1 Introduction

A SNP is a single nucleotide change in the genome sequences. SNPs often modify amino acids in a gene resulting in the diversity of individual responses to stimuli, e.g. drugs. They can even cause serious diseases.

Thus, many countries and international consortiums have been carrying out a large scale survey on SNPs data. In the meantime, databases are available on-line, e.g. 

JSNP(http://snp.ims.u-tokyo.ac.jp/),
HGVbase(http://hgvbase.cgb.ki.se/),

However, their data structure and access methods are heterogeneous. Therefore, DNA Data Bank of Japan (DDBJ) and Japan Biological Informatics Consortium (JBIC) have analyzed their systems and developed Polymorphism Markup Language (PML) based on XML. It is now possible to assemble and analyze SNPs data from distributed heterogeneous database in XML documents described by PML.

2 Model and Method

2.1 International Activity

In 2003, JBIC organized the 1st International Biodata Interoperability Conference to have international discussions on SNPs data standardization. It is an outcome of the conference that JBIC together with European Bioinformatics Institute submitted PML to Life Science Research (LSR) of Object Management Group (OMG) in response to the request for proposal for the standardization of SNPs data. JBIC hosted in June 2004 the 2nd International Biodata Interoperability Conference to improve PML. If you want to more information, please see http://pml.ddbj.nig.ac.jp/. This portal site needs ID and password. You can get it by inquiring of JBIC (uketsuke@jbic.or.jp).
2.2 Model

1. Platform Independent Model (PIM)

PIM is a model independent of the implement platform. PIM is expressed in UML. We built PIM, considering the use case of exchange of polymorphism data. Figure 1 expresses the overview of PIM.

2. Platform Specific Model (PSM) – PML

We developed Polymorphism Markup Language which is a model depending on the platform of XML as PSM. The platform specific model for XML derived its architecture according to the XML as suggested by W3C document “Extensible Markup Language (XML) 1.0 (http://www.w3.org/TR/REC-xml/)”.

2.3 Method

We revised and improved the data model and the original PML by face-to-face meetings, e.g. meetings of working groups and two international conferences that the supporters of the proposal to OMG attended. An electronic mailing list was also heavily used.

3 Discussion

We have proved that existing databases can be expressed in PML format. We'll discuss the following three points further. We're going to extend PML with a view to expanding its application in healthcare and pharmaceutical industry.

1. The proposal shall discuss design decisions relating to the overall performance and efficiency of the query mechanism.
2. The proposal shall discuss interoperation with existing systems and representations.
3. Proposals shall discuss how the proposed model will represent information in each of the following use cases:
   - Download data from a SNP data repository.
   - Upload data to a SNP data repository.
   - Exchange data between researchers.

4 Acknowledgments

This work was supported by European Bioinformatics Institute, Karolinska Institute, Sweden, Helsinki University, Finland, National Center for Biotechnology Information (NCBI), USA, Cold Spring Harbor Laboratory (CSHL), USA, Stanford University, USA, Yale University, USA, Shanghai Center for Bioinformation Technology, China, Shanghai Information Center for Life Sciences, China, School of Medicine, Tsinghua University, China, Institute of Genomics & Integrative Biology (IGIB), India, National Institute of Genetics (NIG), Japan, Japan Science and Technology Agency (JST), Japan, National Cancer Center Research Institute (NCCRI), Japan, Tokyo Institute of Technology, Japan, Ministry of Economy, Trade and Industry, Japan.

5 References