

SciNetS Search : Inference Search over an Integrated Life-sciences Database Based on the Semantic Web

Norio Kobayashi¹, Manabu Ishii¹, Yuko Yoshida¹, Yuko Makita¹, Akihiro Matsushima¹, Yoshiki Mochizuki¹, and Tetsuro Toyoda¹

1. Bioinformatics And Systems Engineering (BASE) division, RIKEN, Japan



Integration of various omics databases is a fundamental approach for comprehensive understanding of life phenomena. Scientists' Networking System (SciNetS) is a Semantic Web-based data repository that realises database integration by collecting metadata of databases described as a set of triples consisting of two bio-items such as gene and ontology terms and a semantically defined relationship between these bio-items. These triples are curated from various life-sciences databases including gene-publication relationships manually curated from MEDLINE and gene-phenotype relationships extracted from published databases by computer programs. SciNetS internally forms a metadata network by maximally concatenating triples.

SciNetS Search realizes a cross-search over the integrated databases by traversing the metadata network managed by SciNetS. SciNetS Search allows a user to discover bio-items included in the ontological category specified by the user. The results include bio-items not only having a user's keyword but also semantically inferred from these bio-items using metadata relationships, and are ranked using statistical scores computed on the basis of the metadata network. SciNetS Search is available via a user's web browser and provides search service against over 30 million bio-items of 100 thousand categories and 100 million relationships curated from over 100 public omics databases of various species including human, mouse, rice and Arabidopsis.

1. Database integration with Virtual Laboratories as a Service (LaaSS)

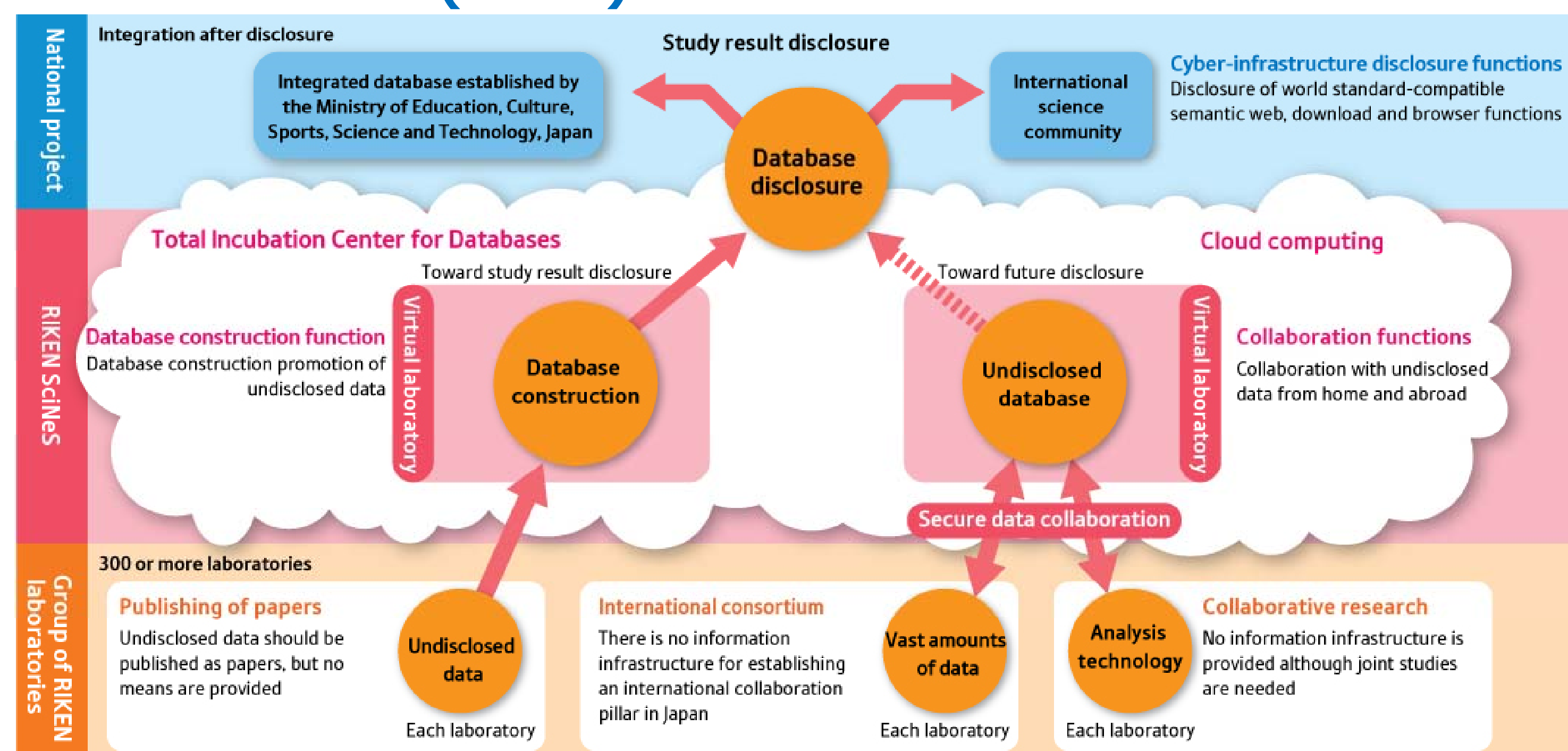


Fig. 1 Concept of LaaS. A database is curated as a Semantic-Web based metadata in a virtual laboratory (VL). Those data items are correlated with data items and ontology terms in a different VLs by semantic links: This mechanism realises semantic data integration among all databases.

2. SciNetS: an implementation of LaaS as a Semantic Web based Cloud Computing Service

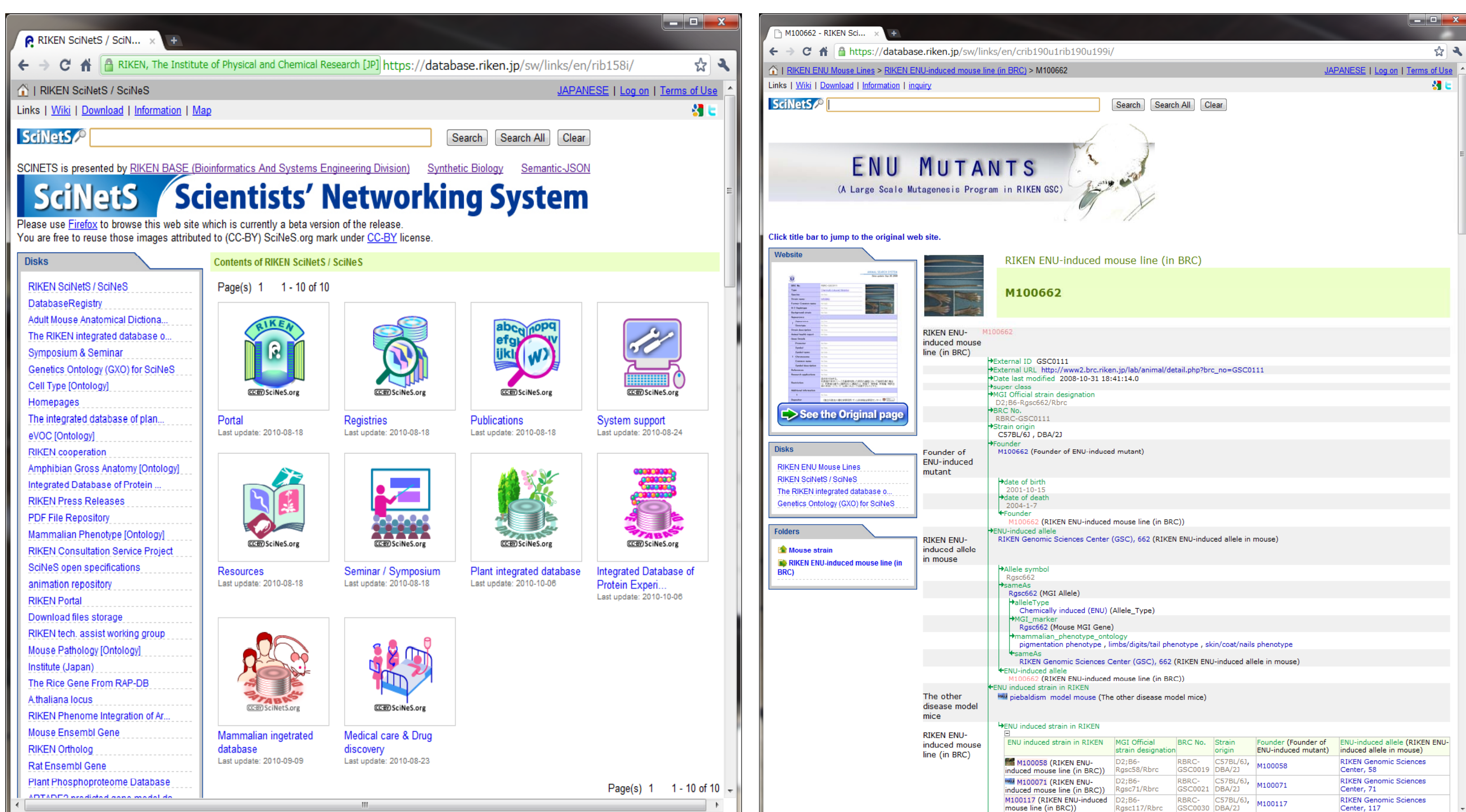


Fig. 2 Snapshots of SciNetS. (Left) The top page displaying Virtual Laboratories categorised on the bases of ontology. (Right) a data record of an ENU mouse mutant presented in a Semantic Web format.

3. Ontological inference search at SciNetS

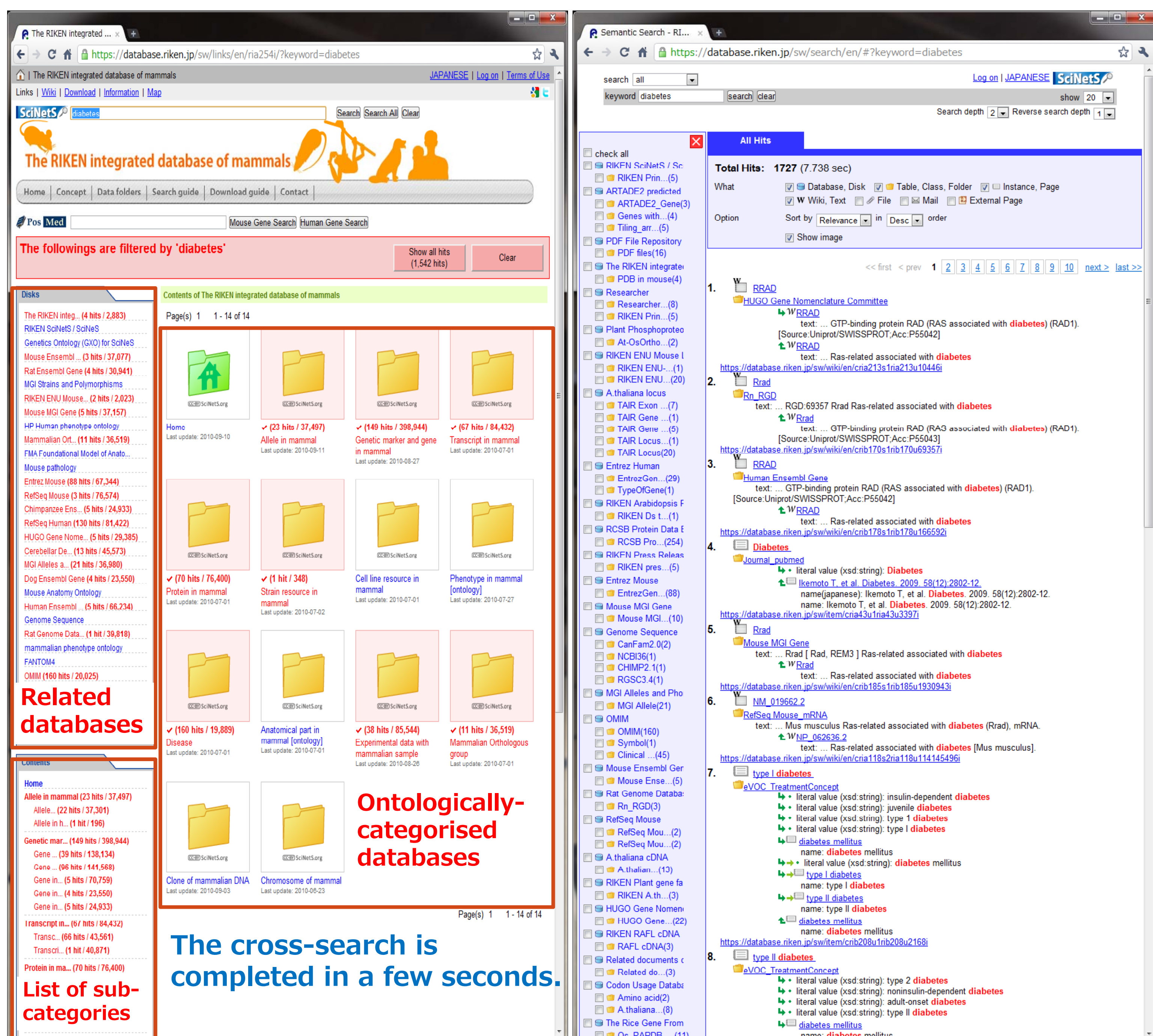


Fig. 3 Inference search results using ontological concept of hierarchy relationships. (Left) A keyword cross-search over the RIKEN integrated database of mammals categorised on the bases of ontology. (Right) A data item search. The list of data items includes not only the one having the keyword but also the one related with a data item having the keyword via a semantic link.

4. Document-based Inference Search at SciNetS

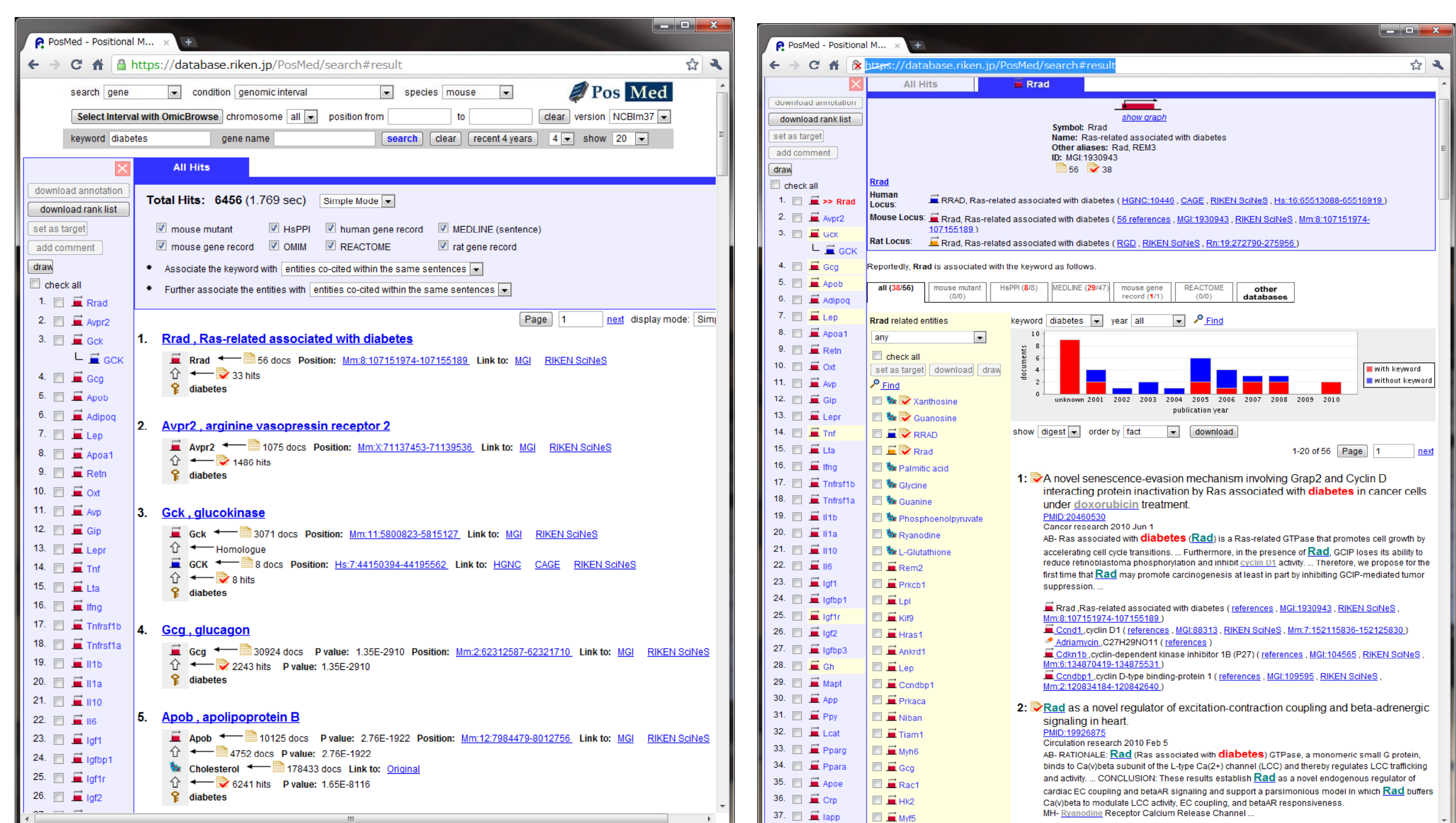


Fig. 4 A search result of Mouse Gene. This search employs human curated gene-MEDLINE relationships. The resultant mouse genes are ranked by statistical analysis based on the numbers of MEDLINE documents having a keyword. (Left) Ranking of mouse genes associated with keyword "diabetes". (Right) MEDLINE documents related with gene "Rad" ranked first in the left figure.

References

- Kobayashi N. and Toyoda T., Statistical search on the Semantic Web. *Bioinformatics*. 24(7): 1002-10. 2008.
- Yoshida Y, Makita Y, Heida N, Asano S, Matsushima A, Ishii M, Mochizuki Y, Masuya H, Wakana S, Kobayashi N, Toyoda T., PosMed (Positional Medline): prioritizing genes with an artificial neural network comprising medical documents to accelerate positional cloning. *Nucleic Acids Res*. 37(Web Server issue):W147-52. 2009.
- Makita Y, Kobayashi N, Mochizuki Y, Yoshida Y, Asano S, Heida N, Deshpande M, Bhatia R, Matsushima A, Ishii M, Kawaguchi S, Iida K, Hanada K, Kuromori T, Seki M, Shinozaki K, Toyoda T. PosMed-plus: an intelligent search engine that inferentially integrates cross-species information resources for molecular breeding of plants. *Plant Cell Physiol*. 50(7):1249-59. 2009.

SciNetS: <http://database.riken.jp>