Author’s response to reviews

Title: Distribution of immunodeficiency fact files with XML - from Web to WAP

Authors:

Mr Jouni Valiaho (jouni.valiaho@uta.fi)
Pentti Riikonen (pentti.riikonen@it.utu.fi)
Prof Mauno Vihinen (mauno.vihinen@uta.fi)

Version: 2 Date: 16 Jun 2004

PDF covering letter
Dear Sir/Madam,

we hereby submit a revised version of our manuscript entitled “Distribution of immunodeficiency fact files – from Web to WAP” (MS: 5876490813157291). We have taken into account the constructive criticism of the Reviewers. A paragraph is added to describe in more detail how the knowledge base (IDR), based on the IDML markup language, is maintained and constructed.

One of the Reviewers is critical about the contents of the manuscript. He expects IDML to be a standard approved by a standardization organization. By standard we mean the systematic description of the genetic, biochemical, clinical and other features related to immunodeficiencies. Bioinformatics is a field where there are not many standardization office approved standards, however, there are numerous de facto standards for example the sequence formats of GenBank and EMBL. By the way, the word “standard” appears together with “bioinformatics” in 220 articles in PubMed and Google finds over 265 000 such articles in Internet. We feel that a systematic, methodological markup is a standard way to present data.

This referee is concerned about how well defined the elements are, especially the “Clinical Description”. This element is first of all written by our team, there are no submissions from outside. In addition, we use expert curators for every single individual disease. Our group takes care that the texts are of the same style and vocabulary throughout the service. Immunodeficiencies form a large and heterogeneous group of very different diseases. Therefore the description has to be flexible.

IDML development was started already on 2001 and it was released by 2002. There are some other clinically oriented markup languages, as well. IDML fits well to our purpose, it is also so generic that it can be applied to other diseases. Currently, IDML has not been used to build up other services. This was asked also by the other Reviewer.

This Referee considers the paper to have poor organization. The article is about IDML. We do not understand the unjustified statement that the markup were bad. IDML produces XML-acceptable code and the IDML is perfectly suited for our purpose to collect and distribute data on PIDs. All the fact files are freely available in our server, so there is no need to include them in the article. In addition, they are modified at least on a weekly basis.

If there is a problem with the title of the article, we could change it. However, the title was written to describe the actual contents and the goal of the work.

We have added a paragraph of the IDR to the text to address the point 3 (difficult to understand). As stated in the added text, the data items in the service are collected from many sources by using Perl scripts and some features are encoded manually. This manuscript is about IDML, how the markup language (on top of which the IDR is constructed) has been designed. IDR contains fact files for practically all known immunodeficiencies. As new forms of diseases are described new entries are added to the service.
The other Reviewer had just mainly technical comments and questions.

As said in the added paragraph, IDR is maintained by our group.

Some of the comments seem to indicate that the Reviewer has been looking also to some other services maintained and distributed by us. We provide several other services and databases in addition to IDR. Currently there are 82 immunodeficiency mutation databases and IDdiagnostics service about genetic and clinical tests for immunodeficiencies. IDML is used only for the IDR, so there are e.g. no direct submissions.

We have a close collaboration with clinicians and experts on these diseases. They are in fact in charge of the validation of the information in the service. However, only our group has access to the information encoded to IDR.

The data in IDR is stored into an XML database, which is accessible via the IDR web service. Several pages are provided by extracting information on demand from the fact files. The IDML schema sets a minimum requirement for the data elements of fact files. It is a necessary tool for the validation process and ensures data integrity at minimum level for interoperation and mediation between applications.

The data is collected and distributed by us and further validated by the disease experts. The users of the IDR service can search different types of data, which is available either in our service or provided by others. Depending on the database user’s needs different items are presented.

We hope that the manuscript could be accepted with these modifications to BMC Medical Informatics and Decision Making.

Yours sincerely,

Jouni Väliaho
Mauno Vihinen