DATABASE
The Journal of Biological Databases and Curation

Database, Vol. 2010, Article ID baq035, doi:10.1093/database/baq035

Original article

The annotation and the usage of scientific databases could be improved with public issue tracker software

Giovanni Marco Dall'Olio[1], Jaume Bertranpetit[1,2] and Hafid Laayouni[1,2,*]

[1]IBE, Institute of Evolutionary Biology (UPF-CSIC), CEXS-UPF-PRBB, Carrer Doctor Aiguader 88, 08003, Barcelona, Catalonia and [2]CIBER Epidemiología y Salud Pública (CIBERESP), Spain

***Corresponding author:** Tel: +34 9333160845; Fax: +34 933160901; Email: hafid.laayouni@upf.edu

Submitted 22 October 2010; Revised 1 December 2010; Accepted 6 December 2010

Since the publication of their longtime predecessor *The Atlas of Protein Sequences and Structures* in 1965 by Margaret Dayhoff, scientific databases have become a key factor in the organization of modern science. All the information and knowledge described in the novel scientific literature is translated into entries in many different scientific databases, making it possible to obtain very accurate information on a biological entity like genes or proteins without having to manually review the literature on it. However, even for the databases with the finest annotation procedures, errors or unclear parts sometimes appear in the publicly released version and influence the research of unaware scientists using them. The researcher that finds an error in a database is often left in a uncertain state, and often abandons the effort of reporting it because of a lack of a standard procedure to do so. In the present work, we propose that the simple adoption of a public error tracker application, as in many open software projects, could improve the quality of the annotations in many databases and encourage feedback from the scientific community on the data annotated publicly. In order to illustrate the situation, we describe a series of errors that we found and helped solve on the genes of a very well-known pathway in various biomedically relevant databases. We would like to show that, even if a majority of the most important scientific databases have procedures for reporting errors, these are usually not publicly visible, making the process of reporting errors time consuming and not useful. Also, the effort made by the user that reports the error often goes unacknowledged, putting him in a discouraging position.

Introduction

An issue tracker software, known more informally as a bug tracker, is an application designed to keep track of all the problems and errors related to a software project or service. Compared to a simple mailing list, this software allows a finer organization on the discussions relative to each single detail to implement and error to solve, facilitating the development of the software.

In most open source projects, issue trackers are also the place where users and testers can contact the developers to suggest feature improvements. All the information on a report is generally shown publicly, informing other users of any inconvenience they may encounter and with the advantage that the discussion on how to solve a problem or improve a component is open.

Moreover, error tracker software can also be used as a way to determine the health and the status of the development of a software package, and to discern between projects that are actually discontinued from those that are active. If the error reporting application of a software contains a lot of reports and feature requests, it means that the software is used by an active community, the developers are facilitated in finding funding to maintain it active and in general the project is more likely to remain active in

Database, Vol. 2010, Article ID baq035, doi:10.1093/database/baq035

the long term. If the authors answer to bug reports and questions quickly, it means that the development is active and the code is likely to be of better quality.

Most of the problems faced by open software projects and that are solved by the use of a public bug tracker affect biomedically relevant databases as well. The data annotated in a scientific database can contain errors that may be spotted and reported by users, as well as unclear annotations.

Finally, the proliferation of scientific databases in the recent years has led to the accumulation of abandoned or discontinued resources (1, 2); and the introduction and usage of an error tracker would be a useful indicator to distinguish between abandoned databases from active ones.

Databases and research

Scientific databases form an essential part of the modern scientific community. The first example of database was the *Atlas of Protein Sequences and Structures* by Margaret Dayhoff, almost 50 years ago. Nowadays, the knowledge described in the scientific literature is reorganized by databases like UniProt (3–5), GenBank (6), KEGG (7–9) or Reactome (10, 11), and many others described yearly by the Nucleic Acid Research Database Issues (12, 13). The service made by a scientific database is to transform the information discovered in the scientific literature into a biological entity (a gene, protein, pathway, molecule), in order to facilitate access. While reading the literature is always an important task for a researcher, many large-scale studies on genomes and organisms would not be possible without a quicker way to access to all the information on each entity of the study. For example, it would be a time-consuming task to manually read all the literature on each protein in the human species in order to carry out an analysis of the complete human proteome.

However, given the complexity of the process of annotating information related to a biological entity, a percentage of erroneous, outdated or unclear data are expected even in the databases with the best annotation practices. Cases of errors in the annotations have been reported previously in the literature (14–19). Even if most databases follow rigorous procedures for annotating data, and often collaborate to experts in a field for manual annotation or review, a scientist with a deep knowledge of his field of specialization is in a better position to find errors or discrepancies in the information annotated by other people. Moreover, even for the cases where the annotation is made with the help of an expert, the data can become outdated or unclear over time.

Currently, most scientific databases already have a well-defined procedure to communicate with users, but it is generally based on private mails or personal communications. In our opinion, this procedure has a series of disadvantages, mostly related to the fact that they are not transparent. In thisarticle, we will show some examples on reporting errors in scientific databases related to the genes of the *N*-glycosylation pathway, with the double scope of showing that errors or unclear annotations may be present in any database, and that private communications fail to acknowledge the effort made by the user reporting the error, making the whole process more complex and time consuming than necessary.

We present here a testable hypothesis of the possible advantages of using an issue tracker to report errors in biological databases. This work is based on our experience reporting errors to many biological databases and do not relay on research data showing the effectiveness of the use of issue tracker for this kind of databases. The goal of this article is to promote the discussion about the importance of reporting errors and the tools available for it and to lay the ground for testing the usefulness of issue tracker for biological databases.

Disclaimer

By the date of publication, the errors and missing data described in this report have already been notified to the corresponding maintainers, and in most of the cases have already been fixed. A table with links to all the bug reports opened during the writing of this article is available in a Supplementary Data. We would like to state that the cases and examples described in this work should not be used for any interpretation on the quality of the annotation in the databases studied. The cases described here are provided only as examples of errors that can be found in a scientific database, in order to illustrate the process of reporting such incongruences to the correspondent maintainers.

Materials and methods

The pathway studied—biosynthesis of the *N*-glycan precursor pathway

N-glycosylation is one of the most important forms of protein post-translational modification. A search in the current Uniprot database shows that almost half of the transmembrane proteins known to date are potentially *N*-glycosylated, and an earlier work showed the same percentage in all the known proteins (20). *N*-glycosylation is important in order to achieve proper folding for most of the proteins in the secretory pathway, making this pathway very important for the fitness of unicellular and multicellular eukaryotes; moreover, knowledge of the steps involved in *N*-glycosylation is also important in the pharmaceutical industry and for the biotechnological production of drugs (21).

In short, the pathway of *N*-glycosylation is a good model to study the status of its annotation in existing databases

because it is well described in the literature, and most of the reactions described have not seen major revision in recent years. In this work, we have taken into consideration only the first step in the pathway of *N*-glycosylation, the synthesis of the common *N*-glycan precursor, since it is the most documented part. This part of the pathway constitutes one of the first biological processes to have been defined at the gene level, described in some reviews as early as the 1980s (22–24). This pathway is also well described in the book *Essentials of Glycobiology* (25), which has kept a complete annotation on this pathway for many years and which is also indexed in different databases. Thus, the structure of this part of the pathway is well established and the components and genes involved are known.

Databases studied

The databases described in the present work represent sample of the resources that would be used to study a pathway or a set of genes (Supplementary Table S1). Gene Ontology is widely used to study the function, localization and involvement in biological processes of a set of genes of interest. Uniprot is a useful resource for annotations on protein entities. String is a database of electronically inferred protein–protein interactions. Kegg-pathways and Reactome contain manually annotated pathways of processes of biological interest.

Results

We present the errors that we found in some databases during the curation of the N-linked glycosylation pathway and we will describe the process necessary to report them. Details of the error reports submitted and errors found for each database are listed in the Supplementary Tables S2–S4.

KEGG pathways

Kegg Pathways (79) (Kyoto Encyclopedia of Genes and Genomes, http://www.genome.jp/kegg/pathway.html) is a well-known database for pathways with high-quality annotations. The pathway for *N*-glycan precursor biosynthesis was annotated inside the entry hsa:00510. While the annotation was good and clean, the figure on the Kegg web page contained some simplifications and did not correspond exactly to the entry in the database (Supplementary Figure S1). For example, the output of the reaction catalyzed by the DPM complex, Dolychyl-P-Mannose, is used as substrate by the genes *ALG3*, *ALG9* and *ALG12*, while the entry in Kegg implies that it is used only by *ALG3*. Similarly, it is not shown that the *GANAB* genes actually catalyze two consecutive reactions, while the intermediate of these two is involved in a very important and complex mechanism called the Calnexin/Calreticulin cycle. In certain cases, an edge between two nodes represented a single reaction, while in others an edge corresponded to multiple ones. In brief, the figure representing the pathway on KEGG presented some simplifications which would be very difficult to spot for a user without good knowledge of the pathway of *N*-glycosylation. Apart from this inconvenience, the Kegg's user support center is very keen at answering any doubt, and answers are given quickly.

Reactome

Reactome (10, 11) (http://www.reactome.org) is an open source database for manually curated pathways, known for being especially open to submissions by users and for having a peer-review system for accepting new entries. In our case, we discovered that the *N*-glycan precursor synthesis pathway was only annotated for a small portion in Reactome and we proposed to the maintainers to submit a new entry.

The process for proposing a new pathway entry in Reactome is well defined and is assisted by a maintainer, who can explain the procedure and can respond to doubts. All the annotations submitted must be justified by an article showing evidence for the reaction and the data must receive the approval of a reviewer before it may be published to Reactome. In the final version released in the Reactome web site, every reaction of every pathway is provided with a button to a link where it is possible to send feedback and comments. Unfortunately, the discussion is not public, and is it not possible to see if a pathway or a reaction received comments.

GeneOntology

The Gene Ontology project (26) (http://www.geneontology.org) is aimed at standardizing the terminology used to describe genes and gene products in the scientific literature. Its purpose is to reduce the usage of synonyms and spelling-errors in the terms used when describing a gene.

Gene Ontology is very well annotated and complete, and it is one of the most actively maintained resources among the ones that we used. Their 'GO Requests' tracker on SourceForge (http://www.geneontology.org/GO.requests.shtml), where proposals for new terms are made, gets at least three or four new entries every day, all of which are answered within a few days. As a result of this efficiency, we were able to propose new terms and to report annotation errors without much delay, and most of the changes will be included in the next Gene Ontology release.

The biggest discrepancy we found in the Gene Ontology annotation was the case of the term 'N-linked glycosylation', which was used ambiguously in two different contexts. This term was associated both with genes which participate to the N-linked glycosylation process and with those which are targets of the *N*-glycosylation process but are not responsible for any reaction within it. After

Database, Vol. 2010, Article ID baq035, doi:10.1093/database/baq035

reporting the case to the GO maintainers, the error was fixed and explained publicly: 10 erroneously annotated genes were removed from the term, 33 were kept and 21 more were added (error report: http://sourceforge.net/tracker/index.php?func=detail&aid=2945847&group_id=36855& atid=605890).

We sent other reports involving suggestions for new synonyms of a term, addition of associations between some genes and a term and small refinements. In all cases, the response was quick, and in our opinion, Gene Ontology is a good model for handling user reports publicly.

String

String (27) (http://string.embl.de) is a database and a web application for protein–protein interactions. A complex algorithm merges the results from different databases and predictors for protein–protein interactions and calculates a *P*-value for each possible interaction. The String web site provides an interface for these results and allows the user to navigate through networks of interactions.

String is a good example of a metadata database, where all the annotations are derived from external sources and the annotation is inferred electronically. One could argue that such a database will not need a bug tracker for errors, as the original data derives from external sources that are out of the jurisdiction of String's authors. However, we think that even in this case it would be useful to have public reports on errors; as the meta-clustering algorithm produces many false positives and it is difficult to evaluate its effectiveness. If users were able to annotate which of the automatically inferred results are wrong and if these annotations were publicly visible, then it would be easier to use this database and the data in it.

To give an example of false positives that could be encountered in a metadata database, one of the most striking discrepancies we found was the case of a gene that was merged with another with a similar name. The information on String for *ALG2*, a gene that participates in *N*-glycosylation, was merged with the annotations for *PDCD6*, a gene involved in apoptosis and formerly known as *ALG-2* (Apoptosis-Linked Gene 2), so that the resulting predicted interactions were mixed (Supplementary Figure S2). We discovered that there was no way to report the error to the String database maintainers or to communicate to the other users a possible source of errors. Another error was that of a false negative result, in which an interaction between the *ALG1* and *ALG11* genes, which is even described in the title of an article (28), was not present in the database. A further point of confusion is the definition of the term 'interaction', which, after looking at the results of the clustering algorithm, appears to assume different meanings. As an example, again for the gene *ALG2*, different types of interactions were shown with the same symbol: metabolic interactions like the one linking *ALG2* with *ALG1*, a potential protein–protein interaction like *ALG2* with *ANXA11* and *PEF1*, and genes simply involved in the same pathway like the interaction between *ALG2* and *DPAGT1*. In this case, it would be good to have a public place where one may ask for clarifications from the authors and where different users may discuss the proper way to interpret an interaction in String. Figure S2 from Supplementary Data illustrates the false positives and negatives that we encountered for the *ALG2* entry.

UniProt

The Universal Protein Resource (UniProt, http://www.uniprot.org) (3–5) is a comprehensive resource for protein sequence and annotation data, originating in 2002 from the merge of three different centers for protein annotation. The majority of the sequences in UniProt are derived from the translation of DNA and RNA sequences deposited to DDBJ, EBML and GenBank, after a manual curation. It is one of the best resources for finding annotation on a protein, since its reviewing process is very well defined.

We found very few annotation errors in UniProt and those that we found were mostly small corrections that we made to the generic description of some proteins. For example, gene *ALG9* was described as associated with a bipolar affective disorders, an association originally described in ref. (29), but later retracted by the same author (30) (error report: http://www.uniprot.org/comment/Q9H6U8). Another point that required clarification was the naming of three different genes, *MAN1A1*, *MAN1A2* and *MAN1C1*, which in the literature appear differently. The UniProt interface allows users to leave comments and send feedback; however, the presence of a different procedure, one for leaving comments and the other for reporting errors, is a bit confusing. It is not clear why reporting an error related to a pathway should be done on a private communication, and it is not clear which kind of comments should be submitted as public. In a recent publication, Uniprot reported that they received only 9 comments on more than 46 million page views (we are probably the authors of three of the comments cited in the report). In our opinion, beside the good overall quality of the data in Uniprot, a possible explanation for this failure lies in the confusing distinction between comments and private reports, which is disorienting for the user and in the lack of acknowledgement for reporting errors.

Discussion

The scientific community can learn from how issues and errors are dealt in open source communities

Nowadays, many open source software are developed by communities of programmers who establish a channel of communication with the users of the software in order to

decide which features to implement and which are not needed. The concept of open source software has been developed by Richard Stallman in the 1970s, but the approach of developing a software among a community of programmers, that communicate over mails, usenet and later Internet, has been greatly innovated by Linus Torvalds for the development of the Linux kernel (31). These software projects usually communicate with their users through mailing lists and bug trackers, the latter being more appropriate to organize the discussion on several independent details. We believe that the scientific community could learn from how open source communities handle communication between programmers and users.

Although not always reported in the literature, on the recent years many efforts and discussion have been carried out about improving the feedback from researchers to scientific databases. A most recent report surveyed 50 researchers who had previously published papers characterizing genes and proteins, to ask them whether they would be interested in providing contributions to databases (32). Other results have been published in the GMOD Annotation Satellite Conference (33), where, among other issues, the lack of recognition for contributing to a public resource has been discussed. However, rarely has the discussion on Open Annotation has been directed toward the adoption of a tracker applications, which is a system that is already adopted with success by the open source community and that will be easy to implement even in databases that do not embrace Open Annotation practices.

A comparison of different issue-tracking systems can be found in the corresponding Wikipedia page (http://en.wikipedia.org/wiki/Comparison_of_issue-tracking_systems) and in Ref. (34).

Error reports should be public

The present work shows that even if most of the databases presented here provide a user interface to report errors, only in few cases the process of error reporting is public and accessible to the public.

The annotations on public scientific databases could be improved with the help of the community of scientists who use them. However, in order to obtain the best results and collaboration, the process for reporting errors and proposing features should be as transparent as possible and should recognize the effort of the contributor.

The lack of a public error tracker in a scientific database is unfair toward the users wishing to report errors. First, it makes the process more difficult, because there is no way to know whether a certain error has already been reported and not fixed yet. Second, if the process of reporting errors remains internal to the corresponding database, then the efforts made by the reporter are not recognized publicly. Reporting errors is a very time-consuming task and a researcher may need to justify the time spent on it to the founders; this is not possible without a publicly accessible link. For example, a young master or PhD student may wish to include a link of all reported errors sent in his annual fellowship report or even in his curriculum vitae. Finally, an error tracking application, such as a mailing list, represents a good place where users can propose improvements, request new features or discuss how to interpret the data shown.

Besides the issues for people wishing to contribute to improve the quality of the data, the absence of a public reporting system is also a problem for the people using the database. A database may contain errors already identified by other researchers, but not yet fixed in the actual data release; public reports will allow people to become aware of errors that are still in the evaluation phase. This problem would be especially evident in the case of databases that do not get updated frequently or that have been abandoned completely. In theory, the data annotated in databases that are not maintained anymore could still be of use with the support of a public bug tracker, where known errors that cannot be fixed because of lack of maintenance can still be reported.

Finally, we wish to remark that there is more than one way to interpret the annotations in a scientific database. If the discussion on how to interpret them is not public, it is likely that different researchers will interpret the same data differently. This issue is intrinsic to the problem of annotating data, and even with a well-specified ontology it is unavoidable. An effective way for the users to be aware of a possible alternative interpretation for an annotation is to have a publicly accessible space, where questions and doubts are clarified to any possible user of the data. However, we wish to note that it is not clear what the response of scientific communities that use biological databases would be if an issue tracker is made available by Biological databases, Nonetheless, the effectiveness of this tool should be tested giving the opportunity to users to take advantage of it.

Supplementary Data

Supplementary data are available at *Database* Online.

Acknowledgements

We would like to thank Brandon Invergo and Kevin Keys for grammatical revisions and their nice inputs; and Ludovica Montanucci, Pierre Luisi and Marc Pybus for their helpful discussion. We would also thank Maria Teresa Rodriguez Plata for valuable feedback and tips. We would like to thank the community at http://biostar.stackexchange.com/ for useful discussion. We are grateful to three anonymous reviewers for their comments and suggestions that greatly

improved the article. Bioinformatics services were kindly provided by the Genomic Diversity node, Spanish Bioinformatics Institute (http://www.inab.org).

Funding

This research was funded by grants SAF2007-63171 and BFU2010-19443 (subprogram BMC) awarded by Ministerio de Ciencia y Tecnología (Spain) and by the Direcció General de Recerca, Generalitat de Catalunya (Grup de Recerca Consolidat 2009 SGR 1101). G.M.D. is supported by a PhD fellowship from the Programa de becas FPI (BES-2009-017731) del Ministerio de Educación y Ciencia, Spain.

References

1. Wren,J.D. and Bateman,A. (2008) Databases, data tombs and dust in the wind. *Bioinformatics*, **24**, 2127–2128.
2. Merali,Z and Giles,J. (2005) Databases in peril. *Nature*, **435**, 1010–1011.
3. The Uniprot Consortium. (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.*, **38**, D142–D148.
4. Jain,E., Bairoch,A., Duvaud,S. *et al*. (2009) Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics*, **10**, 136.
5. Apweiler,R., Bairoch,A. and Wu,C.H. (2004) Protein sequence databases. *Curr. Opin. Chem. Biol.*, **8**, 76–80.
6. Benson,D., Karsch-Mizrachi,I., Lipman,D.J. *et al*. (2010) GenBank. *Nucleic Acids Res.*, **38**, D46–D51.
7. Kanehisa,M., Goto,S., Hattori,M. *et al*. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, **34**, D354–D357.
8. Kanehisa,M. and Goto,S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, **28**, 27–30.
9. Kanehisa,M., Goto,S., Furumichi,M. *et al*. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.*, **38**, D355–D360.
10. Vastrik,I., D'Eustachio,P., Schmidt,E. *et al*. (2007) Reactome: a knowledge base of biologic pathways and processes. *Genome Biol.*, **8**, R39.
11. Matthews,L., Gopinath,G., Gillespie,M. *et al*. (2009) Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.*, **37**, D619–D622.
12. Galperin,M.Y. and Cochrane,G.R. (2009) Nucleic Acids Research annual Database Issue and the NAR online Molecular Biology (2010) Database Collection in 2009. *Nucleic Acids Res.*, **37**, D1–D4.
13. Cochrane,G.R. and Galperin,M.Y. (2010) The 2010 Nucleic Acids Research Database Issue and online Database Collection: a community of data resources. *Nucleic Acids Res.*, **38**, D1–D4.
14. Schnoes,A.M., Brown,S.D., Dodevski,I. *et al*. (2009) Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comput. Biol.*, **5**, e1000605.
15. Bolchini,D., Finkelstein,A., Perrone,V. *et al*. (2009) Better bioinformatics through usability analysis. *Bioinformatics*, **25**, 406–412.
16. Baumgartner,W.A., Cohen,K.B., Fox,L.M. *et al*. (2007) Manual curation is not sufficient for annotation of genomic databases. *Bioinformatics*, **23**, i41–i48.
17. Naumoff,D.G., Xu,Y., Glansdorff,N. *et al*. (2004) Retrieving sequences of enzymes experimentally characterized but erroneously annotated: the case of the putrescine carbamoyltransferase. *BMC Genomics*, **5**, 52.
18. Jones,C.E., Brown,A.L. and Baumann,U. (2007) Estimating the annotation error rate of curated GO database sequence annotations. *BMC Bioinformatics*, **8**, 170.
19. Brenner,S.E. (1999) Errors in genome annotation. *Trends Genet.*, **15**, 132–133.
20. Apweiler,R., Hermjakob,H. and Sharon,N. (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta*, **1473**, 4–8.
21. Betenbaugh,M.J., Tomiya,N., Narang,S. *et al*. (2004) Biosynthesis of human-type N-glycans in heterologous systems. *Curr. Opin. Struct. Biol.*, **14**, 601–606.
22. Hubbard,C.S. and Ivatt,R.J. (1981) Synthesis and processing of asparagine-linked oligosaccharides. *Ann. Rev. Biochem.*, **50**, 555–583.
23. Huffaker,T.C. and Robbins,P.W. (1983) Yeast mutants deficient in protein glycosylation. *Proc. Natl Acad. Sci. USA*, **80**, 7466–7470.
24. Kornfeld,R. and Kornfeld,S. (1985) Assembly of asparagine-linked oligosaccharides. *Ann. Rev. Biochem.*, **54**, 631–664.
25. Stanley,P., Schachter,H., Taniguchi,N. *et al*. (2009) Essentials of Glycobiology, 2nd edn. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009. Chapter 8.
26. Ashburner,M., Ball,C.A., Blake,J.A. *et al*. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
27. Jensen,L.J., Kuhn,M., Stark,M. *et al*. (2009) STRING 8–a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.*, **37**, D412–D416.
28. Gao,X.-D., Nishikawa,A. and Dean,N. (2004) Physical interactions between the Alg1, Alg2, and Alg11 mannosyltransferases of the endoplasmic reticulum. *Glycobiology*, **14**, 559–570.
29. Baysal,B., Badner,J., Corona,W. *et al*. (2002) A mannosyltransferase gene at 11q23 is disrupted by a translocation breakpoint that co-segregates with bipolar affective disorder in a small family. *Neurogenetics*, **4**, 43–53.
30. Baysal,B.E., Willett-Brozick,J.E., Bacanu,S.-A. *et al*. (2006) Common variations in ALG9 are not associated with bipolar I disorder: a family-based study. *Behav. Brain Funct.*, **2**, 25.
31. Raymond,E.S. (2001) *The Cathedral & the Bazaar*. O'Reilly Media Publishing, p. 256.
32. Mazumder,R., Natale,D.A., Julio,J.A.E. *et al*. (2010) Community annotation in biology. *Biol. Direct*, **5**, 12.
33. Clements,D., Adlem,E, Anders,G. *et al*. Community Annotation September 2010 Satellite. c2010. In: The Generic Model Organism Database (GMOD) wiki [Internet]. http://gmod.org/wiki/Community_Annotation_September_2010_Satellite (12 September 2010, date last accessed), Cambridge (England).
34. Janák,J. (2009) Issue Tracking Systems [dissertation]. Brno, Masaryk University, Faculty of Informatics. Available from: http://is.muni.cz/th/60778/fi_m/thesis.pdf (12 September 2010, date last accessed).