Henri Rzepa: If an "article" processing charge from traditional publisher is up to $5000, what will the business model be to publish data followed by an excellent supplementary article? The DPC (data processing charge is surely now the key)?

APC of 5K is crazy. What is the cost of publishing a paper? But the peer review does not cost 5K. The publishing model cannot be as profitable any more beyond the paywall era. Encourage people not to publish in journals that charge 5K. The cost of publishing, and looking after, data can be high - because of volume, stewardship, long term. Now important to write a data stewardship plan and budget for the publication of the data. Funders have to be prepared to pay for reasonable data stewardship and data publishing fees.

Jofephine Ofeimun: how do you prepare your data for reuse ater publishing it. Is there a model or pathway for that

There is a big difference between storing data in FAIR format and putting the metadata in a FAIR data point. This is not the same as making it available for high performance reuse. To provide this, more and more companies or research institutes set up a data model or FAIR data point to allow visiting by virtual machines. Important to publish data in a format so they can be used in a high performance analytics environment.

Felibp PerezJvostov: How do you engage researchers in the "long tail of science" to provide FAIR data and write nanopublications. In these disciplines data gathering is an enormous effort already, so requesting them to additionally write articles in machine readable format would reduce their time to do actual science. How do we engage them?

Disagrees that this reduces time to do science. In fields with large volume, making data FAIR increases the efficiency of research. 80% of effort goes into post facto data munging. Upstream data stewardship is more efficient.

Henra Muller: Health care research often have strict ethical permissions associated. How does this affect using FAIR data?

Very important question. Important not to confuse FAIR with Open. FAIR data are not necessarily open. The default for public funding is Open, but there are very good reasons for protecting some data. Data should stay where they are: the algorithm comes, gets authorisation and queries the data under certain conditions. This enables ethical protection of the data, without querying personal information. The data are not open. They are accessible under well defined conditions.

Josephine Ofeimun: supposing I was applying for a grant, how do I incooporate data reusability into my grant methodology

All funders in the world should require data stewardship plans and should fund this. For the applicant and the researcher this starts with the data stewardship plan. Should also say will you use other people data, do I need to generate data? GO FAIR has a data stewardship wizard. Institutions increasingly need to have data stewardship competence centres to support researchers in this. Push your institution to start a data stewardship competence programme.

Henry Rzepa: If the costs of data stewardship accrues to the funder of the grant, what happens to research which is not funded (speculative research?) There are no funds for the data
In the end it is the responsibility of the data generator to perform proper data stewardship. The good news is now that most funders will consider these costs allowable costs, but for example, in an internally funded project (highly innovative research) in our own institute (making FAIR digital twins of tumours) we budgeted 10% for datastewardzhip from internal funds.

Daureen Nesdill: In your model, is the data peer reviewed?

Data is generally not peer reviewed before publishing. Open peer review should be the new model. Once the data is available, it can be reviewed, assessed and critiqued by the scientific community and by machine analysis. The best way of maintaining quality is to have the data open.

Dominic Farace: What is your position on Data Papers? Are your Excellent Supplementary Articles a type of data paper?

The concept of a ‘data paper’ may be a reasonable transitional step, but the term itself really conveys paper-centric thinking. Data should be published in their own right as ‘fist class publication citizens’ and papers (although permalinks to the relevant data) are just another component of modern publishing practice. They are meant first and foremost for human consumption, whilst (large and complex) data are first and foremost ‘food for machines’ (simply too complex and luminous to be comprehended by humans at first glance, so machines need to discover patterns first), so I am sceptical in the long term about the concept of a ‘data paper’

Elias Mwkilama: (Elias-Malawi) In your opinion, what are some of the better strategies that our governments (especially developing countries) can handle COVID-19 data to support its citizens to make it more accessible and re-usable for well informed policy and decision making?

Please look at the VODAN initiative, where African countries play a co-leading role, you can find the distributed analytics concepts there and how FDP’s in Africa already start to be part fo a global network, and, again, co-leading it.

Link to slides ni PDF: https://www.dropbox.com/s/i987wm1foh77dr7/seminar%20data%20publsihing.pdf?dl=0