Perspectives in Science (2016) xxx, xxx—xxx

ScienceDirect
journal homepage: www.elsevier.com/pisc

Latest developments in Semantic Web technologies applied to the glycosciences☆

Kiyoko F. Aoki-Kinoshita[a,b,*], Nobuyuki P. Aoki[a], Akihiro Fujita[a], Noriaki Fujita[b], Toshisuke Kawasaki[c], Masaaki Matsubara[d], Shujiro Okuda[e], Toshihide Shikanai[b], Daisuke Shinmachi[a], Elena Solovieva[b], Yoshinori Suzuki[b], Shinichiro Tsuchiya[a], Issaku Yamada[d], Hisashi Narimatsu[b]

[a] *Faculty of Science and Engineering, Soka University, Tokyo 192-8577, Japan*
[b] *Glycoscience and Glycotechnology Research Group, AIST, Ibaraki 305-8568, Japan*
[c] *Research Center for Glycobiotechnology, Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan*
[d] *The Noguchi Institute, Tokyo 173-0003, Japan*
[e] *Niigata University Graduate School of Medical and Dental Sciences, Niigata 951-8510, Japan*

Received 4 January 2016; accepted 6 May 2016

KEYWORDS
Semantic Web;
Glycan repository;
Glycan text representation

Summary The Integrated Life Science Database Project of Japan funded a group of glycoscientists to carry out a project to integrate glycoscience databases using Semantic Web technologies. As a continuation of the previous project period, the Japan Consortium for Glycobiology and Glycotechnology Database (JCGGDB) developed several glycoscience-related databases. The GlycoProtDB database is among those being integrated, providing an important resource to understand protein glycosylation. Another database being integrated is GlycoEpitope, a comprehensive database of carbohydrate epitopes and antibodies. In the current project period, we started the development of GlyTouCan, the international glycan structure repository providing unique accession numbers to all glycan structures. Although such databases are sufficiently important in and of themselves, their integration with other—omics data such as the protein information in UniProt will be crucial to bring glycosciences to the forefront of life sciences. However, to integrate such disparate sets of data among different fields in a way such that future maintenance costs are minimal, standardized ontologies and formats must be established.

* Corresponding author at: 1-236 Tangi-machi, Hachioji, Tokyo, Japan 192-8577, Faculty of Science and Engineering, Soka University, Tokyo 192-8577, Japan.
E-mail address: kkiyoko@soka.ac.jp (K.F. Aoki-Kinoshita).

http://dx.doi.org/10.1016/j.pisc.2016.05.012

Please cite this article in press as: Aoki-Kinoshita, K.F., et al., Latest developments in Semantic Web technologies applied to the glycosciences. Perspect. Sci. (2016), http://dx.doi.org/10.1016/j.pisc.2016.05.012

Our latest project has attempted to define the minimal standards that are necessary to enable this integration. The technical challenges to integrate all these databases and the technologies to overcome these challenges will be described.

Introduction

Integrated Life Science Database Project

The purpose of the Integrated Life Science Database Project of Japan (http://biosciencedbc.jp/en/), sponsored by the Japan Science and Technology Agency (JST), is to integrate all life science databases both within and outside Japan. It is apparent that a plethora of data is currently available on the Internet, and it can be bewildering for students and newcomers who come from different life science fields. One of the problems with the development of life science databases is that the data are often left around or taken down after publication or the developer/student has left the organization/university. Therefore, JST had specifically included a requirement in this latest project; all funded database projects must be sustainable even after funding ends. As a result, the currently funded projects are now being developed in a way that the data is available and all developed source code is open source.

In terms of technical sustainability, considering that the previous project focused initially on web services as the main technology, it was found that web services were still limited due to the following:

- Each web service can implement only one query, having a specific input, predefined parameters and a specific output.
- Changing the data often forces changes in the web service implementation.
- Documentation of each and every web service was cumbersome, but required for others to use them.
- Different web service application programming interfaces (APIs) needed to be developed for different programming languages (although Representational State Transfer (REST;) was becoming mainstream).
- Integration with other databases required knowledge of other databases' APIs.

Semantic Web technology

Consequently, participants started looking into the Semantic Web as a means to integrate data among different databases. In contrast to web services, it was found to be simpler to implement using the Resource Description Framework (RDF), which consisted of triples of data (**subject**, **predicate** and **object**) where the **subject** and **objects** could be either a literal (i.e., a text string or numeric, for example) or Uniform Resource Identifiers (URIs) that pointed to data on the Internet. The **predicate** performs an important function as it adds the semantics to the data. **Predicates** are defined by an *ontology* which specifies the relationship between different classes of data. For example, if a particular protein X is glycosylated at position P by an *N*-linked glycan, then, the ontology would need to specify classes of data called *proteins* and *glycans*. Moreover, it would need to define the concept of *glycosylation* as a **predicate** that takes *proteins* as **subjects** and *glycans* as **objects**. Furthermore, the glycosylation position would need to be encapsulated by another **predicate** *glycosylated_at* whose **objects** are classes of data called *aminoacids*, for example. The *aminoacid* class could also serve as a **subject** of other **predicates**, such as *positionnumber* and *aminoacidtype*.

Note that such rules to specify the semantics of data are all stored in an ontology that must be consistently used by data providers for the data to be accurately linked to each other. If one database provider uses *glycosylatedAt* and another uses *glycosylated_at*, then these would be considered different predicates, and their meanings would be considered different from one another.

As a result, as long as a standardized ontology could be developed, data no longer needed to be maintained in relational databases and could be instead provided as URIs, which meant that data could be modified as needed as long as it could be accessed from the same URI consistently (Aoki-Kinoshita et al., 2015b). Thus, as a part of the Integrated Life Science Database Project, the Glycoscience Team also started looking into the Semantic Web and RDFizing existing databases, and we were able to show a proof-of-concept that it was possible to do so with minimal effort (Aoki-Kinoshita et al., 2013a; Katayama et al., 2014). From this, a new glycan ontology was developed, called GlycoRDF (Ranzinger et al., 2015), which is now used by the Carbohydrate Structure Database (Toukach and Egorova, 2015), MonosaccharideDB, UniCarbKB (Campbell et al., 2014), Glycoepitope, GlycoProtDB (Kaji et al., 2012) and GlycomeDB (Ranzinger et al., 2009). By using this ontology, it is now possible to integrate these different databases simply by referencing the URIs of other database entries using RDF.

To implement such integration, data must be stored in a specialized database for RDF data, called a triplestore. In the current work, Virtuoso is used as the triplestore data management system (Erling and Mikhailov, 2009).

Glycoscience Team project

The Glycoscience Team of the Integrated Database Project was granted funds to develop an International Glycan Structure Repository which we call GlyTouCan based on WURCS (Web 3.0 Unique Representation of Carbohydrate Structures), a new text representation of glycan structures. These two subprojects started in 2014 in addition to the

Please cite this article in press as: Aoki-Kinoshita, K.F., et al., Latest developments in Semantic Web technologies applied to the glycosciences. Perspect. Sci. (2016), http://dx.doi.org/10.1016/j.pisc.2016.05.012

continuation of the integration of glycoscience databases worldwide. In particular, databases developed previously as a part of JCGGDB (Japan Consortium for Glycobiology and Glycotechnology Database) are being RDFized to enable this integration. These databases include GlycoEpitope (Kawasaki et al., 2006; http://www.glycoepitope.jp), GlycoProtDB (Kaji et al., 2012), PacDB (carbohydrate-binding pathogen database), and GDGDB (glycogene disease database), among others, and are being released as a part of ACGGDB (Asian Community of Glycoscience and Glycotechnology Database; http://acgg.asia/db/). In this work, we focus on describing GlyTouCan and WURCS by introducing Semantic Web technologies.

Materials and methods

WURCS

The original purpose of developing WURCS started with discussions regarding existing formats to use for representing and implementing the registration of glycan structures in the International Glycan Structure Repository (GlyTouCan). This will be discussed in the next section. The major existing glycan structure representations include GlycoCT (Herget et al., 2008), IUPAC (IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN), 1982), KCF (Aoki-Kinoshita, 2009), LINUCS (Bohne-Lang et al., 2001) and LinearCode (Banin et al., 2002). Although each representation has its own strengths and weaknesses, none could satisfy all the requirements of the repository, which were:

1. The necessity to represent ambiguous glycan ''structures'' including monosaccharide compositions, repeating structures and incompletely determined structures.
2. The necessity to represent glycans on the Semantic Web, meaning that it must be a linear text string, which could be used as a URI.

Based on these requirements, it became clear that the existing representations were insufficient. GlycoCT was the strongest candidate, but the use of a text library for representing monosaccharide modifications and the fact that it was not a linear text string prevented us from selecting it. KCF was also a multi-line format. LinearCode was limited in that it did not cover all possible monosaccharides; and IUPAC used a human-readable format for representing monosaccharides, which would be difficult to parse by computer.

Thus, we began by first considering how to represent monosaccharides uniquely without having to use human-readable text. The reason for this was that such human-readable text was the source of problems for many representations of monosaccharides, such as deoxy-Galactose and Fucose, which are essentially the same monosaccharide. This took us to MonosaccharideDB, which uses the concept of a stereocode to represent monosaccharides uniquely. We adopted a similar concept in WURCS, representing each hydroxyl group of a monosaccharide by a code indicating whether it lies on one side or the other of the carbon ring. Then each monosaccharide in a glycan was sorted in such a way so that it would always be uniquely ordered in the string. Finally, glycosidic linkages were indicated based on this order. Details of the latest WURCS representation can be found at http://www.wurcs-wg.org/ (Tanaka et al., 2014).

GlyTouCan

The original motivation behind developing WURCS came from the development of a glycan structure repository, which had been deemed necessary by the community as a means to apply identifiers, or accession numbers, to glycan structures, especially in the literature and in curated databases (Aoki-Kinoshita et al., 2013b). The Glycoscience Team of Japan was suggested to take the lead in this development, supported by a team of glycoinformaticians in the United States. This resulted in the development of GlyTouCan version 1.0 at http://www.glytoucan.org/ (Aoki-Kinoshita et al., 2015a).

The initial system architecture consisted of a relational database management system (RDBMS; Postgres) and a software called GlySpace to interact with the database, developed by those in the United States, and a triplestore (Virtuoso 7), developed by the Japanese group. The workflow was as follows:

1. User interface (GlyTouCan) receives glycan structure to register.
2. GlyTouCan calls the GlySpace software to register the structure into the RDBMS.
3. GlyTouCan processes the data to store in Virtuoso.

The GlyTouCan user interface (UI) was designed to be as user-friendly as possible, especially for biologists and glycobiologists to intuitively search and register glycan structures. This involved some work in integrating the GlySpace software into the UI.

As a part of the Integrated Database Project of Japan, it was necessary for the data in GlyTouCan to be linked with other databases, and thus, GlycoEpitope was first selected. Having a fully-curated database of carbohydrate epitopes linked with GlyTouCan would allow links to be made to proteomics and lipidomics data as well. Moreover, GlycoEpitope was already RDFized, so the link with glycan structures in GlyTouCan would be fairly straightforward. Following GlycoEpitope, the RDFized data of BCSDB and GlycomeDB were also linked with GlyTouCan data, followed by UniCarb-DB (Hayes et al., 2011) and UniCarbKB.

Results

WURCS

WURCS version 1.0 has been fully refactored in version 2.0 in a way that it can be more easily parsed by GlyTouCan. It is more compact and more consistent with GlycoCT. Fig. 1 is an example of WURCS 2.0. To make the format more compact, the concept of ''unique residues'' was incorporated so that the same residues need not be listed repeatedly. Thus, the general format is as follows: WURCS = <***Version***>/<***Unit count***>/<***UniqueRES list***>/<***RES***

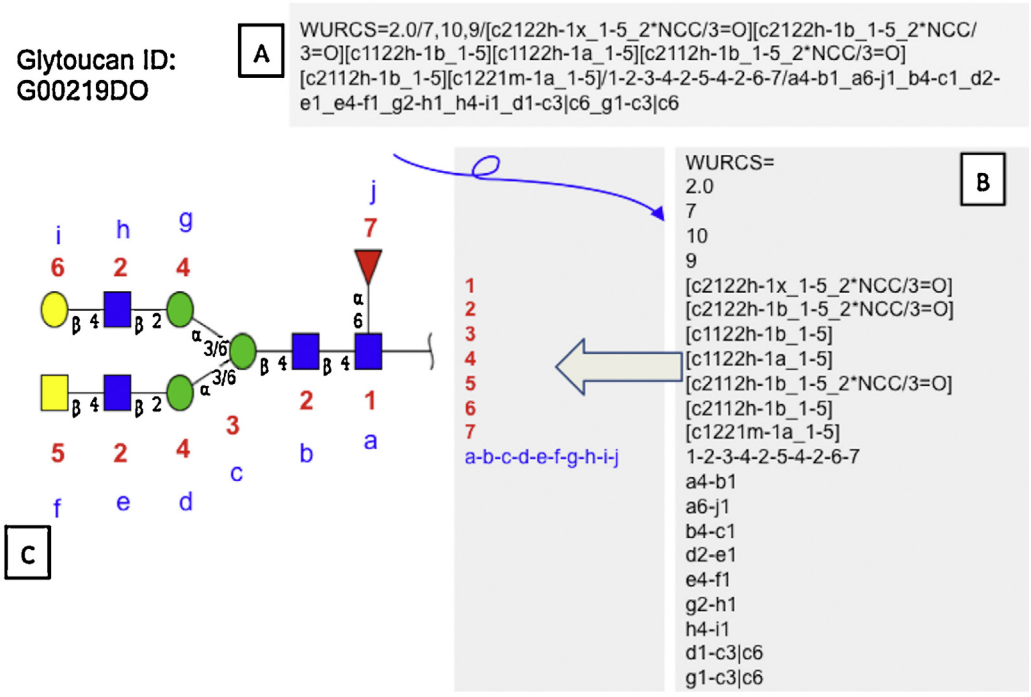

Figure 1 An example of a glycan structure in WURCS 2.0 format. After the version number, the number of unique residues, number of total residues, and number of linkages are listed, delimited by commas. Then the stereocode-like text for each unique residue is listed, followed by linkage information, labelling each distinct residue by alphabetical characters. See text for details.

sequence>/<***LIN list***>, where <*Unit count*> lists the number of unique residues, the total number of residues, and the number of glycosidic linkages, separated by commas. The <*UniqueRES list*> lists the stereocode-like string for each monosaccharide residue. This format has also been extended in version 2.0 such that the anomer is listed directly and no longer needs to be inferred from the stereocode conformation. The backbone and modification information are also included. The <*RES sequence*> section lists the ordering of the unique residues based on a predefined rule such that the string remains unique. The <*LIN list*> then uses alphabetical characters corresponding to the residues ordered in the <*RES sequence*> such that their glycosidic linkages can be listed.

GlyTouCan functionality

The GlyTouCan web interface (Fig. 2) provides four main functionalities, which include:

1. Browsing the structures:
   (a) The ''View All'' menu contains a ''Glycan List'' option which displays the full content of the repository. From this list, filters are presented to help the users to find a particular glycan based on specific parameters, namely structures that:
      (1) contain a particular glycan motif (e.g., Lewis or lactosamine structures),
      (2) have a particular mass (or within a mass range), or
      (3) contain specific monosaccharide residues.

      This view can also be sorted to display structures according to accession number, contributor name, mass, or registered date.
   (b) The same ''View All'' menu also contains a ''Motif List'' option, which displays a list of the structures considered as ''glycan motifs'' which are well-known glycan structure patterns.
2. Searching for particular structures:
   (a) Graphic input: The GlycanBuilder interface (Damerell et al., 2012) is displayed for users to draw their glycan structure query of interest. As a result, if the specific structure is registered, the accession number is displayed, along with all other structures in GlyTouCan that contain the query structure.
   (b) Text input: A glycan structure specified in GlycoCT, LinearCode, or KCF can be inputted into the textbox as a query structure. The results are the same as the graphic input.
   (c) By motif: A list of the motifs registered in GlyTouCan is given; and when one is clicked, it is used as a query structure. The results are the same as the graphic input.
3. Viewing the details (including linked data information) for a particular structure; the glycan entry page provides a number of sections of information:
   (a) Overview: The graphical representation of the specific glycan, its accession number, calculated mass, registered date and user name, as well as WURCS and GlycoCT representations of the text are provided. The form of graphical representation to display can be selected from the preferences menu.

Figure 2 The main web interface of GlyTouCan. The top menu bar provides the main functionality, including glycan data registration, search, browse and visual preference selection.

(b) Related data: The list of motifs contained in the specific glycan, its monosaccharide composition, and a visualization tool to display its relationship to other glycan structures in GlyTouCan are provided.
(c) Linked DB: If the structure is linked with other RDFized databases, the list of databases and their corresponding entry information will be displayed. This is where the data of external databases can be accessed. Currently available are BCSDB, GlycoEpitope, GlycomeDB, UniCarbDB and UniCarbKB.

4. Registering structures: To register their own structure, the users first register as a user by signing in using their Google account. Registration can then be made via:
   (a) Graphic input using the GlycanBuilder interface, similar to the graphic search option.
   (b) Text input using GlycoCT.
   (c) File upload using a file containing a list of two or more glycans in GlycoCT format.

There are also options to change the display of the user interface based on personal preferences, including the language used (e.g., Chinese, Japanese, Russian, French and German), and the symbol notation to use (Essentials of Glycobiology or University of Oxford notation). These options can be selected under the preferences menu.

Discussion

The first international glycan structure repository utilizes Semantic Web technology to show that it is possible to use this new technology to not only integrate various databases, but also to develop web interfaces that can display the data in a user-friendly way. Open source software has been developed to enable this technology, which will hopefully encourage the further integration of other life science databases.

As a major part of GlyTouCan development, the WURCS representation was essential. By building it on work of existing glycan text representations, a robust, linear and unique representation that can capture glycan structure ambiguity has been developed. The software library for WURCS is also open source and will be made available for those interested in using it. This library also includes software for RDFizing glycan structures from WURCS, which is now used in performing the queries in GlyTouCan when searching for glycan substructures. For future work, we will continue to RDFize the available glycoscience databases that have yet to be converted to integrate and link them with GlyTouCan.

Conclusion

Although still fairly new, Semantic Web technologies have shown great potential to integrate life science databases, especially considering the wide range of fields that this encompasses. Glycoscience is an important aspect of the life sciences that can potentially link proteomics with metabolomics, lipidomics and genomics. Software such as triplestores and user interfaces for the Semantic Web are gradually developing, and it is expected that the current bottlenecks in terms of speed of data processing will be overcome in the near future.

Please cite this article in press as: Aoki-Kinoshita, K.F., et al., Latest developments in Semantic Web technologies applied to the glycosciences. Perspect. Sci. (2016), http://dx.doi.org/10.1016/j.pisc.2016.05.012

Acknowledgements

The authors would like to acknowledge Mitsuko Yamazaki of Soka University and Madoka Soyama of the National Institute of Advanced Industrial Science and Technology (AIST) for their contributions to this project. This work has been supported in part by the Database Integration Coordination Program (DICP) under the Japan Science and Technology Agency and the Ministry of Education, Culture, Sports, Science and Technology.

References

Aoki-Kinoshita, K.F., 2009. Glycome Informatics: Methods and Applications. CRC Press.

Aoki-Kinoshita, K.F., Bolleman, J., Campbell, M.P., Kawano, S., Kim, J.-D., Lütteke, T., Matsubara, M., Okuda, S., Ranzinger, R., Sawaki, H., Shikanai, T., Shinmachi, D., Suzuki, Y., Toukach, P., Yamada, I., Packer, N.H., Narimatsu, H., 2013a. Introducing glycomics data into the Semantic Web. J. Biomed. Semant. 4, 39, http://dx.doi.org/10.1186/2041-1480-4-39.

Aoki-Kinoshita, K.F., Sawaki, H., An, H.J., Campbell, M., Cao, Q., Cummings, R., Hsu, D.K., Kato, M., Kawasaki, T., Khoo, K.-H., Kim, J., Kolarich, D., Li, X., Liu, M., Matsubara, M., Okuda, S., Packer, N.H., Ranzinger, R., Shen, H., Shikanai, T., Shinmachi, D., Toukach, P., Yamada, I., Yamaguchi, Y., Yang, P., Ying, W., Yoo, J.S., Zhang, Y., Narimatsu, H., 2013b. The fifth ACGG-DB meeting report: towards an international glycan structure repository. Glycobiology 23, 1422—1424, http://dx.doi.org/10.1093/glycob/cwt084.

Aoki-Kinoshita, K., Agravat, S., Aoki, N.P., Arpinar, S., Cummings, R.D., Fujita, A., Fujita, N., Hart, G.M., Haslam, S.M., Kawasaki, T., Matsubara, M., Moreman, K.W., Okuda, S., Pierce, M., Ranzinger, R., Shikanai, T., Shinmachi, D., Solovieva, E., Suzuki, Y., Tsuchiya, S., Yamada, I., York, W.S., Zaia, J., Narimatsu, H., 2015a. GlyTouCan 1.0—the international glycan structure repository. Nucleic Acids Res. 44 (D1), D1237—D1242, http://dx.doi.org/10.1093/nar/gkv1041.

Aoki-Kinoshita, K.F., Kinjo, A.R., Morita, M., Igarashi, Y., Chen, Y.-A., Shigemoto, Y., Fujisawa, T., Akune, Y., Katoda, T., Kokubu, A., Mori, T., Nakao, M., Kawashima, S., Okamoto, S., Katayama, T., Ogishima, S., 2015b. Implementation of linked data in the life sciences at BioHackathon 2011. J. Biomed. Semant. 6, http://dx.doi.org/10.1186/2041-1480-6-3.

Banin, E., Neuberger, Y., Altshuler, Y., Halevi, A., Inbar, O., Nir, D., Dukler, A., 2002. A novel Linear Code(R) nomenclature for complex carbohydrates. Trends Glycosci. Glycotechnol. 14, 127—137, http://dx.doi.org/10.4052/tigg.14.127.

Bohne-Lang, A., Lang, E., Förster, T., von der Lieth, C.-W., 2001. LINUCS: linear notation for unique description of carbohydrate sequences. Carbohydr. Res. 336, 1—11, http://dx.doi.org/10.1016/s0008-6215(01)00230-0.

Campbell, M.P., Peterson, R., Mariethoz, J., Gasteiger, E., Akune, Y., Aoki-Kinoshita, K.F., Lisacek, F., Packer, N.H., 2014. UniCarbKB: building a knowledge platform for glycoproteomics. Nucleic Acids Res. 42 (D1), D215—D221, http://dx.doi.org/10.1093/nar/gkt1128.

Damerell, D., Ceroni, A., Maass, K., Ranzinger, R., Dell, A., Haslam, S.M., 2012. The GlycanBuilder and GlycoWorkbench glycoinformatics tools: updates and new developments. Biol. Chem. 393 (11), 1357—1362, http://dx.doi.org/10.1515/hsz-2012-0135.

Erling, O., Mikhailov, I., 2009. RDF support in the virtuoso DBMS. Stud. Comput. Intell. 221, 7—24, http://dx.doi.org/10.1007/978-3-642-02184-8_2.

Hayes, C.A., Karlsson, N.G., Struwe, W.B., Lisacek, F., Rudd, P.M., Packer, N.H., Campbell, M.P., 2011. UniCarb-DB: a database resource for glycomic discovery. Bioinformatics 27, 1343—1344, http://dx.doi.org/10.1093/bioinformatics/btr137.

Herget, S., Ranzinger, R., Maass, K., von der Lieth, C.-W., 2008. {GlycoCT}—a unifying sequence format for carbohydrates. Carbohydr. Res. 343, 2162—2171, http://dx.doi.org/10.1016/j.carres.2008.03.011.

IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN), 1982. Abbreviated terminology of oligosaccharide chains. Recommendations 1980. Eur. J. Biochem. 126, 433—437.

Kaji, H., Shikanai, T., Sasaki-Sawa, A., Wen, H., Fujita, M., Suzuki, Y., Sugahara, D., Sawaki, H., Yamauchi, Y., Shinkawa, T., Taoka, M., Takahashi, N., Isobe, T., Narimatsu, H., 2012. Large-scale identification of *N*-glycosylated proteins of mouse tissues and construction of a glycoprotein database, GlycoProtDB. J. Proteome Res. 11, 4553—4566, http://dx.doi.org/10.1021/pr300346c.

Katayama, T., Wilkinson, M.D., Aoki-Kinoshita, K.F., Kawashima, S., Yamamoto, Y., Yamaguchi, A., Okamoto, S., Kawano, S., Kim, J.-D., Wang, Y., Wu, H., Kano, Y., Ono, H., Bono, H., Kocbek, S., Aerts, J., Akune, Y., Antezana, E., Arakawa, K., Aranda, B., Baran, J., Bolleman, J., Bonnal, R.J., Buttigieg, P.L., Campbell, M.P., Chen, Y.-A., Chiba, H., Cock, P.J., Cohen, K.B., Constantin, A., Duck, G., Dumontier, M., Fujisawa, T., Fujiwara, T., Goto, N., Hoehndorf, R., Igarashi, Y., Itaya, H., Ito, M., Iwasaki, W., Kalaš, M., Katoda, T., Kim, T., Kokubu, A., Komiyama, Y., Kotera, M., Laibe, C., Lapp, H., Lütteke, T., Marshall, M.S., Mori, T., Mori, H., Morita, M., Murakami, K., Nakao, M., Narimatsu, H., Nishide, H., Nishimura, Y., Nystrom-Persson, J., Ogishima, S., Okamura, Y., Okuda, S., Oshita, K., Packer, N.H., Prins, P., Ranzinger, R., Rocca-Serra, P., Sansone, S., Sawaki, H., Shin, S.-H., Splendiani, A., Strozzi, F., Tadaka, S., Toukach, P., Uchiyama, I., Umezaki, M., Vos, R., Whetzel, P.L., Yamada, I., Yamasaki, C., Yamashita, R., York, W.S., Zmasek, C.M., Kawamoto, S., Takagi, T., 2014. BioHackathon series in 2011 and 2012: penetration of ontology and linked data in life science domains. J. Biomed. Semant. 5, 5, http://dx.doi.org/10.1186/2041-1480-5-5.

Kawasaki, T., Nakao, H., Takahashi, E., Tominaga, T., 2006. GlycoEpitope: the integrated database of carbohydrate antigens and antibodies. Trends Glycosci. Glycotechnol. 18, 267—272, http://dx.doi.org/10.4052/tigg.18.267.

Ranzinger, R., Frank, M., von der Lieth, C.-W., Herget, S., 2009. Glycome-DB.org: a portal for querying across the digital world of carbohydrate sequences. Glycobiology 19, 1563—1567, http://dx.doi.org/10.1093/glycob/cwp137.

Ranzinger, R., Aoki-Kinoshita, K.F., Campbell, M.P., Kawano, S., Lütteke, T., Okuda, S., Shinmachi, D., Shikanai, T., Sawaki, H., Toukach, P., Matsubara, M., Yamada, I., Narimatsu, H., 2015. GlycoRDF: an ontology to standardize glycomics data in RDF. Bioinformatics 31, 919—925, http://dx.doi.org/10.1093/bioinformatics/btu732.

Tanaka, K., Aoki-Kinoshita, K.F., Kotera, M., Sawaki, H., Tsuchiya, S., Fujita, N., Shikanai, T., Kato, M., Kawano, S., Yamada, I., Narimatsu, H., 2014. WURCS: the Web3 unique representation of carbohydrate structures. J. Chem. Inf. Model. 54, 1558—1566, http://dx.doi.org/10.1021/ci400571e.

Toukach, P.V., Egorova, K.S., 2015. Carbohydrate structure database merged from bacterial, archaeal, plant and fungal parts. Nucleic Acids Res. 44 (D1), D1229—D1236, http://dx.doi.org/10.1093/nar/gkv840.

Please cite this article in press as: Aoki-Kinoshita, K.F., et al., Latest developments in Semantic Web technologies applied to the glycosciences. Perspect. Sci. (2016), http://dx.doi.org/10.1016/j.pisc.2016.05.012