dcc conference
edinburgh, scotland

radical sharing: transforming science?

john wilbanks
creative commons / science commons
1. the “alternative” to radical sharing is sharing inefficiently.
science is not unlike wikipedia...
science is not unlike wikipedia...
it is based on publishing (make public)
it advances via individual action
The p53 regulatory gene MDM2 is a direct transcriptional target of MYCN in neuroblastoma

Andrew Slack,* Zaowen Chen,* Roberto Tonelli,† Martin Pule,* Lisa Hunt,* Andrea Pession,† and Jason M. Shohet*†

*Center for Cell and Gene Therapy, Texas Children’s Cancer Center, Baylor College of Medicine, 1102 Bates Street, Houston, TX 77030, and †Department of Pediatrics, University of Bologna, Santa Orsola Hospital, Via Massarenti, 11-40138 Bologna, Italy

†To whom correspondence should be addressed. E-mail: jmshohet@texaschildrenshospital.org.

Edited by Robert N. Eisenman, Fred Hutchinson Cancer Research Center, Seattle, WA

Received July 28, 2004; Accepted November 19, 2004.

This article has been cited by other articles in PMC.

it advances discrete edit by discrete edit
science is not unlike wikipedia...
science is not unlike wikipedia...

...except authenticated, and expensive.
given **trust ratings** by peer review
Serial Expenditures (+273%)
Serial Unit Cost (+188%)
Monograph Unit Cost (+77%)
CPI (+73%)
Monograph Expenditures (+63%)
Serials Purchased (+42%)
Monographs Purchased (-9%)

*Includes electronic resources from 1999-2000 onward.
science is not unlike wikipedia...

...except authenticated, and expensive.

(and slow!)
and what of the curator?
necessarily involves tangible objects of some sort
from a technical perspective
from a technical perspective
inefficient and expensive ecosystem of processes to **peer-produce and review** scholarly content
disruptive processes can’t be planned in advance.
disruptive processes can’t be planned in advance.

planned innovation tends to be incremental, and slow.
disruptive processes can’t be planned in advance.

planned innovation tends to be incremental, and slow.

...and not innovative.
PHILOSOPHICAL TRANSACTIONS: GIVING SOME ACCOUNT OF THE PRESENT Undertakings, Studies, and Labours OF THE INGENIOUS IN MANY CONSIDERABLE PARTS OF THE WORLD.

Vol I. For Anno 1665, and 1666.

In the SAVOY,
Printed by T. N. for John Martyn at the Bell, a little without Temple-Bar, and James Allestry in Duck-Lane, Printers to the Royal Society.
2.

stable systems are **resistant to change on multiple levels**.
creative expression
the container, not the facts.
the container, not the facts.

but © locks the container.
IGFBP-5 plays a role in the regulation of cellular senescence via a p53-dependent pathway and in aging-associated vascular diseases.
IGFBP-5 plays a role in the regulation of cellular senescence via a p53-dependent pathway and in aging-associated vascular diseases.
LICENSE AGREEMENT
Schedule 1.2(a)
General Terms and Conditions

RESTRICTIONS ON USAGE OF THE LICENSED PRODUCTS/ INTELLECTUAL PROPERTY RIGHTS

GTC1 The Licensed Products may not be used (directly or indirectly) for any of the following purposes: substantial or systematic reproduction; redistribution, reselling or sublicense in any manner including in connection with fee for-service use; supply or distribution in any form to anyone other than Authorized Users, except as may be expressly set forth otherwise in this Agreement; substantial or systematic retention of the Licensed Products. Subscriber shall not use spider or web-crawling or other software programs, routines, robots or other mechanized devices to continuously and automatically search and index any content accessed online under this Agreement.

GTC2 Neither Subscriber nor its Authorized Users may modify, adapt, transform, translate or create any derivative work based on the Licensed Products, or otherwise use same in a manner that would infringe the copyright or other proprietary rights therein. Copyright notices, other notices or disclaimers included in the Licensed Products may not be removed, obscured or modified in any way. With respect to permitted uses of Excerpts, Subscriber shall acknowledge the source, title, author and publisher (if not already included in Excerpts).

GTC3 Subscriber will not reverse engineer, decompile or disassemble any software included in the Licensed Products. However, in the event that a court determines that Subscriber has such right under its local law, which right cannot be waived, this right shall be restricted solely to the study of matters such as interoperability and shall not be used in such a way as to affect its commercial value.

GTC4 Subscriber acknowledges that all rights save for the limited and non-exclusive licenses or other privileges granted under this Agreement are reserved to and remain the exclusive property of the Licensor or its licensors, and nothing in this Agreement shall transfer any rights to Subscriber. Subscriber acknowledges that the Licensed Products contain valuable and proprietary data and information, and that the unauthorized distribution of such material could materially harm the business and prospects of the Licensor or its licensors.

GTC5 The laws and regulations of the United States restrict the export and re-export of certain commodities and technical data of United States origin, including any software included in the Licensed Products. Export or re-export of such software in any form requires appropriate United States government export licenses and may require licenses from other governments.

USE A NOTE. GO TO JAIL.

WELL, NOT QUITE

Music clearance problems can cost you time, money and sometimes a few sleepless nights. Let's face it, who needs the headaches?

At EMG, music rights clearance is our business - our only business.

>> More
what do these ideas mean in a world of integrated data?
<table>
<thead>
<tr>
<th>Compound</th>
<th>Formula</th>
<th>Molecular Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>14301</td>
<td>C₆H₂₀N₂O₁₂P₄</td>
<td>436.1242</td>
</tr>
<tr>
<td>142930</td>
<td>C₆H₂₃N₃O₁₂P₄</td>
<td>453.1548</td>
</tr>
<tr>
<td>143228</td>
<td>C₆H₂₀K₄N₂O₁₂P₄</td>
<td>592.5153</td>
</tr>
<tr>
<td>145157</td>
<td>C₆H₁₅N₂Na₅O₁₂P₄</td>
<td>546.0334</td>
</tr>
</tbody>
</table>

*creative work?*
40 minutes per year
"Behind one door is tenure - behind the other is flipping burgers at McDonald’s."

Copyright © 2003 David Farley, d-farley@ibiblio.org
Overview

The NIH Public Access Policy ensures that the public has access to the published results of NIH funded research. It requires that peer-reviewed journal manuscripts that arise from NIH funds be submitted to PubMed Central upon acceptance for publication. The public has access to these papers on PubMed Central no later than 12 months after publication.

How to Comply

Address Copyright

Before you sign a publication agreement or similar copyright transfer agreement, make sure that the agreement allows the NIH to make the manuscript available on PubMed Central in accordance with the Public Access Policy. See "Whose approval do I need to submit my final peer-reviewed manuscript?" for more information.

Submit Papers

Authors may submit a paper to the journal of their choice for publication. There are four methods to ensure that a manuscript is deposited on PubMed Central in compliance with the NIH Public Access Policy.

Method A: Publish in a journal that deposits all NIH-funded final published articles in PubMed Central (PMC) without an embargo.
Method B: Make arrangements to have a publisher deposit a specific final published article in PubMed Central.
Method C: Deposit the final peer-reviewed manuscript in PMC yourself via the NIH Manuscript Submission System (NIHMS).
Method D: Complete the submission process for a final peer-reviewed manuscript that the publisher has deposited in PubMed Central.

These methods vary in the version of the paper submitted, and the actions undertaken by the author and publisher. Please see "Whose approval do I need to submit my final peer-reviewed manuscript?" for more information.

Cite Papers

When citing their NIH-funded papers in NIH applications, proposals or progress reports, authors must include the PubMed Central identifier in their citation. For more information see the Frequently Asked Questions, especially:

C6. How do I include the PubMed Central reference number in my citations?
C7. What do I do if the PubMed Central reference number (PMCID) has not been assigned yet?
C8. Do I have to include a PMCID for every paper that I cite in an NIH application, proposal or progress report?
Office for Scholarly Communication

The goal of university research is the creation, dissemination, and preservation of knowledge. We collectively take this to be a good. It is an essential part of our duties as faculty members to distribute the fruits of our scholarship as widely as possible.

—Steven E. Hyman, Provost of Harvard University

Steven E. Hyman, Provost of Harvard University, has charged the Harvard University Library (HUL) with creating an Office for Scholarly Communication (OSC). The goal of the new Office for Scholarly Communication is to enable individual faculty members to distribute their scholarly writings in keeping with the University's long-standing policy that "when entering into agreements for the publication and distribution of copyrighted materials individuals will make arrangements that best serve the public interest."

The new Office for Scholarly Communication will be under the aegis of the distinguished historian Robert Darnton, who serves as Harvard's Carl H. Fierzheimer University Professor and Director of the University Library. Working in close collaboration with HUL's Office for Information Systems, the new OSC will oversee an open-access repository for current research.

According to Professor Darnton, "the open-access repository at Harvard is meant to promote openness in general. It will make the current scholarship of Harvard's faculty freely available everywhere in the world, just as the digitization of the books in Harvard's library will make learning accumulated since 1638 accessible worldwide. Taken together, these and other projects represent a commitment by Harvard to share its intellectual wealth. They belong to a cause that has gathered force over the centuries—the democratization of knowledge—and that now can be realized on a global scale, thanks to the progress of information technology."

"Put less grandly," Darnton stated, "the repository will implement the unanimous vote by the Faculty of Arts and Sciences on February 12, 2008, to transfer nonexclusive copyrights of their scholarly articles to the President and Fellows of Harvard. The articles will be stored, preserved, and made freely accessible in digital form. Faculty members will retain the rights to their articles and will be able to make individual arrangements for their publication with peer-reviewed journals. And by taking advantage of an opt-out provision, they may choose not to share the rights to a particular article. The policy is meant to be collective but not coercive."

In a related move announced on May 7, the Harvard Law School faculty unanimously voted to make each faculty member's scholarly articles available online for free, making HLS the first law school to commit to open access. "The Harvard Law School faculty produces some of the most exciting, groundbreaking scholarship in the world," said Dean Elena Kagan. "Our decision to embrace 'open access' means that people everywhere can benefit from the ideas generated here at the Law School."

Stuart M. Shieber, Harvard's James O. Welch, Jr., and Virginia B. Welch Professor of Computer Science, and the author of the motion accepted by
i can has repository staff?
Minature disasters and minor catastrophes
Part of: DSpace.

KT Tunstall’s wonderful song is playing on Pandora as I type this, and it’s just so fitting I have to use it as this post title!

This is a tale of beating DSpace and OS X with many, many rocks until they sorta-kind'a work. I present it here in hopes of sparing someone else considerable annoyance.

One of my best clients emailed me with a “please fix this link in my HTML item” request. Simple enough, right?

The said HTML item is nested in folders three deep. This means that DSpace’s regular exporter breaks, because it’s not smart enough to create intermediate folders. Joy.

So I kicked that up to the dspace-tech list, and got a kind response from Larry Stone of MIT: “use the METS packager export instead.” I did, and lo! it worked.

So I twiddled the file needing twiddling, zipped up the whole, and tried to put it back. First the METS ingester barfed because I’d zipped the folder containing all the files, not the files themselves. Okay, durr, I felt stupid and zipped the files properly.

Then the METS ingester barfed because unbeknownst to me, Mac OS X’s native zip utility adds OS X-specific junk into the zip file. Quite properly, the ingester said primly, “Your METS manifest doesn’t match your actual files. Go forth and fix it.” The solution to this little difficulty
the existing system is **robust against disruption**
3. reports from the front lines: building a commons is really, really hard - and can’t be done without sharing.
Open Access Content
“running code”
Creative Commons

This page is available in the following languages:
Afrikaans български Català Dansk Deutsch Ελληνικά English (CA) English (GB) English (US) Esperanto Castellano Castellano (AR) Español (CL) Castellano (CO) Español (Ecuador) Castellano (MX) Castellano (PE) Euskara Suomksi français français (CA) Galego hrsvaki Magyar Italiano 日本語 한국어 Macedonian Melayu Nederlands Norsk Sesotho sa Leboa polski Português română slovenski jezik srpski srpski (latinica) Sotho svenska 中文 華語 (台灣) isiZulu

---

Creative Commons

Attribution 3.0 Unported

---

You are free:

- to Share — to copy, distribute and transmit the work
- to Remix — to adapt the work

Under the following conditions:

Attribution. You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

- For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to this web page.
>1000 journals under CC
Scholar's Copyright Addendum Engine

The Scholar's Copyright Addendum Engine will help you generate a PDF form that you can attach to a journal publisher's copyright agreement to ensure that you retain certain rights.

(get started)

Description

Each addendum gives you non-exclusive rights to create derivative works from your Article and to reproduce, distribute, publicly perform, and publicly display your article in connection with your teaching, conference presentations, lectures, other scholarly works, and professional activities. However, they differ with respect to how soon you can make the final published version available and whether you can authorize others to re-use your work in various ways. Below is a summary of the available options.

Science Commons / SPARC Addendum

Access - Reuse:
You retain sufficient rights to grant to the reading public a Creative Commons Attribution Non Commercial license or similar license that allows the public to re-use or re-post your article so long as you are given credit as the author and so long as the reader's use is non-commercial. (This is a joint offering from Science Commons and SPARC and represents a new version of the former SPARC Addendum.)

Other Options From Science Commons

Immediate Access:
You retain sufficient rights to post a copy of the published version of your article (usually in pdf form) online immediately to a site that does not charge for access to the article. (This is similar in many ways to the MIT Copyright Amendment below)

Delayed Access:
You also have the right immediately to post your final version of the article, as edited after peer review, to a site that does not charge for access to the article, but you must arrange not to make
Scholars Copyright Integration

This document describes integrating the Scholars Copyright Addendum Generator into your website. For general information on the Scholars Copyright project, see the project page at Science Commons.

Basic Integration

The generator can be integrated with your site using a link or an <iframe> tag. For example:

<iframe src="http://scholars.sciencecommons.org/iframe?partner_id=example"></iframe>

The value supplied for partner_id should be a short, descriptive name for your site. For example, Big Example University might supply partner_id=example_u.

Default Agreement Selection

You may specify the default agreement selection by providing a default parameter in the query string. The value of the default parameter is the agreement identifier. For example, to display the generator with the Access-Reuse agreement selected:

<iframe src="http://scholars.sciencecommons.org/iframe?partner_id=example&default=retaincc"></iframe>

The following agreements are available:

- **noembargo** Immediate Access
- **embargo** Delayed Access
- **retaincc** Access-Reuse
- **mit** MIT Amendment
running policy code
(w. SPARC)

OPEN DOORS AND OPEN MINDS:
What faculty authors can do to ensure open access to their work through their institution

COMPLYING WITH THE NATIONAL INSTITUTES OF HEALTH PUBLIC ACCESS POLICY:
Copyright considerations and options
<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Protocol for Implementing Open Access Data

Status of this Memo

This memo provides information for the internet community interested in distributing data or databases under an "open access" structure. There are several definitions of "open" and "open access" on the Internet, including the Open Knowledge Definition and the Budapest Declaration on Open Access; the protocol laid out herein is intended to conform to the Open Knowledge Definition and extend the ideas of the Budapest Declaration to data and databases.

This memo does not specify an Internet standard of any kind, but does specify the requirements for gaining and using the Science Commons Open Access Data Mark and metadata, by using legal tools and norms that conform to the protocol specified. This memo is available under the Creative Commons Attribution 3.0 (unported jurisdiction) license and will be submitted to the World Wide Web Consortium for consideration.

The terms MUST, MUST NOT, and SHOULD are used herein as defined in RFC 2119 ("Key words for use in RFCs to Indicate Requirement Levels").

1. Intellectual foundation for the protocol

The motivation behind this memorandum is interoperability of scientific data.

The volume of scientific data, and the interconnectedness of the systems under study, makes integration of data a necessity. For example, life scientists must integrate data from across biology and chemistry to comprehend disease and discover cures, and climate change scientists must integrate data from wildly diverse disciplines to understand our current state and predict the impact of new policies.

The technical challenge of such integration is significant, although emerging technologies appear to be helping. But the forest of terms and conditions around data make integration difficult to legally perform in many cases. One approach might be to develop and recommend a single license: any data with this license can be integrated with any other data under this license.

But this approach, which implicitly builds on intellectual property rights and the ideas of licensing as understood in software and culture, is difficult to scale for scientific uses. There are too many databases under too many terms already, and it is unlikely that any one license or suite of licenses will have the correct mix of terms to gain critical mass and allow massive-scale machine integration of data.

Therefore we instead lay out principles for open access data and a protocol for implementing those principles, and we distribute an Open Access Data Mark and metadata for use on databases and data available under a successful implementation of the protocol.

a protocol, not a license
conflicts with the protection instinct
conflicts with the protection instinct

the protection instinct is sometimes an instinct to protect "freedom"
solves the legal problem
but not the container problem.
building a web for data: the “semantic web”
making computers understand links between documents
drinking coffee causes feel awake

making computers understand relationships between concepts
use the web to integrate information from different places and different names

http://ontology.foo.org/coffee
(too much work for coffee)
(radical sharing starts to look pretty good)
web 2.0, science 3.0, what about making Google work better?
over 200 years at one paper/day
what you want is a *list of genes*.

not a list of *documents*.
Open Source
Data Integration
a repository of ontologies, namespaces, and integrated databases.
DRD1, 1812 adenylate cyclase activation
ADRB2, 154 adenylate cyclase activation
ADRB2, 154 arrestin mediated desensitization of G-protein coupled receptor protein signaling pathway
DRD1IP, 50632 dopamine receptor signaling pathway
DRD1, 1812 dopamine receptor, adenylate cyclase activating pathway
DRD2, 1813 dopamine receptor, adenylate cyclase inhibiting pathway
GRM7, 2917 G-protein coupled receptor protein signaling pathway
GNG3, 2785 G-protein coupled receptor protein signaling pathway
GNG12, 55970 G-protein coupled receptor protein signaling pathway
DRD2, 1813 G-protein coupled receptor protein signaling pathway
ADRB2, 154 G-protein coupled receptor protein signaling pathway
CALM3, 808 G-protein coupled receptor protein signaling pathway
HTR2A, 3356 G-protein coupled receptor protein signaling pathway
DRD1, 1812 G-protein signaling, coupled to cyclic nucleotide second messenger
SSTR5, 6755 G-protein signaling, coupled to cyclic nucleotide second messenger
MTNR1A, 4543 G-protein signaling, coupled to cyclic nucleotide second messenger
CNR2, 1269 G-protein signaling, coupled to cyclic nucleotide second messenger
HTR6, 3362 G-protein signaling, coupled to cyclic nucleotide second messenger
GRK2, 2898 glutamate signaling pathway
GRIN1, 2902 glutamate signaling pathway
GRIN2A, 2903 glutamate signaling pathway
GRIN2B, 2904 glutamate signaling pathway
ADAM10, 102 integrin-mediated signaling pathway
GRM7, 2917 negative regulation of adenylate cyclase activity
LRP1, 4035 negative regulation of Wnt receptor signaling pathway
ADAM10, 102 Notch receptor processing
ASCL1, 429 Notch signaling pathway
HTR2A, 3356 serotonin receptor signaling pathway
ADRB2, 154 transmembrane receptor protein tyrosine kinase activation (dimerization)
PTPRG, 5793 transmembrane receptor protein tyrosine kinase signaling pathway
EPHA4, 2043 transmembrane receptor protein tyrosine kinase signaling pathway
NRTN, 4902 transmembrane receptor protein tyrosine kinase signaling pathway
CTNND1, 1500 Wnt receptor signaling pathway
### Bundles

The Neurocommons RDF distribution is organized into modules or "bundles". Following is list of what's provided. Each has its own page of documentation.

<table>
<thead>
<tr>
<th>Bundle</th>
<th>Description</th>
<th>Documentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Derived from MeSH:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mesh/mesh-skos</td>
<td>MeSH hierarchy represented using SKOS</td>
<td>/mesh/mesh-skos</td>
</tr>
<tr>
<td>mesh/qualified-headings</td>
<td>MeSH qualified headings - defines one URI for each valid major/minor heading combination</td>
<td>/mesh/qualified-headings</td>
</tr>
<tr>
<td>Derived from Medline:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>medline/subject-headings</td>
<td>Medline: NLM MeSH subject headings for all articles</td>
<td>/medline/subject-headings</td>
</tr>
<tr>
<td>medline/titles-years</td>
<td>Medline: title and year of publication for each article</td>
<td>/medline/titles-years</td>
</tr>
<tr>
<td>Ontologies:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>obo/all</td>
<td>All OBO ontologies</td>
<td>/obo/all</td>
</tr>
<tr>
<td>galen</td>
<td>Galen ontology</td>
<td>/gelden</td>
</tr>
<tr>
<td>skos</td>
<td>W3C SKOS (Simple Knowledge Organization System) ontology</td>
<td>/skos</td>
</tr>
<tr>
<td>mesh-eswc06</td>
<td>Supporting ontology for conversion of MeSH hierarchy to RDF</td>
<td>/mesh-eswc06</td>
</tr>
<tr>
<td>sciencecommons</td>
<td>Ad hoc Science Commons ontology</td>
<td>/sciencecommons</td>
</tr>
<tr>
<td>bams</td>
<td>BAMS (Brain Architecture Management System)</td>
<td>/bams</td>
</tr>
<tr>
<td>senselab</td>
<td>Senselab</td>
<td>/senselab</td>
</tr>
<tr>
<td>nci-thesaurus</td>
<td>NCI thesaurus</td>
<td>/nci-thesaurus</td>
</tr>
<tr>
<td>Derived from NCBI:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ncbi/goa</td>
<td>NCBI Gene Ontology annotations</td>
<td>/ncbi/goa</td>
</tr>
<tr>
<td>ncbi/homologene</td>
<td>NCBI Homologene selection</td>
<td>/ncbi/homologene</td>
</tr>
<tr>
<td>ncbi/gene-info</td>
<td>NCBI Gene gene synonyms extraction</td>
<td>/ncbi/gene-info</td>
</tr>
<tr>
<td>ncbi/gene-pubmed</td>
<td>Links from NCBI Gene to Medline</td>
<td>/ncbi/gene-pubmed</td>
</tr>
<tr>
<td>Other:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>addgene</td>
<td>Addgene plasmid catalog</td>
<td>/addgene</td>
</tr>
<tr>
<td>neurocommons-text</td>
<td>Neurocommons text processing pilot</td>
<td>/neurocommons-text</td>
</tr>
<tr>
<td>aba</td>
<td>ABA (Allen Brain Atlas)</td>
<td>/aba</td>
</tr>
</tbody>
</table>
we can transform complex queries into links
we can transform complex queries into links
we can transform complex queries into links

Show me all signal transduction genes on the cell surface in pyramidal neurons
we can help scholars “remix” queries

prefix go: <http://purl.org/obo/owl/GO#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix mesh: <http://purl.org/commons/record/mesh/>
prefix sc: <http://purl.org/science/owl/sciencecommons/>
prefix ro: <http://www.obofoundry.org/ro/ro.owl#>

select ?genename ?processname
where
{ graph <http://purl.org/commons/hcls/pubmesh>
    ?article sc:identified_by_pmid ?paper.
    ?gene sc:describes_gene_or_gene_product_mentioned_by ?article.
  }

  graph <http://purl.org/commons/hcls/goa>
    ?res owl:onProperty ro:has_function.
    ?res2 owl:onProperty ro:realized_as.
    ?res2 owl:someValuesFrom ?process.
  }

  graph <http://purl.org/commons/hcls/20070416/classrelations>
  { ?process rdfs:subClassOf <http://purl.org/obo/owl/obo#part_of> go:GO_0006610
    union
    { ?process rdfs:subClassOf go:GO_0006610 .
      ?res3 owl:hasValue ?gene.
    }

  graph <http://purl.org/commons/hcls/gene>
  { ?gene rdfs:label ?genename }
  graph <http://purl.org/commons/hcls/20070416>
  { ?process rdfs:label ?processname }"
we can build a corpus of queries as links

Show me all signal transduction genes on the cell surface in pyramidal neurons

Show me all ribosomal protein-coding genes in cancer development
we can re-use cultural tools for scholarship
4. radical sharing is a smart choice.
simple + open = WIN
World Wide Web

The WorldWideWeb (W3) is a wide-area hypermedia information retrieval initiative aiming to give universal access to a large universe of documents.

Everything there is online about W3 is linked directly or indirectly to this document, including an executive summary of the project, Mailing lists, Policy, Asked Questions.

What's out there?
Points to the world's online information, subjects, W3 servers, etc.

Help
on the browser you are using

Software Products
A list of W3 project components and their current state. (e.g. Line Mode, X11, Viola, NeXTStep, Servers, Tools, Mail robot, Library)

Technical
Details of protocols, formats, program internals etc

Bibliography
Paper documentation on W3 and references.

People
A list of some people involved in the project.

History
A summary of the history of the project.

How can I help?
If you would like to support the web.

Getting code
Getting the code by anonymous FTP, etc.
content

code

physical
knowledge
content
code
physical
open copyright, balanced incentives, and distributed workloads
the infrastructure for this is very, very shaky.

Show me all signal transduction genes on the cell surface in pyramidal neurons.
prefix dc: <http://purl.org/dc/elements/1.1/>
prefix skos: <http://www.w3.org/2004/02/skos/core#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix sc: <http://purl.org/science/OWL/sciencecommons/>
prefix foaf: <http://xmlns.com/foaf/0.1/>
what are the odds that the organizations making the namespaces will be here in 50 years? 100 years?
and what about ontologies?
extension (quality control: spam and junk)

remix (brand confusion, loss of integrity and attribution)

formats (failure to adhere to common protocols or technology)

persistence (the transient nature of all Web things...
Digital curation

From Wikipedia, the free encyclopedia

Digital curation is the curation, preservation, maintenance, and collection and archiving of digital assets[1][2].

Digital curation is the process of establishing and developing long term repositories of digital assets for current and future reference by scientists, and historians, and scholars generally.

Aspects of digital curation

Digital curation entails:
- Collecting verifiable digital assets
- Providing digital asset search and retrieval
- Certification of the trustworthiness and integrity of the collection content
- Semantic and ontological continuity and comparability of the collection content

Challenges faced by digital curation

Significant[3] and major challenges faced by digital curation are:
- Storage format evolution and obsolescence[4]
- Rate of creation of new data and data sets
- Broad access and searching flexibility and variety
- Comparability of semantic and ontological definitions of data sets[4]
“it’s complicated”
“In any case, it is clear that a library containing all possible books, arranged at random, is equivalent (as a source of information) to a library containing zero books.”

exponential content growth
our brain capacity

- 1990
- 1994
- 1998
- 2002
but if we can work together...
conclusion?
don’t wait.
Investigation Site

Class: http://purl.obofoundry.org/obo/OBI_0000225

(and role
(disjoint-with nutrient-role study-personnel-role patient-role regulatory-role drug-role study-participant-role vector-role reference-role))

definition: Investigation site is a role borne by a site realized in an investigation which is located at the site

curation status: metadata-incomplete

preferred term: Investigation site

eexample of usage: A field, a laboratory, a medical institute, a pharmaceutical company

definition source: source pending

editor note: solution2: site is related to trial used located_in relation – site can bear the role

editor note: solution1: is a physical location, should maybe go under processual context, and then be used in conjunction with the located relation

definition editor: Jennifer Fostel

Subject of: location_of, participates_in, is_proxy_for, proper_part_of, derives_from, relationship, transformed_into, has_proper_part, is_output_of, is_realized_as, part_of, has_integral_part, derived_into, is_input_of, ObsoleteProperty, agent_in, has_part, integral_part_of, contains, transformation_of, has_improper_part, improper_part_of, located_in, contained_in, adjacent_to

Object of: location_of, is_proxy_for, proper_part_of, derives_from, relationship, transformed_into, has_proper_part, part_of, has_participant, has_integral_part, derived_into, has_output, ObsoleteProperty, has_part, has_input, integral_part_of, contains, transformation_of, is_realization_of, has_improper_part, improper_part_of, located_in, contained_in, has_role, adjacent_to, has_agent

use existing systems.
Article Authoring Add-in v1.0 for Microsoft Office Word 2007

The Article Authoring Add-in enables authors and editors to open and save Microsoft Office Word files in the National Library of Medicine's NLM XML format, a file format that is used in the publishing and archiving of scientific and technical articles. Beyond its core file format capabilities, the add-in enables additional metadata to be captured at the authoring stage and enables semantic information to be preserved through the publishing process, which is essential for enabling search and semantic analysis once the articles are archived at information repositories. The add-in also aims at simplifying the authoring, submission, and interaction process between authors and journals.

- Download the Article Authoring Add-in for Microsoft Office Word 2007
- Read about this project in Pablo Fernicola's MSDN blog
  Pablo Fernicola is a Group Manager in Live Labs who focuses on varying aspects of work related to scientific and scholarly communication, publishing, and knowledge dissemination.
- Watch a video to see the add-in at work (youtube.com)

Creative Commons Add-in v1.0 for Microsoft Office

This add-in for Microsoft Office Word 2007, Office PowerPoint 2007, and Office Excel 2007 enables individuals to embed a Creative Commons license directly into their Microsoft Office documents. The add-in allows an author of a Microsoft Office document to choose a Creative Commons license from those available on the Creative Commons Web site (by using the Creative Commons Web service). The embedded license links directly to its online representation on the Creative Commons Web site while a machine-readable representation is stored in the Office Open XML document. By using Creative Commons licenses, you can express your intentions regarding how others may use your work.

- Download the Creative Commons Add-in
- Learn more about Creative Commons (creativecommons.org)
- Learn more about the choices among the Creative Commons Licenses (creativecommons.org)

enlist everyone to curate.
create new ways to measure.
invest in curation.
free as in speech
free as in speech
free as in beer
free as in speech
free as in beer
free as in a puppy
Average Cost Of 100 Pound Dog Over A Year

Good Quality Dog Food
$70 x 12 = $840
Dog Accessories (collar, leash, etc.)
$30
Dog Toys
$30 - $50
Vaccines
$35
Flea, Tick, & Heartworm Prevention
$320
Dog Treats
$200
Boarding
$100 - $200 (at $15 - $20 a day)
Emergency Costs
$0 - $2500 or more
Total
$1375 or much more
thank you

wilbanks@creativecommons.org

http://sciencecommons.org