Reuse and Open Access

30 March 2006
Berlin 4 Conference

Except where otherwise noted, this work is licensed under a Creative Commons Attribution 2.5 License
What’s the point?

● Scientific innovation is directly related to the freedom to reuse and transform scientific knowledge…it is built on Open Access.
  ● Science Commons and Creative Commons: focus on enabling reuse and demonstrating the impact of reusable knowledge.
  ● The Neurocommons project: a proving ground.
primer: Creative Commons licenses
Make the policy clear to all copyright holders

**Attribution 2.5**

You are free:
- to copy, distribute, display, and perform the work
- to make derivative works
- to make commercial use of the work

Under the following conditions:

**Attribution.** You must attribute the work in the manner specified by the author or licensor.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the [Legal Code (the full license)](https://creativecommons.org/licenses/by/2.5/legalcode).
Hide the complexity of the law from the user
Make the legal code “portable” to other languages and legal systems

<table>
<thead>
<tr>
<th>Argentina</th>
<th>Australia</th>
<th>Austria</th>
<th>Belgium</th>
<th>Brazil</th>
<th>Bulgaria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>Chile</td>
<td>Croatia</td>
<td>Finland</td>
<td>France</td>
<td>Germany</td>
</tr>
<tr>
<td>Hungary</td>
<td>Israel</td>
<td>Italy</td>
<td>Japan</td>
<td>South Korea</td>
<td>Malaysia</td>
</tr>
<tr>
<td>Mexico</td>
<td>Netherlands</td>
<td>Poland</td>
<td>Slovenia</td>
<td>South Africa</td>
<td>Spain</td>
</tr>
<tr>
<td>Sweden</td>
<td>Taiwan</td>
<td>UK: England and Wales</td>
<td>UK: Scotland</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
十一种授权选择

“创作共用”协议机制提供了由4个最常见的授权选择的组合方式，任何作品都可以通过选择这性地组合声明自己的作品授权，实际上可以有11种常见的组合方式。这些组合方式构成了从“松”到“紧”的授权限制，给作品的创造者更加灵活便利的选择。为网络上的数字作品提供了新的“护身符”。

署名
署名-非派生作品
署名-非派生作品-非商业用途
署名-非商业用途
署名-非商业用途-保持一致
署名-保持一致
Make the transaction machinable

```xml
_LICENSE rdf:about="http://creativecommons.org/licenses/by/2.5/">
  <permits rdf:resource="http://web.resource.org/cc/Reproduction" />
  <permits rdf:resource="http://web.resource.org/cc/Distribution" />
  <requires rdf:resource="http://web.resource.org/cc/Notice" />
  <requires rdf:resource="http://web.resource.org/cc/Attribution" />
  <permits rdf:resource="http://web.resource.org/cc/DerivativeWorks" />
  <permits rdf:resource="http://web.resource.org/cc/CommercialUse" />
</LICENSE>
```
Leverage the Web

Google Advanced Search

- Find results with all of the words, with the exact phrase, with at least one of the words, or without the words.
- Language: any language
- File Format: any format
- Date: anytime
- Occurrences: anywhere in the page
- Domain: not filtered by license
- Usage Rights: free to use or share, free to use or share, even commercially, free to use share or modify, free to use, share or modify, even commercially
- SafeSearch: No filtering, Filter using SafeSearch

Page-Specific Search

- Similar: Find pages similar to the page
- Links: Find pages that link to the page
Backlinks to creativecommons.org via Yahoo! and All The Web

3/1/06: 50,000,000

- alltheweb
- Yahoo!
London Underground bombing, trapped

People trapped in the tube.

This work is licenced under a Creative Commons Licence.

image taken by Adam Stacey
Unintended consequences

our users started using metadata to track genealogy...
open metadata has become a metric of reuse
Back to the point: reuse

- Tracking through copyright licensing lets us track the reuse
- Reuse leads to creativity the original authors did not envision
- Reuse is a fascinating **metric** of the impact of a creative work…
Reusing open text: The Neurocommons project
The Neurocommons Project

- Public domain graph of neurological facts built on Open Access knowledge and public databases
- Standard software infrastructure
  - Text mining
  - Relationship storage
  - Metadata storage
- Community of practice - must develop beyond Science Commons…
- Operations manuals and project documentation

http://sciencecommons.org/data/neurocommons
Knowledge explosion

Growth of GenBank
(1982 - 2004)

Key MEDLINE® Indicators

<table>
<thead>
<tr>
<th></th>
<th>FY2004</th>
<th>% Change</th>
<th>FY2003</th>
<th>% Change</th>
<th>FY2002</th>
<th>% Change</th>
<th>FY2001</th>
<th>% Change</th>
<th>FY2000</th>
<th>% Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citations</td>
<td>571,000</td>
<td>+9%</td>
<td>526,338</td>
<td>+5%</td>
<td>502,056</td>
<td>+8%</td>
<td>463,014</td>
<td>+5%</td>
<td>442,000</td>
<td>+2%</td>
</tr>
<tr>
<td>Indexed for MEDLINE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PubMed® Searches</td>
<td>677 Million</td>
<td>+35%</td>
<td>503 Million</td>
<td>+32%</td>
<td>380 Million</td>
<td>+15%</td>
<td>330 Million</td>
<td>+35%</td>
<td>244 Million</td>
<td>----</td>
</tr>
</tbody>
</table>

Semantic Web (Berners-Lee 2001)

image courtesy of w3.org
Semantic Web (Berners-Lee 2001)

image courtesy of w3.org
The problem

- WWW just needed *technical* interconnection to link pages
- Semantic Web needs *knowledge* interconnection to link resources
- Most of that rich interconnection is proprietary
  - Locked up behind restrictive contracts based on copyrights
  - Publisher can encrypt down to clickstream…
Extracting knowledge

<p53><activates><icam-1 expression>

Natural language processing (NLP) can extract these relationships

Republishing knowledge

\(<p53> <\text{activates}> <\text{icam-1 expression}>\)

Republishing using Semantic Web creates open knowledge sub-graphs.
Open Access: More rights, more impact...

Research article

Select estrogens within the complex formulation of estrogens (Premarin(R)) are protective against neural implications for a composition of estrogen therapy to function and prevent Alzheimer's disease

Li Qin Zhao and Roberta D Brinton


Published 13 March 2006

Abstract (provisional)

Background

Results of the Womens Health Initiative Memory Study (WHIMS) raised concern that the complex formulation of hormone interventions. Conjugated equine estrogens (CEE), used in the WHIMS trial, is a complex formulation containing multiple estrogens, including ovaries, as well as other biologically active steroids. Although the full spectrum of estrogens present in CEE has not yet been resolved, 10 estrogens have been identified and determined which estrogenic components, at concentrations commensurate with following a single oral dose of 0.625 mg CEE (the dose used in the WHIMS trial), and whether combinations of those neuroprotective estrogens provide added computer-aided modeling analyses, to investigate the potential correlation of conferring estrogen neuroprotection with estrogen interactions with the estrogen
Year: 2004

Number of records: 30.

Conference or Workshop Item


What about copyright?

- Applies to *creative expression*, not *facts of nature*
- One major copyright problem is the way firewalls and contracts are used to prevent technical reuse
- Another is the prevention of translation into local languages - but Neurocommons *can be translated*...
Maheswaran S, Englert C, Bennett P, Heinrich G, Haber DA.
Graphing knowledge

<WT1 GP> <stabilizes> <p53>
<p53> <activates> <icam-1 expression>
<WT1 GP> <inhibits> <p53-mediated apoptosis>
Copyright Status

Government information available from this site is within the public domain. Public domain information on the National Library of Medicine (NLM) Web pages may be freely distributed and copied. However, it is requested that in any subsequent use of this work, NLM be given appropriate acknowledgment.

This site also contains resources such as PubMed Central, Bookshelf, OMIM, and PubChem which incorporate material contributed or licensed by individuals, companies, or organizations that may be protected by U.S. and foreign copyright laws. All persons reproducing, redistributing, or making commercial use of this information are expected to adhere to the terms and conditions asserted by the copyright holder. Transmission or reproduction of protected items beyond that allowed by fair use (PDF) as defined in the copyright laws requires the written permission of the copyright owners.
EBI Terms of Use of the EBI Services

1. The public databases of the EBI are freely available by any individual and for any purpose.

2. Software which can be run at the EBI may be used by any individual for any purpose unless specific exceptions are stated. The only exceptions with respect to software will be the prohibition of downloading, redistributing or re-engineering the software, and these exceptions only apply if explicitly stated.
Reuse of knowledge

WT1 Gene Product
- co-occur
- co-occur
- co-occur
- stabilizes apoptosis

p53
- downstream mediates
- activates icam-1 expression

HD
- downstream

Wilbanks
- activates

Wellcome

build additional metrics for impact
Status

- Funded by Teranode (venture backed software company, Seattle WA)
- Launched January 2006
  - Partnerships with major research hospitals, neuroscience infrastructure, W3C HCLSIG, Open Access publishers
  - Working group of neuroinformaticists, neuroscientists, Semantic Web, NLP, legal theorists
- Alpha release: 4th Quarter, 2006
SB21673

Why was this found?
WNT Pathway plays a role in Diabetes Type 2
WNT Pathway is a primary target of SB21673
Where we’re going

- Data sets and relationships from data
- Extend FOSS infrastructure
  - Annotation, markup and query
  - Publishing - automatic additions to the graph
- International partnerships - “leapfrog” the established systems of publishing
- Coordinate with other scientific domains
“Steal this book”

- Back to the point: it’s about reuse…

- If you want this infrastructure in your domain, all our work will be available
  - Graph: public domain
  - Software written by SC: under OSI licenses to open APIs
  - Can use corporate partner software, or use Free Software / Open Source Software wired to the open APIs
Promise to practice...

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities

Preface

The Internet has fundamentally changed the practical and economic realities of distributing scientific knowledge and cultural heritage. For the first time ever, the Internet now offers the chance to constitute a global and interactive representation of human knowledge, including cultural heritage and the guarantee of worldwide access.

We, the undersigned, feel obliged to address the challenges of the Internet as an emerging functional medium for distributing knowledge. Obviously, these developments will be able to significantly modify the nature of scientific publishing as well as the existing system of quality assurance.

In accordance with the spirit of the Declaration of the Budapest Open Access Initiative, the ECHO Charter and the Bethesda Statement on Open Access Publishing, we have drafted the Berlin Declaration to promote the Internet as a functional instrument for a global scientific knowledge base and human reflection and to specify measures which research policy makers, research institutions, funding agencies, libraries, archives and museums need to consider.
Thanks to MIT, CSAIL, the Omidyar Network, Teranode & the High Q Foundation

**HIGH Q FOUNDATION**

**ABOUT HIGH Q FOUNDATION, CHDI AND MRSSI**

The High Q Foundation, Inc. is a private philanthropic foundation that was established in 2002 with the mission of bringing together academia, industry, governmental agencies, and other funding organizations in the search for Huntington's disease (HD) treatments. The Foundation supports numerous projects related to HD, including basic research, a drug-discovery program, and clinical studies.