Licensing in an open access environment: the publishing view

Deborah Dixon

VP, Medical Publishing Director

January 2013

WILEY Open Access

Quality Open Access Research

CC-BY vs CC-BY-NC

- Hypothesis: NC allows researchers and funders to meet their goals without undermining the interests of publishers and learned societies

Value of Publishers ?

Value of Publishers?

Why Does CC-BY Undermine Publishers' Business?

- Article reprints (print and electronic)

- Third Party Rights Income – eg licensing use of content, figures, etc, to commercial companies

- Investment in content that can be better commercialised by richer competitors – eg Under CC-BY a large publisher could scoop up every paper published and become the monopoly source for scholarly literature. Or a big software company could create new services around content

- Societies – Less revenue from reprints and licensing; high investment content (such as Cochrane Reviews) could be incorporated into high value tools with no return for the original work

Researchers' Goals

- Data Mining and Data Sharing

- Finding Commercial Partners

- Teaching

Funders' Goals

- Open and Reusable Science

- Innovation

- Governmental requirements

- Contribute to National Economy

Why CC-BY? (Butterflies By Corey Leopold)
http://wwww.flickr.com/photos/ No…..)

But it's About the Data

And Data is often Not Available

- ? A role for librarians?

- More funding by pharma in medical journals than RCUK etc

- Conflicted by also desire to hide data – that does not support their drugs and from competitors

Cochrane: Trouble Getting the Data

But Data is Useful

Pfizer is stepping up its efforts to get more information from existing clinical trial data. The company is turning to sophisticated data mining techniques to help improve the design of new trials, to better understand possible new uses for existing drugs, and to help examine how drugs are being used after they have been approved.

- "We want to milk as much out of the data as possible," says Mani Lakshminarayanan, director/statistical scientist, at Pfizer

Open Linked Data: Activity Increasing

- Schema.org, new tools evolving and Data.gov offering almost 400,000 open data sets - the time could be ripe for commercialization.

- Linked open data project consortium released next generation of open data tools. Recommends RDF (resource description framework) where facts take the form of triples and infer new facts, eg:

1) MichaelJordan isBornIn Brooklyn

2) Brooklyn isPartOf NewYorkCity

Inference: : MichaelJordan livedIn New York City

- Lead to better information search, improved analytics, superior system integration and automation, and new sorts of aggregated databases.

- Isolated and interesting business cases of open linked data, but so far no large-scale commercial success. So what is the hold up?

WILEY Open Access

Quality Open Access Research
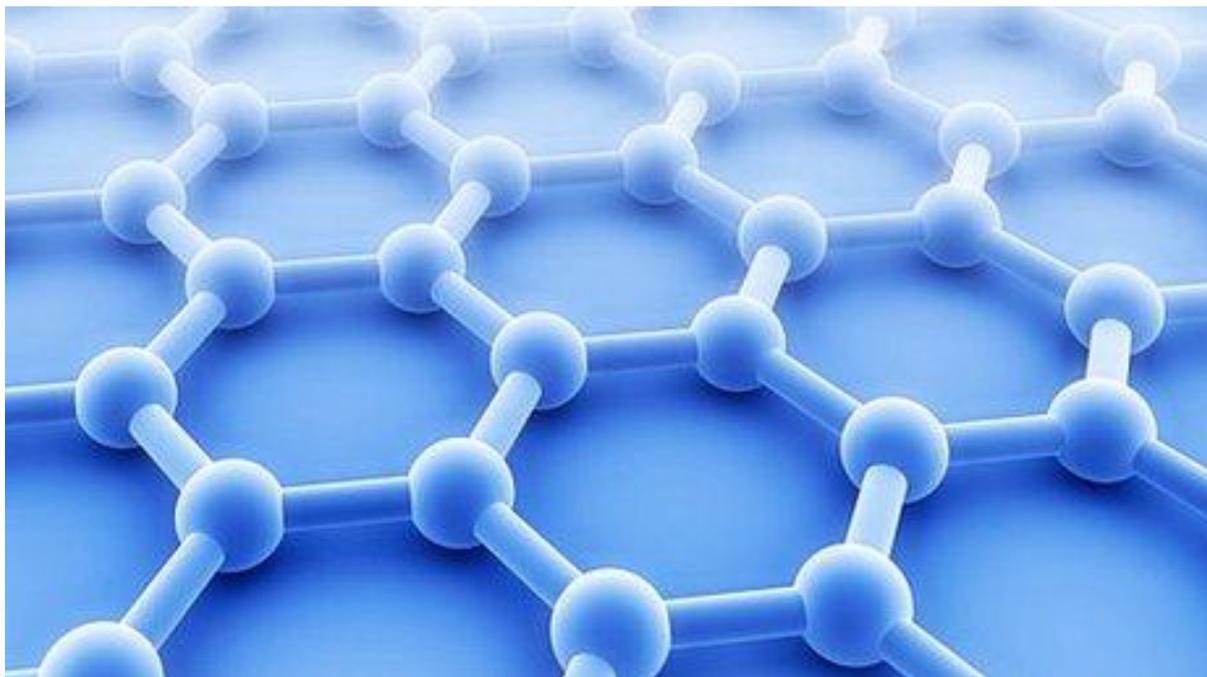
Barriers to Commercialisation

- Approaches for machine readable data, whether manual or automated, are either too expensive, too inaccurate, or too narrow in scope

- Open data schema are difficult to use; often ambiguous and, difficult to use for webmasters and content producers

- So when will we know that the open data cloud is finally taking off commercially? - observe how enterprise applications (EA) vendors such as Siemens are adopting linked open-data export functionality

- Artificial Intelligence (AI) solutions come to the fore, eg IBM's Watson

IBM Watson: as data grows by 800% in next 5 years

IBM Research Project (2006 –) → Jeopardy! Grand Challenge (Feb 2011) → Watson for Healthcare (Aug 2011 –) → Watson for Financial Services (Mar 2012 –) → Watson Industry Solutions (2012 –)

R&D → Demonstration → Commercialization → Expansion → Cross-industry Applications

[Watson after *Jeopardy!*](#)

UK loses out to China and US

The Economy: Quick to Patent?

- Graphenes – global race for patents; UK loses out to China

- HPV vaccine

- Does an NC licence prevent researchers finding commercial partners? It has not prevented researchers in other countries capitalising on the work

Shaping a New Future

• Make content ready for text and data mining (expensive and need incentive)

• Entity extraction – eg linking a chemical compound in a paper to a database with more info on the compound (Wiley's *Functional Chemistry* project)

• Natural Language Processing (NLP) – use in concert with text mining to provide filtering systems

• Discoverability/linking (eg JISC Open Citations project)

• Product development using semantic metadata: Examples: *nanopublications* (http://nanopub.org, *OpenPHACTS* project (under the umbrella of the Innovative Medicine Initiative) http://www.openphacts.org , *Chemspider* arms of the Royal Society of Chemistry, Elsevier's *Clinical Key*

• User centred research for product development

Advantages of CC-BY-NC

- Under CC-BY, without increased research community funding, there is no longer any incentive for publishers to continue to develop and implement these new technologies. Why update your metadata? Why not block bulk downloaders if all they do is cost you money that can't be recovered?

- Under CC-BY-NC text mining and reuse costs are supported by commercial entities. Under CC-BY, the burden falls on the research community

- Economic growth can still be generated even if startups have to pay for access to raw material (eg compulsory licensing)

- Funders outline specifically what activities they require to be free

Lack of control with CC-BY

Graf, K, Thatcher, S. (2012). Point & Counterpoint: Is CC BY the Best Open Access License? *J Librarianship and ScholarlyCommunication* 1(1):eP1043

- CC-BY makes it impossible for an author or publisher to exert any control over how translations are prepared and published

- Authors are worried about how their articles/ chapters
  are reproduced alongside other content (Eg, a conservative author reproduced in a Marxist work; a Cochrane Review on a pharmaceutical marketing site

- History journals interpret as infringement of IP

⊛WILEY Open Access   Quality Open Access Research

Does CC-BY-NC allow researchers to achieve their goals?

- Hargreaves review says researchers can mine data (fair dealing). Authors rarely doing research for profit

- Cochrane has synthesised clinical trials data without CC licences

- Have extensive rights for own work for teaching. Extended for institutions but substantial use requires a licence – Can NC be adapted to meet needs?

Mixing it Up

- Additional cost for CC-BY (vs –NC)

- NC and Compulsory Licensing (eg IBM Watson, Pharma Companies). Encourages QUALITY of curation

- NC becomes CC-BY after a period of time while publisher/authors can make money from the content (like patent model)?

Publishers' Policies

- Expedited by RCUK/WT requirements
- CC-BY for OA journals
- Some exclusion of medical titles
- Author choice, funder compliance for hybrids (Wiley)
- Differential pricing for CC-BY
- OASPA – allows CC-BY-NC (not SA or ND) because of loss of reprints revenues

CC-BY vs CC-BY-NC

- Hypothesis: NC allows researchers and funders to meet their goals without undermining the interests of publishers and learned societies

Need for a New "Fit For Purpose" NC Licence

- Describes exactly what rights authors/funders require:

  Data Mining

  Teaching

  Application for commercial licensing

In Summary…..

Discussion