figure out what is wrong with this battery. But that is what they do, and that of course, the cost the other way, the error in the other direction is worst, Boeing would probably be out of business, if they had three crashes for one type of plane the costs would be so high. Actually this is a nice example that even for somebody with a high risk and a high cost of doing this openly it makes sense. You can imagine the same thing for computing. An area of contention is security, where we often, and some of my colleagues, find problems with an allegedly secure system, like chip and pin for banking recently. A couple of years back some of my colleagues found a weakness in the implementation of a chip and pin protocol which meant there were attacks on the smartcard payment system. The worst thing about these attacks was that you could not actually assign a liability. And they warned the banks about this and the bank said: no, no, do not disclose this, because that would be terrible, and they were like, no, we must disclose this because people should be aware that. Not only are their transactions not safe but they actually will not even know who should be liable and this has to be fixed. It does not matter which direction it is fixed. At least you know where the liability is so you can make choices based on information. That is a nice example of where there is an argument in security. You do not publish weaknesses because you might expose more people to the attack. On the other hand, the bad guys probably already know the attack and you are just obfuscating; you are on their side if you do not reveal this to people. I think that it is kind of similar to the safety argument in the airline industry.

KS: Yes, really good example, thank you. What has the internet changed about how we work in science?

JC: Yes, clearly the collaborations are faster and wider. I have written papers with people I have never met. I mean I have been doing this for a long time and somebody gets in touch with an idea or they read something I have written in a blog or something and say hey, there is this thing we could do, and we end up writing papers. Or we go and present it, or it is in a journal submission or whatever and that is one aspect. The scale of the net means you can reach most people in science and technology work anywhere on the planet. But then the other thing is, as music pirates discovered, it costs virtually nothing to copy stuff. So there is no excuse, unless, there are a few areas were that is not quite true. There was a nice TV programme in England a couple of weeks ago, a science documentary series called Horizon. It had a programme about big data and had a guy talking about the array of radio-telescopes they are building in South Africa and the data from that will be something like three orders magnitudes more than the human genome database, this is not something you are going to copy around very often. But that does not mean that it will not be open. Actually, the nice thing with astronomy is, they are the most open people, I mean they are obsessed with them. There is this huge number of semi-professional amateur astronomers, who actually have optical equipment to look at the sky or they can go look at the radio-telescope data. People come up with names and share it. But they do not have to copy the data anywhere, they just provide access to the data over the net, so anyone who has got access, or can go to a library

and get access can go look at it. That is a very nice example. Of course, at the moment, nobody thought of how to monetise astronomy data. They are not at risk from that someday somebody is going to figure out, which asteroids we can mine, and then it will start. Probably, there will be a pressure on them to be more closed about that, but for now they have this nice world where it is all open. Most stuff the Ph.D. students do or junior researchers do in small teams, they are unlikely to generate gigantic amounts of data, so that it is almost always easy to just put the data in a public repository or allow anyone to download it, which clearly makes life a lot simpler.

KS: You said in the keynote that pretty much most of the work that you do or a lot of the work that you do you make available for other people, but still the question again, how open is your work, how much do you share with other people?

JC: As far as I know, every paper I have ever written is publicly available, if it is not it is just an accident, like I forgot or something. The last fifteen years we were working on mobile. The datasets that we gather we put on a repository called Crawdad, which is run by Dartmouth College in New Hampshire in America. I'm trying to think what other datasets we have got. There are a few social graph datasets that I do not think we have published the data but we have made it available to anyone who asked, and the reason for that is, we are not confident about the anonymisation of graph data. We are worse than not confident, we are confident that you cannot anonymise graph data. We do not want to put that up until we can think of some better solution. There is one paper, two years ago by someone from Santa Barbara, he has got a technique which you can capture the properties of a graph by generating a graph that has similar properties, and even that could potentially let you infer things about the original graph and the nodes on the original graph. So we are a bit worried about that. We have got some data that is not available without asking us, but I think, anyone who asked us that came from an institution where we understood what they were doing we would be fine with that.
I have talked to people about not making data available where, this is kind of unfortunate, some folks I know in one group, a big research lab, had some very nice data which they made available publicly. A few people used it and did not cite where they got the data from, which is bad practice by those people, I mean it is just rude socially; it is ridiculous. If you are using a dataset that has been gathered by some smart people it is to your benefit to cite you have got it from there, because that makes other people confident you are getting goods data, probably, or it is from people who know their methodology. I am sort of surprised by that, but I mean it is definitely the case that there are people who do not go through the correct protocol. This is you use some work by somebody else, whether you are just using the result from their paper, you cite their paper, using their data, clearly, you would not just cite it, you probably put a big acknowledgement, saying thank you to blah blah blah lab's data providing. Whatever, you acknowledge your funding source. But this is even better, it is free. Why would you not acknowledge that.

KS: Yes, exactly. And where do you see the boundaries of openness?

JC: Well, it is interesting. Right now, the biggest barriers to open access to publications have come from high profit margin academic publishers, who are pressure groups against open publication, and so they are a barrier that should just go away, because we know the true cost of publication and they can be born lots of different ways, and that is a discussion, but we do not have to pay these guys when we write the papers, we review the papers, we edit the journals, and then why do we have these charges that are an order of magnitude more than we know what the true costs are. And we have various different organisations we can compare, we can compare the market and see ACM costs are I think two-hundred dollars per article for life or whatever, not nine thousand something like that, or whatever number they claim. So that is a barrier and that is a problem because, they have got money, they are a pressure group. There are also groups that have proprietary systems that they do not want to reveal, so clearly if you have got a software system, a big company, Apple or Microsoft or whatever, you are not necessarily going to publish that software, not at least until it has been around a long time and you have made most of your profit, the problem is that because software is so cheap you will just lose a lot of revenue. Even though you might go out and catch all the people who copied your software, you probably will not, they might be in different jurisdictions, so that is a barrier to open publication of software. It does not mean that you cannot publish an algorithm. So these very same companies have published papers where they are very open about the algorithm. I have a paper with Microsoft people and actually we have a patent on a technique for security where there is a seventy page paper which includes all the necessary bits of the code for anyone to go have a look at how this works. It does not include the actual code that runs on Microsoft systems, and that has been rewritten a few times anyway, but it is completely sufficient there, so there is not a completely open system but there is a reason for it, at the moment, and that is partly to do with fear of loss of revenue. It may not be correct, I would say that by analogy with online music, we do not get piracy when there is affordable online music, so people pay to subscribe to Spotify, they do not have advertisements, they do not have to pay very much and that is sufficient to stop them doing illegal downloading. That is really interesting and that kind of suggests that it might work for many things. You just have a reasonable margin, you can shift the way things work. Then this is open, I can listen to any music, there is a kind of weird thing, there is no DRM on it either, so the barriers are basically, the problem has been over the period were the internet emerged and technology grew, computing technology grew, and in other worlds, like pharmaceutical, it is probably the same story. I am not an expert, the people with the biggest margin have the most money and that means they often have very powerful lobbies to argue against open access and that. I am not an expert in economics, it is a political economic argument, but it is clearly the case that, you have these organisations who really should not have a place at the table in a democracy. They are a commercial academic publisher, or commercial software house or a commercial pharmaceutical company, and the place where people are discussing open access is in academic government tax payer funded research labs, they are not paying tax in the country they are getting money in, so why are they at the table in that discussion. They are a big barrier. It is difficult to see how you completely overcome that, there are active people in this world. I am a very active, but active people in this world, overcome it by enlisting as many people's support as they can. Making it visible, perhaps to the taxpayer, because in the end the taxpayer is going to end up paying for this. If you say well actually there is this world where

we could have more innovation, and these people have lower margins, still be wealthy, and you will get more software cheaper or better drugs cheaper, why is that not bad. There is a / you can make that argument and some people will put in the time to support that.

KS: So for openness to work in science, or openness to work in in academia, what do we need? Are there any requirements that come to mind?

CJ: I am not quite sure I get –

KS: Say, for open science, or openness in research to work is there anything that, like any points that you say, we have to have this or we have to have that…

CJ: Ah, it is not free, it is not free, so you have to teach people about a particular cultural way of doing things. A lot of people, a lot of cultural norms for research are siloed and people work in this world and they do not look outside at all. There is an educational side to it. That is largely about showing the benefit, if your work is more visible and you are any good at that then you are going to get famous and you get a career. There is a selfish reason to be giving your work to other people.
There is a cost to institutions of open access, it is not free, like I say, Elsevier has this price point. I do not want to keep bashing them, other people have price points, Springer, whoever and the ACM, IEEE and other have other price points. Some organisations have a price point but still make everything openly available, in my world Usenix, which is a trade organisation mostly in the US that runs a number of conferences and journals, and all of their stuff is publicly available, for free. But they charge for it as well. So how does that work? Well, it seems to work. That is kind of interesting. But there is somebody paying something somewhere. Institutions have to figure that out. Then of course there is an issue, which is that a lot of people's work is quite incremental and average, and you publish incremental average work openly and a lot of people look at it and go, well, a few people look at it, and go: that is incremental average, I mean that should not be disappointing to people. They do some work and it is perfectly fine, it is just not cited eight-hundred times. They will still benefit by confirming or refusing results, that are out there, and other people would benefit by that. But it is kind of difficult for people in this sort of world, there is a certain amount, certainly in research /it is not like rock music, but it is a bit like some other areas where there are people that are kind of stars and the other people might feel that. They are trying to compete in that. That might be a barrier to people, psychologically perhaps, on the other hand hiding you work is not going to make any difference, it will make even less difference, maybe just making it a cultural norm that you do open work would just mean then, people would not be so concerned about that. And in fact, you probably discover there are an awful lot of really good people out there doing great work. It is just, the stars are stars because they have just got personalities. It is not actually that their work is any better. That is, you can separate to those things out, which is probably important.

KS: Although, personality does impacts the quality of your work, probably. I would have thought.

JC: Yes… The thing is, there is an unfair feedback loop, which is the rich get richer. It is a completely standard social phenomena, and if your work is visible and you get on with people, you get more people to work with and you get, you know, the Paul Erdos factor. In social graph theory, Erdos was this guy who wrote papers with everyone he ever met, and he, is the most central mathematical writer ever, because every time he met somebody he would just have an idea every day and they would work on it and finish the paper and put his name on it. He has gone this amazing centrality in the graph of authorship. The point is, the more that happens the more that happens more, because somebody would go: oh you wrote a paper with Paul Erdos, I have got this idea, how do I get in touch?

KS: Okay. To swing around to the more technical side, which online tools do use for your work?

JC: Online tools? For my research. For actually doing my research? I am not quite sure about this question, where it is going. We build systems which we then measure the performance of. There are two things we need to do, one is, if we are working with other people, typically the systems are quite big, so we have got project at the moment involving more than ten people over three plus years. We had to share sourcecodes, we do distributed development using sourcecode repositories, and these are open source, we use GitHub or, SVN type systems, that is an online tool for sharing software development across groups. It is sort of what any large organisation would use. Something like that even if it is internal proprietary development. So we do that a lot, and we used the same thing for writing papers together actually, these are the same typical repositories. It is probably a bit clunky for non- / so for computing people it is second nature because you use repositories for software development, and you say I'll I put I will latex for our paper in there as well. For somebody used to doing version control in Word, you just email everyone the latest version and it got the tracking changes stuff on it, so that it is fine for people. I increasingly notice in European projects people use things like Google Docs for sharing documents, and they do development in the cloud, as long as it does not proprietary. In fact, we had an interesting discussion last week, where we are doing projects where we have serious privacy concerns about some of our data. We are sharing all the methodology and all the documentation in a place we do not mind people seeing. But Google Docs is not a secure enough repository for the data that identifies people. We have to build a separate repository for that, which is encrypted and has access control. Then we just use all the stuff that is there. I am not a particular instant messaging person but some of my colleagues use Twitter and other things like that. Just for keeping track of where everyone is, but I just use email because it works and it is fine. It is a tool that is built, we keep email lists, project webpages, wikis. We have wikis for most of our projects. I actually use blogs as well, it is a little bit complicated, we have a research group blog, so every time somebody goes on a trip, to a conference, or meeting, or whatever, they are supposed to produce a trip report, so tell everyone back home what they did with the taxpayer's money. There is a public log of what they did. All the stuff that is around, we just use anything and everything there is really. We also use open source apps, so not that there online particularly, they are statistics packages that are open source and free and are very good, like R. There is a whole bunch of them. Clearly, you just get the latest version to those for some statistical problems, fire up

something, correlate things, whatever. There are quite a few things like that, there are cluster analysis tools, loads and loads of tools out there.
Because we do systems work, we need to use, we build systems, and often they have other people using them, we build them so we build cloud software and we just run that in clouds, on data centres, and have people using it, then we measure what happens. That is quite nice, because there is no human user element to that, it is just big programs running on this cloud servers. We are measuring performance from those programs running. We do not have too much of an ethics, or an anonymisation problem with that. But then we are using sort of the fact, that there are affordable, big, data centre resources, like Amazon's cloud services and others, which you can get what used to have to get all your institution have to go and spend millions to buy big high-performance cluster computing systems. You can now get a relatively small part of a research project grant and rent for whenever you need it. So I guess that is a kind of online –
I know lots of other scientists who use those kind of things now. Rather than spending thirty million for a big supercomputer, you just say we will spend three thousand for a few hundred hours for Amazon EC2 time. So that is kind of a change I guess.

KS: So it is like not only data sharing but also tool sharing.

JC: Yes. Yes, absolutely.

KS: Interesting. Another point that I am interested in, is collaborative writing. You have written papers with other people, I am wondering, what is the process of writing with someone else?

JC: It completely changes. There are a lot of completely different ways of doing that. People I know what their skillset is and I worked with them, we typically would break a document into pieces. You have an online meeting or a real meeting and you write the skeleton of the document, saying: okay who is going to do which bit, and then there will be bits where the results drop in to various places of the story and then you have to have another meeting to re-factor the document if the results are different from what you expect or whatever.
With people I am less familiar with, or their different skills, the most common thing is / just this morning, I was sitting at the back here, with a report from a seminar, twenty page report, and basically what we are doing, we are four of us, four authors in four different countries, two different continents, we are doing serial editing. We had a bunch of notes, one person in Germany edited them into a pretty good first pass, and then another person in Sweden has just done a big tidy up, then I was just adding a few bits I thought would be missed and then a bit more of a tidy up and then I fired it back to everyone said: okay, whoever has got the token, not me, anyone else can have it. I have done that with a guy, again, actually this morning, last night this guy sent me a journal version of a conference paper, he is at Delft University, a really really clever paper. This is a longer version with a lot of additional stuff in and what I have done previously on that was, he had missed some related work plus his English was terrible. Not that surprising given that he is in the Netherlands, he must speak a bit of Dutch, but he is not from the Netherlands either. He is from Bulgaria? I am not sure. Anyway, you

see, it is probably in his third language, so it is pretty impressive, he did a really good job, so I completely rewrote all the English. That was not any technical contribution, but I did a bit of related work, and I also went through the maths to check it was okay, and maths formatting, you have to get it right otherwise people who read it get really upset. It was not that big a deal, but there are all kinds of different ways you can drop in, whether it is serial work or it is parallel things you can do and then you have to reallocate.
Then there is a couple of tricks, like I mentioned, whether you can actually distribute access to the document and people will actually edit it and you just use some kind of version tracking repository or whether you have to actually synchronise everyone depending on what tools you are using, what word processor or other things you are using, because, you can get inconsistencies if you are not careful. Although it does not happen very much, this was fear back in the day, twenty-five years ago when people were talking about collaborative authoring they thought that consistency was a big problem. But it turns out, people that are going to deliver a successful document probably have a pretty clear idea of who has got what role or roughly who is doing what at any given time. Even if you do get inconsistent, unless you have got a completely broken way of thinking about documents, you can have the previous versions and you resolve the inconsistency, it may in fact be important that there is an inconsistency, here we go this way and that actually turns out to be: oh we better discuss that because that was something neither of us had a clear idea of what was supposed to happen. It is not like there is not a human facing bit of document, it makes it obvious.

KS: The process you described with the four authors, where each of you get the text after the other, do you do one round of this or do you do a second-round where everyone reads like –

JC: Yes, I have done it. In this case, I think, we will do another round. But I think it will be, because we want to see what the last person does everyone will go look at it, hopefully, assuming we have a consistent idea of what we will do. There will not be any more changes after that, but we will see, we have a deadline for that so I mean, deadlines often, kind of make you do this deadline driven. Most paper work that we actually have a standard pattern for systems work. This is, we / I have got a first year Ph.D. student who has a really nice idea for some work, he is doing the work already, there is a deadline around September were there is a really good conference which would be a good fit. So we work back from that, he has got to do system work, he has got to do some eval[uation] and then somewhere in that period we will take his notes and turn them into a first draft of a paper. We can work back from that deadline and figure out what is going to happen, there are three, at least the authors and that for sure (…), But anyway that deadline thing is quite important because that kind of –

KS: …frames the structure.

JC: Yes. (…)

KS: Anything that you want to add to this that I did not ask you about, any thoughts that came to mind? Anything that is missing.

JC: There is one thing that I did not mention, which is, there is an incredibly different culture about publication and conferences in different areas. Obviously conferences are a space where you see very public / it is public talks, you should see the people who have done the work usually and you see the proceedings of the conference. But there are absolutely different processes. For example, most medical conferences, anything submitted is accepted. Often they just have everything as a poster and there are a few presentations and keynotes and stuff, whereas in computer science and electrical engineering, and to some extent in physics, there is this hugely selective process. Really hypercritical about submissions, and a fraction of papers, and in some cases incredibly small fraction of submissions are accepted, could be as low as ten percent, which means that some work that is definitely of value is not accepted. There is some weird social process going on there and that is not particularly open, right? The open thing to do would be to accept everything but pick what everyone who attends thinks is the best stuff to be presented and then have everything else be on posters so you can go around and look at it. Like I said, they do that in some disciplines but we certainly do not do it in computing much. That seems a bit weird. I think there is a big discussion going on in the community of computing at the moment about: we seem to be doing something a bit peculiar and we could actually make better use of being open and online, by just having everything submitted online just appear, uncritically and then gradually have people add reviews to that material and then have people revise it and show the revisions. At some point people might say, oh, we would quite like to hear you talk about that and way don't you show up at this conference or that. You invite people, or they ask to go and then people say: okay would like to hear about these since they are a really hot topic. But those ones, only 10% of the people on the conference want to hear about that then a poster is fine. Or if 25% of the people want to hear all about this then we will have a parallel session for that. You can imagine doing the system much more reactive, then it would be more up-to-date and more open. In fact reviews in this process are almost always blinded. They are not signed, which means it is not an open process because you have no idea if the reviewer is conflicted, could be indirectly conflicted because they have another submission themselves and they are trying to be negative about other people's papers. Not that it happens very often, but you can imagine somebody.

KS: But can you imagine the process that you just described to actually work in practice?

JC: I do not know. I mean somebody should try it.

KS: Yes, I was about to say, it is a really good thing to try. I guess parts of it were already tried out, like the open peer review, Nature did a trial on that.

JC: Well, in the physics community and the maths community have ArXive, and so people submit papers there as preprint and people are aware of them. When they have got a really cool version they send it to Nature. It gets a really serious, thorough review, the very cool thing about Nature is their turnaround is extremely fast even though it is a print journal. In six weeks you have got all the stuff back saying yes or no and then you are out and that is great because that is an amazing process, and if you send it to a physics conference they all say: oh yes it is in ArXive, we have all read that, that is nice, yes please present it. They have got a

pretty good approximation which appears to actually not be broken. So it seems like maybe we should go somewhere towards that. I don't know.
The reason some people object to this is there are tenure tracks procedures in American universities which this causes problems for, maybe. That is one objection, there are maybe others. But, there are reputational issues, which is interesting. It is not completely obvious you can have a totally open system for that, and open process. The results are reasonably transparent. Anyone can say they have submitted something and it got rejected, anyone who gets in you see what they submitted so, that is clearly not a bad bit of the process there. But there could be a long that process, could be improvements, maybe.

KS: Okay, great.

JC: But, I cannot think of anything else.

KS: Okay, well, I guess we leave it at this.

***

* This interview was conducted as part of my research at the Humboldt Institute for Internet and Society. The interview was recorded and transcribed. The transcription was carried out with the best intentions of accuracy but can nevertheless contain unintended mistakes. This version of the interview has been slightly edited for better readability without any substantial changes to the content.

Suggested citation:
Crowcroft, Jon. Interview by Kaja Scheliga. *The Open Science Interviews*. Brussels, April 2013. DOI: 10.5281/zenodo.12343