Semantics of the Integrated BioMedical Database Project - A Japanese National Project -

Jun Nakaya[1)*], Saburo Sakota[2)], Riichiro Mizoguchi[2)], Kouji Kozaki[2)], Kaei Hiroi[1)], Keisuke Ido[1)], Michio Kimura[3)], Hiroshi Tanaka[1)]
1) Tokyo Medical and Dental University (TMD)
Yushima 1-5-45
Bunkyo, Tokyo 102-0094 Japan
2) Osaka University, The Institute of Scientific and Industrial Research (ISIR)
3) Hamamatsu University, School of Medicine
* Responsive Author

***Abstract*- To utilize dispersed information effectively in Omics medicine that stands on a massive amount of Clinical Omics information, the first hurdle to overcome is to establish a social IT infrastructure that has Electronic Health Record (EHR) and integrated biomedical database with standardized technologies derived from HL7/CEN/ISO. In preparation for Omics Medicine, Japan started the Integrated biomedical database project as a National project of Japan on 2007. This project tries to establish semantic integration, semantic navigation, and semantic interoperability with the Clinical Omics ontology which integrates clinical ontology and omics ontology. This paper introduces the project, its semantic bases, and the status of it.**

I. Introduction

Clinical omics informatics is a kind of informatics which tries to support clinical omics medicine. Here the clinical omics medicine is a kind of clinical medicine based on omics information. "Omics" is a coinage of "ome" and "ics". Roughly it means whole molecular information in a human. It includes genomics, transcriptomics, proteomics, metabolomics, and so on. In this post genomic era, researchers are trying to connect all omics information to phenomics which is essential expression results in a human. To approach it, many scientific domains are challenging to integrate these hierarchical omics information. Additionally, in order to apply these information in clinical medicine, we need information about abnormal state information such as disease, diagnosis, treatment, and causal environment [1].

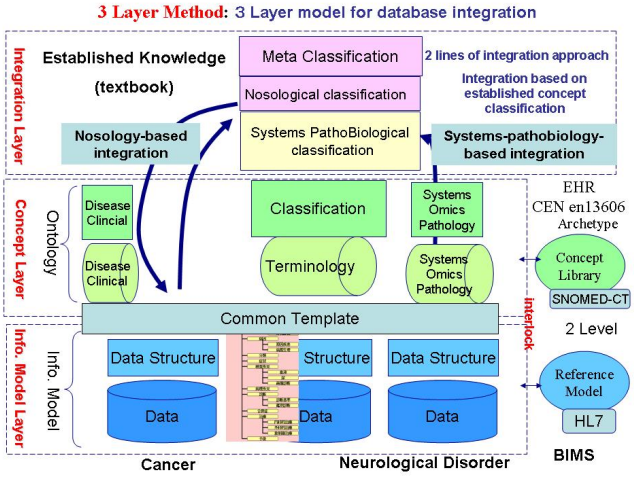
Here, what Omics Medicine needs as social informational infrastructure are standardization, security, and translational research informatics (TRI) [2]. The standardization is essential for sharing, exchanging, and effective use of Omics information that has enormous amount of information. The security is essential to protect personal information, considering that Omics information is a kind of ultimate personal information. The TRI is also essential to support Translational Research which is a social pipeline to make experimental medicine such as omics medicine real [3]. The important two technologies are omics Electronic Health Record (EHR) and integrated biomedical database; these must be coupled and these must be liaised effectively. As elemental techniques, three elemental techniques such as the ontology, the data exchanging format, and the information models are important. Two social foundations such as the standardization and the security are essential to work out this infrastructure, those are derived from ISO/CEN/HL7. Here information itself must be utilized actively, and this information must be in social informational circle consisting of social facilities such as hospitals, research organizations, insurance organizations, and governmental offices.

As a part of a social IT infrastructure, this project tries to construct the Japanese integrated biomedical database for researches.

II. The Integrated BioMedical Database

The Integrated BioMedical Database project started at 2007 in Japan as national project of Ministry of Education, Culture, Sports, Science and Technology (MEXT). This project tries to integrate BioMedical Databases in Japan as a national project.

Figure 1. 3 Layer Methodology for biomedical database integration.



The Organization

The project has an established national organization. The central core institution controls integration of databases of all biological species. Under the central core institution, Clinical Omics Database Center (CODC) is responsible for integration

of clinical databases and omics databases for human. Major members of CODC are TMD University and Osaka University. As associate centers, Tokyo University group is responsible for SNP related databases, and Kyoto university group is responsible for pharmaceutical databases and metabolism related databases. CODC has liaison council with Japanese national ontology project, Japanese EHR project, and ISO GSVML project [4]. The project also has ethics panels in core institution, TMD group, and each facilities.

Semantic Integration

The integration is established virtually with intelligent data format which consist of disease ontology, disease data exchange format, and disease information model. Based on semantic interoperability with meta ontology that can map domain ontologies, we integrate many distributed biomedical databases virtually in semantic way.

To achieve above semantic integration, we developed 3 layer methodology having 3 hierarchical layers such as the information model layer which can correspond to HL7 information models, the concept layer which can correspond to en13606 and/or SNOMED-CT, and the integration layer (Fig. 1). Based on the 3 layer methodology, we integrate databases semantically with 2 types of integration approach such as the nosology based approach and the systems pathobiology based approach. The nosology based approach is premised on Japanese disease terminology and classification which can correspond to SNOMED-CT. We use common template for both data format and ontology in whole disease knowledge space (all clinical fields).

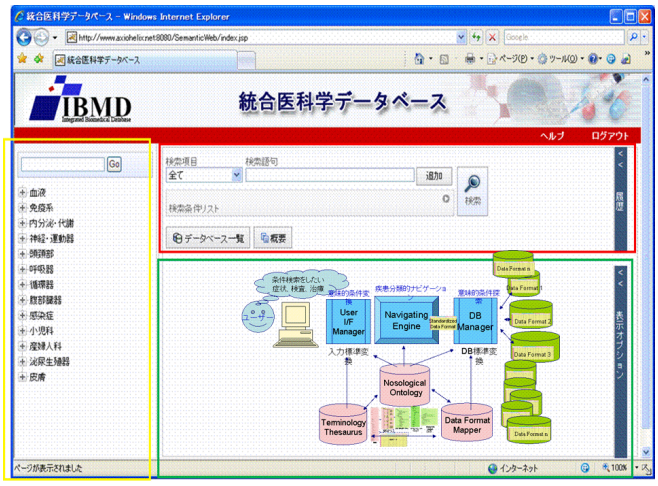


Figure 2. Sample of Semantic Navigation System.

A test disclosure of an early stage version of the integrated biomedical database to the specified assessment persons started at April 2008. The element databases of the early stage version are Cancer Omics database of TMD and Neurological Disorder database of Osaka University. Total clinical cases of the early stage version are 200 clinical cases at year 2008. The cases will increase every year.

Semantic Navigation

We developed a prototype of the semantic navigation system based on the clinical omics nosology. This system provides a semantic navigation to the targeted databases (Fig. 2). The User Interface Manager converts a user input to standardized words with the dictionaries of the terminology and the thesaurus. The Navigating engine navigates users to pertinent databases semantically based on the Nosology. The Data format mapper absorbs format differences of the elemental databases. Currently a prototype is disclosed to the specified persons for assessment.

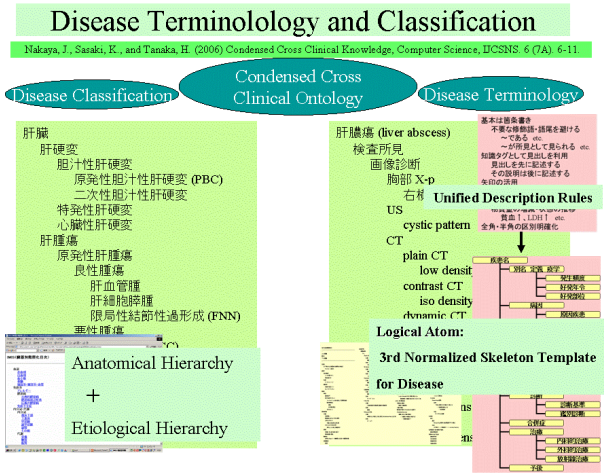


Figure 3. Disease ontology based on EBLA.

Semantic Interoperability

This project also tries to get over Language differences and domain differences with semantic interoperability. Basically this project start with Japanese language as an entry language, in future the ontology mapping with English ontology such as SNOMED-CT or ICDs will be challenged [5].

III. Clinical Omics Ontology

Based on our clinical ontology [6], by applying our ontological architecture for integration named as Evidence Based Logical Atomism (EBLA) [7], we developed a framework of clinical omics ontology. This architecture is also an ontological base of ISO-GSVML and Japanese EHR project. EBLA architecture defines atomic core (conceptual atom) and its evidences at each hierarchy. Fig. 3 is a sample of our disease ontology based on EBLA. Disease classification is based on a combination of an anatomical hierarchy and an etiological hierarchy. Disease terminologies are described with the 3rd normalized skeleton template. Based on EBLA,

we also developed a surgical ontology framework and its intelligent data format for robot-assisted surgery [5].

IV. Discussion

In the post genomic era, too explosive amount of information needs processing huge mass of concepts in computers for effective processing. In many post genomic projects including our national projects, semantic processing is getting to be an essential solution for smooth and efficient computerization.

Here one inevitable major fundamental issue of Clinical Omics semantic approaches is how to integrate ontologies. Basically these hurdles must be on the differences of the recognition systems, and it must be summarized in 3 points such as the grading scale, the scope difference, and the thoughtway mismatch between the clinical ontology and the omics ontology. EBLA tries to get over these points by introducing hierarchically modified logical atomism [8].

To utilize this integrated biomedical database in public, the major issue is the ethical issue of disclosing the personal data. The border of the personal and the public is not clear, and the limitation of the personal data is not easy, especially in omics field. Through this project, we will keep discussing about this issue and will keep trying to get close to the solution of it.

Acknowledgment

This project is funded by the Japanese Government (MEXT). We much oblige to the members of MEXT and the central core institution of the Integrated BioMedical Database project who advised us and supported us.

References

[1] Jun Nakaya, Clinical Genome Informatics (CGI) and its Social Information Infrastructure, IJCSNS, 7(1), 55-59, 2007

[2] Nakaya J, Shimizu T, Tanaka H, Asano S : Current Translational Research in Australia and Translational Research Supporting Center (TRSC) in Japan, Chem-Bio Informatics Journal, Vol. 5, No. 2, 27-38, CBIJ1595, 2005

[3] Jun Nakaya, The Translational Research Informatics (TRI) (Leadoff Article), International Journal of Computer Science and Network Security, 6(7A), 117-122, 2006

[4] Jun Nakaya, Hiroi, K., Yang, W., Ido, K., Kimura, M. Genomic Sequence Variation Markup Language (GSVML) for Global Interoperability of Clinical Genomics Data (Best Paper Award) APAMI2006 A01: 1-8. 2006

[5] Jun Nakaya, Hiroshi Tanaka. Intelligent Data Format for Robot Assisted Surgery -Chellnges and Solutions in software control -, Stanford University Lecture Series: US-ATMC "Advanced Technologies for Bio-med Applications" seminar series, Palo Alto, 2007

[6] Nakaya, J., Sasaki, K., and Tanaka, H. Condensed Cross-Clinical Knowledge. Computer Science, International Journal of Computer Science and Network Security. 6 (7A) 6-11. 2006

[7 Jun Nakaya, Tetsuo Shimizu: Knowledge Architecture based on Evidence Based Logical Atomism for Translational Research, International Journal of Computer Science and Network Se-curity, vol. 6-2, ISSN:1738-7906, 175-179, 2006

[8] Jun Nakaya, Shimizu, T., and Tanaka, H. Knowledge Platform to use BioMedical Semantic WEBs individually: Bridging Gene Ontologies and Clinical Ontologies at individual site (Leadoff Article) International Journal of Computer Science and Network Security 6 (2) 1-9, 2006