Scientific publications on Web 3.0

SCIENTIFIC PUBLICATIONS ON WEB 3.0

Sudeshna Das[1,3]*; Mark Goetz* [1]*; Lisa Girard*[2]*, Tim Clark*[3,4]*.*

[1]Initiative in Innovative Computing, Harvard University
60 Oxford Street, Cambridge, MA 02138, USA
e-mail: sudeshna_das@harvard.edu; goetzma@umich.edu
[2]Harvard Stem Cell Institute, Harvard University
e-mail: lisa_girard@harvard.edu
44 Church Street, Cambridge MA 02138, USA
[3]MIND, Massachusetts General Hospital
Building 114, 16th Street, Charlestown, MA 02129, USA
[4]Harvard Medical School, 25 Shattuck Street, Boston, MA 02115
e-mail: tim_clark@harvard.edu

Abstract

The advent of new technologies and paradigms is constantly changing the landscape of scientific publications. The use of online journals is rapidly rising and most researchers prefer online materials to print. The Internet has also given rise to open-access, online-only publications. There are several advantages to such journals - most importantly, the articles published can become a starting point for online community-based discourse on the subject. Researchers, given the right web environment, can discuss the articles published online, and such collaboration on the World Wide Web is the hallmark of Web 2.0. Another emergent trend is Web 3.0, in which the web becomes the medium for data, information and knowledge exchange through the use of shared semantics.

We have developed the Science Collaboration Framework (SCF), a lightweight software framework that scientific communities can use to create open-access, online, scientific publications. The software uses Web 3.0 technologies (social web, semantic web, text-mining) and thus allows interoperability with other Web 3.0 sites. The software allows communities to publish complex scientific articles, annotate them with controlled vocabularies or ontologies, register research interests of members and conduct discussion forums. The software can integrate with other knowledge repositories and the site knowledge is available as linked data. The software is modular, so different communities can install and enable different features as well as contribute modules back to the main framework; thus creating a software community as well.

The first site based on our software, *StemBook* (www.stembook.org), an online open access peer-reviewed collection of invited review chapters covering a range of topics related to stem cell biology, went "live" in September 2008. Several other sites are under development, including a new web community for Parkinson's disease researchers, *PD Online*, and a re-engineered version of the popular Alzheimer Disease research community *Alzforum* (www.alzforum.org). The sites developed on the SCF platform are interoperable with each other and with other sites on the Semantic Web. In this new paradigm, there is a significant reduction in artificial barriers between research disciplines, and a much more dynamic and agile approach to information exchange.

Keywords: semantic web; electronic publishing; open access; text mining.

1. Introduction

Advances in electronic publishing are changing the world of scientific publications. The use of print journals is declining; biomedical scientists and scholars generally prefer online materials to print; and use the Internet to search, access and read relevant literature [1]. Browsing the university library collections is rapidly becoming an uncommon activity – the university library these days is far more the realm of librarians than of researchers. Scientists increasingly use hyperlinks to view referenced or related materials and are thus changing how science is being conducted [2].

Another important trend has been an emergence of a parallel world of open access, online publications. Notable examples of such online journals are PLoS ONE [3], WormBook [4], Nature Reports Stem Cells [5] and the journals published by the open access publisher Biomed Central [6]. These publications have a peer review process and a Digital Object Identifier (DOI) available for citations, just as in traditional journals. PLoS ONE has a streamlined electronic production workflow coupled with a rigorous peer-review process. The articles are available for commenting and debate. It also has an inclusive scope, allowing the research to reach a large audience. WormBook is a comprehensive collection of peer-reviewed articles on topics related to the biology of Caenorhabditis elegans (C. elegans) and up-to-date descriptions of technical procedures used to study this animal. Nature Reports Stem Cells provides the readers with the latest news and science behind stem cell research and tries to reach a large audience of interdisciplinary researchers. Biomed Central is a publisher of research articles in science, technology and medicine and has pioneered the open access publishing model.

Online publications offer numerous advantages that are not available to a print journal – (i) the review and publication process is much streamlined allowing for a quick publication process, (ii) readers can download tools and data from the website and (iii) the format allows multimedia elements to be embedded in the article, (iv) the articles can be kept up-to-date, and finally (iv) the conversation does not end with the article – researchers can participate in web-based discourse on the subject matter of the article, making it a live discussion topic. Even though there are several advantages of online publications, it is likely that information exchange will continue through both the online and print media in the near future [7].

Advances in Internet technologies have also changed the nature of online publications. Web 2.0, a term made popular by Tim O' Reilly, refers to set of practices, design and technology that aims to facilitate interaction, collaboration, information sharing and communication on the World Wide Web [8]. It includes technologies for blogging, forums, tags, linking and so on. Web 2.0 concepts have led to the development and evolution of web-based social communities. Highly popular examples are *FaceBook* [9] & *My Space* [10].

We also see Web 2.0 in various forms in the world of scientific publications and broadly scientific communities. Examples of scientific social networking include *Nature Network* [11] and *SciLink* [12]. These sites allow researchers to find, connect and share information with each other. *SciLink* also aims to link research scientists to science teachers, has received positive feedback from the teachers and has helped in the advancement of science education [13]. Another category of Web 2.0 sites includes social bookmaking sites such as *Connotea* [14] developed by the Nature publishing Group and *CiteULike* [15]. These sites help researchers to search, organize and annotate scholarly publications. Some of these tools also make the structured meta-data of the citation available to the user. Connotea is becoming popular and its features and usage are described in detail by Lund et al [16]. Wikis [17] enable collaborative writing and discussions and good examples in the biomedical community include *Open Wetware* [18] & *WikiProtein*s [19]. Finally, there is an emergent class of Web 2.0 moderated social scientific communities (moderated SSCs). The most notable example is *Alzforum* [20,21]. *Alzforum* is a community of over five thousand registered Alzheimer's researchers who are collaborating on the Word Wide Web to find a cure for Alzheimer's disease. In Alzforum, researchers can discuss papers and news spontaneously and participate in live discussions. Researchers are also invited to provide perspectives on key research news and comment on papers of the week. Compendia of genes, antibodies, animal models and protocols are also available on the site [20,21].

Until recently there has been no reusable software to develop a scientific community such as Alzforum, while at the same time there was a clear need to "clone" the successful Alzforum model for use by other groups of disease researchers. Thus, we were inspired both by Alzforum, and by the needs of geographically distributed interfaculty initiatives at our University, to develop a reusable framework for building online scientific communities [22]. At the same time we wanted to leverage emergent Web 3.0 technologies [23] to enable interoperability between scientific communities and other knowledge repositories.

Web 3.0, also known as the Semantic Web, could be defined as Web 2.0 with linked data and shared meaning. It incorporates a number of new standardized web technologies based on the vision of Semantic Web [23, 24]. A simple way to think about Web 3.0 is this: your web page now has a schema, just like a database. The schema is a graph model and it can be published and shared. That is, web pages can become web services and/or data repositories. In the Semantic Web, the semantics of data and services are well defined and information is expressed in a machine-interpretable format and thus can be processed by software agents. The Semantic Web allows interoperability between heterogeneous sites and thus information and knowledge exchange can be carried out successfully using the Web as a medium. Information is available as RDF (Resource Description Framework) [25] "triples", which allows the expression of relationships between objects and their attributes and between two objects. The meaning of these triples is shared by being declared in a web-standard schema language, RDF Schema (RDFS) [26], or ontology language, Web Ontology Language (OWL) [27]. All resources on Web 3.0 are identified with a URI (Uniform resource Identifier). Another emergent Semantic Web paradigm is Linked Data – it refers to a method of exposing, sharing, and connecting data via dereferenceable URIs on the Web [28]. In Linked Data, HTTP URIs are used to locate and dereference objects, structured useful information is available when the resource is dereferenced and links to other resources are available as URIs.

The Semantic Web and Linked Data have allowed for interoperability and knowledge exchange on the Web. However, these are infrastructural features that must be packaged in applications oriented to specific kinds of end-users. Thus, in the case of scientific publications on the web, tools to aid in the annotation process that incorporate the ability to produce and consume linked data are necessary for the transition to Web 3.0. Semantic annotation tools facilitate the process of annotating and linking web resources with other resources or controlled vocabularies and ontologies. There have been several efforts to annotate web resources with controlled vocabularies or ontologies [29, 30, 31]. Shah et al [29], developed a system for ontology based annotation and indexing of biomedical data that processed the text metadata of diverse biomedical resource and annotated them with terms from medical ontologies. Using this system, researchers are able to retrieve biomedical data resources related to particular ontology concepts.

Although text mining is an active area of research; the precision of searches (percentage of retrieved elements that are relevant to the search) is usually poor.

Comparative studies have shown that precision improves with the addition of contextual elements. In a study by Moskovitch et al, precision is between 31% to 53% (at 50% recall) when semi-structured clinical guideline documents are retrieved from a digital library [32]. To improve precision and recall simultaneously, we use a hybrid approach. We use automated text-mining to locate all relevant resources, and thus obtain high recall. The output is then manually curated to obtain high precision. Such semi-automated methods leverage the ability of machines to process large amounts of information and at the same time take advantage of the high quality, expert information processing. Such semi-automated text-mining efforts have been previously used in quite different contexts [33, 34]. We have adapted and enhanced them for our use case.

Thus, we have developed a software system, the Science Collaboration Framework (SCF) that combines the powerful approaches of social web, Semantic Web (Web 3.0) and semi-automated text-mining. SCF is a lightweight software framework that scientific communities can use to create open-access, online, scientific publications. The software enables researchers to author complex scientific articles, annotate them with biomedical resources or ontologies and conduct discourse and debate around the topics of the articles. The software allows researchers to publish research interests and expand biomedical resources. The first site based on our software, an online journal of stem cell research publications, published by the Harvard Stem Cell Institute (www.stembook.org), went "live" in September 2008. Several other sites are under development, including a new web community for Parkinson's disease researchers, PD Online [35], and a re-engineered version of the popular Alzheimer disease research community, Alzforum (http://www.alzforum.org) [20,21]. The knowledge represented within the site is available as RDF and thus allows interoperability between various interdisciplinary communities.

2. Software

Our software, the Science Collaboration Framework (SCF) is freely available under the GNU public license (www.sciencecollaboration.org) and is based on the popular open source content management system, Drupal [36]. Drupal is a content management system based on the programming language PHP and a relational database (MySQL is most commonly used). Drupal has thousands of registered developers and thus has a large and growing number of custom modules. Thus, developing SCF on such a richly featured and dynamic content management system allows us to leverage the work of a large number of software developers.

SCF allows scientific communities to publish research articles, reviews, interviews, news items; engage in discussion forums; create member profiles; register research interests and integrate with linked data on the web [22]. The software has a modular architecture – different communities can install a different set of modules based on their individual requirements. Thus, we are able to satisfy the needs of diverse scientific communities with a single framework. The software is geared towards life science researchers and makes common biomedical resources such as genes, antibodies and model organisms available to the researcher for annotation and comments. The modular architecture and integration with linked data is described in detail in Das et al [22]. We have further developed the software and added numerous features. The most significant enhancement is the development of tools for semi-automated annotation of content with controlled vocabularies and ontologies. The software enhancements are described in the following sections.

2.1. Authoring Scientific Articles

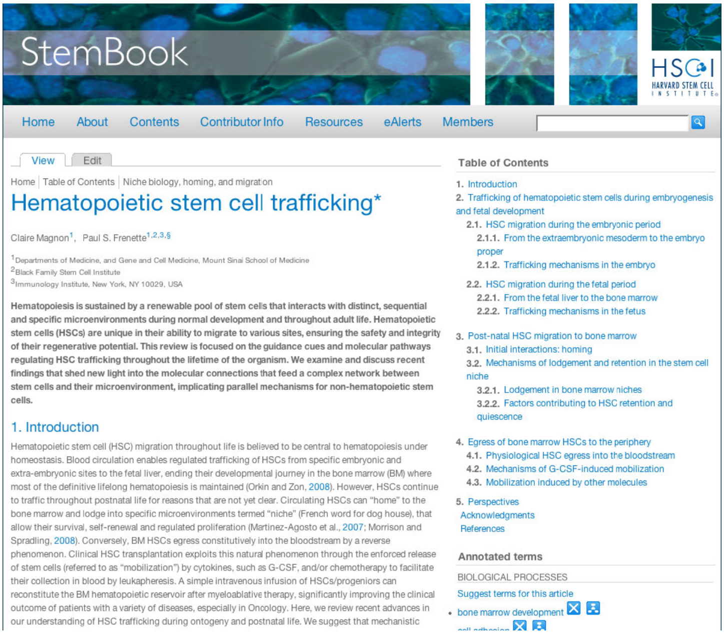
The most important component of a scientific publication is the ability to author scientific articles. We were faced with the dual need of the ability to publish complex articles with hyperlinks and embedded multimedia elements as well as quick "blog" style articles authored spontaneously on the Web. We took a two-prong approach to address these diverse needs as described in the next two sections.

2.1.1. Complex Articles

We needed the ability to render complex HTML (Hypertext Markup Language) of the scientific abilities and simultaneously minimize production costs and time. We chose to format the articles with XML (Extensible Markup Language) and adopted the National Library of Medicine (NLM) DTD (Document Type Definition) for this purpose. We chose the NLM DTD because our audience is primarily biomedical communities and the format is widely used in the community (such as the preeminent open access publisher - Public Library of Science, PLoS). The NLM DTD supports complex structures within the article and also links to external resources.

The article is authored by the scientists using a rich text editor of their choice. Then the document is converted to an XML format by an external content transformation service provider. The XML of the article is uploaded into the site by the editors and processed using XSLT (Extensible Stylesheet Language Transformation) to create XHTML and PDF files. The creation of XML documents is a manual process and is the primary bottleneck in the production process. In the future, we would like to adopt the Open Document Format (ODF) [37] that is currently supported by a number of word processors (e.g. Google Docs, Open Office). A project to develop an ODF plugin for Microsoft Word is also underway.

A



Home | Table of Contents | Niche biology, homing, and migration

Hematopoietic stem cell trafficking*

Claire Magnon[1], Paul S. Frenette[1,2,3,§]

[1]Departments of Medicine, and Gene and Cell Medicine, Mount Sinai School of Medicine
[2]Black Family Stem Cell Institute
[3]Immunology Institute, New York, NY 10029, USA

Hematopoiesis is sustained by a renewable pool of stem cells that interacts with distinct, sequential and specific microenvironments during normal development and throughout adult life. Hematopoietic stem cells (HSCs) are unique in their ability to migrate to various sites, ensuring the safety and integrity of their regenerative potential. This review is focused on the guidance cues and molecular pathways regulating HSC trafficking throughout the lifetime of the organism. We examine and discuss recent findings that shed new light into the molecular connections that feed a complex network between stem cells and their microenvironment, implicating parallel mechanisms for non-hematopoietic stem cells.

1. Introduction

Hematopoietic stem cell (HSC) migration throughout life is believed to be central to hematopoiesis under homeostasis. Blood circulation enables regulated trafficking of HSCs from specific embryonic and extra-embryonic sites to the fetal liver, ending their developmental journey in the bone marrow (BM) where most of the definitive lifelong hematopoiesis is maintained (Orkin and Zon, 2008). However, HSCs continue to traffic throughout postnatal life for reasons that are not yet clear. Circulating HSCs can "home" to the bone marrow and lodge into specific microenvironments termed "niche" (French word for dog house), that allow their survival, self-renewal and regulated proliferation (Martinez-Agosto et al., 2007; Morrison and Spradling, 2008). Conversely, BM HSCs egress constitutively into the bloodstream by a reverse phenomenon. Clinical HSC transplantation exploits this natural phenomenon through the enforced release of stem cells (referred to as "mobilization") by cytokines, such as G-CSF, and/or chemotherapy to facilitate their collection in blood by leukapheresis. A simple intravenous infusion of HSCs/progeniors can reconstitute the BM hematopoietic reservoir after myeloablative therapy, significantly improving the clinical outcome of patients with a variety of diseases, especially in Oncology. Here, we review recent advances in our understanding of HSC trafficking during ontogeny and postnatal life. We suggest that mechanistic

Table of Contents

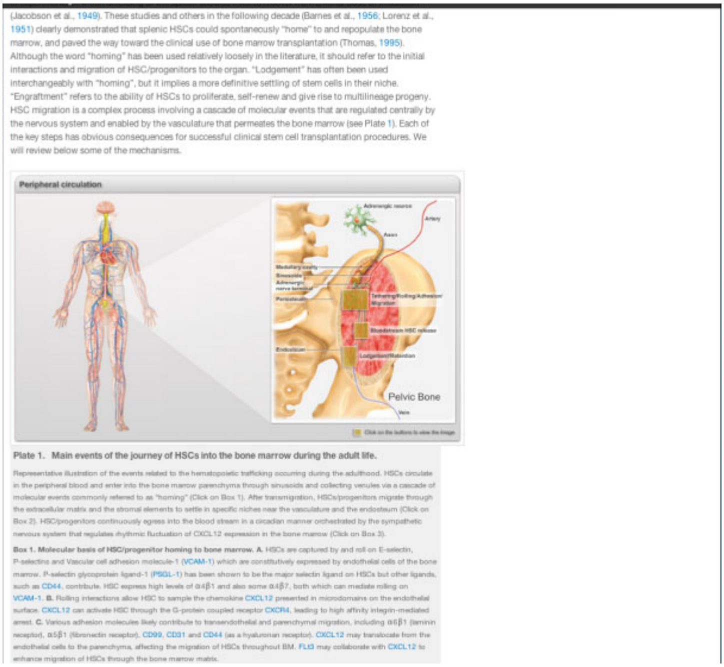
Annotated terms

BIOLOGICAL PROCESSES

Suggest terms for this article

- bone marrow development ☒
- cell adhesion ☒

B



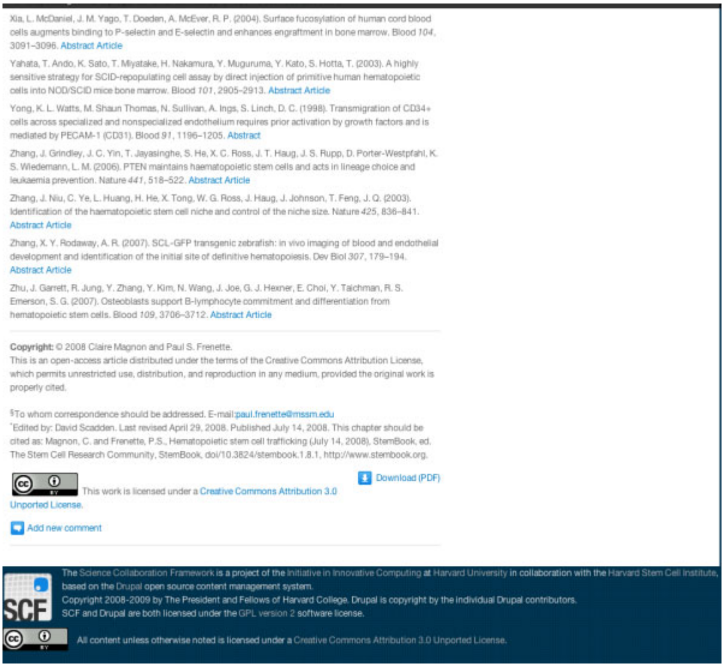
(Jacobson et al., 1949). These studies and others in the following decade (Barnes et al., 1956; Lorenz et al., 1951) clearly demonstrated that splenic HSCs could spontaneously "home" to and repopulate the bone marrow, and paved the way toward the clinical use of bone marrow transplantation (Thomas, 1995). Although the word "homing" has been used relatively loosely in the literature, it should refer to the initial interactions and migration of HSC/progenitors to the organ. "Lodgement" has often been used interchangeably with "homing", but it implies a more definitive settling of stem cells in their niche. "Engraftment" refers to the ability of HSCs to proliferate, self-renew and give rise to multilineage progeny. HSC migration is a complex process involving a cascade of molecular events that are regulated centrally by the nervous system and enabled by the vasculature that permeates the bone marrow (see Plate 1). Each of the key steps has obvious consequences for successful clinical stem cell transplantation procedures. We will review below some of the mechanisms.

Plate 1. Main events of the journey of HSCs into the bone marrow during the adult life.

Representative illustration of the events related to the hematopoietic trafficking occuring during the adulthood. HSCs circulate in the peripheral blood and enter into the bone marrow parenchyma through sinusoids and collecting venules via a cascade of molecular events commonly referred to as "homing" (Click on Box 1). After transmigration, HSCs/progenitors migrate through the extracellular matrix and the stromal elements to settle in specific niches near the vasculature and the endosteum (Click on Box 2). HSC/progenitors continuously egress into the blood stream in a circadian manner orchestrated by the sympathetic nervous system that regulates rhythmic fluctuation of CXCL12 expression in the bone marrow (Click on Box 3).

Box 1. Molecular basis of HSC/progenitor homing to bone marrow. A. HSCs are captured by and roll on E-selectin, P-selectins and Vascular cell adhesion molecule-1 (VCAM-1) which are constitutively expressed by endothelial cells of the bone marrow. P-selectin glycoprotein ligand-1 (PSGL-1) has been shown to be the major selectin ligand on HSCs but other ligands, such as CD44, contribute. HSC express high levels of α4β1 and also some α4β7, both which can mediate rolling on VCAM-1. **B.** Rolling interactions allow HSC to sample the chemokine CXCL12 presented in microdomains on the endothelial surface. CXCL12 can activate HSC through the G-protein coupled receptor CXCR4, leading to high affinity integrin-mediated arrest. **C.** Various adhesion molecules likely contribute to transendothelial and parenchymal migration, including α6β1 (laminin receptor), α5β1 (fibronectin receptor), CD99, CD31 and CD44 (as a hyaluronan receptor). CXCL12 may translocate from the endothelial cells to the parenchyma, affecting the migration of HSCs throughout BM. Flt3 may collaborate with CXCL12 to enhance migration of HSCs through the bone marrow matrix.

C



Xia, L. McDaniel, J. M. Yago, T. Doeden, A. McEver, R. P. (2004). Surface fucosylation of human cord blood cells augments binding to P-selectin and E-selectin and enhances engraftment in bone marrow. Blood *104*, 3091–3096. Abstract Article

Yahata, T. Ando, K. Sato, T. Miyatake, H. Nakamura, Y. Muguruma, Y. Kato, S. Hotta, T. (2003). A highly sensitive strategy for SCID-repopulating cell assay by direct injection of primitive human hematopoietic cells into NOD/SCID mice bone marrow. Blood *101*, 2905–2913. Abstract Article

Yong, K. L. Watts, M. Shaun Thomas, N. Sullivan, A. Ings, S. Linch, D. C. (1998). Transmigration of CD34+ cells across specialized and nonspecialized endothelium requires prior activation by growth factors and is mediated by PECAM-1 (CD31). Blood *91*, 1196–1205. Abstract

Zhang, J. Grindley, J. C. Yin, T. Jayasinghe, S. He, X. C. Ross, J. T. Haug, J. S. Rupp, D. Porter-Westpfahl, K. S. Wiedemann, L. M. (2006). PTEN maintains haematopoietic stem cells and acts in lineage choice and leukaemia prevention. Nature *441*, 518–522. Abstract Article

Zhang, J. Niu, C. Ye, L. Huang, H. He, X. Tong, W. G. Ross, J. Haug, J. Johnson, T. Feng, J. Q. (2003). Identification of the haematopoietic stem cell niche and control of the niche size. Nature *425*, 836–841. Abstract Article

Zhang, X. Y. Rodaway, A. R. (2007). SCL-GFP transgenic zebrafish: in vivo imaging of blood and endothelial development and identification of the initial site of definitive hematopoiesis. Dev Biol *307*, 179–194. Abstract Article

Zhu, J. Garrett, R. Jung, Y. Zhang, Y. Kim, N. Wang, J. Joe, G. J. Hexner, E. Choi, Y. Taichman, R. S. Emerson, S. G. (2007). Osteoblasts support B-lymphocyte commitment and differentiation from hematopoietic stem cells. Blood *109*, 3706–3712. Abstract Article

Copyright: © 2008 Claire Magnon and Paul S. Frenette.
This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

[§]To whom correspondence should be addressed. E-mail:paul.frenette@mssm.edu
[*]Edited by: David Scadden. Last revised April 29, 2008. Published July 14, 2008. This chapter should be cited as: Magnon, C. and Frenette, P.S., Hematopoietic stem cell trafficking (July 14, 2008), StemBook, ed. The Stem Cell Research Community, StemBook, doi/10.3824/stembook.1.8.1, http://www.stembook.org.

This work is licensed under a Creative Commons Attribution 3.0 Unported License.

The Science Collaboration Framework is a project of the Initiative in Innovative Computing at Harvard University in collaboration with the Harvard Stem Cell Institute, based on the Drupal open source content management system.
Copyright 2008-2009 by The President and Fellows of Harvard College. Drupal is copyright by the individual Drupal contributors.
SCF and Drupal are both licensed under the GPL version 2 software license.

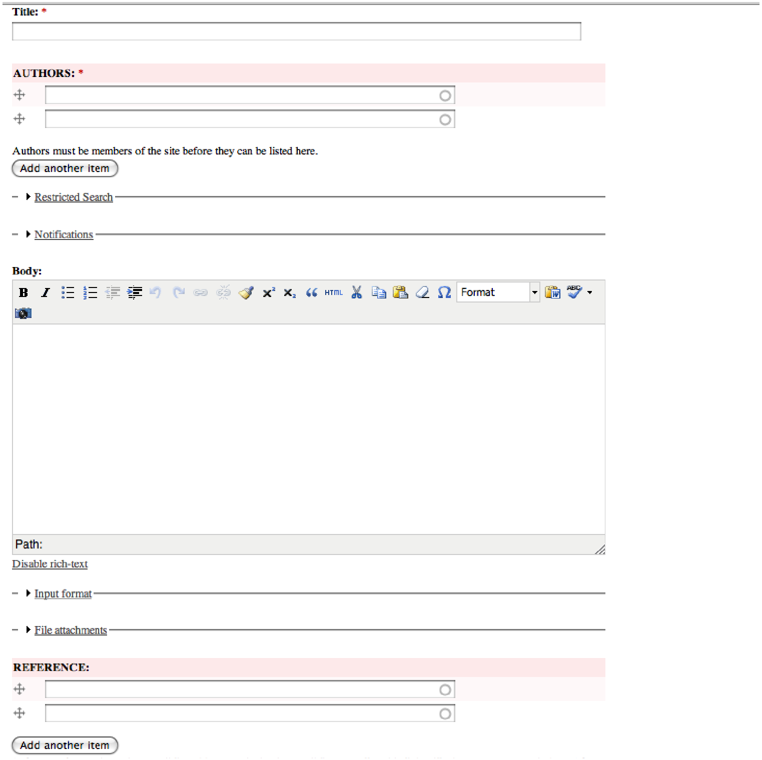
All content unless otherwise noted is licensed under a Creative Commons Attribution 3.0 Unported License.

Figure 1: Example article produced using the Science Collaboration Framework.
1A) Elaborate table of content;
1B) Figures and hyperlinks to genes;
1C) Copyright & DOI (Digital Object Identifier).

An example article authored using our software is shown in Figure 1. The software automatically produces an elaborate Table of Contents. The articles can contain complex figures, tables and embedded multimedia elements. The articles have digital object identifiers (DOI) and can be cited by other researchers. The articles are copyrighted under the Creative Commons attributions license.

2.1.2. Short Articles

Shorter articles are authored on the site using WISIWIG editors. Forms based interface is used to specify the title, authors and bibliographic references (explained in further detail in the next section). These are "blog" style writings but adapted to the scientific use case. They allow multiple authors and addition of bibliographic references. Moreover, the articles can be tagged using biomedical ontologies (Section 2.2). Figure 2 shows an example interface for authoring shorter articles.



Authors must be members of the site before they can be listed here.

Figure 2: Interface for authoring short articles.

2.1.3. Citations

One of the important components of a scientific article is its bibliographic record. We need to be able to represent each record as a separate entity as well as capture its relationship to the article, to mine the knowledge in the rich citations network. We chose to capture each record as a Drupal node and its relationship to the article node. Thus, we are able to capture and exploit the citation relationships and our representation is more explicit than a HTML page with a list of citations. We have also integrated searches from the United States National Library of Medicine pubmed service (http://www.ncbi.nlm.nih.gov/pubmed/), which provides access to almost all major biomedical articles dating back to 1948. Such explicit knowledge representation and user interaction is the essence of Web 3.0. The last step is to make this knowledge available in the language of the Semantic Web such as Resource Description Framework (RDF), and we are working towards that.

2.2. Semantic Annotation

Developing elaborate knowledge bases involves large, time consuming and expensive manual methods. While capturing biology of the articles in a semantic form is hard, huge enhancements in search and browsing capabilities can be gained with semantic annotation. Such annotation, while not providing answers to open biomedical problems, will enable researchers to find facts and resources that in turn will help them solve the biomedical problems.

2.2.1. Ontology Representation

The first step towards semantic annotation is the representation of common biomedical vocabularies/ontologies within the system. We have developed methods to load ontologies such as the Gene Ontology (GO) [38] using a standard format such as the Open Biomedical Ontology, OBO format [39]. The Gene Ontology project provides descriptions of genes and proteins using controlled vocabularies. Each term in GO has a unique identifier, a term name and term definition. Many terms have synonyms and relationships to other terms. The GO terms are assigned to one of the three ontologies, molecular function, cellular component or biological process. An example of GO record is shown in Table 1.

```
id: GO:0045739
name: positive regulation of DNA repair
namespace: biological_process
def: “Any process that activates or increases the frequency, rate or
  extent of DNA repair.” [GOC:go_curators]
subset: gosubset_prok
synonym: “activation of DNA repair” NARROW []
synonym: “stimulation of DNA repair” NARROW []
synonym: “up regulation of DNA repair” EXACT []
synonym: “up-regulation of DNA repair” EXACT []
synonym: “upregulation of DNA repair” EXACT []
is_a: GO:0006282 ! regulation of DNA repair
is_a: GO:0048584 ! positive regulation of response to stimulus
is_a: GO:0051054 ! positive regulation of DNA metabolic process

relationship: positively_regulates GO:0006281 ! DNA repair
```

Table 1: Example Gene Ontology Term in OBO format.

The next step will be to develop flexible methods to load any ontology in OWL (Web Ontology Language) [27] or OBO format and export ontologies authored on the site in OWL or OBO formats. Often vocabularies evolved in the community do not have a formal representation, and we are developing methods to use the Simple Knowledge Organization System, SKOS [40] to assert that a term in a community vocabulary is a conceptual resource and to capture the semantic relationships to other concepts.

2.2.2. Semi-automatic Text-mining

Once a vocabulary or ontology is represented in the framework, the software enables editors or authors of the published articles to annotate the text with terms from the vocabulary/ontology. Development of efficient text-mining tools is always a challenge and one has to balance the trade-off between precision and recall [34]. One strategy is to use a two-step process – use automated tools to come up with all possible suggestions for annotations (and thus obtain high recall) and then manually review the suggested terms to pick the appropriate ones (and thus obtain high precision). Such a strategy leverages both the capacity of machines to process large amounts of information and experts to provide high quality information.

We have adopted such a strategy in our software to provide semi-automatic annotations. Since there is always a manual review, we can in principle obtain 100% precision and recall, by design. Hence, it is not meaningful to measure the precision and recall of the system. The performance of the automated part of the text-mining system does determine the volume of tasks performed by the editor and should be tuned to achieve optimal output. In our case, we provide a simple automated text-mining tool that suggests all possible matches. And then, we primarily focused on developing a richly featured graphical ontology browser to assist the editor in dealing with the output of the text-mining tool. The annotations, meta-data of the articles and other knowledge on the website is available on the Semantic Web using the SWAN ontology [41, 42] and is described in further detail in Das et al [22]. Knowledge representation in Web 3.0 enables enhanced searches and integration with content from other similar sites and in the future we would like to develop such search and integration methods as part of our software.

2.2.2.1. Automated Text-Mining Tool

The software contains a rather simple text-mining engine that enables it to find whether a Gene Ontology (GO) term occurs within the text of a journal article, the total number of matches, and the full sentence in which the match occurs. The tool parses out the words in the sentence, stems the word (obtains the root word using the Porter stemming algorithm [43]) and counts the number of matches with the words in the names of the GO terms. If the number of matches exceeds a preset threshold, the term is suggested as a possible annotation. The steps of the algorithm are show in Table 2.

```
Split the article body into component sentences and headings
For each sentence and heading {
     If the exact term occurs within the sentence {
           register a match
     }
     Split the term definition and name into individual words and
        discard stopwords
     Split the sentence into individual words and discard stopwords
     Convert each remaining word in the term and sentence into stems
     If more than one-half of the words in the term occur in the
        sentence {
           register a match
     }
Return the list of all matching sentences
```

Table 2: Steps of text-mining algorithm

An example output of the algorithm is shown in Table 3 below. The tool identifies a match to the term "embryonic hemopoiesis", even though the exact name does not appear in the sentence.

Sentence	Researchers tracking down the cause of anemia in mutant zebrafish embryos have discovered a protein that guides the formation of new blood cells.

Term	`id: GO:0035162`
	`name: embryonic hemopoiesis`
	`namespace: biological_process`
	`def: "The stages of blood cell formation that take place within the embryo."`

Table 3: Text mining tool identifies match to GO term whose name does not appear in the sentence.

2.2.2.2. User Interface for review of text-mining results

To improve the precision and recall of the text-mining system, we developed an elaborate visual browser application integrated with the site to assist the review of the output of the automated text-mining tool. This section details the capabilities of this browser and the process used to develop it. Example features of the browser are shown in Figure 3. The text-mining results are presented to the user as shown in Figure 3A. Users can click on the graph icon next to each suggested term to further explore the context as described below. Within the graph browser application, users can see all sentences within the article in which the current term occurs. The application supplies a "View context" button for each matching sentence (Figure 3B). This button will reload the article in the browser, with the matching sentence highlighted and focused. The browser enables readers to explore related terms in the Gene Ontology.

Users can search for any term in the Gene Ontology, or look up terms that occur within the article they are reading. The browser focuses on an individual term, and displays all terms up to the parent term (either biological process, cellular component or molecular function), and all terms that are children of the current term (Figure 3D). This enables users to see the full context of where they are in the ontology, as well as provides them with the ability to navigate farther up or down the tree. The browser graph highlights any terms that have content associated with them. Any term that the current article is tagged with will have a brighter green background, whereas any term with any other content will be colored with a more moderate shade of green (see Figure 3C). The effect is to draw the user's eye towards terms that are well represented on the site.

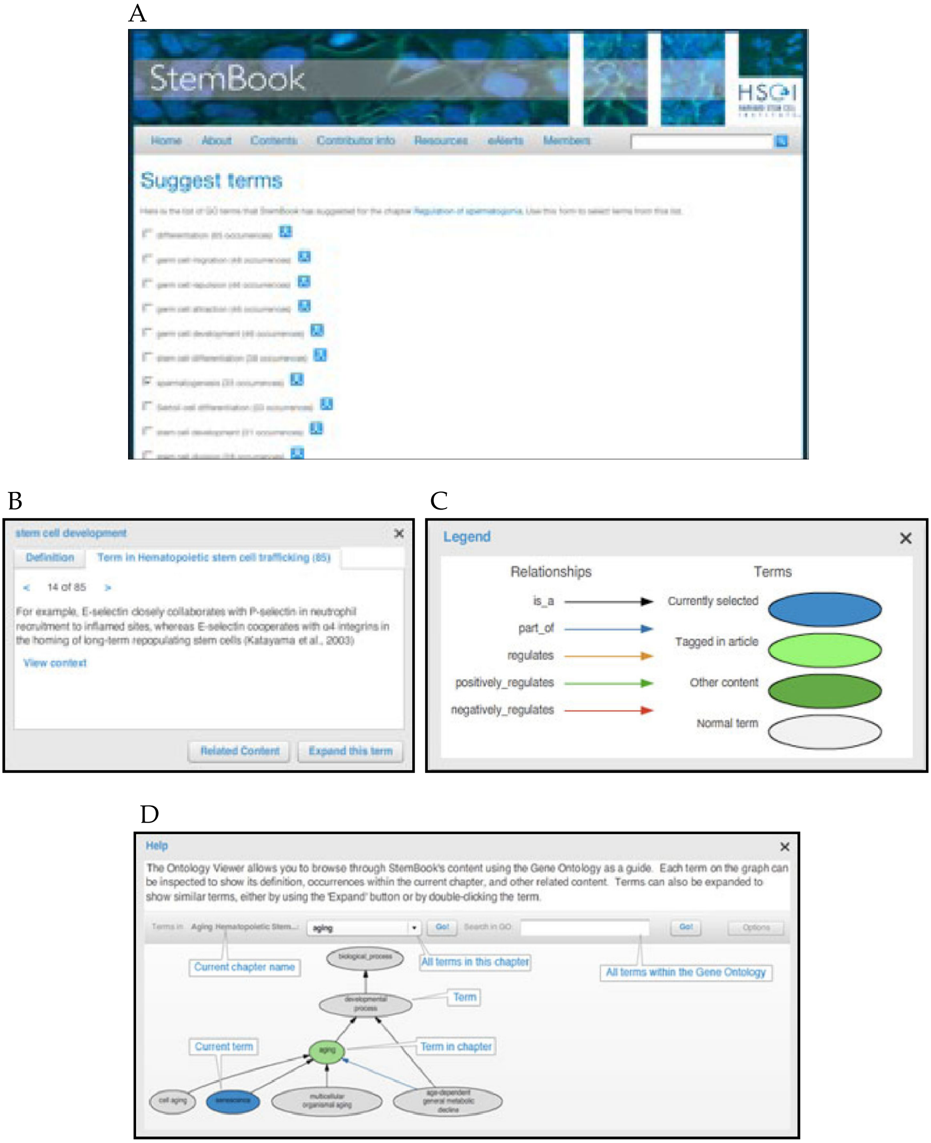
Any currently visible node can be expanded to re-center the graph around that node. This will reload the graph, and the nodes in the current graph will reposition themselves or appear or disappear as needed. The transition is animated over a span of 1.5 seconds to help the user understand how the new graph relates to the old one. When a term is clicked in the graph, the browser automatically loads and displays the term's Gene Ontology definition in a new panel. The browser also provides a "Related Content" button, which loads the list of all articles and content on site that has been tagged with that same term (see Figure 3A).

The automated text-mining tool may miss some terms (as it does not produce 100% recall). Since the Gene Ontology contains close to 26,000 terms, it is infeasible for the site editors to search through the entire ontology to find the perfect term to annotate an article with. To compensate for this, the application contains an auto-suggest feature which runs the text mining engine over the entire ontology and reports the terms that match the most in the article. Any auto-suggested terms can be automatically added to the article, or explored in the graph browser to learn more about how they relate to the article and other nearby terms that may also be applicable.

2.2.2.3. Technical Overview

The following section details how the visual ontology browser was built, including the technologies we considered and how it was finally implemented. In the course of developing the browser, we found that the most challenging task was to render the Gene Ontology graph in a fashion that was interactive and reflected the complex relationships between terms. We considered a few existing technologies and sites to explore the problem space.

- GraphViz (http://www.graphviz.org/) is an open-source network graph layout program developed by AT&T Labs. GraphViz works by taking in the structure of a network graph as input, in a format consisting of the graph's nodes and edges, and performs layout algorithms to position each element on the screen. The result can be rendered in image formats such as SVG, GIF, and JPG. GraphViz was the only layout technology we found that is capable of rendering the full structure of the Gene Ontology, which is a directed acyclic graph. The



Here is the list of GO terms that StemBook has suggested for the chapter Regulation of spermatogonia. Use this form to select terms from this list.

For example, E-selectin closely collaborates with P-selectin in neutrophil recruitment to inflamed sites, whereas E-selectin cooperates with α4 integrins in the homing of long-term repopulating stem cells (Katayama et al., 2003)

The Ontology Viewer allows you to browse through StemBook's content using the Gene Ontology as a guide. Each term on the graph can be inspected to show its definition, occurrences within the current chapter, and other related content. Terms can also be expanded to show similar terms, either by using the 'Expand' button or by double-clicking the term.

Figure 3. Graphical browser. A) Terms suggested by text-mining tool B) Viewing matched sentences and context within article C) Legend D) Viewing graph and context.

directed nature of GO means that ordinary network visualization tools are inappropriate, especially since there is a definitive root node in the structure that needs to be highlighted. However, tree renderers cannot be used since terms may have several paths to the root. On the other hand, GraphViz's default renderers can only generate a static image, which was found to be unsatisfactory for our purposes.

- AmiGO is the official Gene Ontology Consortium-developed application to navigate the Gene Ontology. Users can search for any term in the Gene Ontology and see all the details for that term, including its definition, synonyms and all related terms. It offers a graphical view that we modeled the browser application after; however, this graphical view is not sufficiently interactive for our purposes.
- Canviz is an open-source layout engine that renders GraphViz graphs in a web browser. GraphViz graphs are laid out on a web server and then sent to the browser via AJAX, where they are rendered using the HTML Canvas element. Canviz's use of GraphViz was very promising, as GraphViz can capture the structure of the Gene Ontology in a network form; however, at the time when the application was being developed, Canviz did not feature any interactivity and was poorly documented.
- SpringGraph is a free force-directed graph component for Adobe Flex. It is intended to visualize large sets of network data, and allow users to browse through the set by showing a small, portion of the data set near the user's current location at any time. As the user browses through the graph, items in the graph can be hidden or revealed. SpringGraph provides a great deal of interactivity, but it cannot handle a data structure such as the Gene Ontology, which is a directed acyclic graph with many possible paths from any node to the root node.

Based on what we learned from the technology review, we decided to use GraphViz, like Canviz, but greater interactivity is provided using Flex. As no Flex library for rendering GraphViz graphs was found, we decided to build our own. This renderer, which we named VizierFX, accepts GraphViz output in DOT format from a web service, then renders the graph in an element based off Flex's default Canvas element. Each node in the graph can be moused over and clicked on, providing the interactivity we sought. VizierFX has been released as a free, open-source library for other projects to use at (http://markandrewgoetz.com/vizierfx/).

2.2.2.4.Testing Methodology

In order to assess the utility of the system to potential users, we conducted two rounds of usability testing. The first round used a paper prototype of the system before development of the system had commenced, and the second round used a live, functional version of the system after development had completed.

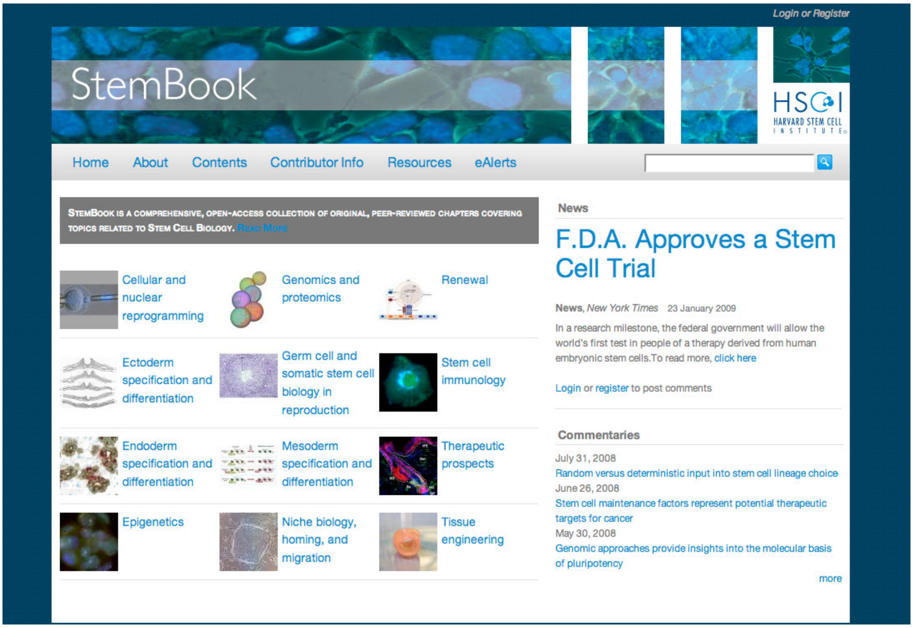
To test the overall concept of the graph browser, we conducted a paper prototype test using low-fidelity mockups created in OmniGraffle that had been printed out. Two users were asked to perform a single task – to find a annotated term in an article, remove the annotation and add a more specific annotation. Users were requested to follow the think-aloud protocol to help uncover mental models, and were asked several follow-up questions about their experience after completing the task.

Once the full system had been built, a full-scale usability test was conducted to determine if all features of the browser were discoverable and usable. One pilot user and one real user were asked to perform five tasks using the system; search for a term, view content that had been annotated with a term, expand a term, suggest terms for an article, and view a term in the context of an article. Users were requested to follow the think-aloud protocol to help uncover mental models, and were asked several follow-up questions about their experience after completing the task. Our software was iteratively refined as a result of these user tests.

3. StemBook - a Case Study

StemBook (www.stembook.org) [front page shown in Figure 4] is the first instance of an open access, online publication and web community based on our innovative Science Collaboration Framework. Since its launch in September 2008 with 12 chapters, StemBook has grown to 39 chapters and has a current total of 89 commissioned chapters. StemBook is an editorial-board reviewed compilation of articles on stem cell biology.

Although the Harvard Stem Cell Institute publishes StemBook, it is truly a product of the international stem cell research community. As an example of this, half of StemBook's editorial board is composed of non-Harvard associated experts in the stem cell field (14 Harvard-affiliated stem cell experts and 13 non-Harvard). Similarly, of the 39 StemBook chapters already published, 10 are written by Harvard affiliates and 29 articles authored by non-Harvard contributors from around the world, each selected for their noteworthy track record in a particular subspecialty. StemBook is divided into 12 different sections (see Figure 4), each representing a different facet of stem cell biology. Each chapter in StemBook has links from the gene names mentioned to a relevant gene page with additional information about the gene. Additionally, references in the text are linked to their abstract in PubMed and full-text article when available. In addition to figures, StemBook's online infrastructure (SCF) allows us to include movies, animation, and other complex images that enhance the chapters. The online infrastructure also allows us to keep StemBook up to date, a feature that would not be possible in a print version and a parameter particularly critical given the fast pace at which the stem cell field advances. Usage statistics show an upward trend; the site has already been accessed by over 8500 unique visitors and has 30-200 visitors/day.


Login or Register
StemBook
HSCI
HARVARD STEM CELL INSTITUTE
Home
About
Contents
Contributor Info
Resources
eAlerts
Cellular and nuclear reprogramming
Genomics and proteomics
Renewal
Ectoderm specification and differentiation
Germ cell and somatic stem cell biology in reproduction
Stem cell immunology
Endoderm specification and differentiation
Mesoderm specification and differentiation
Therapeutic prospects
Epigenetics
Niche biology, homing, and migration
Tissue engineering
more

StemBook is a comprehensive, open-access collection of original, peer-reviewed chapters covering topics related to Stem Cell Biology. Read More

News

F.D.A. Approves a Stem Cell Trial

News, *New York Times* 23 January 2009

In a research milestone, the federal government will allow the world's first test in people of a therapy derived from human embryonic stem cells.To read more, click here

Login or register to post comments

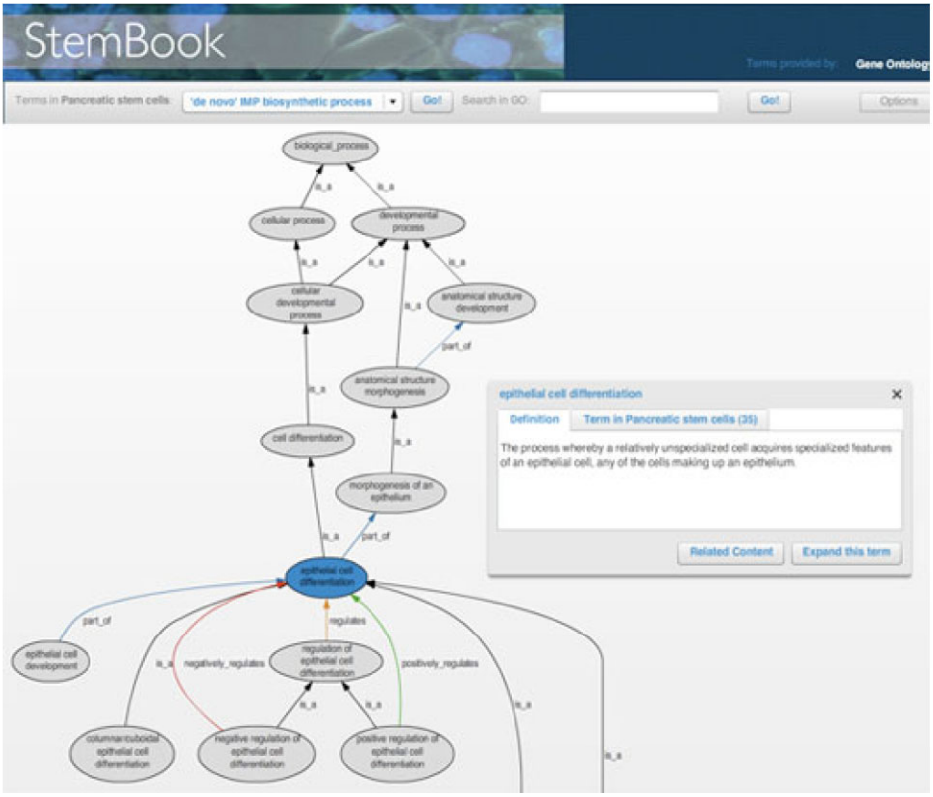
Commentaries

July 31, 2008
Random versus deterministic input into stem cell lineage choice
June 26, 2008
Stem cell maintenance factors represent potential therapeutic targets for cancer
May 30, 2008
Genomic approaches provide insights into the molecular basis of pluripotency

Figure 4. StemBook covers a wide range of topics related to stem cell biology.

Articles are invited by the editorial board and then written by the authors using any rich text editor (usually Microsoft Word). An external service provider then formats these in XML following the National Library of Medicine (NLM) DTD. The bibliographic records are also provided as structured XML (using the EndNote XML DTD). The turn around time for this activity is typically 3-5 days and the production costs are quite low. The formatting is reviewed by the editors and authors and revisions are made as necessary. The finished XML package (compressed files with XML, figures, movies etc) is then uploaded on the site by the editor. The package is processed by the software to produce the HTML files.

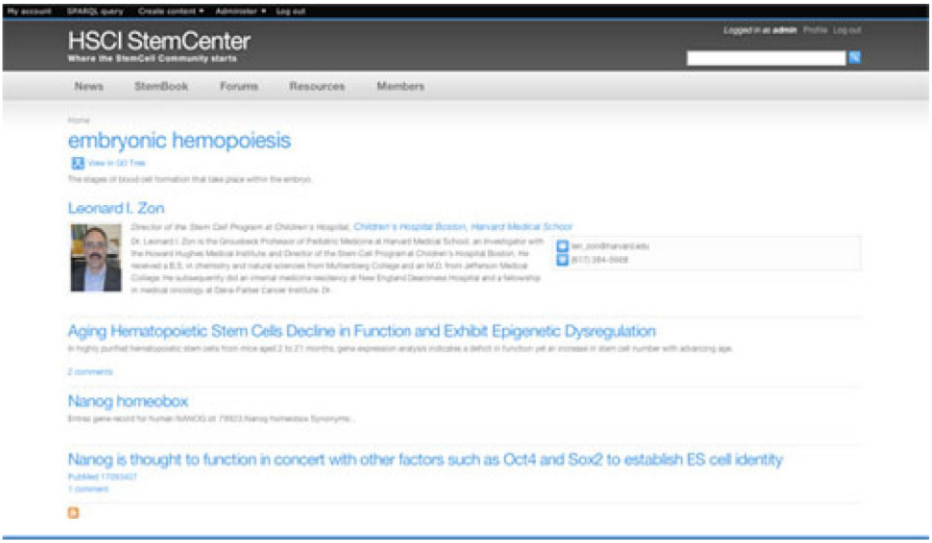
After the articles are uploaded on the site, they are indexed with terms and definitions from the Gene Ontology [38] using the semi-automatic annotation methods described in the previous section by the StemBook editors. The biological processes, cellular components and molecular functions in the ontology can be explored using a graphical browser. An example use case of how the text-mining tools and graphical browser aid in the annotation process is described. The article "Pancreatic Stem Cells" is processed using the text-mining tool. One of the suggested terms is "epithelial cell differentiation". The editor, who is familiar with the subject matter of the article, is able to choose the more specific term "regulation of epithelial cell differentiation", as shown in Figure 5.


StemBook
Terms provided by: Gene Ontolog
Terms in Pancreatic stem cells:
'de novo' IMP biosynthetic process
Go!
Search in GO:
Go!
Options
biological_process
is_a
is_a
cellular process
developmental process
is_a
is_a
is_a
cellular developmental process
is_a
anatomical structure development
part_of
anatomical structure morphogenesis
is_a
cell differentiation
is_a
morphogenesis of an epithelium
is_a
part_of
epithelial cell differentiation
part_of
regulates
epithelial cell development
is_a
negatively_regulates
regulation of epithelial cell differentiation
positively_regulates
is_a
is_a
is_a
is_a
columnar/cuboidal epithelial cell differentiation
negative regulation of epithelial cell differentiation
positive regulation of epithelial cell differentiation
epithelial cell differentiation
×
Definition
Term in Pancreatic stem cells (35)
Related Content
Expand this term

The process whereby a relatively unspecialized cell acquires specialized features of an epithelial cell, any of the cells making up an epithelium.

Figure 5. Review of suggested annotations allows selection of more specific or general terms as necessary.

Such semantic annotation allows readers to find articles on the same and related concepts. In fact, users can locate all resources that have been annotated with the same term. For example, researchers who have registered a research interest using the same term or genes that have been annotated with the term can be located. The resources (member, article, gene and research statement) annotated with the term "embryonic hemopoiesis" is displayed in Figure 6.



embryonic hemopoiesis

View in GO Tree

The stages of blood cell formation that take place within the embryo.

Leonard I. Zon

Director of the Stem Cell Program at Children's Hospital, Children's Hospital Boston, Harvard Medical School

Dr. Leonard I. Zon is the Grousbeck Professor of Pediatric Medicine at Harvard Medical School, an investigator with the Howard Hughes Medical Institute and Director of the Stem Cell Program at Children's Hospital Boston. He received a B.S. in chemistry and natural sciences from Muhlenberg College and an M.D. from Jefferson Medical College. He subsequently did an internal medicine residency at New England Deaconess Hospital and a fellowship in medical oncology at Dana-Farber Cancer Institute. Dr. ...

len_zon@harvard.edu
(617) 384-5668

Aging Hematopoietic Stem Cells Decline in Function and Exhibit Epigenetic Dysregulation

In highly purified hematopoietic stem cells from mice aged 2 to 21 months, gene expression analysis indicates a deficit in function yet an increase in stem cell number with advancing age.

2 comments

Nanog homeobox

Entrez gene record for human NANOG id: 79923 Nanog homeobox Synonyms: ...

Nanog is thought to function in concert with other factors such as Oct4 and Sox2 to establish ES cell identity

PubMed 17093407
1 comment

Figure 6: Resources annotated with the Gene Ontology term “embryonic hemopoiesis”.

The relationships between terms can also be exploited for searches. Moreover, it is a low cost and more efficient method of indexing (in comparison, pubmed uses a large team of curators to create the Medical Subject Heading (MeSH) indexes). The articles are also linked to genes using the Uniform Resource Identifier assigned to the gene. Again such explicit, machine readable content can be exploited for efficient searches in the future. Finally, such presentation of knowledge on Web 3.0 allows for interoperability with other sites, giving rise to a community of communities.

4. Discussions & Future Directions

The Web has qualitatively transformed the research information lifecycle by creating flexible and highly convenient alternative methods of publishing and accessing scientific articles and discussion. With the emergence of Web 2.0, the “social web”, user-supplied content could be directly submitted on the web for online discussions of scientific articles. Web 3.0, the social+semantic web, enables content including discourse, concepts, terms and resources, to be linked across sites in a way that explicitly recognizes them as computer-processable data and establishes shared context through shared vocabularies. Emerging and established scientific web communities such as Alzforum, StemBook, and PD Online are now beginning to make use of these Web 3.0 technologies (through the Science Collaboration

Framework), which we expect will soon begin to be adopted in the world of scientific publishing generally. In this new paradigm, there will be a significant reduction of artificial barriers between research disciplines, and a much more dynamic and agile approach to information exchange. In turn, we expect this dynamic, agile and barrier-reduced content exchange to significantly enhance the ability of scientists to collaborate across disciplines and to achieve progress in complex research problems.

Our work to produce scientific publications using Web 3.0 is currently in the proof-of concept stage and plenty of future enhancements and work are necessary to obtain interoperable communities that advance the current state of research. Enhancements planned for the text-mining tool will provide a web-services architecture and interface that can support different mining algorithms and vocabularies other than the Gene Ontology. When the interface for reviewing the mining output is further tested and refined, the review can be done by the community at large. Such crowd sourcing strategies to obtain and improve knowledge are already in use by Wikipedia and Encyclopedia of Life [44]. The manually curated body of annotation can be used as training set by machine learning algorithms to further improve the performance of the text-mining tool. We are also planning to add flexible tools to integrate more knowledge repositories (such as antibodies and model organisms) and make them available for annotation. The ontologies and RDF graphs are currently being enhanced to support larger knowledge repositories and cross-community search strategies.

With the advent of linked data using common formal vocabularies, and the incorporation of text mining driven by these vocabularies, it now becomes possible to envision and build dynamically linked "communities of communities" on the web, with shared data and discourse, in neuroscience and other complex fields, such as oncology. Our belief is that such communities can significantly speed trans-disciplinary and translational research, resulting in more rapid cures for many currently intractable diseases. This is not a distant vision, but something *that is immediately practicable* over the next few years.

Acknowledgements

The authors would like to acknowledge Thomas Green, Stéphane Corlosquet and Ben Melançon for their work on the Science Collaboration Framework. We would also like to acknowledge the Initiative in Innovative Computing at Harvard and the anonymous Foundations for their support.

Notes & References

[1] DE GROOTE, S.L. and DORSCH, J.L. Measuring use patterns of online journals and databases. J Med Libr Assoc., 91 (2), Apr. 2003, p. 231-241.

[2] EVANS, J.A. Electronic publication and the narrowing of science and scholarship. Science, 321(5887), Jul 2008, p. 395-9.

[3] PLoS ONE, (Journal published by Public Library of Science (PLoS), a non-profit organization), <http://www.plosone.org/>.

[4] WormBook, (open-access online journal),<http://www.wormbook.org>.

[5] Nature Reports Stem Cells, (stem-cell journal published by Nature Publishing Group), <http://www.nature.com/stemcells/>.

[6] BioMed Central, (open-access publisher), <http://www.biomedcentral.com>.

[7] Free Market Science. Nat Cell Biol. 2007 Jul; 9(7), p. 721.

[8] O'REILLY, T. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. Communications & Strategies, First Quarter 2007 No. 1, p. 17.

[9] *FaceBook*, (application), <http://www.facebook.com/>.

[10] *My Space*, application, <http://www.myspace.com/>.

[11] *Nature Network*, website, <http://network.nature.com>.

[12] *SciLink*, website, http://www.scilink.com.

[13] ANDERSON, N.D. SciLink: An innovative project linking research scientists and science teachers. Journal of Science Teacher Education, 4 (2), May 2007, p. 44-50.

[14] *Connotea*, (application), <http://www.connotea.org/>.

[15] *CiteULike*, (application), <http://www.citeulike.org/>.

[16] LUND, B., HAMMOND, T., FLACK, M. and HANNAY, T. Social Bookmarking Tools (II) - A Case Study - Connotea. D-Lib Magazine, 11(4), Apr 2005.

[17] Wikis, (How wikis work), http://computer.howstuffworks.com/wiki.htm.

[18] OpenWetWare, (wiki to support research, education, publication, and discussion in biological sciences and engineering), <http://openwetware.org>.

[19] MONS, B., ASHBURNER, M. and CHICHESTER, C, et al. Calling on a million minds for community annotation in WikiProteins. Genome Biol., 2008, 9(5), p. R89.

[20] CLARK, T and KINOSHITA, J. Alzforum and SWAN: the present and future of scientific web communities. Briefings in Bioinformatics, 2007, 8(3), p. 163-171.

[21] KINOSHITA, J. and CLARK, T. Alzforum. Methods Mol Biol. 2007, 401, p. 365-81. Review.

[22] DAS, S., GIRARD, L., GREEN, T., WEITZMANN, L., LEWIS-BOWEN, A, and CLARK, T. Building Biomedical Web Communities Using a Semantically-Aware Content Management System. Briefings in Bioinformatics. Brief Bioinform., Mar 2009, 10(2), p. 129-38.

[23] BERNERS-LEE, T., HENDLER, J. and LASILLA, O. The Semantic Web. Sci Am., 2001, 284(5), p. 34-43.

[24] FEIGENBAUM, L., HERMAN, I., HONGSERMEIER, T., NEUMANN, E. and STEPHENS S. The Semantic Web in Action. Scientific American, December 2007.

[25] RDF, (W3C recommendations), http://www.w3.org/RDF/.

[26] BRICKLEY, D., GUHA, RV. and MCBRIDE, B. (eds). RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation, 10 February 2004, <http://www.w3.org/TR/rdf-schema/>.

[27] BECHHOFER, S., VAN HARMELEN, F., HENDLER, J., HORROCKS, I., MCGUINNESS, D., PATEL-SCHNEIDER, P.F., STEIN, L.A., DEAN M., SCHREIBER G. eds. OWL Web Ontology Language Reference. W3C Recommendation 10 February 2004. http://www.w3.org/TR/owl-ref/.

[28] BIZER, C., HEATH, T., IDEHEN, K. and BERNERS-LEE, T. Linked Data Linked Data on the Web. In Proceedings WWW2008, Beijing, China.

[29] SHAH, N.H., JONQUET, C., CHIANG, A.P., BUTTE, A.J., CHEN, R. and MUSEN, M.A. Ontology-driven indexing of public datasets for translational bioinformatics. BMC Bioinformatics, 10 (2):S1, Feb 2009.

[30] SPASIC I., ANANIADOU S., MCNAUGHT J. and KUMAR A. Text mining and ontologies in biomedicine: making sense of raw text. Brief Bioinform, 6(3), Sep 2005, p. 239-51.

[31] SHAH N.H. et al. Ontology-based Annotation and Query of Tissue Microarray Data. AMIA Annual Symposium. Washington, DC 2006.

[32] MOSKOVITCH R., MARTINS S.B., BEHIRI E., WEISS A. and SHAHAR Y. A comparative evaluation of full-text, concept-based, and context-sensitive search. J Am Med Inform Assoc, 14(2), Mar-Apr 2007, p. 164-74.

[33] YANG H., NENADIC G. and KEANE J.A. Identification of transcription factor contexts in literature using machine learning approaches. BMC Bioinformatics, 11(9): S11, Apr 2008.

[34] CAMON E.B., BARRELL D.G., DIMMER E.C., LEE V., MAGRANE M., MASLEN J., BINNS D. and APWEILER R. An evaluation of GO annotation retrieval for BioCreAtIvE and GOA. BMC Bioinformatics, 6(1) :S17, May 2005.

[35] ROGAN, M.T., DAS, S. and CLARK, T. Swarming for the Cure: An Experiment in Community-Guided Funding of Parkinson's Disease Research. Neuroscience 2008; November 15-19; Washington, DC.

[36] Drupal, (an open source content management system), http://www.drupal.org.

[37] Information Technology - Open Document Format for Office Applications (OpenDocument) v1.0, International Organization for Standardization, ISO/IEC 26300:2006.

[38] GENE ONTOLOGY CONSORTIUM. The Gene Ontology (GO) project in

2006. Nucleic Acids Res. 2006 Jan 1; 34 (Database issue): D322-6.

[39] SMITH, B., ASHBURNER, M., ROSSE C., et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol., 25(11), 2007, p. 1251-5.

[40] MILES A., ROGERS N. and BECKETT D. SKOS Core 1.0 Guide. World Wide Web Consortium, 2004.

[41] CICCARESE, P. et al. 2008. The SWAN Biomedical Discourse Ontology. Journal of Biomedical Informatics. 4, 2008, p. 1739–751.

[42] CICCARESE, P. et al. 2009. The SWAN Ontology Ecosystem, Version 1.2 Release candidate. http://swan.mindinformatics.org/ontology.html

[43] PORTER M.F. An algorithm for suffix stripping, Program, 14(3), 1980, p. 130-137.

[44] WILSON EO. The encyclopedia of life. Trends in Ecology & Evolution, 18(2), Feb 2003, p. 77-8

June 2009
Printed on demand
by *"Nuova Cultura"*
www.nuovacultura.it

Book orders: ordini@nuovacultura.it