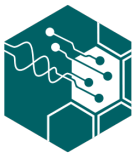

Junior Professorship
Computational
Life Science

RWTH AACHEN UNIVERSITY

Proposal

My Proposed Contribution to Science

Elouën Corvest

Contents

1 Models are the Past, Present, and Future

Mathematical modelling has been a staple of sciences for centuries [1]. From using simple geometry to construct temples and calculate the distance of the Sun to the Earth, to simulate neutron transport in nuclear fission [2]. This approach has also not been overlooked in the biological sciences, where models have been used to understand population dynamics [3], enzyme kinetics [4], and many more. Over time, these mathematical models have become more and more complex, which entail a difficulty in understanding them. Fortunately, humanity has developed more and more ways to communicate with complex mathematics by using actual computational power. This advancement has allowed modern sciences to create computational models that help with scientific advances.

There are many different types of computational models used in biological studies. In photosynthesis alone, some may be used to understand a specific part of the mechanism [5, 6]. In contrast, others may be used to predict the behaviour and "output" of a plant [7, 8]. Some are built using kinetic knowledge, while others are based on stochastic processes. They all have key differences that make each unique and suitable for different purposes. While variety is the spice of life, this also introduces a significant complexity when trying to choose, understand, and use these models.

This "problem" has already been in the focus of the scientific community, therefore some proposed solutions already exist. These solutions tackle the problem from different angles, which can be separated in three different categories: Creation (Past), Presentation (Present), and Sharing (Future).

1.1 Creation (Past)

There are many different ways to create a computational model in biological studies. Some may be built completely from scratch, using programming languages like Python, or MATLAB. Others may be built using more specialised software developed for this kind of tasks, like MxLpy [9], COPASI [10], or Tellurium [11]. Whatever the method, all have their own respective costs and benefits, however, one thing is common to all: the annotation and documentation of the model is based on the author of the model.

This annotation and documentation is key to understanding the model, as it allows the author and others to understand how, why, and what the model represents. However, as there are many different ways to create a model, there are also many ways to document it. There have been attempts to standardise this process, one of the most notable being the Systems Biology Markup Language (SBML) [12]. SBML is an XML-based format for representing computational models in systems biology. It allows for the sharing and exchange of models between different software tools, making it easier for researchers to collaborate and build upon each other's work. Since its first publication in 2003 [12], the SBML language has been updated and extended several times [13]. It has become a common way to represent biological computational models, with many more also available.

While the idea and reception of such standardised languages have been positive, they still carry some issues. Their implementation is not always straightforward and with every new version, new features are added that may not be backward compatible.

1.2 Presentation (Present)

Reusability and reproducibility of work has always been a core principle of scientific work, however even the sharing of data has fallen victim to poor practices. An analysis conducted by Roche et al. [14] by examining a dataset of 100 nonmolecular evolutionary or ecological articles between 2012 and 2013 that have been published in seven different leading journals found that 56% of the articles archived incomplete data and furthermore 64% of the articles were not fully reproducible. This has sparked the creation of the Findable, Accessible, Interoperable, and Reusable (FAIR) Guiding principles [15], which has since been extended to the Credible, Understandable, Reproducible, and Extensible (CURE) Guidelines [16].

These guidelines propose a set of principles to ensure good scientific work that also extends to computational models. They encapsulate a varying range of aspects, ranging from how the model is created and validated, to how it is provided and documented. The FAIR guidelines have already been widely adopted (see Fig. 1), which is shown by the 7841 citations [17] (as of October 2025). Solely in 2024 the comment was cited 1408 times (as of October 2025), in which 69.2% are articles. As the CURE guidelines are more recent, they have not yet seen the same adoption rate, but they are gaining traction in the community.

1.3 Sharing (Future)

While computational models in biological studies rise of popularity is at surface level a very positive change, it does bring its own troubles with. The more models available to the public, the larger the pool of choices. As every model is released by using a publication, someone that wishes to use a model will have to sift through numerous publications to find a model that suits their needs. Solely in 2024, 1475 publications were published when querying for "photosynthesis model" [18] and the trend seems to continue upward every year (see Fig. 2). Due to this trend and accessibility issue, an idea for databases arose. The most notable one, called BioModels [19], already includes 1096 manually curated models and 1661 non-curated ones, totaling to 2757 different models.

BioModels does not restrict what type of model can be submitted to the database. Therefore, the assortment of models is very diverse and ranges between species, from *Homo Sapiens* to *Saccharomyces cerevisiae*, modelling approaches, from ordinary differential equation (ODE) to constraint-based, and modelling formats, from SBML to Python. What is restricted, however are some aspects of model publication. BioModels has constructed their guidelines called Minimal Information Required In the Annotation of Models (MIRIAM), which focuses on consistent annotation and curation of computational models [20]. To further their goal of easier access to models, the team behind BioModels also manually curate models to a general standard, with a prerequisite that these submitted models are written in SBML. This effort has proven to be popular in the scientific community to many extents, as it not only makes sharing easier, but also encourages modellers to publish their model into the public domain, as BioModels does not take ownership of any submitted model. Even though this resource is hugely valuable to science, it still has some flaws.

A study done by Tiwari et al. [21] has analysed the reproducibility of 455 models taken from the BioModels database, which equates to over 20% of literature-based models available on BioModels at the time of writing. They found out, that 49% of the published models are not directly reproducible, even with the

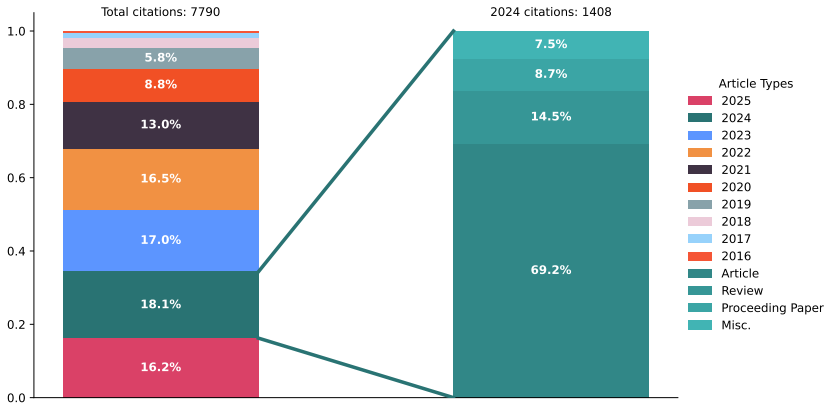

Figure 1: The usage of the FAIR Guidelines found by citations.
The citations of the original Findable, Accessible, Interoperable, and Reusable (FAIR) guiding principles comment [15], from the years 2016 to 2026, were found. Additionally, the citations of the year 2024 have been separated into the different types of publications. Articles, Review, and proceeding papers have their own category, while others (editorial material, early access, data paper, etc.) are grouped together. Data taken from Web of Science on 29th October 2025. [17]

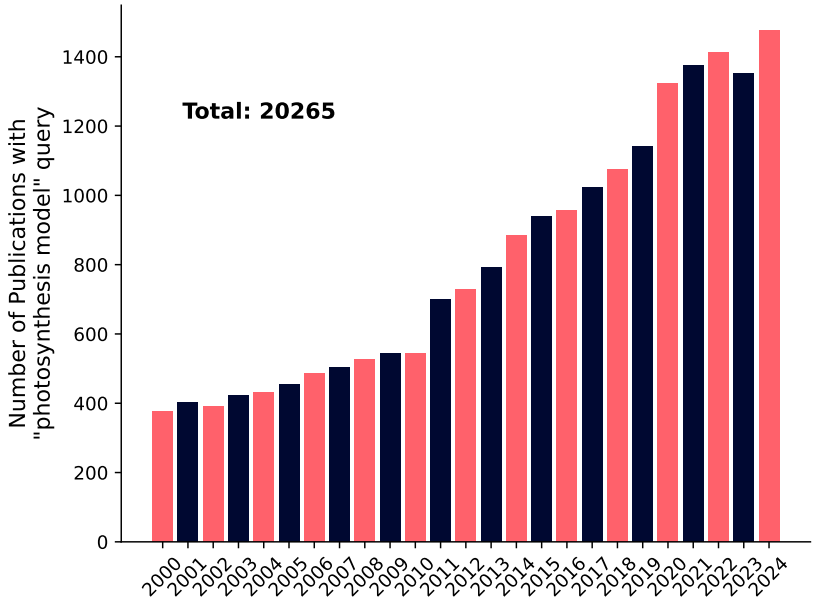

Figure 2: Number of publications found by using "photosynthesis model" as a query.
The number of publications found by querying "photosynthesis model" from the years 2000 to 2024, totaling to 20265 publications. Data taken from Web of Science on 11th November 2025. [18]

MIRIAM-following prerequisite of uploading to Biomodels. From these directly irreproducible models, 25% (54 models, which is 12% from the original 455 models) could be saved by manual empirical correction or author support. At the end, about 37% of the total analysed models (equates to 169 models) from the BioModels database could not be reproduced at all. The common issues that sometimes could be manually repaired, were the missing of critical model information, wrongly described equations, or other small, easily overseen, mistakes, that can completely destroy the description of the model. To combat these problems and errors, the authors of the study have proposed a reproducibility scoring card, that should be implemented when publishing a model, either in the BioModels database or otherwise. It has to

be noted, that this study has been done by the BioModel team themselves who have opened the discussion of the flaws in their database and science reproducibility in general. This shows, the need and desire of the community to make model sharing as easy as possible.

2 A Real Solution

It is clear that there is a problem with computational models right now and while there are solutions proposed they still did not have enough traction to truly change this scientific field. This is where **GreenSloth** comes into play. The main aspect of **GreenSloth** is creating a bridge between scientists that create models and those that will actually try to use them for experimental work, as these two groups often do not coincide. To create this bridge, **GreenSloth** will encompass many aspects of computational work that are already being tackled on as mentioned prior but with one key advantage. It will minimise its scope onto kinetic models of photosynthesis.

While models come in many different shapes and sizes, all showing different things, We believe it is important to first reduce the problem at hand to a manageable feat, by only encompassing a single, yet incredibly important, biochemical process. The importance of this process has also not escaped the eyes of the modelling community, as seen by the incredible amount of publication every year (see Fig. 2). Sifting through these amounts of publications is hard for both the modellers, that may want to expand their model with other interpretations, and the experimentalists that may want to build upon their work by using the new, popular and useful method of computational modelling.

The **GreenSloth** project can be seperated into four main branches (see Fig. 3). These branches are Model Creation Pipeline, Model Comparison & Validation, Database, and Website. Each branch has its own set of tasks and goals that are integral to **GreenSloth** to create a comprehensive platform for kinetic models of photosynthesis. All these branches will be explained in the following sections.

2.1 The Promise

As shown with the large amount of publications on kinetic models of photosynthesis every year (see Fig. 2), and a large amount that cannot be directly reproduced (see Tiwari et al. [21]), model creation is not a simple and straightforward task. You start by having a question on real systems that you specify into a problem. From there, you gather knowledge on your system to design a conceptual model. With this conceptual model, you can formalise a mathematical model, where you sometimes have to make assumptions based on your knowledge of the system. You can now either calculate analytical solutions, or more commonly program your model into a computational format to simulate it. Either way, you arrive into results, which you have to validate with your initial or other real systems. If the results do not depict your system well enough, you reevaluate and create a new problem to start the process again. If the results are satisfactory, you have solved your real questions with a mathematical model (see Fig. 4).

While **GreenSloth**'s intention is solely based on a model that has already reached the solved stage, it still encapsulates several aspects of the model creation process. The design of a conceptual model, the formulation into a mathematical model, the programming into a computational format, examples of model validation are all part of the **GreenSloth** project. Having these aspects semi-standardised, will ensure the quality of models which will help the modellers with the publishing of their models,

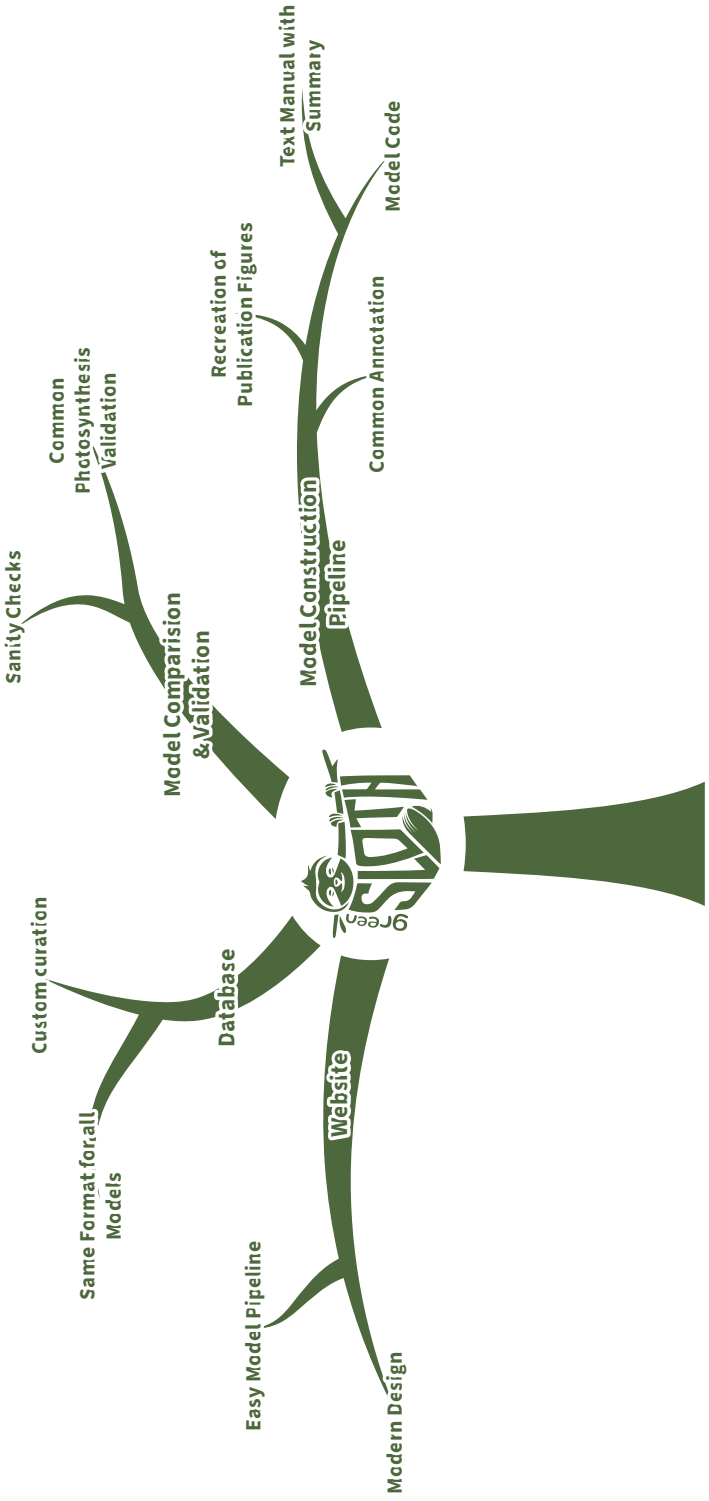

Figure 3: The tree of GreenSloth.
A schematic representation of the different aspects of the **GreenSloth** project and how they relate to each other. The four main branches are Model Creation Pipeline, Model Comparison & Validation, Database, and Website. Each branch has its own set of tasks and goals that will be tackled to create a comprehensive platform for kinetic models of photosynthesis.

and keeping them under the CURE guidelines. Furthermore, it will also help users of models, understand the process of mathematical modelling better, and therefore be able to use models more effectively.

To help the users of models more,

GreenSloth is the central hub to find photosynthesis kin etics models. It provides a way to compare models, see their capabilities and help the user with deciding which model is best suited for their needs. This shoudl eliviate the problem of finding the right model, and opens up the field of mathematical models of photosynthesis to a wider audience. Therefore, **GreenSloth** represents the bridge between a model author and a model user, helping both sides with their respective problems.

2.2 Model Creation Pipeline

To ensure a standardised model format, a pipeline on how to create a model will be created. This pipeline will emcompass a guide to formatting the code itself, automatically export the model infromation, for example the ODE sytem, parameters, etc., into tables and math expressions, force the user to provide a summary and a recreation of the publication figures with code. All this information will then be stored in one directory, with a Markdown file as its front page. This directory can then be easily uploaded to the **GreenSloth** database. This pipeline will be provided as a Python package under the name **GreenSlothUtils**.

2.2.1 Model Code

To ensure a common format for the model code itself, a set requirement for it, will be using the MxlPy package [9]. This package enables a building block approach to building ODE models, however by keeping the familiar linear flow of Python. Furthermore, it allows for easy simulation, plotting, parameter fitting, and many more functionalities that are crucial while using a model. Because of that, and continuing support from the authors, it is the perfect candidate to be the base of **GreenSloth** models.

On top of that, the actual formatting of the code will be standardised to a template that will be provided in the **GreenSloth** model creation pipeline. This template will ensure that all necessary information is provided, and that the code is easy to read and understand. It will be continously updated to reflect the best practices in coding and modelling, for example by having the same names for common biological variables, such as ATP, NADPH, etc.

2.2.2 Glossaries

To ensure that all models are properly annotated, a set of glossaries will be created. These glossaries will be created for the reactions, the variables, the parameters, and the derived versions of the latter two. These glossaries will provide a standardised way of naming and describing these entities, ensuring that they are easily understandable and comparable between models. Furthermore, they will provide a way to link these entities to external databases, such as Kyoto Encyclopedia of Genes and Genomes (KEGG) [22]. To facilitate this process, the **GreenSlothUtils** package will provide functions to automatically extract the information from the MxlPy model. Furthermore, it will alert the user if all the information that the model entails can be found in the glossaries.

These glossaries will come in two forms. One tied with the model itself, as many parameters and variables may be model-specific, due to assumptions made during the model creation. These glossaries will follow how the publication of the model presents the existing information. The second form will be a global glossary, that will provide a standardised way of naming and describing common entities between models. This global glossary will be continuously updated as more models are added to the **GreenSloth** database.

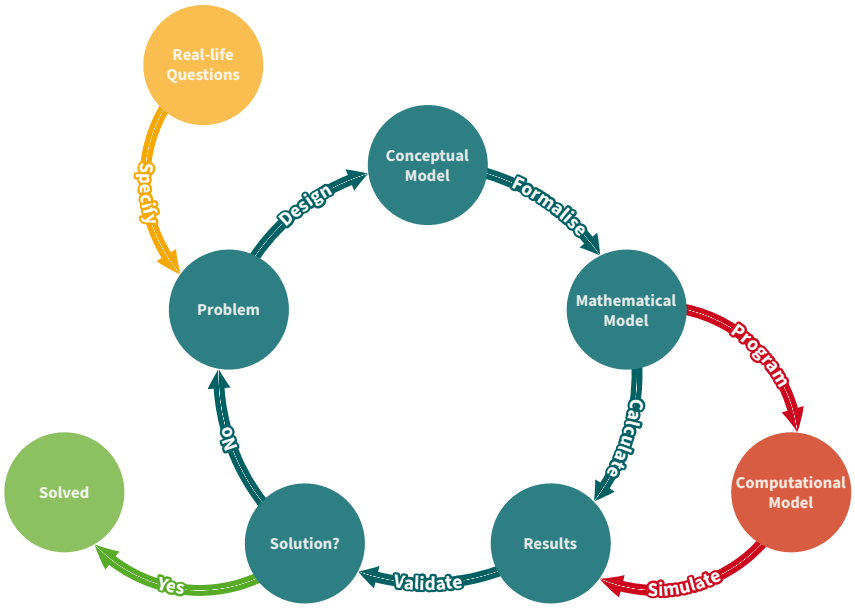

Figure 4: A simple scheme of mathematical modelling.
A schematic representation of the mathematical modelling process, starting from a real system with a question, to finding a solution using a model. After specifying the problem, knowledge is gathered to design a conceptual model. This should then be formalised into a mathematical mode, by adding assumptions if needed. From there, analytical or computational solutions can be found to arrive at results. These results need to be validated with real systems, and if they are not satisfactory, the process starts again from the problem definition. If they are, the real question has been solved using a mathematical model.

2.2.3 Gather Information

To make the model information as easily accessible as possible, the **GreenSlothUtils** package will provide functions to automatically create a markdown file that will summarise the model. It will automatically create tables from the glossaries, and let the user be able to automatically generate the ODE system and other needed mathematical expression of the model in LaTeX format. Furthermore, it will provide space for the model validation to remind the user to include them.

This markdown file is a prerequisite for any model in the database, and shall facilitate the process of publishing the models to the CURE standards.

2.2.4 Packaging it up

All the information gathered from the previous steps will then be packaged into a single directory, that will contain the model code, the glossaries, the markdown file, and all the code used for the model validation. This directory will then be easily uploadable to the **GreenSloth** database and also directly be transformed into a Python package that can be installed and used by anyone. This enables easy sharing and reusing of models, as well as ensuring that all the necessary information is provided.

2.3 Model Comparison & Validation

With the information gathered from the model creation pipeline, the first step to comparing models is already done. The standardised format of the models, as well as the tables and mathematical expressions of critical model information already show what different models were made to show. However, to truly compare models, they need to be validated the same way. Therefore, a set of standardised validation protocols will be created, that will test each

model to common practices in ODE models and photosynthesis work. These protocols will be provided as part of the **GreenSlothUtils** package, and will be applicable to any model in the database.

The first step in validation is to ensure that the model is mathematically sound. This includes checking for mass balance, energy balance, and other common checks that are applicable to ODE models. One crucial aspect, for example, is to ensure that the model does not produce negative concentrations, as this is not physically possible. This step can colloquially be called "Sanity Checks", as it ensures that the model behaves in a realistic and sane manner.

The second step in validation, is the recreation of the publication figures. This step is crucial to ensure that the model behaves as expected and presented, and that the code provided is correct. This means that the code used to create these figures needs to be manually curated to ensure true results. With the recreation of these figures, a score can be set to see how well the model performs compared to the publication. Depending on the model curated and the information provided, help from the authors may be needed to ensure the best possible recreation of the figures. However,if additional information is needed to recreate the figures in the publication, this is proof that the model does not follow the CURE principles, and therefore the version on will entail the revisions needed and give the community the best possible version of the model.

The third and final step in validation, is to create a set of standardised tests that are applicable on a wide variety of photosynthesis models. These tests will be based on expert knowledge, from modellers to experimentalists, and will show the extensive simulation possibilities of a model. In this step, a model may not be able to pass all the tests, as it may not have been designed to do so. Therefore, a rating on how "good" a model is, cannot be done using this step as it can with the former mentioned. However, this will give a good overview of the model capabilities, and will allow for easier comparison between models and should help users determine which model is best suited for their needs. One example for these tests include "injecting" the simple Farquhar, von Caemmerer, and Berry (FvCB) model [7] into the model and see how it simulates the commonly used relation of carbon assimilation (A) and intercelluar CO_2 concentration (C_i).

2.4 Database

Having a centralised database to store and share models is crucial to ensure that they are easily accessible and reusable. This database will be the backbone of **GreenSloth**, as it will provide a way to store the models. In the beginning, this database will be a simple file-based system, where each model is stored in a separate directory. This can easily be hosted on GitHub, which provides version control and collaboration features. Furthermore, it is a well-known and used platform that strives on an open-source nature. Exactly that, what **GreenSloth** is about. This openness, also allows access to all the models inside the database, for users, but also other parts of **GreenSloth**, like the website (see Section 2.5 Website). However, as the number of models grows, this system may become unwieldy. Therefore, a more robust database system will be implemented, that will still let the website access all models easily.

2.5 Website

Sharing models is no easy task. Only using publications can be tedious, as the code may not be available, or the model is not properly

explained. Therefore, a centralised platform to collect and present models, would be beneficial to the scientific community. We believe that the easiest way to create such platform is in a web-based format, where anyone can access it from anywhere. This platform will present models in a standardised format, that includes a summary, a table of model information, and the recreation of the publication figures. Furthermore, having a central location with all the models allows for easier comparison between them.

Web development and design has become an integral part of modern communication, boasting around 74% of the global population being internet users in 2025 [23]. From social media to spreading knowledge, from entertainment to education, and many more, websites shape how people interact with their politics, their art, their community, and themselves [24]. Therefore, a great deal of research has been done to optimize and cater web pages to the public. Good we design can be separated into four categories [25]: ease of navigation, site speed, clarity and conciseness of content, and visually appealing design. There are many resources that help web developers deal with these aspects, such as books and guidelines [26], further updates to existing basic web development languages (HTML, CSS, JavaScript), additional web development frameworks (React, Angular, Vue.js) that help create better web design, and many more. Web development is as popular as never before.

Therefore, giving access to **GreenSloth** through a well-designed website is crucial to its success. This rise to popularity of web development can be seen in science, as more and more scientific databases and tools are being made available through web interfaces. A few examples are Basic Local Alignment Search Tool (BLAST) [27], KEGG [22], Biochemical, Genetic and Genomic (BiGG) [28], or the aforementioned BioModels, with many more examples in various scientific fields. This rise in popularity also shows the advantage of web interfaces over other platforms, such as applications or software packages. They are much more accessible and easier to use for the public, as they do not require any installation or specific operating system. Furthermore, they advocate an open-source approach, and therefore align with the FAIR and CURE principles, making scientific data and tools more accessible and reusable for everyone.

3 Conclusion

The **GreenSloth** project is a promising step into having a more transparent bridge between model authors and model users. By providing a platform where models can be shared, explored, and understood in depth, it fosters a more collaborative and efficient scientific community. The starting blocks of this project, have already been laid out.

References

[1] Sandip Banerjee. *Mathematical Modeling: Models, Analysis and Applications*. 2nd ed. Boca Raton: Chapman and Hall/CRC, 2021. ISBN: 978-1-351-02294-1. DOI: 10.1201/9781351022941.

[2] Christophe Demazière. *Modelling of Nuclear Reactor Multi-Physics: From Local Balance Equations to Macroscopic Models in Neutronics and Thermal-Hydraulics*. London: Academic press, 2020. ISBN: 978-0-12-815070-2.

[3] Peter J. Wangersky. „Lotka-Volterra Population Models". In: *Annual Review of Ecology and Systematics* 9 (1978), pp. 189–218. ISSN: 0066-4162. JSTOR: 2096748.

[4] Athel Cornish-Bowden. „One Hundred Years of Michaelis–Menten Kinetics". In: *Perspectives in Science* 4 (2015), pp. 3–9. ISSN: 22130209. DOI: 10.1016/j.pisc.2014.12.002.

[5] Anna Matuszyńska et al. „A Mathematical Model of Non-Photochemical Quenching to Study Short-Term Light Memory in Plants". In: *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1857.12 (2016), pp. 1860–1869. ISSN: 00052728. DOI: 10.1016/j.bbabio.2016.09.003.

[6] Meng Li et al. „Impact of Ion Fluxes across Thylakoid Membranes on Photosynthetic Electron Transport and Photoprotection". In: *Nature Plants* 7.7 (2021), pp. 979–988. ISSN: 2055-0278. DOI: 10.1038/s41477-021-00947-5.

[7] Susanne Von Caemmerer. „Steady-state Models of Photosynthesis". In: *Plant, Cell & Environment* 36.9 (2013), pp. 1617–1630. ISSN: 0140-7791, 1365-3040. DOI: 10.1111/pce.12098.

[8] Xin-Guang Zhu et al. „*E*-photosynthesis: A Comprehensive Dynamic Mechanistic Model of C3 Photosynthesis: From Light Capture to Sucrose Synthesis". In: *Plant, Cell & Environment* 36.9 (2013), pp. 1711–1727. ISSN: 0140-7791, 1365-3040. DOI: 10.1111/pce.12025.

[9] Marvin Van Aalst et al. *MxlPy - Python Package for Mechanistic Learning in Life Science*. 2025. DOI: 10.1101/2025.05.06.652335.

[10] Stefan Hoops et al. „COPASI—a COmplex PAthway SImulator". In: *Bioinformatics* 22.24 (2006), pp. 3067–3074. ISSN: 1367-4811, 1367-4803. DOI: 10.1093/bioinformatics/btl485.

[11] Kiri Choi et al. „Tellurium: An Extensible Python-Based Modeling Environment for Systems and Synthetic Biology". In: *Biosystems* 171 (2018), pp. 74–79. ISSN: 03032647. DOI: 10.1016/j.biosystems.2018.07.006.

[12] M. Hucka et al. „The Systems Biology Markup Language (SBML): A Medium Forrepresentation and Exchange of Biochemical Network Models". In: *Bioinformatics* 19.4 (2003), pp. 524–531. ISSN: 1367-4811, 1367-4803. DOI: 10.1093/bioinformatics/btg015.

[13] *SBML.Org: Essential Publications about SBML*. https://sbml.org/documents/publications/.

[14] Dominique G. Roche et al. „Public Data Archiving in Ecology and Evolution: How Well Are We Doing?" In: *PLOS Biology* 13.11 (2015), e1002295. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.1002295.

[15] Mark D. Wilkinson et al. „The FAIR Guiding Principles for Scientific Data Management and Stewardship“. In: *Scientific Data* 3.1 (2016), p. 160018. ISSN: 2052-4463. DOI: 10.1038/sdata.2016.18.

[16] Herbert M. Sauro et al. „From FAIR to CURE: Guidelines for Computational Models of Biological Systems“. In: *ArXiv* (2025), arXiv:2502.15597v1. ISSN: 2331-8422.

[17] *Citations of Comment: The FAIR Guiding Principles for Scientific Data Management and Stewardship – 7,841 – Web of Science Core Collection*. https://www.webofscience.com/wos/woscc/sum 0d91-4cf2-9009-874ff09eaec6-0184c41f53/date-descending/1.

[18] *Citation Report - 1,475 - Web of Science Core Collection*. https://www.webofscience.com/wos/woscc/s 6cfd-4ca6-9a3c-104bc60c1ee6-0187df4b2d/5787b316-f80a-4f4e-ada2-ef631e25ff12-0187df3faf/relevance/

[19] Rahuman S Malik-Sheriff et al. „BioModels—15 Years of Sharing Computational Models in Life Science“. In: *Nucleic Acids Research* (2019). ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkz1055.

[20] Nicolas Le Novère et al. „Minimum Information Requested in the Annotation of Biochemical Models (MIRIAM)“. In: *Nature Biotechnology* 23.12 (2005), pp. 1509–1515. ISSN: 1546-1696. DOI: 10.1038/nbt1156.

[21] Krishna Tiwari et al. „Reproducibility in Systems Biology Modelling“. In: *Molecular Systems Biology* 17.2 (2021), e9982. ISSN: 1744-4292. DOI: 10.15252/msb.20209982.

[22] M. Kanehisa. „KEGG: Kyoto Encyclopedia of Genes and Genomes“. In: *Nucleic Acids Research* 28.1 (2000), pp. 27–30. ISSN: 13624962. DOI: 10.1093/nar/28.1.27.

[23] *Measuring Digital Development: Facts and Figures 2025*.

[24] *DataReportal – Global Digital Insights*. https://datareportal.com.

[25] Fadliyani Nawir and Satya Arisena Hendrawan. „The Impact of Website Usability and Mobile Optimization on Customer Satisfaction and Sales Conversion Rates in E-commerce Businesses in Indonesia“. In: *The Eastasouth Journal of Information System and Computer Science* 2.01 (2024), pp. 15–30. ISSN: 3025-566X, 3026-6041. DOI: 10.58812/esiscs.v2i01.324.

[26] Jason Beaird and Alex Walker. *The Principles of Beautiful Web Design*. Fourth edition. Collingwood: SitePoint, 2020. ISBN: 978-1-925836-36-3.

[27] Stephen F. Altschul et al. „Basic Local Alignment Search Tool“. In: *Journal of Molecular Biology* 215.3 (1990), pp. 403–410. ISSN: 00222836. DOI: 10.1016/S0022-2836(05)80360-2.

[28] Zachary A. King et al. „BiGG Models: A Platform for Integrating, Standardizing and Sharing Genome-Scale Models“. In: *Nucleic Acids Research* 44.D1 (2016), pp. D515–D522. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkv1049.