

The past and present status of integration of life science databases



Nov 2000

Genome Science Committee, Panel on Life Science, Council for Science and Technology
"Our strategy in Genome Informatics" (Nov 17, 2000)
- Strategic proposals based on three factors were presented: human resource development, DB construction, and technological development for information analysis

Apr 2001

JST's Institute for Bioinformatics Research and Development (BIRD) was established.

Aug 2005

Working Group on Strategy for Database Preparation, Life-Science Committee, Subdivision on Research Planning and Evaluation, Council for Science and Technology
"Appropriate state of the strategy for DB preparation for life science in Japan" (May 17, 2006)
- Urgent issues were presented: establishment of a strategic commission, launch of a portal site, technological development for DB integration, and human resource development

Sep 2006

The Integrated Database Project by MEXT was started, in which the Research Organization of Information and Systems plays a central role.

Apr 2006

Integrated DB projects by MAFF and METI were started.


Dec 2008

Integrated Database Task Force, Life Science Project Team, Council for Science and Technology Policy
"Integrated Database Task Force report" (May 27, 2009)
- Unified operation of the Database Center for Life Science (DBCLS) and the BIRD was proposed.

Apr 2011

JST's National Bioscience Database Center (NBDC) was established. The Life Science Database Integration Project was started and 1 subject for the Core Technology Development Program and 10 subjects for the Database Integration Coordination Program (DICP) were adopted.

Dec 2011

The four-ministry joint portal site for life science DBs was launched.
<http://integbio.jp/en/>  **integbio.jp**

Apr 2012

1 subject for the DICP was adopted.

Sep 2013

8 subjects for the Tool Prototype for Integrated Database Analysis (TPIDA) were adopted.

Apr 2014

9 subjects for the DICP were adopted.

Sep 2014

4 subjects for the TPIDA were adopted.

Apr 2015

2 subjects for the DICP were adopted.

May 2015

3 subjects for the TPIDA were adopted.

Inter-ministerial cooperation: steps toward database integration

Integbio DB Catalog

Collected information on DBs from MEXT, MHLW, MAFF, and METI are provided on the Integbio DB Catalog

Cross Search

One-stop search for MEXT, MHLW, MAFF, and METI DBs was made possible with shared index data and standardized search formats

LSDB Archives

DB archives are collections of the data created by MEXT, MHLW, MAFF, and METI based on the common guidelines.

MEXT / NBDC



MHLW / Natl. Inst. of Biomedical Innovation, Health, and Nutrition



MAFF / Natl. Inst. of Agrobiological Sciences



METI / Molecular Profiling Research Center for Drug Discovery, Natl. Inst. of Advanced Industrial Science and Technology

Database reconstruction

Construction of integrated DBs based on semantic web technologies

● MEXT: Ministry of Education, Culture, Sports, Science and Technology ● MHLW: Ministry of Health, Labor and Welfare
● MAFF: Ministry of Agriculture, Forestry and Fisheries ● METI: Ministry of Economy, Trade and Industry

National Bioscience Database Center (NBDC), Japan Science and Technology Agency (JST)

5-3, Yonbancho, Chiyoda-ku, Tokyo 102-8666, Japan

Tel. +81-3-5214-8491 Fax. +81-3-5214-8470

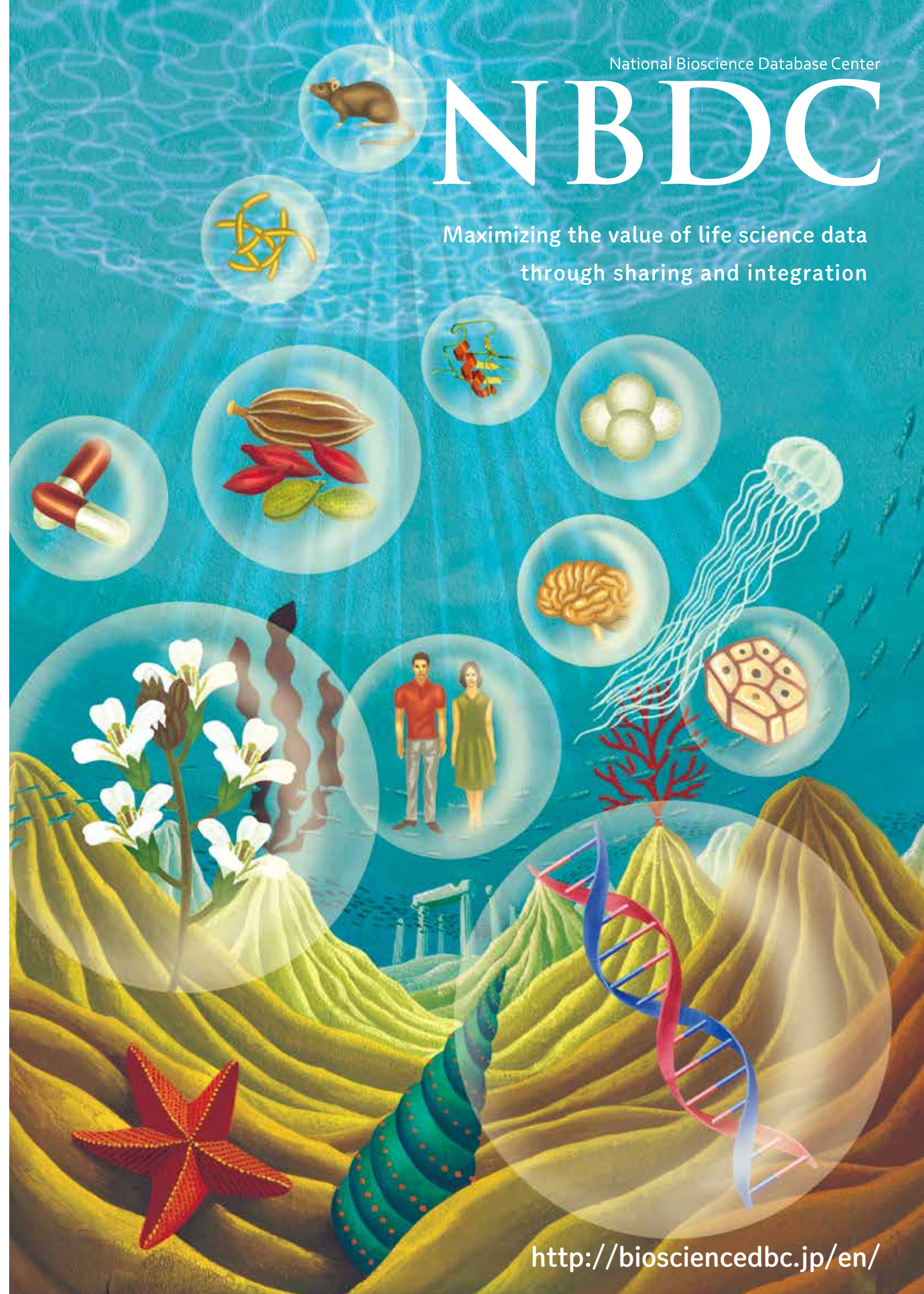


2016.01

National Bioscience Database Center

NBDC

Maximizing the value of life science data through sharing and integration



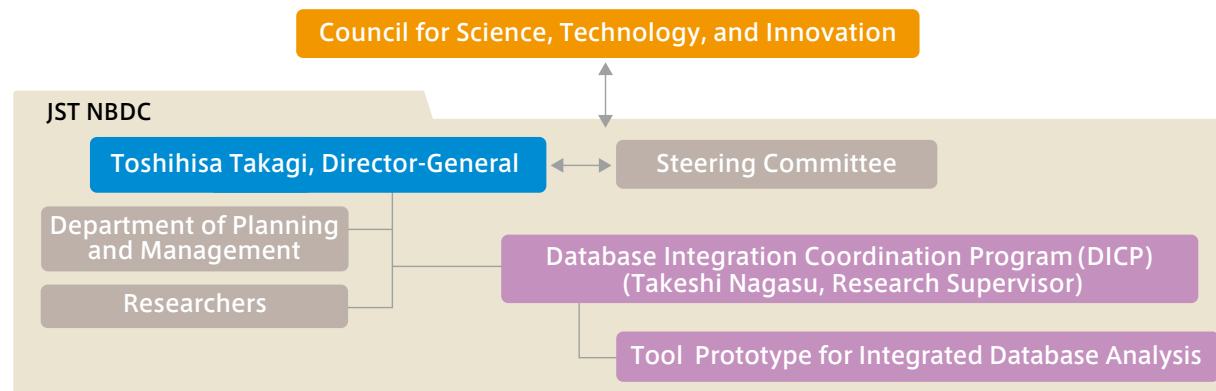
<http://biosciencedbc.jp/en/>

The National Bioscience Database Center (NBDC) is committed to integrating life science databases in Japan.

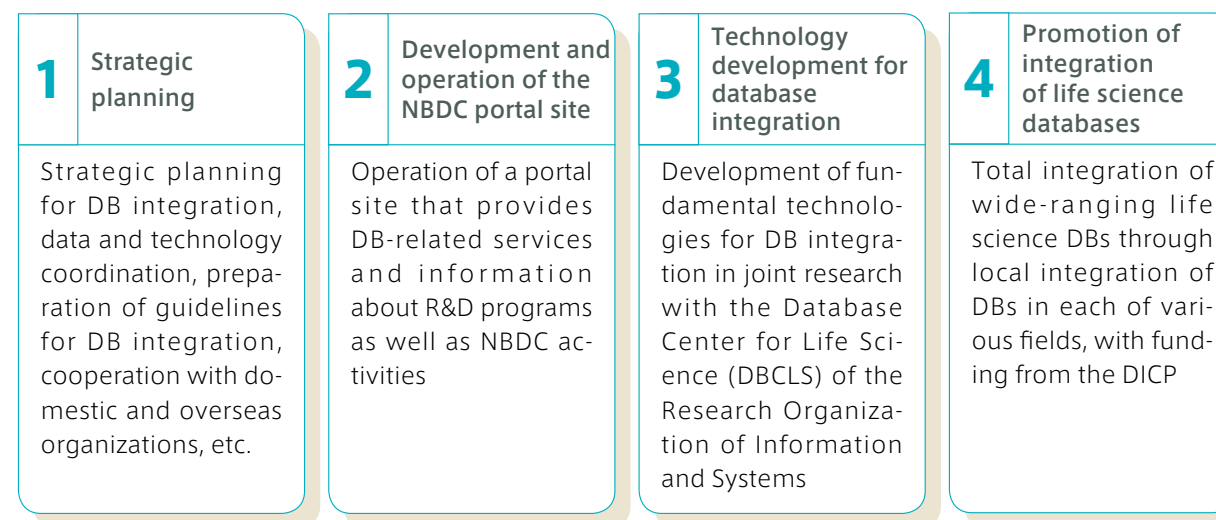
Toshihisa Takagi, NBDC Director-General
(Department of Biological Sciences,
Graduate School of Science, The University of Tokyo)



Organization

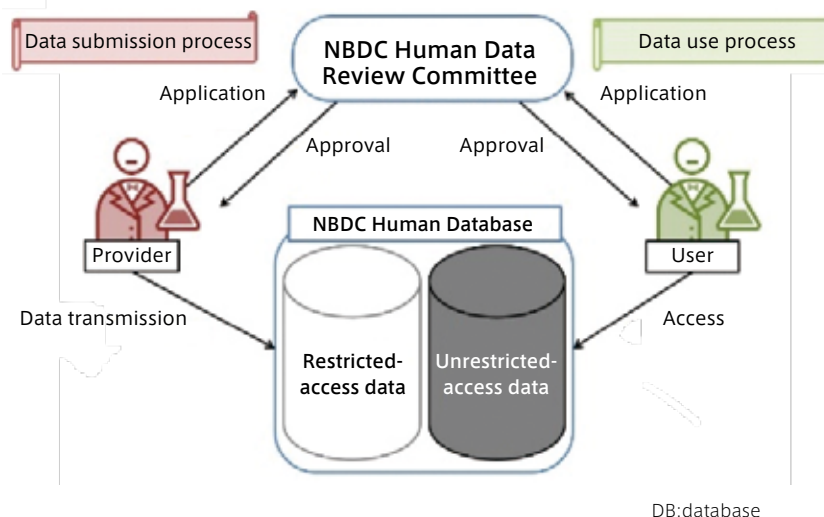


Life Science Database Integration Project



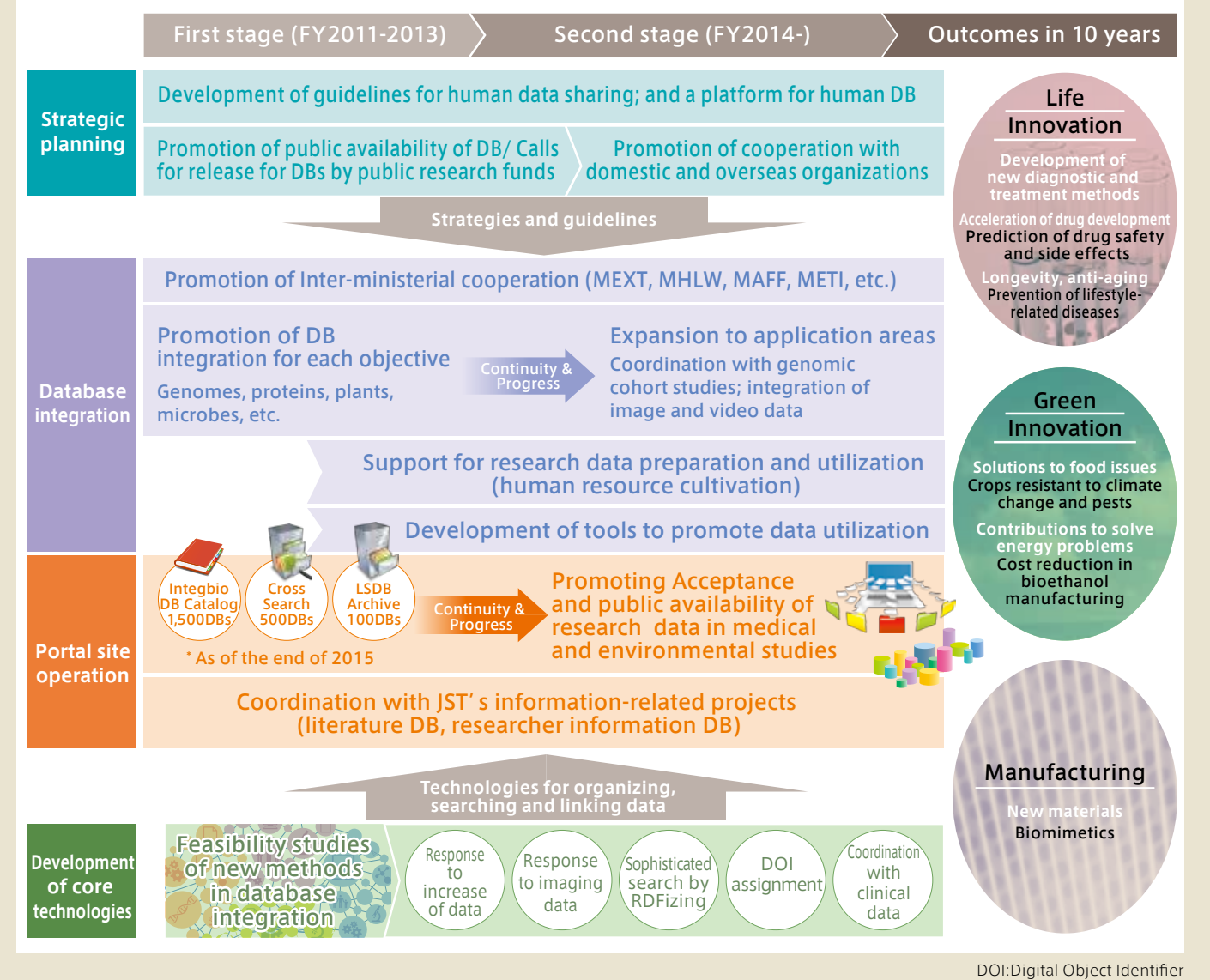
NBDC Human Database <http://humandbs.biosciencedbc.jp/en/>

The NBDC Human Database was established as a platform for promoting the sharing of human data while taking personal information protection into account. It is now operated in collaboration with the DDBJ Center at the National Institute of Genetics (DDBJ): DNA Data Bank of Japan). The NBDC Human Database organizes and stores human data, which is generated in enormous quantities with next-generation sequencers and other analytical technologies, and provides mechanisms, rules, and guidelines for effective use of such data for progress in the life sciences.



DB:database

Roadmap of the Life Science Database Integration Project



DOI:Digital Object Identifier

Public Relations

Symposia and events

- **Symposium of Database Togo (Integration)** <http://events.biosciencedbc.jp/sympo/> (Japanese version only)
October 5th was designated as the Day of Database Togo (Integration) and a symposium is held annually where participants discuss issues concerning life science DB integration.
- **BioHackathon (international developer workshop)** <http://www.biohackathon.org/>
Providers of major life science DBs and researchers and software developers specialized in DB integration technology from Japan and abroad gather at this workshop. They hold discussions and develop technologies to solve current DB-related issues.



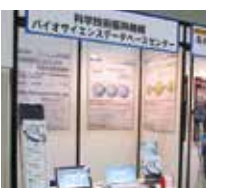
Seminars <http://events.biosciencedbc.jp/training/> (Japanese version only)

We organize seminars for beginners that introduce them to DB integration activities and how to use life science DBs and related tools. Participants engage in hands-on practice to learn how to use DBs and tools.



Exhibitions <http://events.biosciencedbc.jp/exhibition/> (Japanese version only)

At private-sector exhibitions such as BioJapan and exhibitions of academic conferences, we present how to use DBs and relevant services using posters, handouts, and computer demonstrations.



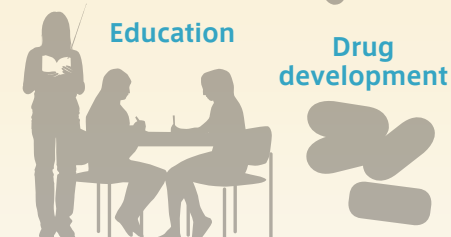
NBDC activities

The National Bioscience Database Center (NBDC) conducts research and development and provides various services to achieve integration of life science DBs in Japan. These efforts are intended to encourage widespread sharing of life science research outputs in the researcher community. It is expected that such sharing will stimulate academic research as well as industrial research.

Various outcomes

It is expected that utilization of the NBDC services for life science research and in medical and educational fields will lead to scientific discoveries, the development of new drugs, and other advances.

Medical applications



Users

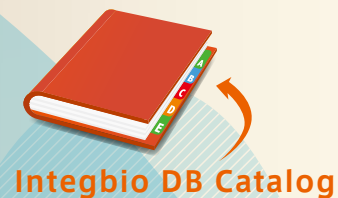


Services

NBDC portal site

<http://biosciencedbc.jp/en/>

Access to information collected from various fields at this one-stop portal site



Integbio DB Catalog

Find the required DBs
<http://integbio.jp/dbcatalog/en/>

Sophisticated search and DB tools



Life Science Database Cross Search

Search across multiple DBs
<http://biosciencedbc.jp/dbsearch/?lang=en>

LSDB Archive

Download DBs

<http://dbarchive.biosciencedbc.jp/index-e.html>



NBDC Human Database

Accept various types of human data and make them publicly available as restricted- or unrestricted-access data
<http://humandbs.biosciencedbc.jp/en/>

Data deposition

Data, DBs, databank information, tools, dictionaries, and ontologies in Japan

Japan



Research and development

Information technology

Build a sophisticated platform for DB search and provide new tools. Develop information technologies supporting RDFization in life science fields.

Search



Organize



Link

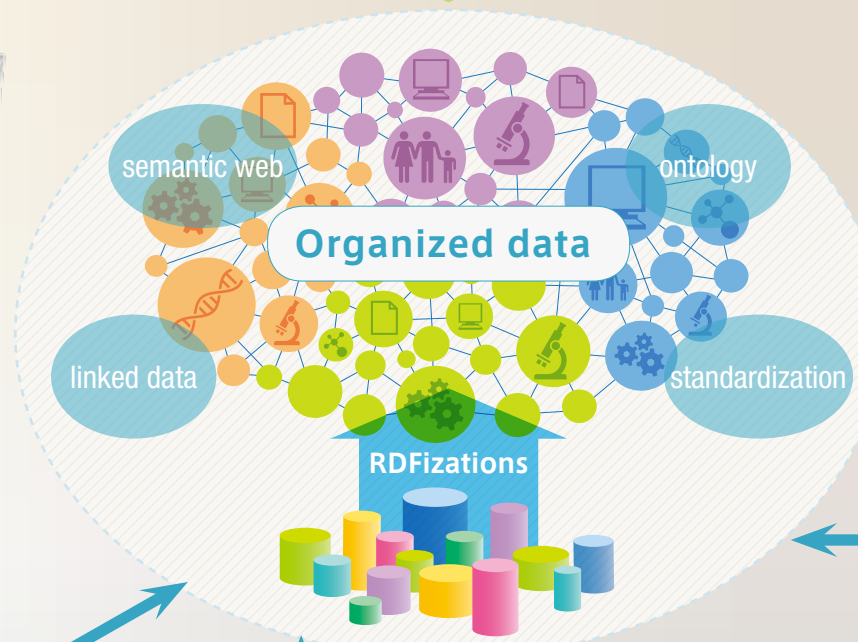


Support for RDFizations, etc.

Extract



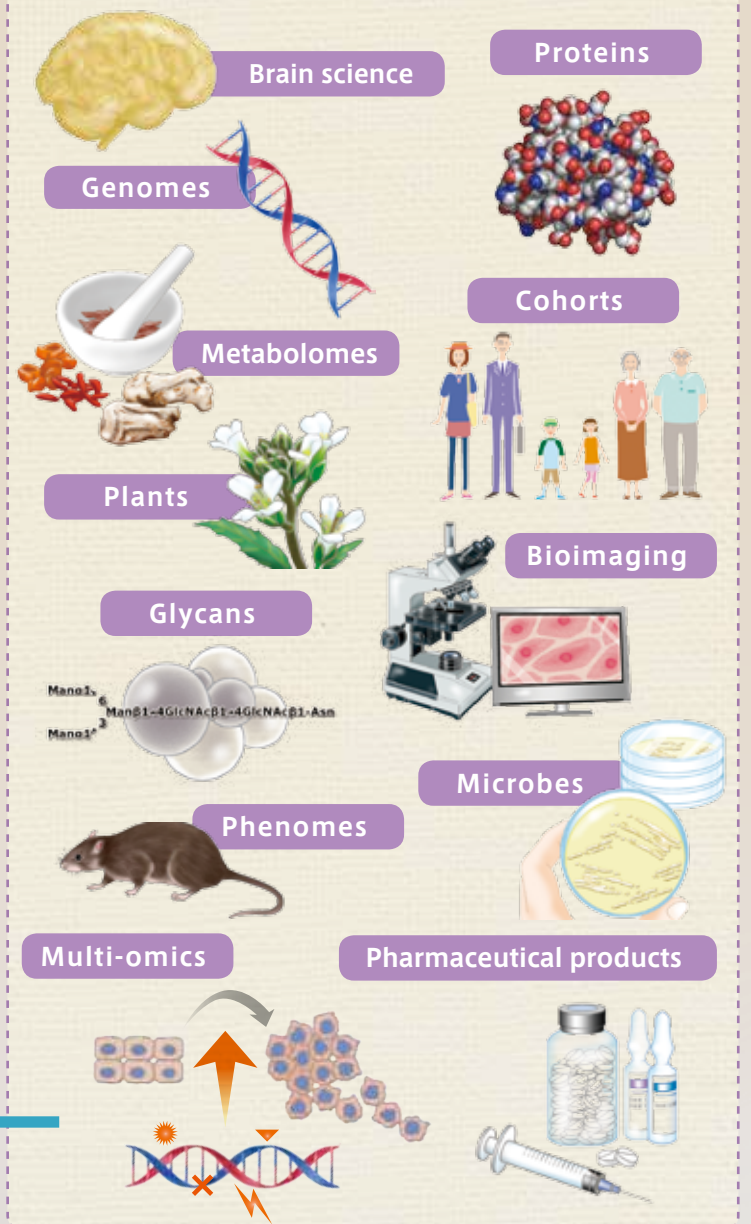
Analyze



RDF : Resource Description Framework

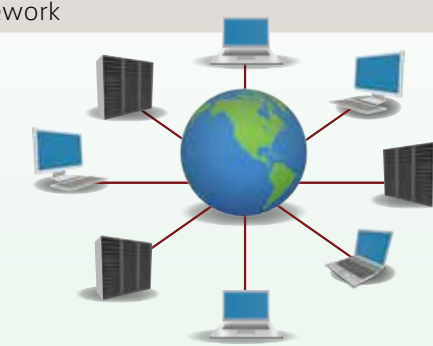
Databases in life science

Knowledge in each domain of the life science is accumulated by comprehensive collection and organization of the information generated in various fields.



International cooperation

DBs and data-sharing projects around the world



R&D programs

List of research directors, subjects, and duration

Life science databases

Database Integration Coordination Program (DICP)

2014 ~

Takeshi Nagasu


Research Supervisor



To integrate life science DBs from a wide range of sources, encompassing various species, objectives, and projects.

Masanori Arita

Team Leader, Center for Sustainable Resource Science, RIKEN



2014 ~

Designing metabolome models for biological species

Shigehiko Kanaya

Professor, Graduate School of Information Science, Nara Institute of Science and Technology

2011 ~ 13

Development of a metabolome database

Bio-MassBank

<http://bio.massbank.jp/>

Yasushi Ishihama

Professor, Graduate School of Pharmaceutical Sciences, Kyoto University




2015 ~

Development of an integrated database for proteomes

Shuichi Onami

Team Leader, Quantitative Biology Center, RIKEN



2015 ~


Integrated database for biological dynamics and images of cell and developmental biology

2012 ~ 14

Integration of databases for systems science of biological dynamics


Systems Science of Biological Dynamics (SSBD) database

<http://ssbd.qbic.riken.jp/>



Minoru Kanehisa

Professor, Institute for Chemical Research, Kyoto University



2014 ~


Integrated database linking genomes to phenotypes, diseases and drugs

2011 ~ 13

Genome-based integrated information resource for diseases, drugs, and environmental substances


KEGG MEDICUS

<http://www.kegg.jp/kegg/medicus.html>



Sumio Sugano

Professor, Graduate School of Frontier Sciences, The University of Tokyo




2014 ~

Integration of transcriptome database with multi-omics data and disease-associated human variations

Ken Kurokawa

Professor, Earth-Life Science Insititute, Tokyo Institute of Technology



2014 ~


Advancement of MicrobeDB.jp with integration of genomic and metagenomic information

2011 ~ 13

Integration of a microorganism database based on genomic and metagenomic information


Integrated database for microbes

<http://microbedb.jp/>



Satoshi Tabata

Director, Kazusa DNA Research Institute



2014 ~


Integration of plant databases based on genome information

2011 ~ 13

Plant Genome DataBase Japan (PGDBj) for integration of plant genome-related resources and information


Plant Genome DataBase Japan

<http://pgdbj.jp/?ln=en>



Katsushi Tokunaga

Professor, Graduate School of Medicine, The University of Tokyo



2014 ~

Human Genome Variation Database towards personalized medicine

2011 ~ 13


Development of a human genome-variation database

Human Genome Variation Database

<https://gwas.biosciencedbc.jp/index.html>

Haruki Nakamura

Professor and Director, Institute for Protein Research, Osaka University



2014 ~


Upgrading and integrative management of Protein Data Bank Japan

2011 ~ 13

International construction and integration of protein structure databank

PDBj (Protein Data Bank Japan)

<http://pd bj.org/>



Visit

<http://biosciencedbc.jp/en/db-link> (R&D results DBs)

to see the results for these programs.

Hisashi Narimatsu

Invited Senior Researcher, National Institute of Advanced Industrial Science and Technology



2014 ~

Development of an international glycan structure repository and integrated glycan database

2011 ~ 13

Development of integrated glycoscience database and research-support tools

JCGGDB

http://jcg gdb.jp/index_en.html



Fumihiko Matsuda


Professor and Director, the Center for Genomic Medicine, Graduate School of Medicine, Kyoto University

2011 ~ 13

Construction of a foundation of integrated information for large-scale genome cohort studies

Human Genetic Variation Browser

<http://www.genome.med.kyoto-u.ac.jp/SnpDB/index.html>



Database Integration Coordination Program (Tool Prototype for Integrated Database Analysis)

To develop data analysis tools for DBs integrated under the DICP as subprograms of the DICP.

	Integrative database for exploiting published ChIP-seq data	Assistant Professor, Kyushu Univ.	Shinya Oki
2015	Development of a mobile application for cross-search of various herbal medicines	Assistant Professor, Toyohashi Univ. of Technology	Tetsuo Katsuragi
	Interaction energy analysis between lectin and glycochain based on quantum mechanical calculation	Assistant Professor, Shizuoka Univ.	Shogo Nakano
	Development of pdbBAM, comprehensive mapping of PDBj protein sequences on human genome	Assistant Professor, Tohoku Univ.	Matsuyuki Shiota
2014	A general framework for interlinking data between RDF stores and its application to ortholog analysis	Research Staff, Natl. Inst. for Basic Biology	Hirokazu Chiba
	An extension of biochemical reaction network analysis environment	Technical staff, RIKEN	Kozo Nishida
	Pipeline of phase-defined complete sequencing for the HLA genes	Visiting Researcher, Natl. Inst. of Genetics	Kazuyoshi Hosomichi
	Fast feature selection method for large-scale analysis of local sequence/structure/function relationships in proteins		
	Network analysis system for genome-scale metabolic models with multi-omics data		
	Integration of KNApSACk and the other chemical databases to understand efficacy of plants on human at the molecular level		
2013	Development of a pipeline for generating ligand - protein binding site prediction tools using machine learning		
	AtMetExpress: a database for mass spectrometry-based metabolite profiling data in <i>Arabidopsis</i> and development of related tools to facilitate omics-data integration		
	Development of a method for detecting functional modules by analyzing co-occurrences of metadata in protein		
	Development of a web tool for the elucidation of glycan binding patterns of protein-glycan interaction		
	Development of a metagenomic analysis pipeline that utilizes the MicrobeDB.jp data		


Information technology

Core Technology Development Program

To develop fundamental technologies for achieving DB integration

Yuji Kohara

Director, Database Center for Life Science (DBCLS), Research Organization of Information and Systems



2011 ~ 13

Development of fundamental technologies related to integration of databases

To achieve completely new "federation-type" DB integration, rather than traditional large-scale centralized DB integration, this program has been developing fundamental technologies necessary to integrate distributed DBs such as the DDBJ, the PDBj, and other domestic and DICP DBs from various fields, and developing technical elements for new integration system after investigating and testing RDF-based technologies. Technology development and service provision to utilize large-scale data sets, including those from next-generation sequencers, are also being conducted. Various activities aimed at increased data use are in being carried out.

TogoTV

<http://togotv.dbcls.jp/en/>

FirstAuthor's (Japanese version only)

<http://first.lifesciencedb.jp/>

GGRNA

<http://ggrna.dbcls.jp/en/>

RefEx

<http://refex.dbcls.jp/?lang=en/>

Allie

<http://allie.dbcls.jp/en/>

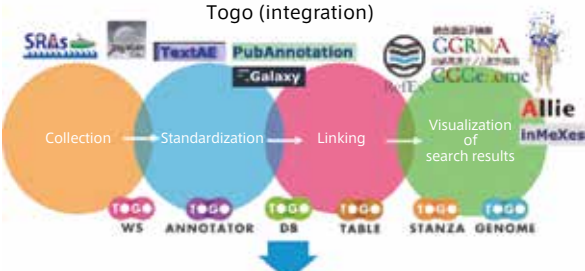
TogoGenome

<http://togogenome.org/>

etc.

DBCLS services for DB integration

Togo (integration)



Integration and discovery of knowledge on life science

* Development of fundamental technologies related to integration of DBs has been continued as a collaborative research project with NBDC since 2014.

National Bioscience Database Center (NBDC), Japan Science and Technology Agency (JST)

5-3, Yonbancho, Chiyoda-ku, Tokyo 102-8666, Japan

Tel. +81-3-5214-8491 Fax. +81-3-5214-8470

2016.01

DB:database