

■タイトル

生命科学系データベースの統合における高度化と統合検索に向けた取り組み

■発表者

○大波純一 1、櫛田達矢 1、三橋信孝 1、熊谷禎洋 2、大久保克彦 2、杉崎太一郎 3、畠中秀樹 1

(1 独立行政法人科学技術振興機構バイオサイエンスデータベースセンター、2 株式会社日立ソリューションズ、3 三井情報株式会社)

■抄録本文（日本語）：

バイオサイエンスデータベースセンター（NBDC; <http://biosciencedbc.jp/>）では 2011 年 4 月の設立以来、散在する生命科学系データベースの統合と利用促進のために、データベース統合の 4 ステップ（カタログ・横断検索・アーカイブ・再構築）に沿って、各種サービスの拡充・高度化と、2014 年度以降に本格化する第 4 のステップで実現を予定する統合検索に向けた検討を進めている。

多数のデータベースや文献・特許を一括検索できる横断検索では、全文検索エンジン **Hyper Estraier** の P2P 連携機能により 4 省の検索インデックスサーバにまたがる検索を実現しつつあり、現在は検索インデックスの更新や検索負荷の分散など運用体制の合理化を進めている。通常のネット検索と異なり全エントリの網羅が重視されるデータベース検索のインデックス作成・更新は、各データベースサイトの構造に合わせて行うため負担が大きく、軽減のための検討を行っている。また、横断検索をベースとして蛋白質関連データベースを対象に作成した統合検索プロトタイプ **TogoProt** では、ターゲットタンパク研究プログラムの検索サイト **ASETT** の各データベースも併せて検索できるように拡充された。

統合検索はデータベースの間を有機的に結びつけて複雑な検索条件で検索するものであるが、多様な連結を正しく表現するためにデータベースの **RDF**（**Resource Description Framework**）形式への変換と再構築を目指している。データベースをダウンロード可能な形で共有するアーカイブでは、統合検索での活用を目的として **RDF** 形式でもダウンロードできるよう整備を進めている。また **RDF** においてはリソースが **URL** で表現されるため、永続性の高い **URL** である **PURL**（**Persistent URL**）サービスの提供を開始して、ユーザが統合データベースを半永久的に利用できるための条件の整備に努めている。

■TITLE

Functional enhancement of life science database integration and approaches to integrated search

■AUTHORS

○Jun-ichi Onami 1, Tatsuya Kushida 1, Nobutaka Mitsuhashi 1, Sadahiro Kumagai 2, Katsuhiko Okubo 2, Taichiro Sugisaki 3, Hideki Hatanaka 1

(1 National Bioscience Database Center, JST, 2 Hitachi Solutions, Ltd., 3 Mitsui Knowledge Industry Co., Ltd.)

■抄録本文（英語）：

National Bioscience Database Center (NBDC; <http://biosciencedbc.jp/>) was established in April 2011 for the purpose of integrating dispersed life science databases and improving their usability. In line with the four-step plan of database integration (catalog, cross-search, archive and reconstruction), it promotes improvement of related services and discussing approaches to the integrated search expected for the fourth step from 2014.

Life Science Database Cross Search provides searching service across more than 320 life science database which contains Japanese literatures and patent publications. This system is achieving searches across databases of four governments with P2P communication of Hyper Estarier search engine. We are now continuing the improvement of the operation with rationalizing its task of updating indexes and balancing its search load. Cross search systems for scientific uses require covering all entries of databases, unlike popular search systems. Also, making and updating of search-indexes are quite laborious due to database structure dependencies, so we are seeking for cost saving.

Integrated search system uses complicated search conditions connecting databases organically, so we plan to convert databases to RDF (Resource Description Framework) format in order to discriminate different types of links. Life Science Database Archive plans RDF downloading expecting sharing databases with integrated searches. In addition, since resources in RDF are expressed as URL, we started PURL (Persistent URL) service, which will contribute to the infrastructure with which users can utilize integrated databases permanently.