 doi:10.20944/preprints202009.0457.v1

Fostering Accessible Online Education Using Galaxy as an e-learning Platform

Authors

Beatriz Serrano-Solano[1*], Anika Erxleben[1*], Cristóbal Gallardo-Alba[2], Helena Rasche[2], Saskia Hiltemann[3], Melanie Föll[4,5,6], Matthias Fahrner[5,7,8], Marc J. Dunning[9], Marcel Schulz[10], Beáta Scholtz[11], Dave Clements[12], Anton Nekrutenko[13], Bérénice Batut[1] and Björn Grüning[1]

[1] Bioinformatics Group, Department of Computer Science, Albert-Ludwigs-University Freiburg, Georges-Köhler-Allee 106, Freiburg 79110, Germany

[2] unaffiliated

[3] Erasmus Medical Center, Clinical Bioinformatics Group, Department of Pathology, Wytemaweg 80, 3015 CN, Rotterdam, The Netherlands

[4] Khoury College of Computer Sciences, Northeastern University, Boston, MA 02115, USA

[5] Institute for Surgical Pathology, Medical Center, University of Freiburg, 79106 Freiburg, Germany

[6] Faculty of Medicine, University of Freiburg, 79110 Freiburg, Germany

[7] Faculty of Biology, University of Freiburg, 79110 Freiburg, Germany

[8] Spemann Graduate School of Biology and Medicine (SGBM), University Freiburg, 79106 Freiburg, Germany

[9] Faculty of Medicine, Dentistry and Health, University of Sheffield, UK

[10] Institute for Cardiovascular Regeneration, Goethe University, Frankfurt am Main, Germany

[11] University of Debrecen, Faculty of Medicine, Dept. of Biochemistry and Molecular Biology, Debrecen, Hungary

[12] Johns Hopkins University, 3400 N Charles Street, Mudd Hall 144, Baltimore 21218, MD, USA

[13] Center for Comparative Genomics and Bioinformatics, Penn State University, 505 Wartik Lab, University Park, PA, 16802, USA

Abstract

The COVID-19 pandemic is shifting the teaching paradigms to an online setting all over the world. The Galaxy framework caters to computational biologists a set of features to facilitate the online learning process and make it accessible to everyone. Besides the high-quality training materials, Galaxy provides easy access to data and the possibility to share the progress and achievements, both student to student and student to instructor. By combining the different features offered by the Galaxy framework and by choosing the adequate communication channels, effective training activities can be designed inclusively, regardless of the students' environments.

KEYWORDS: e-learning; online teaching; Galaxy; TIaaS; accessibility; scalability

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 September 2020 doi:10.20944/preprints202009.0457.v1

Introduction

The current pandemic in which we are engrossed bares the harsh reality that online education has come to stay and hence, universities need to face the challenge of adapting to a new paradigm in a very short period of time. It is well-known that the students' motivation worsens in virtual environments [1] and the collective grief for a world that will never be - for better or worse - as we knew it, makes it even more challenging for everyone to keep focused. More empathy than usual is, therefore, required in assisting the learning process.

Such an unprecedented situation is also rekindling the issue of the equity of opportunities to access education. Yet the uneven access varies extensively across educational stages: the digital divide is much more present at the primary education than at the university level, relying even more on the socio-economic context of the families [2]. No doubt that the technological gap will leave many students behind, although some education centres are alleviating the issue by lending equipment to their enrolled students that do not have access to those resources. These rapidly shifting times are a great opportunity to soften the differences not only between genders but also for other underprivileged collectives.

One of the greatest advantages of online education is its flexibility, a feature that makes education compatible with different responsibilities, such as homeschooling, demanding work schedules or students coping with distracting environments. However, despite its undoubted advantages, learner isolation and loneliness are two key issues reported to be responsible for learner failure or lack of satisfaction in online learning experiences [3]. Thus, the creation of effective online learning communities is critical for preventing those harmful effects on the learning process. Although personal relationships are definitely more enriching in a live scenario - left aside the interactions that take place out of the teaching setting -, online learning communities aim to achieve a similar outcome regarding social interactions. In addition, according to social constructivism theory, setting up a learning community is an essential element for the construction of knowledge [4] [5].

As an attempt to overcome the situation, virtual teaching platforms are replacing traditional training scenarios during this pandemic. Video calls, chats, and shared documents are presented as the combination of tools for a successful online teaching activity [6]. However, the proliferation of tools that students are not familiar with adds cognitive load to the learning process [7]. Moreover, although some level of redundancy in communication channels is useful in case of failure of one of the systems, having different simultaneous ways to interact can be confusing for the students.

In this work, we present Galaxy as a teaching platform that addresses the recommendations and features required for effective learning [8], such as sharing capabilities, formative assessment of the progress and checking exams to keep the students engaged. The Galaxy Training Network (GTN) [9] provides a vast amount of detailed training materials - created by the community - covering many scientific disciplines, from classical omics to imaging, ecology or climate analyses. All the tutorials are accessible to everyone, without any cost. Not only the training materials are free, but also the computational infrastructure to run the analyses. More importantly, the cloud infrastructure is accessible from a browser with no need for special hardware specifications or extra installation of software.

The technical foundation of Galaxy

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 September 2020 doi:10.20944/preprints202009.0457.v1

Galaxy is an open, web-based platform for accessible, reproducible, and transparent computational biological research.

The fact that Galaxy is a web-based platform helps enormously with the accessibility since no specific hardware or software requirements are needed to run an analysis. A High-Performance Computing (HPC) environment runs the tools launched by the users in the backend with the specified data. In order to avoid potential overload in the HPC cluster, a special queue is provided for the training activities scheduled through the Training Infrastructure as a Service (TIaaS)[1] [10]. This service is completely free and unrestricted for any Galaxy tool, including the more than 150 training tutorials available in the GTN [9]. The most prominent Galaxy instances already have all the tools necessary for the GTN available locally. Furthermore, TIaaS enables the lecturers to assess the progress of the students in real-time in a GDPR-compliant way [10].

Sharing as a fundamental concept of open science

The incalculable value of open science is an essential principle to be transmitted to students and, by extension, to the entire society. Scientific teaching is a great opportunity to train new generations of researchers on the transparency values to avoid repeating the same mistakes over time [11]. Not only data must be open and accessible to everyone, but the openness of the software is a key element to tackle reproducibility issues [12]. In this context, Galaxy provides features that make shareability, both for data and software, particularly effortless.

In an online teaching environment, providing students with real-time feedback can be challenging, due to the lack of visual contact and time-shift restrictions. Consequently, a straightforward shareability method becomes extremely important for the progress assessment and final evaluation of the goals achieved.

Galaxy provides a set of features to facilitate the shareability either with only one user (e.g. between peers or with the instructor), with defined groups or publicly for unregistered users. These features comprise data and software in the form of histories, workflows and data visualisations.

On the software side, Galaxy histories are the central feature to ensure data analysis transparency, reproducibility, and shareability. Every data analysis step (tool version, selected parameters, etc.) along with the entire provenance of a dataset are captured into the Galaxy histories and constantly available for the students. In addition to a visual inspection of the results, students can re-run a tool from the Galaxy history to explore a different parameter space or simply for reproducibility purposes.

Several tools can be interlinked to perform a specific task, in which case workflows are created. If a given tool yields an error during the process, the workflow is paused and the successive step would not be executed. In a teaching context, workflows are particularly useful to run a particular analysis that can be exported and shared with the students or the supervisors.

In terms of data, users can upload their own datasets that will live in their Galaxy history. The simplest way to do it is via the web front-end, although alternative mechanisms are available for larger datasets[2]. Nevertheless, for teaching purposes and to reduce the processing times, the GTN data is usually of small dimensions, still yielding some meaningful results for its subsequent scientific interpretation.

[1] https://galaxyproject.eu/tiaas
[2] https://galaxyproject.org/tutorials/upload/

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 September 2020 doi:10.20944/preprints202009.0457.v1

Alternatively, all GTN datasets are stored and shared via the Galaxy Data Libraries[3] on the main, public usegalaxy.* servers[4]. The power of this feature resides in its scalability, allowing data versioning and preventing data duplication across different user Galaxy histories and long waiting times when uploading datasets into Galaxy. To optimally benefit from this feature, trainers can prepare the Data Libraries in advance.

Galaxy extracts the metadata associated with every dataset, including the file format, a crucial element in life sciences data analysis. Although there are many different formats available, students need to learn, at least, how to identify, convert and visualise the most popular ones. Based on our experience, reinforcing the basic concepts and the different file formats is relevant for the students: first, to understand which information is included in which file format; and second, to acquire the scientific vocabulary for a future competent communication with colleagues.

Proposed teaching format within the Galaxy framework

Trainers can bring the potential of all the Galaxy features together to create effective e-learning activities. Still, web-based online teaching based on well-elaborated material and a scalable infrastructure stand in need of communication channels to boost the interaction among the participants.

It is the trainers' responsibility to establish adequate communication channels to engage students and to create a respectful environment in which students feel empowered to share their knowledge and learn from each other [5] [13]. In this sense, students must be guided through the available technology by clearly stating which particular tools will be used and in which way.

The effect of the learner control on the pace of the lesson improves their engagement and satisfaction [14], so a combination of synchronous and asynchronous communication channels need to be jointly considered during the preparation phase of the training (Fig. 1). Asynchronous communication allows a higher degree of self-management on the side of the students [15]. Furthermore, it has been found that it is more useful for task-oriented communication and to facilitate high levels of thinking, such as critical thinking [16] [24]. Synchronous communication, on the other hand, has been shown to be more effective in promoting social interactions [24].

Video calls can be classified as synchronous channels. The basic features are screen sharing, chat and recording - with the consent of the participants for GDPR compliance. Whiteboards, polling or remote control rights can increase participation. Although many commercial tools (e.g. Google Meet, Zoom, Microsoft Teams) offer plenty of these features, open-source resources cover the most important aspects too (e.g. BigBlueButton, Jitsi). It is important to highlight here that some of these options require installation of additional software and/or creation of email accounts with different providers, which could limit the freedom of choice of the students.

One consideration to guarantee inclusiveness is to bear in mind that the synchronous channels provide interactivity only for those students that are able to join the live sessions. Time zone shifts, distracting environments and other limitations, might prevent the attendance of part of the student body. That makes the asynchronous channels especially convenient: regular exchange of emails, posts in forums and shared documents can be used to provide integrated feedback, comments or suggestions and to ask questions in an asynchronous way. Chats serve as a communication channel for those attending the live sessions as well as for those visualising the recordings in a semi-synchronous way.

[3] https://galaxyproject.org/data-libraries/
[4] https://github.com/usegalaxy-eu/shared-data/

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 September 2020 doi:10.20944/preprints202009.0457.v1

Based on the above ideas we suggest four parts for an e-learning approach with Galaxy.

First part: Introductory seminar

To provide students with some context on the scientific background of the training course, a short presentation (30 to 60 minutes) as a general introduction is recommended (Fig. 1A). However, just replicating an online version of a classroom lecture lacks its strengths and pedagogical characteristics [17]. Thus, the materials need to be adapted and the interactivity will have to be promoted in order to increase the engagement of the students [18].

For accessibility and equity of opportunities, the presentation should be recorded (e.g. with *Open Broadcaster Software*) and shared with the students that cannot attend the live session. Ideally, the slides - the GTN is offering introductory slides as well - will be shared with the students in a printable format with enough time ahead of the course.

During this session, questions will inevitably arise. A possible way to address them is to ask the students to write them down - in the chat or a shared document - and answer them at the end of the presentation. Planning frequent breaks will make questions flow more thoroughly. The shared document with questions and answers can stay active for some time so that the students that are only watching the recording can also formulate their questions in equal conditions.

Second part: Showcase

After the presentation of the scientific topic, a live demonstration of the data analysis could take place (Fig. 1B). This demonstration is completely optional as the GTN materials are comprehensive enough to cover every aspect of the procedure. Nonetheless, a detailed demonstration with exhaustive commenting of every step can help ensure that the students can later repeat them autonomously. In case of performing this part, the recommendation for the lecturers is to use two monitors, one for the GTN material and another one for the Galaxy sessions shared with the students, always bearing in mind that the same assumption cannot be made on the students' side.

Third part: Self-study session

A flexible arrangement is imperative for e-learning. It cannot be assumed that every student can find the time for self-study during the day, thus a self-paced session with a few days to repeat the - ideally recorded - demonstration is recommended (Fig. 1C). All the GTN tutorials are designed as self-study material with very clear objectives, detailed step-by-step guided instructions through the data analysis, sometimes even including supporting materials.

Learners need some time to get used to the interface of the tools and to explore the parameter space. According to the constructivist pedagogical approach, students construct their understanding through experiences and reflecting on them afterwards [19][20]. It is encouraged to artificially create short erroneous tasks that are not covered by the GTN materials, with the purpose of provoking errors that force students to hypothesise about the issue and find out solutions themselves rather than being guided [21]. Students would be active participants and no longer passive receptors: reading the error messages and restarting tools with the appropriate parameters will improve their problem-solving and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 September 2020 doi:10.20944/preprints202009.0457.v1

critical thinking skills. In this scenario, sharing Galaxy histories is again especially advantageous to debug, request feedback and to foster discussions with other students.

Even in this student-centred part, the role of the instructor should not be diminished. Teaching presence is relevant to guide students through the different stages of learning: exploration, integration and application [22]. The different communication channels should, therefore, be open over the entire semester. Open video call hours, similarly to the open-door policies are also a valuable mechanism to enable the discussions and encourage questions.

Fourth part: Assessment

The GTN tutorials are exemplary work packages that can be distributed to students as homework (Fig. 1D). To assess the level of understanding during the previous parts of the training, the GTN tutorial can be used with different and more complex datasets. The degree of complexity of the data can vary from real-world datasets or the data from the Personal Genome Project[5]. A more advanced exercise to promote critical thinking can be to hand over an open-access article which data analysis will have to be reproduced [23].

Once the homework has been performed, students can deliver the results, together with the steps in the analysis, by sharing their Galaxy histories with the lecturer or other students. Time-stamps and metadata on every dataset are particularly useful to prove authorship. The role of the instructor is again vital at this stage to create a collaborative environment in which students can learn and help each other; in contrast to a model where the teacher is an authority on the subject [13]. Furthermore, to develop their scientific writing skills, the students can also write a report interpreting the results they obtained from the data analysis in a structured way.

After all the assessments have been collected, a gold-standard Galaxy history can be distributed to the students with the purpose of showing the scientific reasoning behind the proposed homework.

Conclusion

Here we present Galaxy as a teaching platform in response to the recent changes in the education needs due to the COVID-19 world pandemic. Galaxy provides a wide variety of ready-to-share features that facilitate the online learning process. Those features, together with the free materials available at the GTN, can be combined for an integrative teaching design that makes education accessible for many different student profiles.

In terms of addressing the content, an effective format seems to be a combination of a webinar to introduce the main theoretical topic together with practical self-study sessions in which students can follow very detailed training materials. Ideally, the first part could be recorded for flexibility purposes, always with the consent of the presenters to be GDPR compliant. If a demo is shown instead of - or in addition to - a webinar, all the steps need to be very slowly explained and the recordings help in this case especially if another software needs to be side by side with the video. The practical part requires the training materials to be more detailed with the advantage of bringing flexibility to the students and addressing everyone' needs.

[5] https://www.personalgenomes.org/, https://galaxyproject.eu/posts/2020/01/16/pgp/

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 September 2020 doi:10.20944/preprints202009.0457.v1

The uncertainty and complexity of the current pandemic situation complicate the prediction of future trends in education, but clearly, adapting to an e-learning setting will be compelling. Our hope is that the current global situation will make us reflect on the accessibility of education in equal terms for every underprivileged community and, more importantly, act upon it.

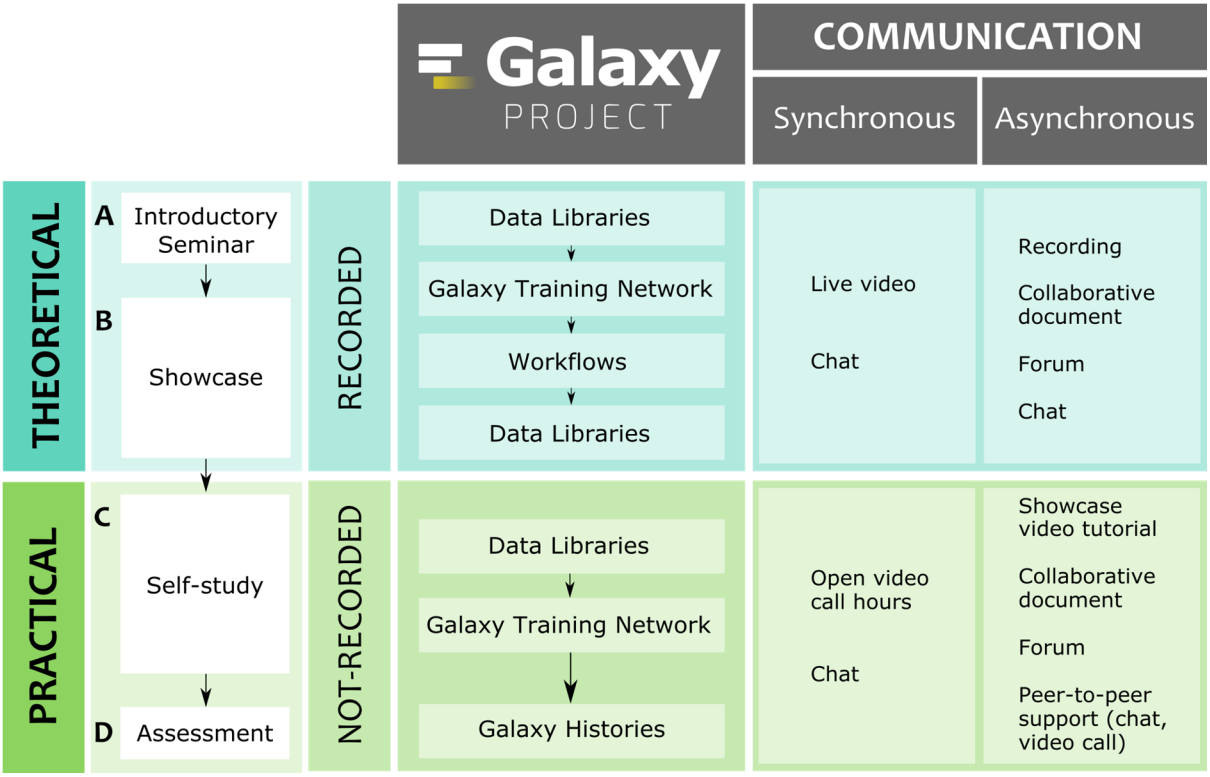

Figure 1. Galaxy features and communication channels for each part of the online teaching process.

Acknowledgement

We would like to thank the Galaxy community for their constant work and support.

References

[1] Selvi, K. (2010). Motivating factors in online courses. Procedia Social and Behavioral Sciences 2 (2010) 819–824.
[2] Wei, Lu, and Douglas Blanks Hindman. "Does the digital divide matter more? Comparing the effects of new media and old media use on the education-based knowledge gap." *Mass Communication and Society* 14.2 (2011): 216-235.
[3] McDonald, Barb, et al. "Breaking down learner isolation: How social network analysis informs design and facilitation for online learning." *AERA, Montreal, Canada* (2005).
[4] Vygotsky, L. S. (1962). Thought and language, trans. E. Hanfmann & G. Vakar. *Cambridge, MA: Massachusetts Institute of Technology*
[5] Lee, S. M. (2014). The relationships between higher order thinking skills, cognitive density, and social presence in online learning. The Internet and Higher Education, 21, 41-52.
[6] Dabbagh, N. (2007). The online learner: Characteristics and pedagogical implications. *Contemporary Issues in Technology and Teacher Education*, *7*(3), 217-226.
[7] Chen, C. Y., Pedersen, S., & Murphy, K. L. (2011). Learners' perceived information overload in online learning via computer-mediated communication. *Research in Learning Technology*, *19*(2).

[8] Gallardo-Alba, C.; Grüning, B.; Serrano-Solano, B. A Constructivist-Based Proposal for Teaching Practices During Lock-Down. Preprints 2020, 2020080532 (doi: 10.20944/preprints202008.0532.v1).
[9] Batut, Bérénice, et al. "Community-driven data analysis training for biology." *Cell systems* 6.6 (2018): 752-758.
[10] Rasche, Helena, and Bjoern Andreas Gruening. 2020. “Training Infrastructure as a Service.” *Bioinformatics*. bioRxiv.
[11] Galaxy and HyPhy developments teams, Anton Nekrutenko, Sergei L Kosakovsky Pond. No more business as usual: agile and effective responses to emerging pathogen threats require open data and open analytics. bioRxiv 2020.02.21.959973; doi: https://doi.org/10.1101/2020.02.21.959973
[12] Jiménez RC, Kuzak M, Alhamdoosh M, et al. Four simple recommendations to encourage best practices in research software. *F1000Res*. 2017;6:ELIXIR-876. Published 2017 Jun 13. doi:10.12688/f1000research.11407.1
[13] Mezirow, J. (1997). Transformative learning: Theory to practice. *New directions for adult and continuing education*, *1997*(74), 5-12.
[14] Palloff, R. M., & Pratt, K. (2003). The virtual student: A profile and guide to working with online learners. John Wiley & Sons.
[15] Hantula, D. A. (1998). The virtual industrial/organizational psychology class: Learning and teaching in cyberspace in three iterations. *Behavior Research Methods, Instruments, & Computers*, *30*(2), 205-216.
[16] Kanuka, H., & Garrison, D. R. (2004). Cognitive presence in online learning. Journal of Computing in Higher Education, 15(2), 21.
[17] Bates, T. (2015). *Teaching in a digital age: Guidelines for designing teaching and learning for a digital age*. Tony Bates Associates.
[18] Scagnoli, N. I., Choo, J., & Tian, J. (2019). Students' insights on the use of video lectures in online classes. British Journal of Educational Technology, 50(1), 399-414.
[19] Matsuura, Koichirō. “COVID-19: 10 Recommendations to Plan Distance Learning Solutions.” UNESCO, 28 Apr. 2020, en.unesco.org/news/covid-19-10-recommendations-plan-distance-learning-solutions.
[20] Linney, Sarah. “What Students Expect from Universities During the Coronavirus Outbreak.” QS, Publisher Name QS Quacquarelli Symonds Publisher Logo, 19 Mar. 2020, www.qs.com/what-students-expect-from-universities-during-the-coronavirus-outbreak/.
[21] Lin, B., & Hsieh, C. T. (2001). Web-based teaching and learner control: A research review. Computers & Education, 37(3-4), 377-386.
[22] Gorsky, P., & Blau, I. (2009). Online teaching effectiveness: A tale of two instructors. The International Review of Research in Open and Distributed Learning, 10(3).
[23] Chakraborty, M. (2017). Learner engagement strategies in online class environment (Doctoral dissertation)
[24] Im, Y., & Lee, O. (2003). Pedagogical implications of online discussion for preservice teacher training. Journal of research on technology in education, 36(2), 155-170.