As the chair of rapidly approaching Force16, it is time to reflect, examine the tea leaves and ponder the future.

Further, the recent NEJM editorial made me MAD.

A little about my FORCE11 history. Last fall I was voted onto the FORCE11 executive board, which recently held a retreat for the passing of the torch from the old members to the new ones. In Beyond-the-PDF2, I won the 1K challenge for ideas about scholarly communication and research data education outreach using the library as a vehicle. I co-led the Research Resource Identification working group to aid reproducibility, and the Attribution working group to advance our understanding of contributorship. I have participated in other working groups to improve data resource, and software citation. Last year at Force15 in Oxford, I was the program chair, where my biggest focus was on bringing global perspectives on how to both consume and share scholarly information from a diversity of countries. This year I am hosting Force16 at Oregon Health & Science University in Portland and we have a truly awesome lineup of speakers and activities. This diversity of activities has provided me numerous, regularly conflicting, perspectives on the state of scholarly communications and what we as a community are doing about them.

In the NEJM article it states: “The aerial view of the concept of data sharing is beautiful.” Myself and very many others are doing everything we can all along the research data lifecycle to promote this. This is my Force mission.

I believe that there is a discrepancy in this day and age between the form our scholarly communication takes in the literature, where we tell stories - stories for which the format has not changed much at all since the first Philosophical Transactions of the Royal Society published 351 years ago (which I most amazingly got to personally examine at FORCE2015), and the research activities that are taking place and the products thereof. The stories are simply our conjectures, our mind wanderings, our hypotheses, and conclusions. The actual research is only marginally referenced therein; such content used to exist (and largely still does) in laboratory notebooks, on instruments, and in autobiographical notes. What better time
than to move these ill-defined homes for the real research to the digital ecosystem. But we don’t seem to agree about how, when, or—as the NEJM article highlights—why we should do that.

Is the idea that we should use the literature solely as inspiration for our work? That the goal of publishing in the literature is tenure and promotion, feeling good about our smartness, advertising our work? Clearly the authors of this NEJM article feel that anyone who is doing anything with the actual research content are #researchparasites.

HERE, I WILL TALK ABOUT WHY I THINK THEY ARE WRONG.

For one, we need to identify and build digital mechanisms that allow for researchers to share their research process, not only their conclusions. This includes automagically gathering metadata about experimental data from instruments and online laboratory notebooks, uniquely referencing important components of the research throughout, providing access to the versionable outputs: data, algorithms, and code used in the research, and—as the NEJM authors seem to want to hide—metadata about the eligibility criteria for the subjects and protocols. Is it their fault that they are averse to sharing? Not really. We have very few tools and very little incentive to help us do better in this regard. How does a clinical study proposer know how well their selection criteria stack up against data collected in another study? How do they know a priori how a systematic review will utilize their data down the road, to actually effect changes in clinical care?

One problem is the lack of incentivization structures. It is much easier to get funding for a shiny new machine than a much smaller dollar amount for using data that has already been collected. There’s a sense that this is somehow recycled or second-hand science, that data reuse is not innovative. It harkens the Monty Python scene in the hospital with the Administrator who loves the machine that goes “ping” but ignores the woman in labor. Innovations and discoveries in science come in many forms - most of them are not really first or new. I’d even conjecture that it is the actual corroboration that is where the real innovation lays. Come to the Force16 session on “Data for the people, by the people” to participate in a conversation where both machines that ping and women in labor are part of conversation (no, let us actually hope not!)

Similarly, authoring tools need to be a lot smarter. As researchers and biocurators, we all know what the key structured data needs to be, but there are not good pluggable mechanisms to get these standards into the authoring workflow. Some disciplines, such as biodiversity, are making great strides thanks to the tools and efforts of Pensoft. My group, Monarch Initiative, is implementing a new standard in journals and patient registries to share structured genotype-phenotype information so as to make it accessible for algorithmic use. Some online laboratory notebooks are starting to capturing data in a more standards-friendly way that better supports downstream publishing. Such activities require coordination across a variety of stakeholders - the editors, the scientists such as the comparative morphologists and clinicians, biocurators that currently manually collect the data, the informaticists that consume the data, and the journal’s publishing technical infrastructure experts. All of these people are represented in Force11. We need to work together to help make capturing of information all along the research road more computable and more accessible.

WHY.

What is science for? Why should we do all this? What is wrong with the current state of affairs? Why ARE the NEJM article authors wrong? There are very many stories and rationales for why
(see Lowe, Shaywitz, Huston, among others) some of which include ethical obligations. We should have a hippocratic oath for research, and we should be especially obligated to share our methods and our data if our tax payers are paying for it. However, even if we are not funded publicly, do we not have an ethical obligation to share our advances from the whole research data cycle, and not just our conclusions, so that humanity may profit? I work on rare disease, where my group spends a lot of effort to take a large number of other people’s data and integrate them to help support disease diagnosis. It is a lot of work, mostly unattributed, to develop the data models and tools. I am therefore the quintessential research parasite referenced in the NEJM article, by definition “people who had nothing to do with the design and execution of the study but use another group’s data for their own ends.” Well my ends are not mine. The total number of undiagnosed patients globally is unknown, but it is a lot, as worldwide there are an estimated 350 million people living with a rare disease. This is WHY. We WILL diagnose disease, despite you, Longo and Drazen. And to all you other research parasites out there, I celebrate you and the hard and tedious work it takes to take other’s data into your custodial care.

AND NOW FOR THE WE.

Who is Force11 to me? The answer is all of us. It is not those that creatively started this organization (though I greatly appreciate their forethought and ingenuity). It is not the president, nor the executive board. It is not the working groups. And it is not even the people who have signed up to be members of Force11 (all 1597 of us now), who admittedly don’t yet know what this means, only that they care about changing all of the above. It is the force within all of us - from a biocurator like myself, helping to deliver better data to the computer scientist so that we can build better tools to diagnose disease; to the publishers at PLoS, helping me define a new place to publish AND evolve best practices in data science; to the citizen scientists’ need to understand the goals and outcomes of the research that their tax dollars pay for and so that they too may benefit and participate; to the people in all the countries of the world that need access to information about how to improve lives and the health of great world, and to share back their own learnings and scholarly advances to the greater whole.

At a time when levels of cooperation and trust in so many walks of life are sadly falling, science and scholarly communications have an unparalleled opportunity to move in the other direction. It’s not just about tools or technology, it’s also about cultivating the culture, value systems, and evolving attribution, accreditation, behavior, and ethics that support and encourage sharing. Next time you are reviewing a grant or are on a tenure committee, look past that Nature paper and look at what the person has actually done, not just the story that they have chosen to tell.

Force11 is not the only place where such things are happening, there are many. But it is is one place where people from all walks of life, in any role, can come together and make a difference. May the force be with you all. Keep it alive, care for it and nurture it wherever you go. Join me at Force16.