



## **Web based Reference Services to Bioinformaticians: challenges for librarians**

**Shubhada Nagarkar**  
Bioinformatics Center  
University of Pune  
Pune, India

### **Meeting:**

**111 — *The challenges and changing roles of the Science and  
Technology Libraries* — Science and Technology Libraries Section**

### **Abstract:**

*Bioinformatics is one of the challenging fields where voluminous open access data sets are available including literature databases, nucleotide and protein sequence databases, genomics and proteomics databases to name a few. Scientists spend a huge amount of time on the Internet searching the relevant information. Out put of the searches is not always proportional to the time spent. Science and technology librarians have the expertise to design and provide personalized information services utilizing various tools and resources that are newly developing. This paper describes how personalized reference services for bioinformaticians are being developed using a portal <https://sites.google.com/site/reftobioinfo/>. Various tools and resources are utilized starting with Google Sites as the platform and then integrating RSS feeds, Google gadgets, Google News for current awareness services and PubMed and Google for selected dissemination services.*

*Initial feedback from scientists is very encouraging. They are satisfied with the information they get and the time saved. Librarians face a number of challenges when developing this type of personalized service. They have to work closely with the scientists as a team member to understand their information needs as well as conduct information literacy programs for scientists. However, librarians should take up the challenge to apply their expertise to the research process and develop tools to help scientists to be more productive.*

### **Introduction**

Availability and accessibility are two major issues related to the bioinformatics information. Available information is vast, diverse and in public domain which includes voluminous open access data sets including literature databases, nucleotide and protein sequence databases, genomics and proteomics databases to name a few. The current 18<sup>th</sup> Database Issue of *Nucleic Acids Research* features descriptions of 96 new and 83 updated online databases

covering various areas of molecular biology (Galperin & Cochrane, 2011). Overlapping contents, dispersal throughout the Internet, growing complexity, and constant updation of ever increasing bioinformatics databases have outpaced the ability of bio-researchers to locate, identify, and develop sufficient expertise in the bioinformatics resources they need (Rein, 2006). For the seamless accessibility to bioinformatics information there is a strong demand for the automated tools to retrieve relevant information in a timely manner. Several automated programs have been developed over the past decade (e.g. BioCatalogue, BioMoby). However, the number of databases and resources is growing too fast to keep track. Apart from molecular biology databases, number of research papers in the literature databases is increasing to a large extent. To access the latest publications, bioinformaticians mostly use Internet search engines such as Google and Google Scholar. PubMed is the most frequently accessed bibliographic database (Kim & Rebholz-Schuhmann 2008).

Internet has changed over the time and has become interactive. It includes Web 2.0 tools viz. RSS feeds, e-mail alerts, Google gadgets, wikis, blogs, online groups, news aggregators, social networking and social book marking tools, etc. With the help of these tools researchers can create, annotate, review, reuse and represent the information in new ways (Currie, 2010; Hey, Tansley, & Tolle, 2010). Moreover, Web 3.0 is heading towards semantics (or the meaning of data) and personalization (e.g. iGoogle) of web (<http://www.labnol.org/internet/web-3-concepts-explained/8908/>). Scientists must take the advantage of these tools to save their time for the core research. The major problem is that scientists are not formally trained in “information retrieval” techniques. They often use the resources they know, get some of the information they need, and move on. As the number and complexity of available resources grow, users generally lose track of what they have done. They need guidance on using various tools under different conditions. Librarians’ role under such a situation should be dynamic. Specially trained librarians could work as embedded librarians need to work with scientists and must establish their reputation as problem solvers (Rein, 2006, Konieczny, 2010; Hoffman & Ramin, 2010).

### **Role of Librarians**

Librarians can make a significant impact by employing online Current Awareness Services (CAS) and Selective Dissemination of Information (SDI) to keep up with the plethora of available resources (Alpi, 2003). Further, librarians who are “Masters of the Info Universe” (Kerith, 2011) should recognize the complexity of the scientists’ queries (Osterbur et al., 2006; Rein, 2006; Tennant, 2005) and design good search strategies for each and every database used. They need to keep constant watch on the new developments of search engines, reference management tools as well as services offered by database providers. The immediate need is to design personalized information services to users to save their time (Youngkin, 2010; Wu WG, 2007) rather than the traditional reference services. Barrett mentions that the use of reference services is declining while the inception of chat and email reference services broaden the patron base (Barrett, 2010). Librarians should take the advantage of Web 2.0 tools to promote the use of library services (Ivie et al., 2011).

This paper discusses the prototype of the information portal “Ref-2-Bioinfo” (<http://sties.google.com/site/reftbioinfo>) developed by librarian at the Bioinformatics Centre of University of Pune to addresses the very specific needs of this user group. The main focus of the portal is to provide personalized information services. Librarians’ traditional services can be delivered via web and “reftbioinfo” is proof of the same. It is important to note that librarians need not be software professionals to develop such portal, but do need to be

involved and aware of current developments in the fields of bioinformatics and information technology.

### **Developing the Prototype -“reftbioinfo”**

Twenty years of interactions with bioinformaticians helped the librarian understand the information needs of users. User profiles in the form of keywords are maintained and regularly updated through email and meetings. These provide insight into what researchers are looking for and what tools they are comfortably using. The design and development of “Ref-2-Bioinfo” was carried out on the basis of these insights and additional assessment of information needs was considered unnecessary.

Researchers primarily use PubMed and Google Scholar for literature searches. The librarian designed SDI and CAS for individual users based on their profiles. The prototype is developed on Google Sites and uses several other Google tools. Table 1 indicates the tools used and the purpose. These tools and services provide the basic content to populate the various categories of the portal.

**Table 1: Tools and purpose**

<b>Tools</b>	<b>Purpose</b>
Google Sites	Portal development
Google Gadgets	Links to various databases and portals
Google Scholar	Citations of research papers
Google News	Links to world wide news related to Bioinformatics and India
Google Docs / form	Online information request form
RSS feeds	Gather the data from databases and news servers
Delicious	Social book marking
Email alerts	Email alerts from databases and publishers sites

### **Main Page**

The Main Page has three primary components: the main body, the right hand current information box and the left hand navigation pane. The main body succinctly explains the purpose of the portal and provides appropriate contact information. Right hand box displays an RSS feed of Google News, which displays daily news on keywords “Bioinformatics AND India”. PubMed related latest news are integrated with the help of RSS feeds. Figure 1 shows the main page of the portal with Google News display.

The navigation page reflects the portal's eleven categories listed in Table 2. The Table also lists the tools used to populate the category.

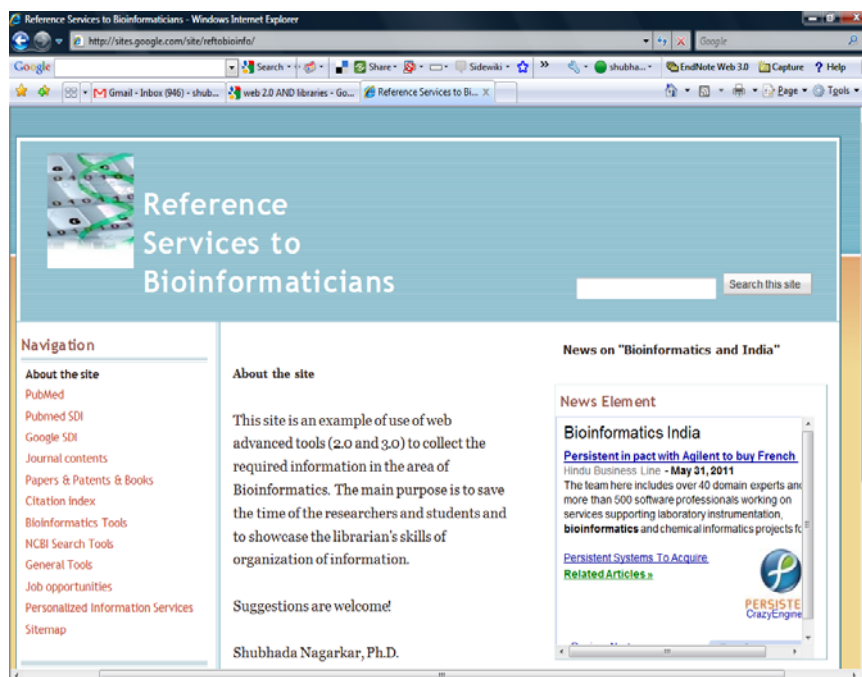


Figure 1: Main page with Google news

Table 2: Category and Services

Sr. no.	Category	Tools
1.	PubMed	<ul style="list-style-type: none"> <li>• Google gadget for PubMed basic search screen</li> <li>• PubMed RSS</li> <li>• Links to PubMed alternative interfaces</li> </ul>
2.	PubMed SDI	<ul style="list-style-type: none"> <li>• RSS feeds of PubMed queries</li> <li>• Google form for queries</li> </ul>
3.	Google SDI	<ul style="list-style-type: none"> <li>• Links to solved queries on Google</li> <li>• Google form for new queries as well as for updation in queries</li> </ul>
4.	Journal contents	<ul style="list-style-type: none"> <li>• Google gadget with RSS feeds of latest issues of core Bioinformatics journals</li> </ul>
5.	Papers, patents and Books	<ul style="list-style-type: none"> <li>• Google gadgets to access Google Scholar search, EPO patents, Customized search for Bioinformatics Journals</li> <li>• Amazon gadget for latest books</li> </ul>
6.	Citation index	<ul style="list-style-type: none"> <li>• Gadget for Google Scholar citation index</li> </ul>
7.	Bioinformatics tools	<ul style="list-style-type: none"> <li>• Google gadgets for popular Bioinformatics tools</li> </ul>
8.	NCBI tools	<ul style="list-style-type: none"> <li>• Links to NCBI Google gadgets</li> </ul>
9.	General purpose tools	<ul style="list-style-type: none"> <li>• Gadgets of Wordnet dictionary and visual thesaurus</li> </ul>
10.	Job opportunities	<ul style="list-style-type: none"> <li>• Google gadget for Job opportunities (for students)</li> </ul>
11.	Personalized Information services	<ul style="list-style-type: none"> <li>• Access to personalized services (email based access)</li> </ul>

## Personalized Information Services

From the Main Page of the portal, a registered user can click onto the Personalized Information Services. This personalized portal is divided into three categories – PubMed results, Google Scholar results and citations links. It further links to homepages of those journals in which the user has published papers. New categories will be added for other databases as demand is demonstrated.

### PubMed results

Figure 2 illustrates PubMed based SDI services with RSS feeds. For each personalized portal, PubMed is searched on given keywords. MeSH terms and Boolean operators are used to design search strategies. The search queries are listed and are hyperlinked with PubMed results page. The user can also modify the search query at any time. RSS feeds are displayed in the right hand column so the user can check the latest information without visiting PubMed site. Due to availability of these links, users need not write and remember the complex search strategies. Figure 2 (lower half) shows the PubMed search strategy used for the given topic “computational vaccine design”.

Table 3 represents few solved search queries. It lists queries, PubMed search strategies written by Bioinformaticians and librarian along with number of reference received.

The screenshot displays a web browser window titled "PubMed-Results - Welcome to Personalized Information Services - Windows Internet Explorer". The address bar shows the URL "https://sites.google.com/site/bioinfouser1/pubmed-results". The page content is organized into several sections:

- Services:** A sidebar menu with options like "PubMed-Results", "Scholar-Results", "Citations of your papers", "Scopus citations", and "List of Journals".
- Pubmed-Results:** A main section titled "Links to PubMed to edit queries" containing a numbered list of 9 search strategies, such as "1. Epitope Prediction", "2. Computational vaccine design", and "9. Structural genomics of viruses".
- RSS feeds:** A section titled "PubMed RSS" with a table of categories: Epitope Prediction, Computational Vaccine Design, mRNA binding site prediction, and Antigen Antibody Reaction. Below this, there are several article snippets with titles like "PubMed: 'Epitopes'[Majr] and..." and "Promiscuous prediction and conservancy analysis of CTL binding epitopes of HCV 3a viral proteome from Punjab Pakistan: an in silico approach".

At the bottom of the page, there is a search bar with the text "Search: PubMed" and a dropdown menu. To the right of the search bar are links for "RSS", "Save search", "Limits", "Advanced search", and "Help". Below the search bar, a search query is entered: "Computational Biology"[Majr] AND "Vaccines"[Majr] AND design. A "Search" button and a "Clear" button are also present.

Figure 2: PubMed SDI

**Table 3: Search strategies**

Topic	Search strategies			
	Bioinformaticians	No. of references	Librarian	No. of references
Epitopes prediction	Epitopes prediction	1161	"Epitopes"[Majr] AND prediction	556
miRNA binding site prediction	MicroRNAs Binding Sites	691	("MicroRNAs"[Majr]) AND "Binding Sites"	426
Scoring Matrices bioinformatics	Scoring Matrices bioinformatics	126	"Scoring Matrices" AND bioinformatics[Majr]	51
Molecular phylogeny methods and viruses	Viruses AND molecular phylogeny AND methods	1841	"Viruses"[Majr] AND "molecular phylogeny" AND methods	10

Above table shows that search strategies written by users are simple and number of references are more than the search strategies by librarians. It indicates that the special training regarding the use of MeSH terms and other tools is necessary.

### Google Scholar results

Google Scholar is searched using given keywords with the help of Boolean, field and advanced search strategies. The URL of the result page is linked to keywords. One click will take users to the results page.

### Citations

Publications of the user are listed on this page of the personalized portal. Figure 3 shows the links to abstract, full text and Google Citations. Web of Science is a paid database and therefore it was not possible to hyperlink the result pages to author name or to individual papers. As it is paid database the search session never lasts for long time. Due to time limit of each search session, the citation page is saved as .mht format, which includes all graphs and yearly analysis of citations. At present it is decided to search web of science monthly and upload the latest .mht file on the portal. Gadget for Google scholar citations is also added to get author h-index and other related information. Link to Scopus "Author View" is provided under the navigation.

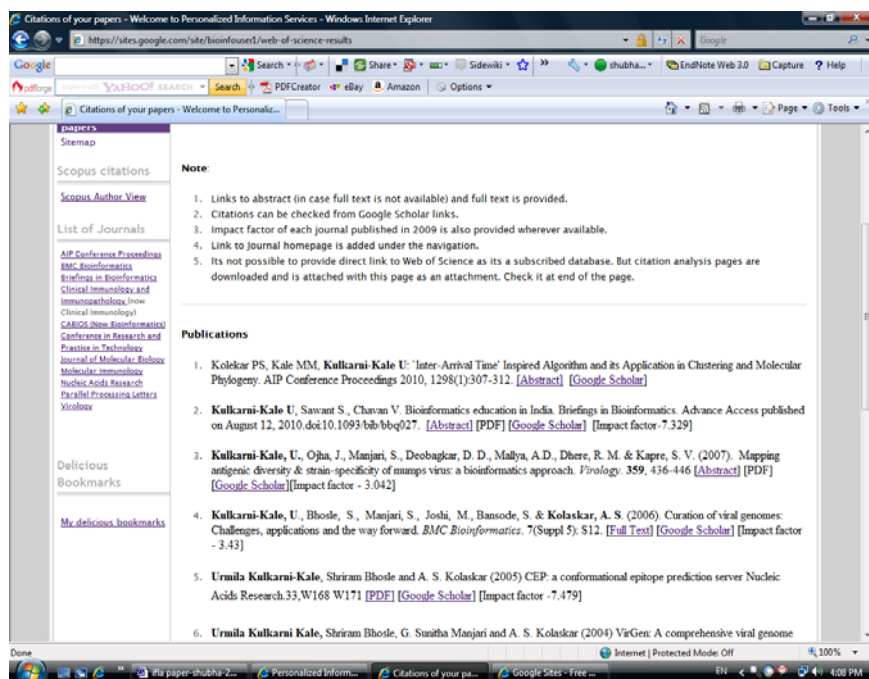


Figure 3: List of publications and links to citations

## Discussion

Response to “Ref-2-Bioinfo” portal by researchers is very positive. They are very happy to have personalized services for them. They mention that such initiatives will not only save their time of research but also teach new techniques to retrieve the relevant information from various sources in an efficient manner. They suggest selecting other databases viz. ScienceDirect, Annual Reviews, Wiley-Blackwell Publishing, etc. which are subscribed by the University of Pune. They further said that the simple hyperlinks to additional databases would be of limited use; however, the present portal is an excellent example of good search strategies developed and implemented. Several meetings as well as discussions over the online chat and email are helping to keep the track of changing areas of research of bioinformaticians in the centre. The researchers suggest developing a “knowledge management” cell in the Bioinformatics center maintained by librarian. They have also added the topics like “Literature database Searches” in the syllabus of masters degree program. They suggested conducting information literacy programs on regular basis which should include the topics like writing search strategies, use of thesaurus and subject headings, and web 2.0 and 3.0 tools.

## Challenges for Librarians

The experiences during the development and usage of Ref-2-Bioinfo suggest that librarians should focus on their role as a liaison between researchers and information and strive to embed them in the research process. They can focus their work in a particular subject area and provide services to users in that discipline. Special effort can be made to understand the basics of the subject. Referring to books, other reference materials and talking to subject experts will help in overcoming this difficulty. Communication could also be the problem in providing information services. It was noticed that initially librarians may found difficulties un understanding the terminologies used by scientists. This could be overcome when they regularly work in a group of scientists as a team member. They need to face the challenges of monitoring ever increasing numbers and types of information containers and tools to extract

information out of it (Nagarkar & Parekh, 2010). Librarians need to refresh and update their domain knowledge and improve their computational skills. Apart from traditional course work, library schools should design special training programs to serve special users and domain experts and to learn the documentation practices in science and technology. Some efforts have been reported worldwide (Lyon, et.al., 2004). Each library should develop a strategy based on the number of people who could use the service, and have a real understanding of where these services make any impact on decision making and in research (Brice & Muir Gray, 2004). Sathe (2007) mentions that human intelligence, remains critical for assessing information and recognizing patterns and connections in information that lead to knowledge. Librarians can contribute this intelligence and can help ensure that the profession continues to be a vital force for informing high-quality health care and biomedical research, education, and policy.

## References

- Alpi, K. (2003). Bioinformatics training by librarians and for librarians: developing the skills needed to support molecular biology and clinical genetics information instruction. *Issues in Science and Technology Librarianship*, 37.
- Barrett, F. (2010). An analysis of reference services usage at a regional academic health sciences library. *Journal of the Medical Library Association: JMLA*, 98(4), 308.
- Brice, A., & Muir Gray, J. A. (2004). What is the role of the librarian in 21st century Healthcare? *Health Information & Libraries Journal*, 21(2), 81-83.
- Currie, J. P. (2010). Web 2.0 for reference services staff training and communication. *Reference Services Review*, 38(1), 152-157.
- Galperin, M. Y., & Cochrane, G. R. (2011). The 2011 nucleic acids research database issue and the online molecular biology database collection. *Nucleic Acids Research*, 39(suppl 1), D1.
- Hey, T., Tansley, S., & Tolle, K. (2010). The fourth paradigm: data-intensive scientific discovery. Seattle: Microsoft Research.
- Hoffman, S., & Ramin, L. (2010). Best Practices for Librarians Embedded in Online Courses. *Public Services Quarterly*, 6(2), 292-305.
- Ivie, T., McKay, B., May, F., Mitchell, J., Mortimer, H., & Walker, L. (2011). Marketing and Promotion of Library Services Using Web 2.0: An Annotated Mediagraphy. *The Idaho Librarian: A Publication of the Idaho Library Association*, 61(1).
- Kerith Page McFadden (2011) Librarians: Masters of the info universe available online at [http://articles.cnn.com/2011-04-12/living/librarians.masters.of.universe\\_1\\_librarians-academic-libraries-national-library-week?\\_s=PM:LIVING](http://articles.cnn.com/2011-04-12/living/librarians.masters.of.universe_1_librarians-academic-libraries-national-library-week?_s=PM:LIVING)
- Kim, J.J., & Rebholz-Schuhmann, D. (2008). Categorization of services for seeking information in biomedical literature: a typology for improvement of practice. *Briefings in Bioinformatics*, 9(6), 452-65
- Konieczny, A. (2010). Experiences as an Embedded Librarian in Online Courses. *Medical Reference Services Quarterly*, 29(1), 47-57.
- Lyon, J., Giuse, N. B., Williams, A., Koonce, T., & Walden, R. (2004). A model for training the new bioinformationist. *Journal of the Medical Library Association*, 92(2), 188.
- Nagarkar, S., & Parekh, H. (2010). *Fungal Species Information*. Berlin: Lap Lambert Academic Publishing.
- Osterbur, D. L., Alpi, K., Canevari, C., Corley, P. M., Devare, M., Gaedeke, N., et al. (2006). Vignettes: diverse library staff offering diverse bioinformatics services. *Journal of the Medical Library Association*, 94(3), 306.



- Rein, D. C. (2006). Developing library bioinformatics services in context: the Purdue University Libraries bioinformationist program. *Journal of the Medical Library Association*, 94(3), 314.
- Sathe, N. A., Jerome, R., & Giuse, N. B. (2007). Librarian-perceived barriers to the implementation of the informationist/information specialist in context role. *Journal of the Medical Library Association JMLA*, 95(3), 270-274. Retrieved from <http://search.ebscohost.com.ezproxy.openu.ac.il/login.aspx?direct=true&db=aph&AN=26019298&site=ehost-live>
- Tennant, M. R. (2005). Bioinformatics librarian: meeting the information needs of genetics and bioinformatics researchers. *Reference services review*, 33(1), 12-19.
- Wu WG, L. J. (2007). RSS made easy: a basic guide for librarians. *Med Ref Serv Q.* , 26(1), 37-50.
- Youngkin, A. (2010). Librarian-Controlled RSS: A Novel Approach to Literature Search Follow-up. *Journal of Hospital Librarianship*, 10(2), 123-131.