Usability and Scientific Software

SugarToo Project Proposal

Lauren Paton
Department of Computer Science
University of Cape Town
Cape Town, South Africa
ptnlau002@myuct.ac.za

Joseph Sidley
Department of Computer Science
University of Cape Town
Cape Town, South Africa
sdljos001@myuct.ac.za

CCS CONCEPTS

- Human Centred Computing → Human Computer Interaction → HCI Design and Evaluation Methods → Usability Testing
- Applied Computing → Life and Medical Sciences → Computational Biology
- Software and its Engineering → Software Creation and Management → Designing Software → Software Design Engineering
- Information Systems → Data Management Systems → Database Design and Models

KEYWORDS

Glycan, Visualisation, Bioinformatics, Databases, Usability

1. PROJECT BACKGROUND

Usability practices within the sphere of scientific software are not well-defined [3, 17]. Researchers have tried to develop usability metrics for the design of this kind of software but even with these metrics, the structure of development teams is not conducive to creating usable software. They are usually small groups of domain experts with little experience with software development and even less experience with usability practices [3, 17]. The specialised nature of scientific software also complicates its design process, as domain knowledge is usually required by developers. Some research into usability testing of scientific software has been done and there are various usability evaluation methods (UEMs) available for use. A usability evaluation method (UEM) performs usability assessments of a system or interface against a set of pre-defined metrics at any stage in the design process [12]. There are also methods for choosing the most appropriate UEM for a project. Some examples of scientific software in which usability practices are not evident are web-tools for 2D and 3D visualisation of carbohydrate molecules.

Carbohydrates are present on the outer shell of pathogenic bacteria and viruses, and can be targeted in vaccine development or other treatments [5]. Carbohydrate study is an important and expanding area of treatment of bacteria borne illnesses and so, having robust and usable software for carbohydrate research is vital. Molecules "express" their function throughout their structure [3] and thus the 3-dimensional qualities of a molecule characterize their biological function. So, structure-building/visualisation are two of the most important activities done in structural glycobiology. Many web-tools for visualisation of carbohydrate molecules exist including glycan sketchers (2D visualisation tools) and glycan builders (3D visualisation tools). Two examples of sketchers include Glyco.ME and DrawGlycan-SNFG (which has a particularly user-friendly and clean interface) [8]. Some available builders include PolysGlycanBuilder and GlyCam-Web [15][11], both with aspects that contribute to their usability. These glycan sketchers and builders, generally have outdated interfaces that are not beginner friendly. It is evident that usabilitywas not a main consideration in their design and none are particularly convenient for rapidly generating images of carbohydrates for use in research. There are no "perfect" tools for 2D and 3D visualisation of carbohydrate molecules. Their flow of control is often difficult to decipher and requires the knowledge of a domainexpert.

However, some good design choices of these tools, when combined,could achieve a more holistically usable platform. For example, the"drag and drop" aspect of PolysGlycanBuilder [15], the organizedarrangement of monosaccharides used in GlyCam-Web [11] andthe simple, beginner-friendly interface of DrawGlycan-SNFG[8]. Glycarbo - developed at the University of Cape Town (UCT)-is a tool for both sketching and building glycans. It utilisesCarbBuilder, an open-source software, also developed at UCT bythe supervisor of this project, Michelle Kuttel. Glycarbo has lotsof potential to become a holistically more usable tool but needs tobe brought up to an industry standard. It is still experiencing issueswith usability and robustness. The incorporation of the above-mentioned aspects into Glycarbo's interface could benefit itsusability. Editing and testing its interface could make Glycarbo oneof the most usable and useful tools for carbohydrate visualisation. Molecular structure databases are also an essential part ofcarbohydrate research. The necessity for carbohydrate structuredatabases is clear, a store of information on any specificcarbohydrate that one intends to study is very helpful in getting afull understanding of that carbohydrate, its origin, its functions andof course its structure - quickly and easily. It also allows for furtherresearch, as there are references and DOI links to the papers thatfully explain and study the carbohydrate.

However, the tools available leave much to be desired. They are often not stable, usable or data-rich enough. Many large and small databases have been built for glycomics research with some being more focused on a specific singular bacterium such as E. *coli*, which EK3D [14] and ECODAB [18] are good examples of. Some other currently available databases are broad and more general with data on a variety of bacteria, such as PolySacDB [2] and the Carbohydrate Structures Database (CSDB) [10] and these are often less in depth. Many of the database's issues are due to the fact that they are too small and specific for a bacterium (such as E. *coli)*. In proteomics, the primary database, the Protein Data Bank (PDB) [6], is the database that most people in the field use forresearch on any protein but there is no equal to this in glycomics, as the general, broader databases are either too small in terms of volume of data. They sometimes do not have all the correct information needed by scientists or are unstable and unusable. Anyattempts at creating an overarching central repository like the PDB, such as the CSDB, have been rife with software issues and bad

usability practice. Through speaking with our supervisor and trying these tools ourselves, we can see that these tools are extremely complex and hard to navigate, even for experts. These databases are unusable due to the fact that usability testing was not evident in any of the database interfaces.

2. PROBLEM STATEMENT

Visualisation tool: There are many imperfect tools available for sketching and building carbohydrate molecules, including Glycarbo. Glycarbo still faces some issues with usability and robustness, as it has not gone through sufficient user testing. It faces issues with memory and time efficiency, its notations need to be standardised and its usability evaluated. Glycarbo needs to go through a process of editing and testing to be brought up to an industry-standard and become a holistically useful and usable tool for carbohydrate specialists all over the world.

GlycanDB: There is no comprehensive database for all bacterial polysaccharides that has the option of user input, visual aspects, a strong focus on usability and a deep enough set of data, such as the Protein Data Bank but for glycomics research.

3. AIMS

Glycarbo: This work aims to perfect both the front- and back-end of Glycarbo and bring it up to industry standard. We aim to improve the usability and robustness of Glycarbo. This includes curating successful aspects of available tools and incorporating them into Glycarbo's interface. This work will emphasize the use of feedback from expert evaluations and user testing to improve Glycarbo's interface. During the development of Glycarbo, it was found that the feedback from expert-testing was the most valuable [*19]*. So, two rounds of expert evaluations will be conducted throughout the process (before development begins and as a final evaluation). Using the suggestions of its previous developers, we aim to edit Glycarbo to allow its use by a less computer-literate audience. Additionally, Glycarbo cannot handle large PDB files, this causes the browser to freeze, so improvements on its time and memory efficiency need to be made. Its back-end needs to be updated when updates to CarbBuilder are made. This is currently happening on the server, by editing the program files manually. A script to check the CarbBuilder download page for new versions and automatically update the program files is necessary.

GlycanDB: In terms of the database, we aim to create a database, called *GlycanDB* with data on carbohydrates present on the surface of potentially any bacterium. GlycanDB should have the ability to have all data necessary for understanding and further researching any polysaccharide contained in it. Its intent is scalability and addition of new antigens and species through further development. Furthermore, we aim to make sure it is better than currently available database options in terms of usability and easy understanding and navigation. It should be extensible through the community and of course bug-free. This means looking at best design practices from other databases, adding visualisation aspects and understanding our user needs through a strong focus on testing and inquiries.

4. METHODS

4.1 Design Principles

In order to incorporate usability into our design decisions, the following principles will be adhered to throughout the course of the project [16]:

Solve the right problem first: Students and academics want quick and easy access to information via a carbohydrate database.

Understanding user motivations: Understanding the users need for antigen 2D structure creation is important. Users will need these images for academic publications and/or to quickly examine structure to better understand function of molecules. Users also want access to fast, reliable, trustworthy data about carbohydrates via a database.

Understanding context of use and user metrics:

We aim to minimise the number of steps it takes to create and download a useful image. We also aim to maximise the utility of a database search by including only necessary data and incorporating only necessary query options.

Clean interfaces do not indicate good design: We want the flow of control to be clear to beginners (not carbohydrate experts, but students), and offer *immediate, detailed feedback* to users. However, it is important to hide unnecessary complexity from users.

Use standard notations throughout the entire system.

Iterate and test often: We want to make sure that input from supervisors, experts and evaluators can easily be added into the codebase**.** Testing will be done with experts and beginners to ensure different needs are met.

4.2. Approach

Our approach to improving both the usability and robustness of Glycarbo is an iterative process of testing and editing. In order to conduct robustness tests, some standard test cases will be developed and either unit or manual testing will be done. Using the prior work done on the platform, Glycarbo's front- and back-end will be improved to remove any basic system errors and to ensure nothing will impede the testing process. Then, an expert evaluation with a single expert in the field of glycoscience will be conducted, as a starting point. This allows us to remove any obvious errors that may impede the testing process. Such errors include the "straightening" of the carbohydrate molecules in the rendering process and the current non-standard notations and colours used. This will hopefully be done prior to the Initial Software Feasibility demonstrations happening during the week of the 25th of July 2022, so that the software can be checked before user testing begins. Two more iterations of testing will be done. The first, an expert evaluation followed by a round task-based user tests. The design of these task-based tests is an integral part of the project. We are aiming to make changes after each round of testing/feedback. Additionally, after each iteration we will evaluate the usefulness of the feedback and whether a different UEM would be more appropriate, and to further study the use of different UEMs on scientific software.

Our approach in the database section of the project will be to build a database in the mould of the Protein Data Bank (PDB), but focused on carbohydrates. GlycanDB will not be a database purpose built for one specific bacterium, but will be broad. Hopefully it will become the overarching database for glycans as there is currently not a single database used throughout the field as the PDB is. The database data fields will be decided through requirements gathering with experts in the field (Michelle Kuttel and Neil Ravenscroft), but we know we want to have 2D and 3D representations of the molecules as data fields. Those representations, as well as the molecule's origin, references and DOI links to the papers are the bare minimum of data that must be supplied in a query.

Query options will also be decided through requirements gathering, however, we have found that being able to have the option of typing

in one's own queries makes for a better user experience than a drop down menu, especially for databases with large data sets, due to the fact we want to minimise scrolling and ease the user experience.
Both the 3D and 2D representation aspects of GlycanDB will be preloaded into the database rather than hyperlinking elsewhere. This was a confusing part of some surveyed databases and will minimise user error. It will also hopefully speed up the process of acquiring useful data.
The software we will be using is the SQL framework PostgreSQL. This database management system is slightly slower than mySQL but it is in the same ballpark of speed for a project such as ours. It can also support many more complex data structures and queries which may be needed depending on how the 3D models are added and stored in the database.
Usability will be a large focus of GlycanDB. An issue that has been identified with many databases we have surveyed is a complex user interface that even experts struggle with. Using design iterations with the input of expert evaluators will help us make sure that the database is as usable as possible. GlycanDB will be designed over three iterations, first after a contextual inquiry and then the next two iterations will be after expert evaluations. This is expanded on in the evaluation section.
The data set that we will start preloading GlycanDB storage with is a data set of *Klebsiella pneumoniae* glycans that was collected at the University of Cape Town, and some *Escherichia coli* glycans, specifically some O-antigens. Further additions may be possible if time allows.
Following on from this, another area of focus is extensibility. The addition of data into the database from users of the database, above and beyond what will be on there at launch, is something that must exist in the application. For GlycanDB to be widely used in research and industry, there must be a large volume of carbohydrates on the database. It would be possible to load some of this data as the developers (such as the *Klebsiella* and *E. coli* data) but the desired size of the database makes it infeasible for all data to come from the developers, as does the constant moving and advancement in this field as more and more carbohydrates are discovered on bacterium surfaces. Thus, a feature for submissions by the community will be added, subject to acceptance by the people who run the database, upon verification.
A visualisation of the breadth of the database and the functionality of the database is important for usability. A landing page explaining the functionality visually will make for a more pleasant and understandable user interface. Visualising how GlycanDB works and what is incorporated in it will help speed up understanding of the database and how to find what a user may need. This is due to the way we represent the data, as well as the aesthetic-usability effect [7]. Many of the other databases we have surveyed have ignored aesthetics entirely, which makes them hard to use, even if functionality is not an issue, and we will not make the same mistake.
Throughout the refactoring process of Glycarbo and the design of GlycanDB, we aim to follow the set of specialized design principles that were noted above.

5. EVALUATION

5.1. Glycarbo Sketcher and Builder

We are going to be doing a combination of expert evaluations and user testing to test the changes made to Glycarbo's interface.
Our testing will follow an agile approach, where the interface will be edited after every iteration.
We are aiming to recruit 2-3 experts via our supervisor, Michelle Kuttel, to perform in-depth usability evaluations. We are hoping to recruit some post-graduates in the Computer Science department with some experience in Human Centred Design and they can hopefully act as consultants throughout the project, and aid in the co-design of the interface. We are hoping for insight into the perceived flow of the interface (whether it is intuitive or not), and whether the output is in the desired format and the overall ease of use.
Following this, editing of the interface will be done again - using the feedback gathered as well as research into other available tools for visualisation of carbohydrate molecules. This includes incorporating elements of the other interfaces that may benefit the usability of the interface.
Next, a phase of task-based user testing will be done and improvements to the interface made. The design of this task-based test is an integral part of the project. Research, in combination with feedback from the expert evaluation will guide the design of the test. We are hoping to gain insight from the perspective of a beginner through the user testing process. This will be conducted with the help of UCT post-graduate chemistry students.
A final expert evaluation will be done, if time permits, using a "benchmarking" design, in which the original interface will be compared to the new, refactored interface. This is a way to ensure no unnecessary changes have been made that could jeopardize the simplicity of the interface or the intentions of the original developers.
Our user-testing approach, including recruitment information is outlined below in *Figure 1*:

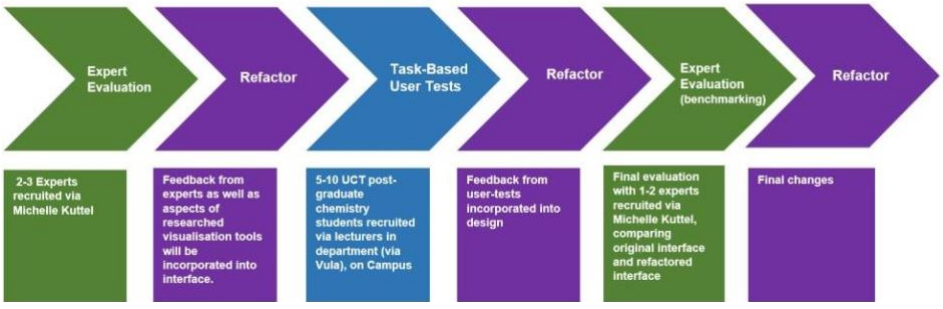

Figure 1: Outline of Testing Process

5.2. GlycanDB

The GlycanDB usability evaluations will consist of two expert evaluations following a contextual inquiry with experts. The experts recruited will be Michelle Kuttel and Neil Ravenscroft. These evaluations will inform the design iterations of the project. On the 16th of June, a contextual inquiry with experts in the field will be conducted in order to set out the guidelines for the needs of GlycanDB through a requirements gathering process. It is important to understand what is needed by the potential users of the database and the supervisors of this project in order to start the project. The implementation of GlycanDB will then start on the 28th of June.

On the 4th of August, after the initial design phase, an expert evaluation will be conducted in order to guide the changes needed for the second iteration of the project. This will then commence, and the changes will be implemented until the 12th of August.

On the 14th of August, a final expert evaluation will be conducted in order to guide the finalisation and perfection of the database ahead of the final code submission on the 5th of September.

This is all fully fleshed out in the gantt chart below *(Figure 3).*

The choice of this iterative approach will maximise utility and usability of GlycanDB's interface.

6. ETHICAL CONSIDERATIONS

The ethical clearance we need in order to find evaluators has been obtained. There are no risks associated with being an evaluator for this project, and thus the only ethical considerations we must make is minimising time wasting and making sure that students who participate in our evaluation have no pressure from their supervisors in order to join the evaluation, and instead have full informed consent. As far as other ethical or legal obligations go, we must make sure that the information in the database is always 100% correct in order to not negatively affect research.

7. Risks and Challenges

7.1. Risks

Tabulated below *(Figure 2)* are our risk, their likelihoods and severity, as well as our plans to manage and mitigate them.

7.2. Challenges

The integration of the two systems is an issue, and understanding how they may work together and how connected they should be is paramount. As the focus is on web-tools for carbohydrate research we think incorporating them as an all-encompassing web tool system would work well, and would increase the utility of each other when bonded together. As far as other challenges go, we are both lacking carbohydrate domain knowledge - this should not be a massive issue for software design however, as we will be helped in understanding by our supervisor. We also have little experience with web-development. We must make sure to upskill ourselves in this area over the next few weeks to ensure we have the skills necessary to make this project a success. A focus on speed in the database and visualisation tool is important as the use of 3D software is often slow which may be a challenge. It will also be a challenge to make sure usability in both tools is fully fleshed out, and we must make sure this is a priority at every stage and iteration of the development. Finding evaluators and making sure we ask the correct questions, as well as getting ethical clearance is a challenge we must take very seriously.

Description	Probability (1-10)	Severity (1-10)	Mitigation Strategy	Monitoring	Management
Illness of team member (COVID-19)	5	8	Plan clearly and well in advance to ensure some leeway. Try to finish work as early as possible.	Monitor own health and symptoms.	Reduce scope (testing iterations or design of database)
Insufficient time to complete project	4	9	Plan carefully and in depth. Support each other throughout process.	Regularly check-in with supervisor with progress reports. Begin write-ups immediately.	Reduce scope of project.
Insufficient skills to complete project	4	9	Check coding requirements early. Ensure we are familiar with development framework prior to coding blocks.	Self-monitoring of task-completion rate/rate of adoption of new framework/skills.	Get assistance from individuals with knowledge of the framework/supervisor.
Impacts functionality of Glycarbo permanently	2	10	Ensure we are using local copies of source code. Save original code locally before project begins.	Regularly check if Glycarbo still working as expected.	Start from scratch and use original saved code.
Failure to recruit enough test participants	5	5	Begin recruitment early and use tools at our disposal (Vula, access to supervisor and thus lecturers in chemistry department)	Use a sign-up method to keep track of participants.	Change tests to be more thorough to get more in-depth feedback from fewer participants.

Figure 2: Risk Matrix

8. TIMELINE

Figure 3 shows a Gantt chart of the members' activities and *Figure 4* shows a table of activity dates and deadlines Both can be seen below.

9. WORK ALLOCATION

Lauren will work on testing and improving the Glycarbo interface. She will conduct the task-based user testing and benchmark testing,

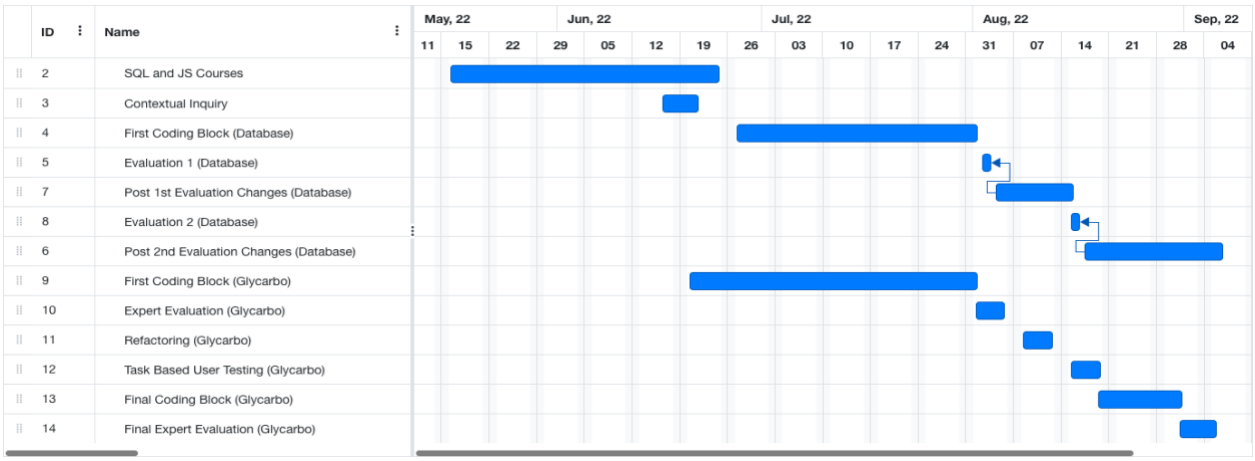

Figure 3: Gantt Chart of Project Activities

as well as the bug fixing on the source code. Joseph will design the GlycanDB and perform expert-testing. He will be building the database from scratch, and will design the backend. Both coders will design the web interface that houses the database and Glycarbo.

10. SUCCESS MEASUREMENT AND ANTICIPATED OUTCOMES

The project will be a success if the software runs without issues, can handle large loads, and can handle complex computations. It is vital to the success of this project that the final product is seen as more usable and more robust than currently available tools. Good feedback from experts during a final inspection will mean it is up to industry standard and will be a useful tool for students and academics in the field. We expect the code to be bug free and robust, and we expect our strong testing and evaluation methods to ensure the software is seen as very usable.

11. POSSIBLE IMPACTS

The increased ease of use of Glycarbo will mean it becomes a more efficient and convenient visualisation tool. It could become a widely used tool for research in academic institutions all over the world. An extensible and usable database such as GlycanDB will mean improved research into glycans on bacteria surfaces for academic and industry purposes. This will hopefully result in better treatments for illnesses and infections caused by these bacteria.

Activity/Deadline	Date
Project Proposal presentations	Tue 24-May, Wed 25-May
Project Proposal due	Fri 27-May
Glycarbo Initial Evaluation	30 May - 3 June
Ethics applications preferred deadline	Fri 10-Jun
Glycarbo Editing (Phase 1)	3 June - 25 July
Database Contextual Inquiry	16 June
Ethics applications final deadline	Mon 11-Jul
Revised Proposal Finalised and uploaded to Vula	Fri 15-Jul
Initial Software Feasibility Demonstration	Mon 25-Jul to Fri 29-Jul
Glycarbo 1st Expert Evaluations	1-5 August
Database 1st Expert Evaluation	4 August
Glycarbo Editing (Phase 2)	6-14 August
Database 2nd Expert Evaluation	14 August
Glycarbo User Testing	15-19th August
Glycarbo Editing (Phase 3)	19-23 August
Complete Draft of final paper due (if not done -10%)	Tue 23-Aug
Glycarbo Final Expert Evaluation (Final Check)	23-26 August
Project Paper Final Submission	Fri 2-Sep
Project Code Final Submission	Mon 5-Sep
Final Project Demonstration	Mon 19-Sep to Fri 23-Sep
Poster Due	Mon 3-Oct
Website Due	Mon 10-Oct

Figure 4: Deadlines and Activities

REFERENCES

[1] Patricia S. Abril and Robert Plant, 2007. The patent holder's dilemma: Buy, sell, or troll? *Commun. ACM* 50, 1 (Jan, 2007), 36-44. DOI: https://doi.org/10.1145/1188913.1188915

[2] Abhijit Aithal, Arun Sharma , Shilpy Joshi et al. 2012. PolysacDB: A Database of Microbial Polysaccharide Antigens and Their Antibodies. PLOS ONE Volume 7, Issue 4, April 2012, 1-4. DOI: https://doi.org/10.1371/journal.pone.0034613

[3] Felipe Albrecht, Peter Ebert and Markus List. (2017). Ten Simple Rules for Developing Usable Software in Computational Biology. *PLOS Computational Biology. Volume 13(1), 1-5.* https://doi.org/10.1021/acs.jctc.1c00169.

[4] Sten Andler. 1979. Predicate path expressions. In *Proceedings of the 6th. ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages (POPL '79)*. ACM Press, New York, NY, 226-236. DOI:https://doi.org/10.1145/567752.567774

[5] Rena Astronomo, Dennis Burton. 2010. Carbohydrate vaccines: developing sweet solutions to sticky situations?. *Nat Rev Drug Discov*, Volume 9**,** 308–324 (Aprin 2010). DOI: https://doi.org/10.1038/nrd3012

[6] Helen Berman, John Westbrook, Zukang Feng et al. 2000. The Protein Data Bank, *Nucleic Acids Research*, Volume 28, Issue 1, 1 January 2000, 235–242 DOI: https://doi.org/10.1093/nar/28.1.235

[7] Nick Cawthon and Andrew Moere. 2007. The Effect of Aesthetic on the Usability of Data Visualization, *2007 11th International Conference Information Visualization (IV '07)*, , pp. 637-648, DOI: https://doi.org/10.1109/IV.2007.147

[8] DrawGlycan-SNFG. 2020. Render glycans and glycopeptides with fragmentation info. using the Symbolic Nomenclature for Glycans [SNFG]. Retrieved from http://www.virtualglycome.org/DrawGlycan/. Accessed May 2022.][Glyco.me SugarBuilder. Retrieved from https://beta.glyco.me/sugarbuilder. Accessed May 2022.

[9] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. DOI:https://doi.org/10.1007/3-540-09237-4.

[10] Ksenia Egorova, Anna Kondakova, Phyl Toukach. 2015. Carbohydrate Structure Database: tools for statistical analysis of bacterial, plant and fungal glycomes. *Database (Oxford)*. Volume 2015, September 2015. DOI: https://doi.org/10.1093/database/bav073

[11] GLYCAM Web. 2020 Complex Carbohydrate Research Center, University of Georgia, Athens, GA. Retrieved from http://glycam.org. Accessed May 2022

[12] Hanan Hayat. 2020. Evaluating Software Usability from Different Perspective: A Framework for Encouraging Usability Evaluations by Focusing on Software Developing Projects. Master's Thesis. Loughborough University, Loughborough, England.

[13] David Kosiur. 2001. *Understanding Policy-Based Networking* (2nd. ed.). Wiley, New York, NY..

[14] Bharathi Reddy Kunduru, Sanjana Anilkumar Nair, Thenmalarchelvi Rathinavelan. 2015. EK3D: an *E. coli* K antigen 3-dimensional structure database. *Nucleic Acids Research*, Volume 44, Issue D1, January 2016, D675–D681. DOI: https://doi.org/10.1093/nar/gkv1313

[15] PolysGlycanBuilder. *A user friendly tool for to build 3D structures of complex glycan and polysaccharides.* http://glycan-builder.cermav.cnrs.fr/. Accessed May 2022.

[16] Lavanya Ramakrishnan and Daniel Gunter. Ten principles for Creating Usable Scientific Software. (2017*). IEEE 13TH International Conference on eScience. October 24-27, 2017, Auckland, New Zealand, 210-218*. doi: 10.1109/eScience.2017.34

[17] Laurisha Rampersaad, Sarah Blyth, Ed Elson and Michelle Kuttel, M. 2017. Improving the Usability of Scientific Software with Participatory Design: a new Interface Design for Radio Astronomy Visualisation Software. In Proceedings of SAICSIT'17, September 26-28, Thaba Nchu, South Africa, 1-9. https://doi.org/10.1145/3129416.3129899

[18] Miguel Rojas-Macias, Jonas Ståhle, Thomas Lütteke, Göran Widmalm, Development of the ECODAB into a relational database for *Escherichia coli* O-antigens and other bacterial polysaccharides, *Glycobiology*, Volume 25, Issue 3, March 2015, 341–345. DOI: https://doi.org/10.1093/glycob/cwu116

[19] Wood, Tristan. 2020. Development of a web tool for 3D visualization of carbohydrate molecules. Final Paper. *University of Cape Town, Cape Town, South Africa.*