

# Revolutions in publishing

Ian Mulvany - Head of Technology - eLife

@IanMulvany

# Caveats

Before going into these topics I want to set out my stall and be very clear about my biases:

# Caveats

Before going into these topics I want to set out my stall and be very clear about my biases:

advocate for open access

# Caveats

Before going into these topics I want to set out my stall and be very clear about my biases:

advocate for open access

no formal background in XML

# Caveats

Before going into these topics I want to set out my stall and be very clear about my biases:

advocate for open access

no formal background in XML

a fan of the open web

# Caveats

Before going into these topics I want to set out my stall and be very clear about my biases:

advocate for open access

no formal background in XML

a fan of the open web

some experience of building publishing workflows from scratch

# Caveats

Before going into these topics I want to set out my stall and be very clear about my biases:

advocate for open access

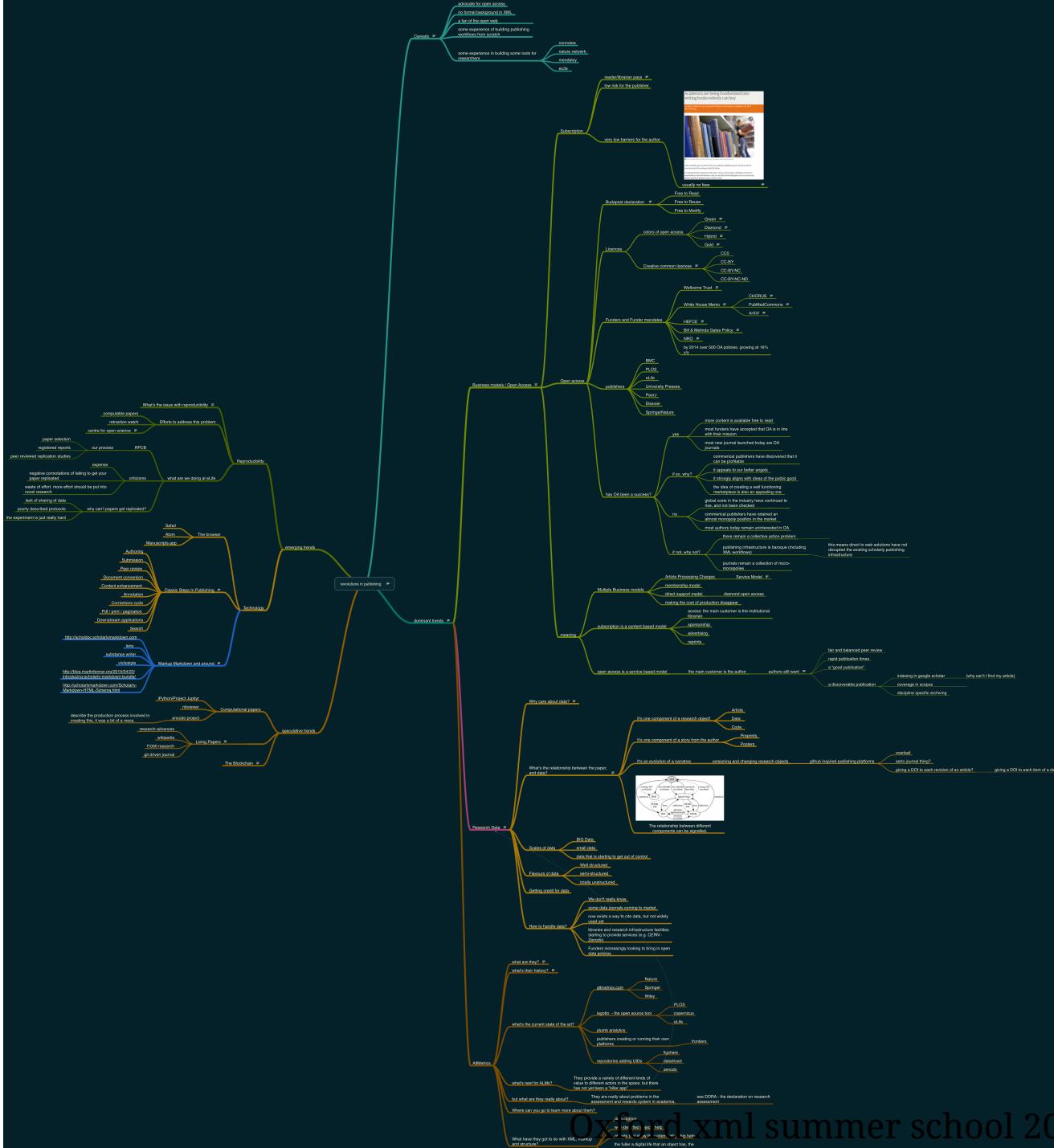
no formal background in XML

a fan of the open web

some experience of building publishing workflows from scratch

some experience in building some tools for researchers

- connotea
- nature network
- mendeley
- eLife



- Open Access
- Research Data
- Object level/Alt Metrics
- Reproducibility
- Webkit and the browser
- HTML5/Markdown/JSON
- Computational papers
- Living papers
- The blockchain

**dominant trends**

**Open Access**

**Research Data**

**AltMetrics**

# Open Access

open access

Oxford xml summer school 2015 - CC-BY 3.0

TODO Insert OA slides here!

# Subscription Benefits

- reader/librarian pays Can be argued that this fairly places the burden of the costs of production on the target market, however IMO the bundling that happens in big deals makes this argument less watertight.
- low risk for the publisher
- very low barriers for the author
- usually no fees

# Open Access History

- Budapest declaration
  - February 2012
  - Meeting funded by the Soros foundation
  - 13 original signatories
- definition of open access
  - Free to Read
  - Free to Reuse
  - Free to Modify

At that time no specific license was mentioned.

(George Soros was a student of Karl Popper, and later a merchant banker)

## colors of open access

- Green authors can deposit their content into a repository, some restrictions may apply around format and rights for re-use.
  - simplest to implement
  - good support for institutional repositories
  - has the backing of CAS
  - does not provide long term financial sustainability
  - under legal threat from Elsevier
- Gold
  - An article processing fee is paid for the article to be published.
  - Scales with submissions
  - Proven to work
  - Can lead to charges of predatory publishing
  - this argument is baseless
  - Applied without waivers and it becomes a barrier
  - The natural price is not yet known

## colors of open access

- Hybrid
  - A journal supports a mixture of subscription based articles and gold open access
  - can transition an existing journal gradually
  - has tended to be a bit of a clusterfuck
  - leads to charges of "double dipping"
  - tracking license through metadata is hard
- Diamond
  - Content goes through the normal publishing process, but
  - fees are supported through public or charitable funds,
  - authors experience no need to arrange payment
  - large question over long term sustainability of this model

# Creative common licenses

- CC0
- CC-BY
- CC-BY-NC
- CC-BY-NC-ND

## Funders and Funder mandates

- (2013) Wellcome Trust
- (2013) White House Memo
  - (1991) ArXiV
  - (2000) PubMedCommons
  - (2013) BioArXiV
  - (2014) CHORUS Launched 2014
- (2014) HEFCE
- (2014) Bill & Melinda Gates Policy
- (2015) NRO

2014 over 500 OA policies, growing at  
16% y/y

# publishers

- BMC
- PLOS
- eLife
- PeerJ
- University Presses

# publishers

- BMC
- PLOS
- eLife
- PeerJ
- University Presses
- Elsevier
- SpringerNature

# has OA been a success?

YES

- more content is available free to read
  - from 2003 to 2013 OA PubMed content grew from 1% to 15%
- most funders have accepted that OA is in line with their mission
- many new journal launched today are OA journals

## if so, why?

- commercial publishers have discovered that it can be profitable
- it appeals to our better angels
- it strongly aligns with ideas of the public good
- the idea of creating a well functioning marketplace is also an appealing one

**NO**

- global costs in the industry have continued to rise, and not been checked
- commercial publishers have retained an almost monopoly position in the market
- most authors today remain uninterested in OA for it's own sake
- it's only captured about 2% of the revenue in the market

## if not, why not?

- there remains a collective action problem
- publishing infrastructure is baroque (including XML workflows)
  -
- direct to web solutions have not disrupted the existing scholarly publishing infrastructure
- journals remain a collection of micro-monopolies

cost of transition to a more efficient system is high

# meaning

OA publishing is a service based model, recouping the costs can be done via:

- APCs the main customer is the author/funder
  - membership model
  - direct support model (diamond)
  - making the cost of production disappear

considered a cost of the research process

OA publishing is a service based model, recouping the costs can be done via:

- APCs the main customer is the author/funder
  - membership model
  - direct support model (diamond)
  - making the cost of production disappear

Subscription publishing is a content based model

- access: the main customer is the institutional librarian
- sponsorship
- advertising
- reprints

considered a cost of the research process

As long as you are creating more value  
then you are capturing, then you are  
doing good for your community.

# What services does the customer want?

- the main customer is the author

author needs can still be served by XML workflows, but the rest of the stack is not serving them well.

they want:

- fair and balanced peer review
- rapid publication times
- a “good publication”
- a discoverable publication
  - indexing in google scholar
  - coverage in scopus/web of science
  - discipline specific archiving
  - software that is respectful of their time

# Research Data

research data

Oxford xml summer school 2015 - CC-BY 3.0

## **Why care about data?**

It's obvious to say this, but the data that research is built on is an integral part of the scientific process.

## What's the relationship between the paper, and data?

A fair criticism of current publishing workflows is that publishers force authors to reduce much of their raw data into the form of graphics or images, and then to compress these images into low resolution versions for inclusion in a manuscript. The reader wants to get the underlying data, but has to go through a lot of hoops to get it.

# Patterns

- don't connect the data to the paper at all
- refer obliquely to to the data set in the body of the publication
- link to the data set in the body of the publication via uri /identifier
- dump the data into supp info
- deposit in data cite, and hope there is a link to the paper
- link to the paper from the dataset
- create a specific section of the paper tagged about data
- cite the data in the reference list
- enhance the metadata of the paper in crossmark pointing to the data
- create a micro publication
- create a meta-paper about the data



research data



Oxford xml summer school 2015 - CC-BY 3.0

- It's one component of a research object!
  - Article
  - Data
  - Code
- It's one component of a story from the author
  - Preprints
  - Posters
- It's an evolution of a narrative
- versioning and changing research objects
  - giving a DOI to each revision of an article?
  - giving a DOI to each item of a dataset?
- The relationship between different components can be signalled.

# Scales of data

- BIG DATA
- small data
- data that is starting to get out of control

# Flavours of data

- Well structured
- semi-structured
- totally unstructured

Where can you put the data, and what can you find out?

Data repo	Metrics	API
EBI/PDG/BGI ...	?	Many
Figshare/projects	views/downloads/shared	REST Oauth
Imeji	web metrics	RDF
Dryad	views/downloads	OAI-ORE/PMH RDF
Datacite affiliated store		REST Basic HTTP Auth, OAI
Zenodo	Altmetric	OAI-ORE/PMH
Lab archives	No	Yes
Dataverse	web metrics	REST Basic HTTP Auth, OAI
Github	pull requests forks following	REST Oauth
Amazon	usage cost	REST keys based
Lab cluster research data	No	No

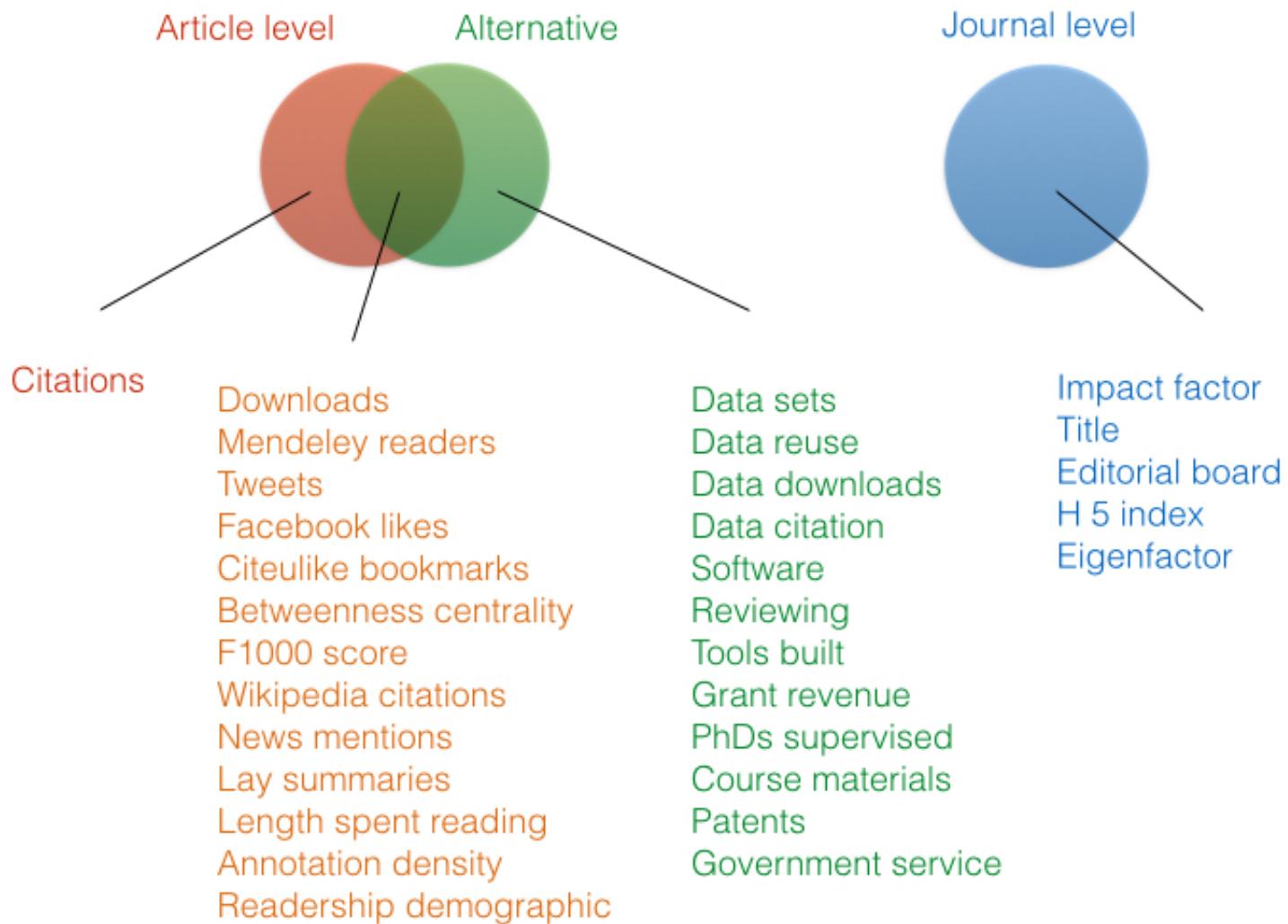
## How to handle data?

- We don't really know
- some data journals coming to market
- now exists a way to cite data, but not widely used yet
- Funders increasingly looking to bring in open data policies

# Getting credit for data

# AltMetrics

# what are they?



## **what's their history?**

Manifesto was written in 2005

Bibliometrics have been around for a lot longer

The impact factor might be the first citation metric to be developed (1950s),  
but it was restricted to journal counts due to technical limitations.

# what's the current state of the art?

- altmetric.com
  - Nature
  - Springer
  - Wiley
- lagotto - the open source tool
  - PLOS
  - copernicus
  - eLife
- plumb analytics
- publishers creating or running their own platforms
  - frontiers
- repositories adding UIDs
  - figshare
  - datadryad
  - zenodo

## what's next for ALMs?

- They provide a variety of different kinds of value to different actors in the space, but there has not yet been a “killer app”
- They allow researchers to create narratives of their work and impact (impactstory).

## but what are they really about?

- They are really about problems in the assessment and rewards system in academia.
  - see DORA - the declaration on research assessment

# What have they got to do with XML, markup and structure?

- data citation
- well identified objects help
- objects that show their interrelationship help
- the fuller a digital life that an object has, the easier it is to automate the gathering of information about that object, that can then facilitate the creation of stories around that object.

emerging trends

Reproducibility

Webkit and the browser

HTML5/Markdown/JSON

# Reproducibility

# What's the issue with reproducibility?

AMGEN result - 88% could not be replicated

2011 - 80% of studies could not be replicated

nature editorial

# Efforts to address this problem

- computable papers
- retraction watch
- centre for open science - Reproducibility project

# Reproducibility Project Cancer Biology

- RPCB
  - our process
  - paper selection
  - registered reports
  - peer reviewed replication studies

## criticisms

- expense
- negative connotations of failing to get your paper replicated
- waste of effort, more effort should be put into novel research

# why can't papers get replicated?

- lack of sharing of data
- poorly described protocols
- the experiment is just really hard

# The role of the publisher

- push for open peer review - [eLife Demo](#)
- make all data open and available
- implement data checking in the review process (e.g. EMBO checks for image manipulation)
- experiment with new ways of presenting papers and data

# The browser

# The browser

Safari

WebKit

Electron

# Demos

Atom

Slack

Pixate Studio

# Classic Steps In Publishing

Most of these steps can now be implemented in the browser, or in a browser based app.

- Authoring
- Submission
- Peer review
- Document conversion
- Content enhancement
- Annotation
- Corrections cycle
- Pdf / print / pagination
- Downstream applications
- Search

- Authoring
  - Manuscripts.app - DEMO
  - substance writer - DEMO
  - gitbook - DEMO
- Submission
  - Overleaf
- Peer review
- Document conversion
  - see lens demo in next session
- Content enhancement
- Annotation
  - [hypothes.is](#)
- Corrections cycle
  - screenshot of tool from Ravi
- Pdf / print / pagination
  - see vivlio style demo later
- Downstream applications
  - HTTP / FTP
- Search

dit ▾ Q Find Add ▾ Cite ▾ Style ▾ Export To PDF ▾  
 I x² x, u E ▾ E ▾ E ▾ E ▾

Michael Cox at the University of Wisconsin, Madison. We incorporated *p*-azido-L-phenylalanine (*p*AzF) (Chin et al., 2001) into RecA WT to site specifically label the protein with Alexa Fluor 488 DIBO alkyne. For the cloning of RecA, we used the pET21a(+) vector, which was a gift from Michael Cox at the University of Wisconsin, Madison. The Phe21 sequence in RecA WT was mutated to F21A with the amber codon via site-directed mutagenesis using Pfu Ultra polymerase (Agilent Technologies). Once the construct was confirmed, pAIR79 was cotransformed with the vector pEVOL-*p*AzF (a gift from the Peter Schultz lab at The Scripps Research Institute, San Diego, CA) into the BLR expression strains (Young et al., 2010). The RecA<sub>F21AzF</sub> protein was expressed using the same standard protocol for RecA WT (Cox et al., 1981). RecA<sub>F21AzF</sub> was labeled with Alexa Flour 488 (RecA<sub>F21AzF</sub> 488) according to manufacturer's instructions (Life Technologies).

## Clut assembly

Modified 45nt ssDNA was covalently attached to Cyanogen-Bromide Sepharose resin according to manufacturer's instructions (Sigma-Aldrich) and transferred to a spin column (Biorad). Briefly, the resin was activated with 1 mM HCl then washed and equilibrated with coupling buffer (0.1 M NaHCO<sub>3</sub>, 0.5 M NaCl). 5' amino-modified DNA (20 nmoles) was coupled to 100 mg resin overnight at 4°C, and unbound oligomers were removed by washing resin 7 times with coupling buffer. Free amino groups were blocked with Ethanolamine (1 M, pH 8; Sigma-Aldrich) to prevent non-specific binding then 1 M NaCl. The concentration of 45nt ssDNA bound to the resin consistently provided about 6 nmoles per 100 mg of resin. RecA\* was assembled by incubating excess RecA or RecA mutants and ATPgS (Roche) or ATP (Amersham) when stated with 0.5 nmole ssDNA-bound resin for 20 min at 37°C. Free RecA and ATPgS were separated from the reaction by gentle centrifugation at 0.1 g for 1 min and collected in the flow through. Washes were repeated with reaction buffer (0.1 M Tris-HCl pH 7.5, 25 mM Sodium Glutamate, 8 mM MgCl<sub>2</sub>, 8 mM DTT, 4% glycerol, 0.1 mM EDTA) until no RecA was detected.

the browser

Changes Media References C

Bertram JG, Bloom LB, Hingorani MM, Beechem JM, O'Donnell M, Goodman MF. 2001. Molecular mechanism and energetics of clamp assembly in *Escherichia coli*: the role of ATP hydrolysis when gamma complex loads beta on DNA. *J Biol Chem* 276: 28413–20.

Bruck I, Woodgate R, McEntee K, Goodman MF. 1996. Purification of a soluble UmuDC complex from *Escherichia coli*: cooperative binding of UmuDC to single-stranded DNA. *J Biol Chem* 271: 10767–74.

Brune M, Hunter JL, Howell SA, Martin SR, Hazlett TL, Corrie JE, Webb MR. 1998. Mechanism of inorganic phosphate interaction with phosphate binding protein from *Escherichia coli*. *Biochemistry* 37: 10370–80.

Burckhardt SE, Woodgate R, Scheuermann RH, Echols H. 1988. UmuD mutagenesis protein of *Escherichia coli*: overproduction, purification, and cleavage by RecA. *Proc. Natl. Acad. Sci.* 85: 1811–5.

Chin JW, Santoro SW, Martin AB, King DS, Wang L, Schultz PG. 2002. Addition of *p*-azido-L-phenylalanine to the genetic code of *Escherichia coli*. *J Am Chem Soc* 124: 9026–7.

Corzett CH, Goodman MF, Finkel SE. 2013. Competitive fitness during feast and famine: SOS DNA polymerases influence physiology and evolution in *Escherichia coli*. *Genetics* 194: 409–20.

Dutreix M, Moreau PL, Bailone A, Galibert F, Battista JR, Walker GC, Devoret R. 1989. recA mutations that dissociate the various RecA protein activities in *Escherichia coli* provide evidence for an additional role for RecA protein in UV mutagenesis. *J Bacteriol* 171: 2415–20.

Eggerl AL, Lusetti SL, Cox MM. 2003. The C terminus of the *Escherichia coli* RecA protein modulates the DNA binding competition with single-stranded DNA-binding protein. The C terminus of the *Escherichia coli* RecA protein modulates the DNA binding competition with single-stranded DNA-binding protein. *J Biol Chem* 278: 16389–96.

Frank, E. G., Hauser, J., Levine, A. S. & Woodgate, R. 1993. Targeting of the UmuD, Umu and MucA' mutagenesis proteins to DNA by RecA protein. *Proceedings of the National Academy of Sciences (USA)*, 90, 8169-8173.

Friedberg, E. C., Walker, G. C., Siede, W., Wood, R. D., Schultz, R. A. & Ellenberger, T. 2005. *DNA Repair and Mutagenesis*, Washington, D. C., ASM Press.

# Markup Markdown/ HTML5

This will go into more detail on some document conversion steps that might be interesting to this group.

- <http://scholdoc.scholarlymarkdown.com>  
<http://blog.martinfenner.org/2015/04/23/introducing-scholarly-markdown-bundle/>
- pandoc
- lens
- substance - lens writer
- paged media standard

(Vivliostyle demo) (<http://www.vivliostyle.com/en/sample/>)

# Lens Demo

markup/markdown/html5

Oxford xml summer school 2015 - CC-BY 3.0

# Slides Demo

markup/markdown/html5

Oxford xml summer school 2015 - CC-BY 3.0

Speculative Trends

Computational Papers

Living Papers

The Blockchain

# Computational papers

- encode project
- iPython/Project Jupyter
- nbviewer
- Bio.JS
- FigShare

# Encode Demo

computational papers

Oxford xml summer school 2015 - CC-BY 3.0

# Living Papers

- wikipedia
- eLife research advances
- f1000 research
- <http://theoj.org>
- ReScience

# eLife Research Advances demo

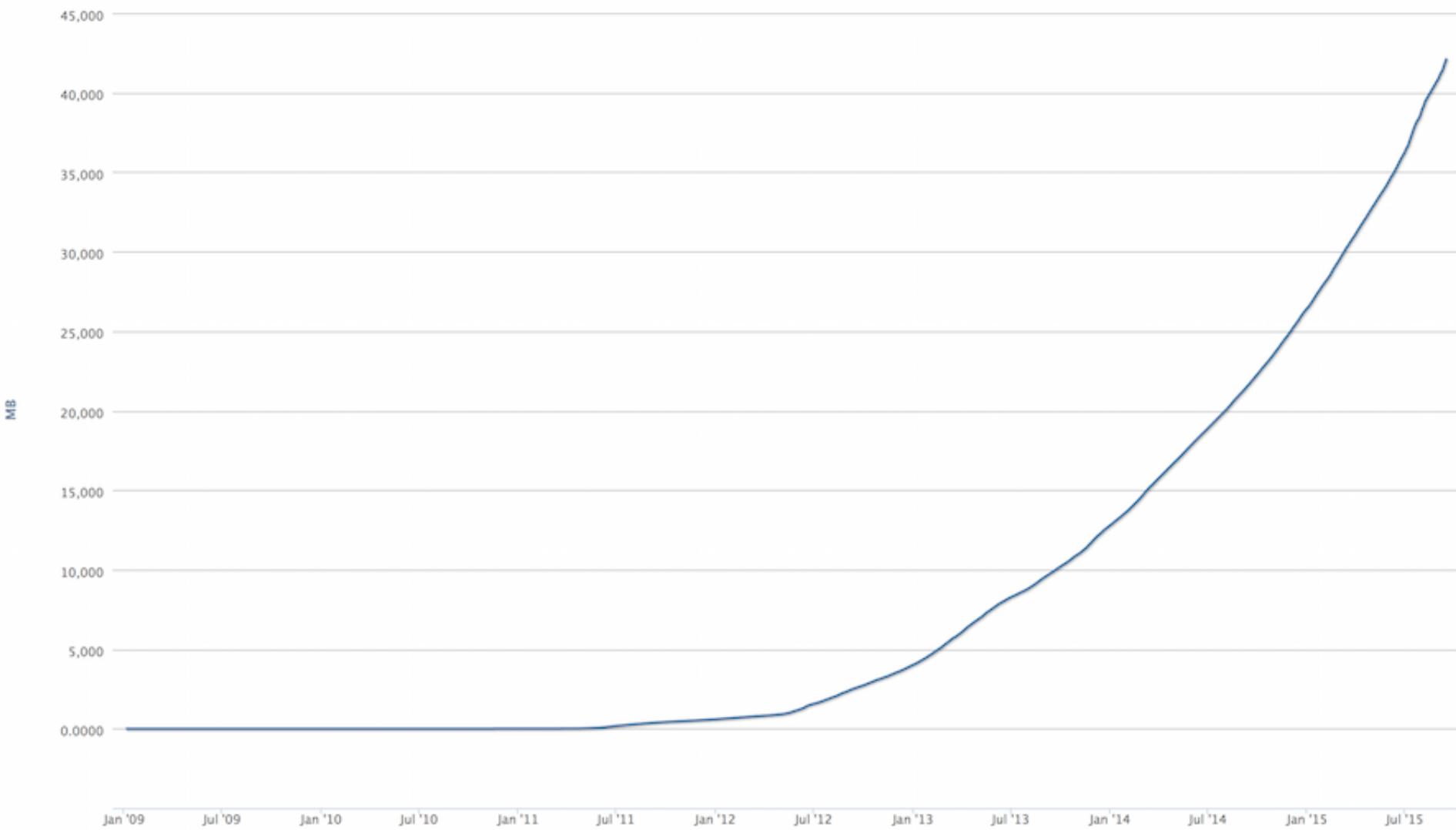
# ReScience demo

living papers

Oxford xml summer school 2015 - CC-BY 3.0

# The Blockchain & Science

Blockchain Size  
Source: blockchain.info



# Resources

These slides, mindmap, and associated slides in PDF, markdown and HTML versions can be obtained from [this git repo](#)

```
$ git clone https://github.com/IanMulvany/2015-09-oxford-xml-summer-school.git
```

The GitBook generated from the markdown notes for this talk can be obtained from [oxford-xml-summer-school](#).

Please direct any questions to me on Twitter [@IanMulvany](#)

## Other eLife resources

review process: <http://vimeo.com/49775707>

research advances: <http://elifesciences.org/Moving-research-forward-eLife-announces-the-Research-Advance>

reproducibility project: <http://elifesciences.org/collections/reproducibility-project-cancer-biology>

Lens: <https://github.com/elifesciences/lens/>

