

NBDCのRDFの活用に向けて

○ 畠中秀樹¹⁾、大久保克彦²⁾、井上圭介³⁾、加藤健弘²⁾、信定知江¹⁾、八塚茂¹⁾

1) バイオサイエンスデータベースセンター 2) 日立製作所 3) 日立公共システム

要旨

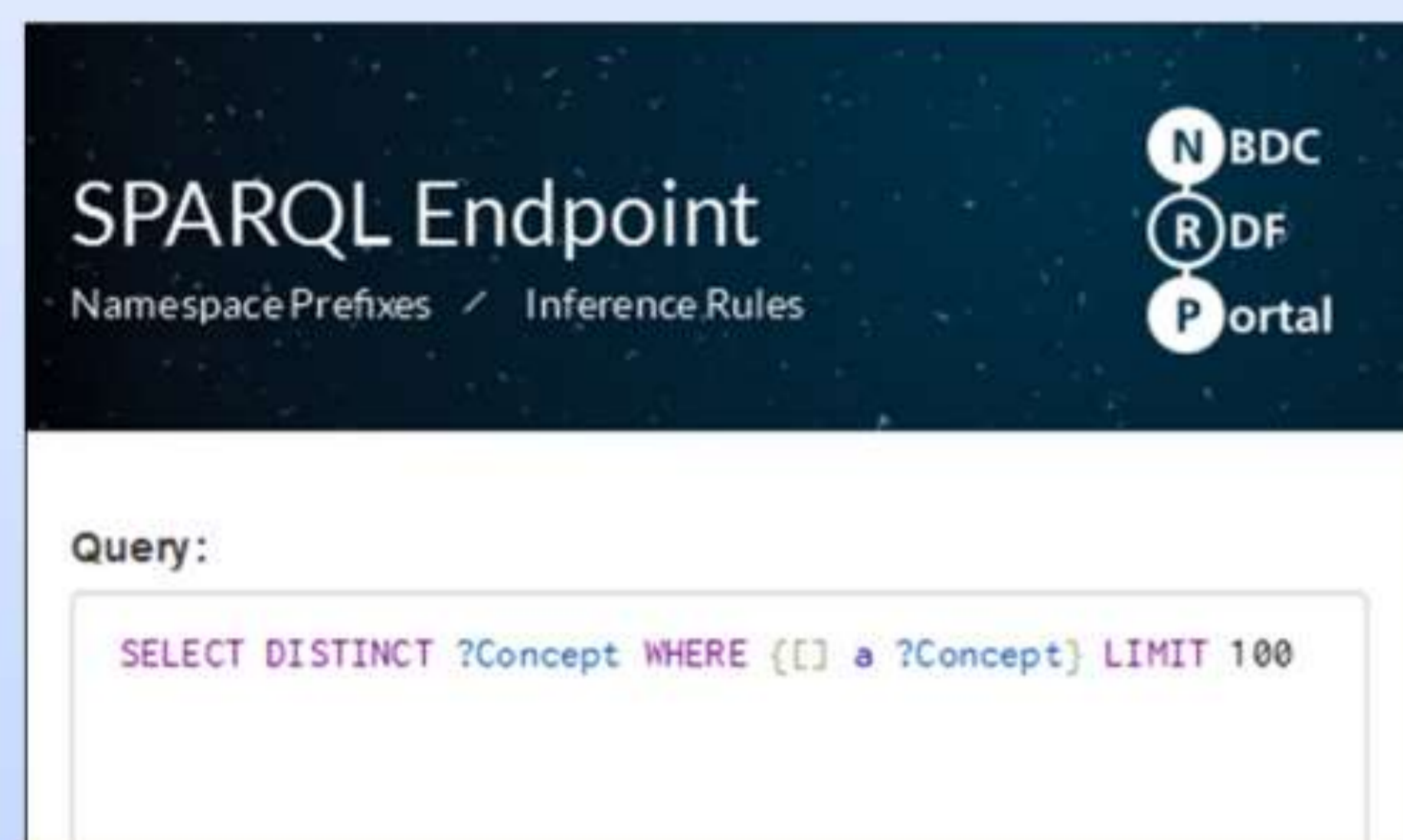
バイオサイエンスデータベースセンター(NBDC)は昨年11月にNBDC RDFポータルを公開し、アーカイブ、HOWDY-R/TogoProtと併せ、3つの大きなRDFストアを有している。これらにSPARQL Builder、LODQA、TogoTableを適用し活用するため、RDFストアの各種整備を進めている。

アーカイブのRDFストアはTogoDBの表形式データを自動RDF化して収納している。各種DBのIDをidentifiers.orgやEDAM

オントロジーでURI化・クラス定義し、既にSPARQL Builderを適用できているが、改修したTogoDBにより、他のカラムを用いたラベルの付与やクラス定義の追加を進める予定である。

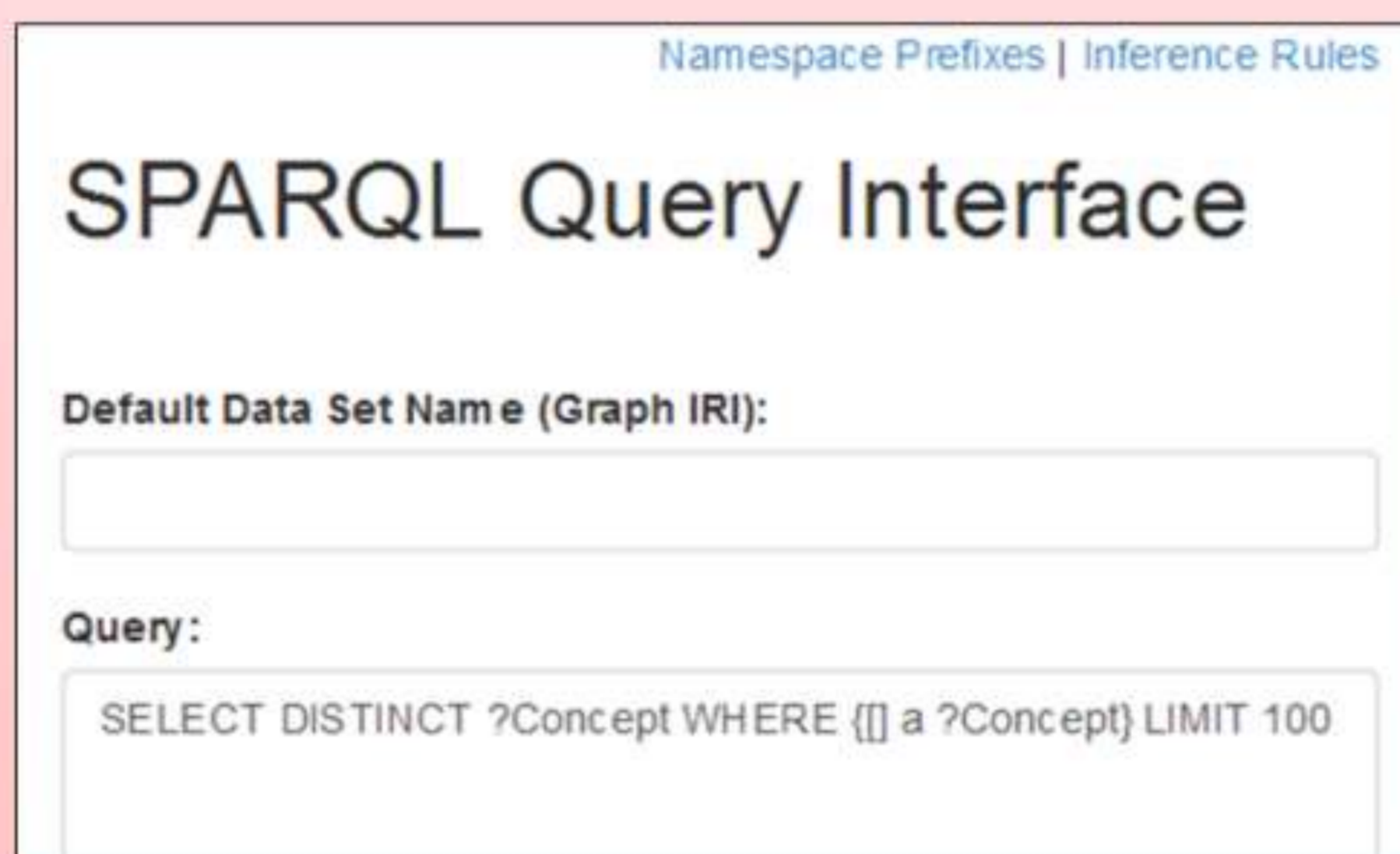
一方HOWDY-R/TogoProtのRDFストアは、Bio2RDFを利用して得たヒト遺伝子関連RDFと、UniProtと蛋白質関連DBを紐付けるRDFを収納しており、こちらもIDのURI化やクラス定義を行ったので、上記3アプリケーションの活用に向けてさらに整備を進めたい。

RDFポータルのRDFストア
http://integbio.jp/rdf/sparql



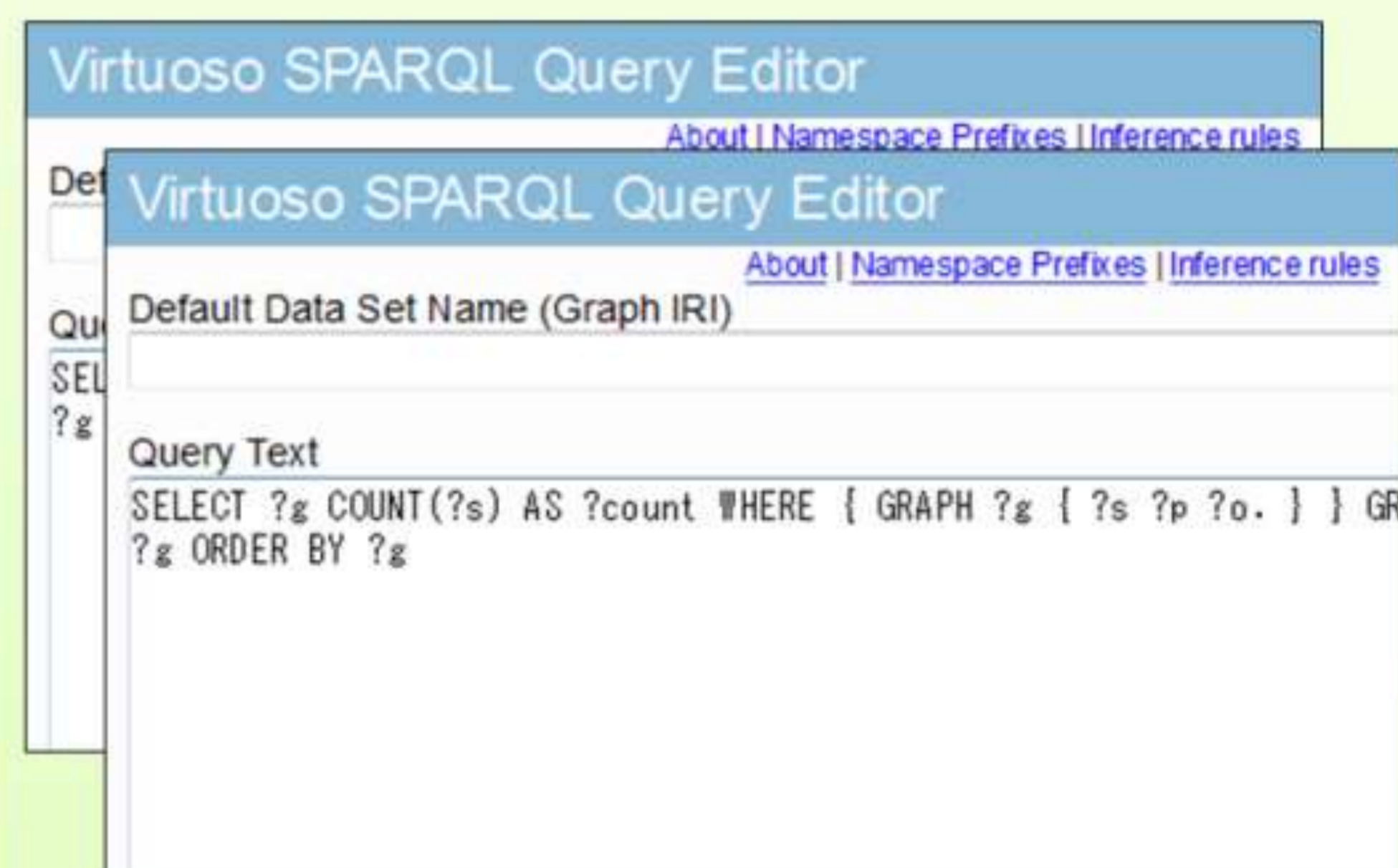
- ・RDF化ガイドラインに従って作成されたRDFを収納
- ・巨大なRDFの追加を順次進めている

アーカイブのRDFストア
http://dba-rdf.biosciencedbc.jp/sparql



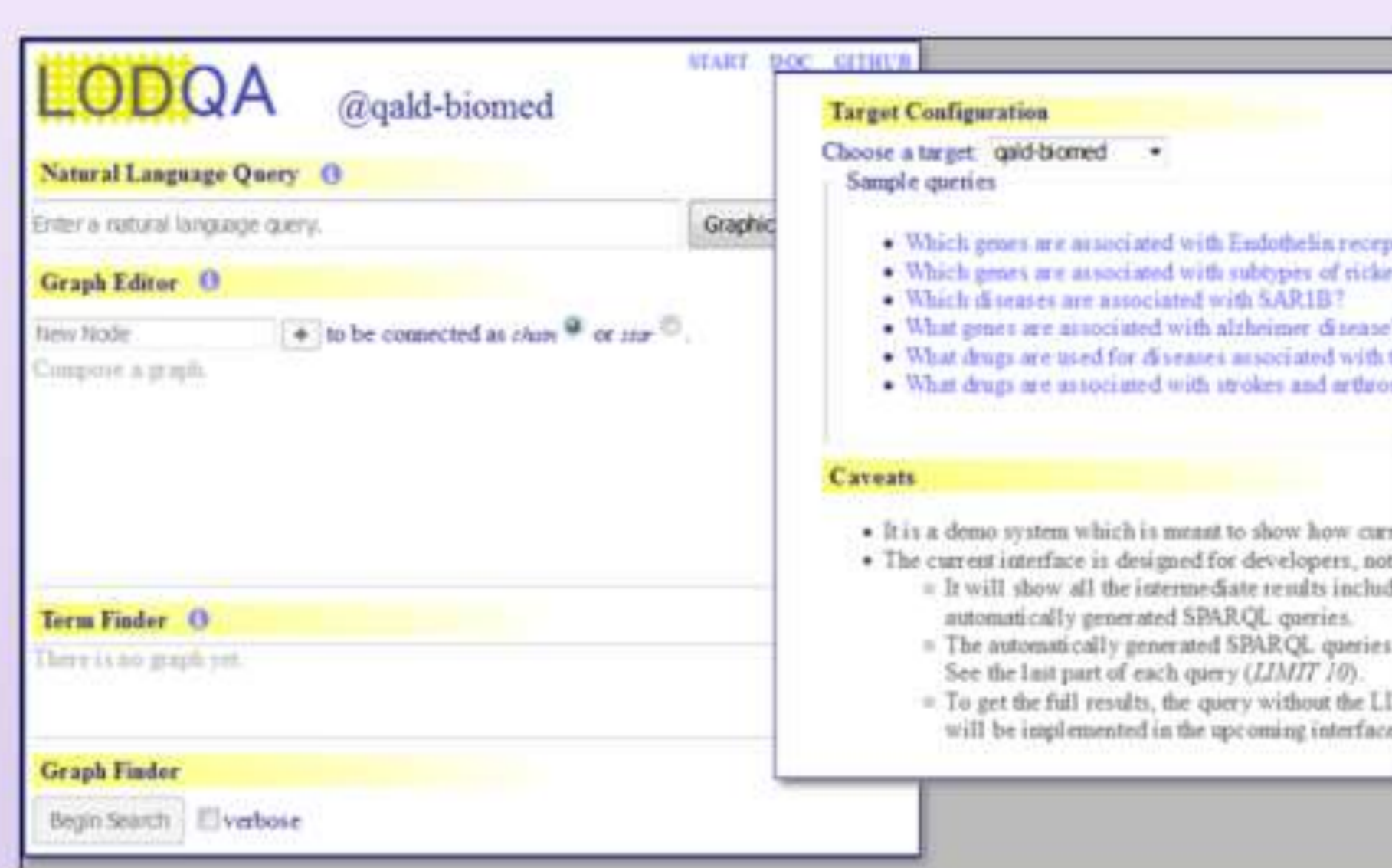
- ・生命科学系データベースアーカイブの簡易検索サイト(TogoDB)に収納された表形式データをRDF化して収納
- ・セルの値が外部DBのIDである場合、identifiers.orgによりURI化し、EDAMオントロジーによりクラスを定義

HOWDY-RとTogoProtのRDFストア
(未公開)



- ・ヒト遺伝子関連DBをBio2RDFを利用して変換したRDF(毎日更新)と、蛋白質関連DBをUniProt ACに紐づけたRDFを収納
- ・アーカイブRDFと同様、外部IDをidentifiers.orgでURI化、EDAMでクラス定義

SPARQL Builder、LODQA、TogoTableなどの適用に向け、さらに整備を進めている



- ・SPARQL Builder(GUIによるSPARQL検索)は既にアーカイブRDFに適用(http://sparqlbuilder.org/dba/)されているが、TogoDBを改修し、subjectへのクラス付与や、外部ID以外のURI化と他カラムによるラベル付与を進めている。またRDFポータルもUmaka経由でのメタデータ抽出が予定されている。
- ・LODQA(自然言語によるSPARQL検索)は、predicateを用いたdomainのラベル付与など各種検討を依頼。
- ・TogoTable(ユーザの表形式データへのカラム追加)はアーカイブやTogoProtからのカラムも追加できるよう、SPARQLクエリを作成中。

謝辞 SPARQL Builder、LODQA、TogoTableとの連携では、それぞれDBCLSの山口敦子さん、金進東さん、河野信さんらにご協力いただいています。