

#solo13digital: What should the scientific record look like in the digital age?

SpotOn Session Participants

8 November, 2013



writel_AT_EX

<https://www.writelatex.com/489577gzhhfv>

<http://bit.ly/solo13digital-tools>

Digital publishing unarguably has brought about many advantages for the dissemination of research findings. The distribution of research papers — a format which has been used to present original data during the past 350 years — has become faster, more affordable, longer lasting and content can be customised by the reader as publications can be assessed individually.

Yet, these are incremental innovations — a pdf remains a linear narrative, contextualised in a manner that demands human interpretation. This fundamental and largely unchanged workflow often leads to sub-optimal presentation of data within publications or even omission of making data publicly available. In addition, the unique potentials of digital technologies in terms of enhancing, curating and commenting content are barely being tapped.

We will discuss the impact of collaborative and dynamic authoring and publishing technologies and contemplate on what we think of as scholarly output of individual researchers. At the same time we will embark on our own collaborative authoring project and produce live session notes using [writelatex.com](https://www.writelatex.com) — everyone's input is more than welcome!

Discussion

How will it move outside of LaTeX — into other disciplines? Rich text will help.

Why not go straight to XML? LaTeX is established. Some people can use it.

Can we go from LaTeX to XML? LaTeX to Markdown to JSON to XML is apparently in development. And we're working on it!

What about commenting and track changes? We should use the open annotation framework to expose the comments.

We are working on it.

What about Word import? There are some solutions for Word to LaTeX, but they are not perfect.

How will you reach the biology / life sciences disciplines? We are teaching some courses. F1000Research is a life sciences publisher, and we are working with them.

Bundling together of assets: code can be shared between multiple objects. A micro-attribution framework can allocate credit to components of research objects. A more robust 'citation fabric' is required. DOIs can be allocated to individual components (old-fashioned? but well-established).

There is a need for standards for data formats and for annotation. There are good standards for mice. Many more are needed.

What does the scientific record look like in 10 years' time? Will it still have the same notion of 'finality'? We will still probably be trying to get tenure at Harvard.

There are already papers that have a large number of authors. Attribution is already 'micro attribution'.

We need to be able to build narratives — stories. We often lose these narratives. We also need to be able to recombine results. 'Fragmented narratives' are possible, but they don't necessarily make good stories. But what do we lose if we only tell one story?

Research objects go beyond a narrative. They are still 'discrete' objects. Research does branch and merge. 'Aggregation objects' can represent branches and merges. Threading is currently done by PhD students who go back through the literature survey sections manually. Can we automate this?

Are small groups very different from large groups in how they should collaborate?

We do need unique and persistent identifiers for objects. A single proprietary system (e.g. impact factors) creates problems.

Can we learn from the Wikitrust system?

The paper is the 'currency' of science. Can we define a finer-grained currency?

It is complicated to mix ontologies — a challenge in the current

semantic web. Namespaces are a form of credit in the semantic web, but do they really help practitioners?

The positive results bias: now that we don't have to worry about length, will this change? It's hard even to publish things that did work. But there are people encouraging this — e.g. figshare and f1000research (waive fee for negative data).

Tweets

We'll collect your tweets here!

The #solo13digital session looks fascinating -
with @CaroleAnneGoble @Lambo @writelatex
<http://nature.com/spoton/2013/11/spoton-london-2013-what-should-the-scientific-record-look-like-in-the-digital-age/>
Anna Sharman @sharmanedit

How do we represent this body of work? Students are still trained to "salami publish" #solo13digital
Lauren Sandhu @LaurenKSandhu

#solo13digital make sci record method based on principals of software engineering - release notes, forks, merges, dependencies
Mithu Lucraft @mithulucraft

.@CaroleAnneGoble explains her vision of Research Objects and how they can be shared
<http://www.wf4ever-project.org/research-object-model>
Neil P Chue Hong @npch

My thoughts: Releasing data, papers like software means we wouldn't get hung up on "final" versions just useful ones
Neil P Chue Hong @npch

@CaroleAnneGoble great vision for resrch objs and digital future! But oft forgotten - the record of ideas and understanding?
Martin Johnson @martwine

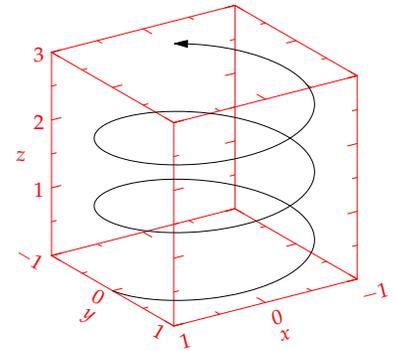


Figure 1: Helix!

It does seem archaic to continue to rely solely on papers as a means of recording science. So many additional options!
MVibhuti Patel @VibhutiJPatel

talking about research objects gets us away from just thinking about formats and thinking about content instead
Mithu Lucraft @mithulucraft

<http://www.writelatex.com> like google docs for science
Isla Kuhn @ilk21

final speaker from <http://www.writelatex.com>
- online collab LaTeX Editor @ruthej we need to add to #research20 programme
Jenny Evans @jennye

.@writelatex is like Google Docs for LaTeX looks like it back ends to @figshare as well which is cool!
Neil P Chue Hong @npch

what is impact of part of research output vs impact of whole of research output/assess/objects?
Isla Kuhn @ilk21

need robust citation track A: DOI - maybe not permanent long term solution, but a good solution now and scientists like it
Isla Kuhn @ilk21

@CaroleAnneGoble say researchers want a doi for research objects. She seems to think this isn't the best kind of identifier
Anna Sharman @sharmanedit

What does @CaroleAnneGoble think is the problem with dois? They are the universal standard for a citable document
Anna Sharman @sharmanedit

just as word not best for docs, excel not best for data!

Martin Johnson @martwine

@martwine but LaTeX is harder to learn than other software, scientists don't have time. Needs to become easier

Anna Sharman @sharmanedit

@martwine ... so it's not just a persuasion job -

LaTeX needs to become more accessible

Anna Sharman @sharmanedit

TCP/IP for science - relevant stuff for #solo13digital from

@CameronNeylon <http://bit.ly/y2xJXE>

Martin Johnson @martwine

narrative vs multiplier of impact - which will have

priority in 10 years time?

Isla Kuhn @ilk21

RT @bazzargh: "@AdamLeadbetter it'd be cool if this was part of it

<http://www.sagemath.org/doc/tutorial/sagetex.html>"

#writelatex #solo13digital

Adam Leadbetter @AdamLeadbetter

Tension between need to tell a story with need to present the data, software etc that can be used in various ways

Anna Sharman @sharmanedit