Japanese databases for the life science field, such as DDBJ (DNA Data Bank of Japan) of the National Institute of Genetics, PDB (Protein Data Bank Japan) of the Institute for Protein Research, Osaka University, and KEGG (Kyoto Encyclopedia of Genes and Genomes) of the Institute for Chemical Research, Kyoto University are highly regarded in the international community.
Although there has been lots of Japanese government aided projects supporting genome and post-genome research and they have produced enormous amounts of raw data and annotation results, the environment to utilize those information has been poor. Accordingly, MEXT (Ministry of Education, Culture, Sports, Science and Technology) launched the Integrated Database Project in FY2006 as a one-year project. From FY2007, the project was reinforced as a four-year project and Database Center for Life Science (DBCLS) was newly founded in the Research Organization of Information and Systems as its core institute.

Here in Japan, we consider the integration of life science databases an endless process, because we have to cope with ever changing needs of users. We do not aim at constructing just a unified database, but we aim to improve productivity in research and development. We also aim to handle unstructured data such as research papers, patents and textbooks with figures and images as well as text data. In this project we consider research in life science and developers in bioindustry, planners and evaluators of life science projects and constructors of life science databases as our users. We are integrating literature and phenotypic data with clinical information as well as molecular data. We are considering how to build a portal site that shows how to find and how to use various databases and analysis tools; to construct an integrated database of molecular data, literature and knowledge; to develop retrieval function that makes perspective acquiring and hypothesis building feasible; to produce indices, dictionaries, standard formats and database construction tools for database developer use; and to make teaching materials for curators, annotators and database engineers; as well as to present various functions in Japanese.

In FY2006, we focussed on four areas of database integration strategy; technology development relating to database integration; portal site prototyping and training methods for curators and so forth. To study database integration strategies, we organized a committee that enables us to get experts’ opinions. To provide adequate information to the committee, we also engage in research on related algorithms, and current status of medical data and genomic annotation as well as a preliminary study of integration of meeting proceedings and construction of a table of contents of databanks to obtain overall perspectives. As for the developing technology, we have produced a few kinds of dictionaries or thesauruses, such as gene name thesaurus, scientific names to a Japanese dictionary and a dictionary on Japanese institutes.
We also developed several kinds of technologies, such as anatomical tagger for animals and plants, anatomy dictionary that maps anatomical terms to 3D human bodies, cut and collection tool for web pages, and knowledge expression study for cancer or gene expression. Portal prototyping offers a life science database portal containing 371 databases with novel classification and brief descriptions and a web resource portal containing 456 analysis tools with functional classification. With regards to training methods, we began to make textbooks for annotators and database managers. In addition, we tried preliminary integration using human gene expression data and plant EST data.

From FY2007, we restarted the project in a more reinforced framework. Three satellite institutions and fourteen research groups in total are participating in this project. DBCLS are mainly engaged in the following researches and developments with seven other research groups: projection of database integration strategy based on researches of trend and users' opinion and examination of intellectual property situation around database; development of technologies, such as the construction of human 3D model, retrieval of figures and drawings, scientific workflows, and so forth; construction of integrated database for human and development of dictionaries and classification cabinets for plants and microbes. In addition, we continue to build the portal site and implement activities on training of human resources. Satellite institutions are engaged in the integration of drugs and chemical compound information, clinical and omics data for specific diseases, and genetic polymorphism and clinical data.

We hope that our activities contribute to the promotion of life science research internationally, especially in Asia Pacific region.

Contact Details:

**Keiichi Nagai**

Address: Database Center for Life Science, Research Organization of Information and Systems, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Email: k-nagai@dbcls.rois.ac.jp, tt@k.u-tokyo.ac.jp

**Toshihisa Takagi**

Address: Database Center for Life Science, Research Organization of Information and Systems, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Email: takagi@dbcls.rois.ac.jp, tt@k.u-tokyo.ac.jp