

EDITORIAL

NUCLEIC ACIDS RESEARCH ANNUAL WEB SERVER ISSUE IN 2010

The 2010 Web Server Issue of *Nucleic Acids Research* is the eighth in a series of annual special issues dedicated to web-based software resources for analysis and visualization of molecular biology data. It is freely available online under *NAR*'s open access policy. The present issue reports on 122 web servers.

Topics

This year's special emphasis is on next-generation sequencing data analysis, molecular network and pathway analysis, and biological text mining. A total of 15 papers deal with these topics. A large number of papers cover two common tasks, alignment for DNA, RNA or proteins (10 papers), and various forms of gene set enrichment analysis (9 papers). A large number of papers cover DNA and RNA sequence and structure analysis (23 papers) and protein analysis, primarily of protein structure (35 papers).

Because very high-volume experimental data have become more common, in particular, with the advent of next-generation sequencing, the 2010 Web Server issue has inaugurated a new section for stand-alone (non-web server) programs that analyze such data. These programs combine the ease-of-use principles of web servers, including simplified installation and a well-designed graphical user interface, with the ability to overcome problems associated with uploading high-volume data across the internet. Five papers are in this category.

The 2010 issue also inaugurates a section for papers describing large collections of popular web services, i.e. automated analyses that can be utilized programmatically rather than through manual interaction with a web browser. Eight papers fall in this category.

Also included is the Bioinformatics Links Directory 2010 update by Michelle Brazas, Francis Ouellette and their colleagues at the Ontario Institute for Cancer Research. The directory, http://bioinformatics.ca/links_directory, is a searchable compilation of web servers published in this and previous Web Server issues together with other useful tools, databases and resources for life sciences research.

Instructions for submissions

To streamline the review process, authors are required to send a one-page summary of their web server to the editor, Dr Gary Benson (narwsrv@bu.edu), for pre-approval prior to manuscript submission. For the 2010 issue, 344 summaries were submitted and 162, or 47%, were approved for manuscript submission. Of those approved, 122, or 75%, were accepted for publication.

Review of a summary includes evaluation of the proposal and extensive testing of web server functionality. The key criteria for pre-approval are high scientific quality, wide interest, the ability to do computations on user-submitted data, and a well-designed, well-implemented and fully functional website. Note that there is a minimum 2-year interval before publication in the Web Server issue for web servers, or essentially similar web servers, which have been the subject of a previous publication, including publication in journals other than *NAR*.

With respect to the website, the following are guidelines for approval. It should have an easy-to-find submission page with a simple mechanism for loading test data and setting test parameters. The preferred method is one-click loading using javascript or a similar mechanism. Also acceptable, but less preferred, is data available through a link next to the data submission box. This requirement simplifies the review process for the editor and the referees, and provides potential users with a quick way to examine and judge a web server's features. Additional mechanisms that assist the user in submitting data should be implemented where appropriate. If the user can submit data that resides in an external database where it can be downloaded programmatically, for example, a pdb structure file or a GenBank sequence file, then the web server should provide automatic download of that data once the user has entered the appropriate identifier.

Output of the web server should be dynamic and rich in detail. Wherever possible, links should be provided to supporting evidence used in calculations and/or external databases containing additional information. Numerical, textual and visual output should be mixed and any visualization tools that add information or increase the user's understanding should be utilized (for example, the java plug-in tools jmol for structure visualization and jalview for sequence alignment visualization). Note that a web server with output which consists merely of a few numerical values, a static spreadsheet or a compressed file will not be approved (although download options for static files should be an option).

Many web servers provide time-consuming analysis and are not able to return results immediately. In that case, a mechanism for returning results to the user should be implemented. Although notification by email is straightforward and can be provided as an option, many users balk at revealing their identity through email. The preferred method is to return

a web link to the results, at the time of data submission, which the user can then copy and access at a later time. This link should ideally report the status of the job (queued, running and finished). Even if the user provides an email address for results, they should be provided on a webpage (in the dynamic form mentioned above) rather than mailed as large files, which might be rejected as spam by email programs.

The website should be supported by an extensive help section or tutorial that guides the user through the submission process, contains details about input file formats and parameters, and importantly, explains the meaning of the output. Whenever possible, the help pages should link to dynamic output examples similar to those provided by the website.

Any proposal for a web server that is *predictive* must include details on validation of predictions from new data not used in training. *N*-fold cross validation methods will not be considered sufficient. Details should include size and composition of the validation data set (number of positive and negative cases), and several measures of predictive performance, including sensitivity, specificity and precision. Proposals are regularly rejected for lack of adequate prediction validation information.

Many summaries are rejected because the websites are clearly not designed to accept user-submitted data. This applies to those established primarily for lookup or exploration in a data set, or serve the function of 'data integrators.' Authors of websites that provide novel data should consider the *NAR* Database Issue as a possible venue (see the instructions at http://www.oxfordjournals.org/our_journals/nar/for_authors/msprep_database.html).

Proposals that describe a novel analysis method are generally not appropriate for the Web Server issue because limited space makes adequate method description and validation problematic. Authors of such methods might instead consider sending their manuscript to *NAR* as a regular computational biology paper (see the instructions for authors at http://www.oxfordjournals.org/our_journals/nar/for_authors/criteria_scope.html#Computational%20Biology).

Special emphasis for 2011

For the 2011 issue, analysis of next generation sequencing data, network and pathway analysis, and biological text mining will remain topics of special emphasis.

Deadlines for 2011

Authors wishing to submit manuscripts for the 2011 Web Server issue must submit their one-page proposal along with the URL address of the fully functional website to narwsrv@bu.edu by 31 December 2010. Detailed instructions and requirements are presented at http://www.oxfordjournals.org/nar/for_authors/submission_webserver.html. This information should be consulted before sending in the summary. The deadline for submission of articles is 31 January 2011.

New requirements for references links

Manuscripts submitted for the 2011 issue must format their References section to include active links to electronic versions of the cited papers, including links to PubMed, PubMed Central and a DOI link. To assist those who submit manuscripts in LaTeX, two resources are available. One is a terrific website, TeXMed (<http://www.bioinformatics.org/texmed/>), created by Arne Muller, which performs a PubMed type search for articles and returns BibTeX entries that include the links. The other is a Perl script written by me (<http://lobstah.bu.edu/bibtolinks.pl>), which adds the links to an existing BibTeX file, also by searching PubMed.

ACKNOWLEDGEMENTS

The Web Server issue would not be possible without the work of the many scientists and programmers, who have provided us with outstanding, freely available web resources, and the conscientious efforts of literally hundreds of reviewers. Thanks to you all. In addition, I would like to thank Drs Brazas and Ouellette for their tireless work on the Bioinformatics Links Directory.

My work was made possible only by the dedicated editorial assistance of Fay Oppenheim. Thank you. Authors' and reviewers' impressions of my good nature are due primarily to her. Thanks also to Martine Bernardes-Silva who has provided frequent, cheerful assistance and especially with regard to adding electronic reference links to manuscripts submitted in MS Word. Thanks also to Claire Hardwidge of *NAR*, and to Claire Bird and the staff at Oxford University Press.

Gary Benson
Executive Editor
Web Server Issue
Nucleic Acids Research
June 8, 2010