Usability Assessment in Scientific Data Analysis: A Literature Review

Fernando Pasquini[a], Lucas Brito[b] and Adriana Sampaio[c]
Faculty of Electrical Engineering, Federal University of Uberlândia, Av. João Naves de Ávila 2121, Uberlândia, Brazil

Keywords: Scientific Data, Data Analysis, Data Visualization, Usability, Literature Review.

Abstract: Big Data has transformed current science and is bringing a great amount of scientific data analysis tools to help research. In this paper, we conduct a literature search on the methods currently employed and the results obtained to assess the usability of some of these tools, and highlight the experiments, best practices and proposals presented in them. Among the 38 papers considered, we found challenges in usability assessment that are related to the rapid change of software requirements, the need for expertise to specify and operate this software, issues of engagement and reterntion, and design for usability that supports reusability, reproducibility, policy, rights and privacy. Among the directions, we found proposals on new visualization strategies based on cognitive ergonomics, on new forms of user support and documentation, and automation solutions for supporting users in complex operations. Our summary thus can point to further studies that may be missing on usability of scientific data analysis tools and then improve them on their efficiency, prevention of erros and even their relationship to social and ethical values.

1 INTRODUCTION

Big Data has transformed current science in its practices, roles and institutions (Hey et al., 2009). With this, new tools and infrastructures are constantly being developed to deal with the increasing quantities of data which are being employed to engender discoveries in many domains of knowledge. With these tools, however, comes the question of their effectivity considering the human factor involved in their operation. The field of human-computer interaction (HCI) has traditionally dealt with questions such as these, however it is now presented with new challenges involving the specificity of practices inside the scientific domains, such as the use of machine learning and scientific visualization techniques (Zhang and Chignell, 2021; Amer-Yahia, 2018). Many studies today in the area of human-data interaction (HDI), for example, are focused in final users and recipents of data applications, which is surely an important topic, however, the usability of data intensive tools in current science remains not so much explored and conceptualized (Xu, 2019; Macaulay et al., 2009). Also, of the studies focusing on usability of scientific visualizations, fewer deal with usability in other phases of the data analysis process, such as infrastructure and modeling (Baca, 2009). Does it seem that scientists and technicians do not need to worry so much with usability since they are just dealing with the ”nuts and bolts” of science and technology and do not have time for having good and pleasurable means – in other words, should we take the interfaces of scientific tools as ”forgivable”? (Machado Paixão-Cortes et al., 2018) This is indeed an old discussion in human factors and human-computer interaction research. And this work has the purpose of showing that research on the usability of current scientific tools can be important — not just at the purpose of increasing efficiency and prevention of errors in scientific practice, but even as a way to better understand these practices and organize them in relationship to social and ethical values.

Therefore, we aim to gather current research on usability of scientific data analysis tools (Swaid et al., 2017) and present them in an integrated way so as to trace the current practices, challenges and prospects. Our search, however, is restrained to only findings that are directly related to the domain of scientific practice and the specific issues that it engenders. We chose not to discuss in detail findings that would amount to usability in general, such as font size and color in graphical interfaces, since these could be discussed on a more general level such as cognitive ergonomics, user experience (UX) or visual analytics. And we also limited ourselves to usability assessment methods that are generally associated with ”first-wave HCI” (Bødker, 2015); i.e., classical methods such as usability test and heuristic evaluation, and still not

[a] https://orcid.org/0000-0002-2259-7229
[b] https://orcid.org/0000-0001-8600-6953
[c] https://orcid.org/0000-0001-5745-5522

Pasquini, F., Brito, L. and Sampaio, A.
Usability Assessment in Scientific Data Analysis: A Literature Review.
DOI: 10.5220/0011665400003417
In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023) - Volume 2: HUCAPP*, pages 63-72
ISBN: 978-989-758-634-7; ISSN: 2184-4321
Copyright © 2023 by SCITEPRESS – Science and Technology Publications, Lda. Under CC license (CC BY-NC-ND 4.0)

engage with more recent and social-oriented methods such as activity theory (Clemmensen et al., 2016) or even distributed cognition (Perry, 2003), although recognize that this is an interesting further direection.

Our research questions (RQ) are three: RQ1) Which methods are currently being employed to assess the usability of scientific data analysis tools? RQ2) What are the usability problems and the challenges encountered? and RQ3) What are common solutions and practices that aim to increase usability in these areas? We employ a literature search procedure and group our findings according to different subjects. The procedure is described on section 2 and the results obtained in section 3, where we also try to answer the RQs according to our findings.

2 METHODS

We conducted a literature search on titles, abstracts and keywords through the databases Web of Science, PubMed, ACM, IEEEXplore and arXiv.org. The search was conducted in september 2022. Two search strings were used and are presented in Table 1: the first one specifying some well-known methods in usability and human factors, and the second specifying common activities in scientific data analysis. Without these specifications, our search would have returned 426 items and the initial analysis and sorting would involve too much human labor for a little noticeable gain.

Exclusion criteria were 1) papers on usability and/or data science that were not on scientific applications, such as business or education, and 2) papers on use of data science methods for increasing usability (which is the opposite counterpart of our approach). Inclusion criteria were papers experimenting and/or discussing usability of scientific data analysis tools.

The three authors read the selected papers and used the Obsidian knowledge base software to collect their findings and organize them through linked notes and tags. The tags were also classified into three groups: methods, challenges and proposals, and were also used to organize the discussion in the next section.

3 RESULTS

3.1 Paper Selection and Classification

The literature search returned in total 167 papers. After reading their abstracts, 69 papers were selected for a thorough reading. Then, of these, 38 were finally selected.

The 38 selected papers were then grouped in three main categories: a) papers presenting experimental results of usability methods, b) papers presenting best practices on usability and c) papers that do not employ experimental methods, but propose features that the authors claim that could increase usability. The papers were also grouped according to scientific domain, and a summary of their count is presented in Table 2.

It can be noticed that we only found papers on best practices (11 in total) in general purpose scientific software (7) and the bioinformatics domain (4). Of these papers, 7 were based on literature reviews (6 on general purpose and 1 on bioinformatics) and 4 on case studies (1 on general purpose and 3 on bioinformatics). Furthermore, 1 of the articles on best practices (Queiroz et al., 2017) and 1 on proposals (Cid-Fuentes et al., 2021) are arXiv preprints and thus are not peer-reviewed.

Of the 10 papers presenting features and proposals that the authors argue to increase usability, 5 were on general purpose scientific software, 3 on bioinformatics and 2 on environmental science. The proposals included visualization strategies, automation of procedures, standardization, documentation and user support, and will be discussed further.

Finally, the 16 articles specifically presenting usability methods and experimentation presented three main types of methods: usability tests, heuristic evaluations and user-centered design procedures. Table 3 presents how these were employed in each of the papers. With this table, we have a clear view on our RQ1.

The next two sections present the main themes found in the papers after their reading and tagging. These are grouped according to usability challenges (answering, thus, RQ2) and directions for usability (answering RQ3).

3.2 Usability Challenges

3.2.1 Specificity and Rapid Change of Requirements

Ahmed and Zeeshan (2014) comment on the difficulties of developing scientific software in face of the open ended questions and unexpected paths that scientists sometimes have to take. As they notice, "the rapid changes in requirements destabilize the systems design engineering processes by causing prompt changes in use cases with newly required modifications (or sometimes cancellations of previous designs

Table 1: Strings used in the literature search.

(”scientific data” OR ”scientific tool” OR ”scientific software” OR ”escience”) AND (”cognitive *load” OR ”mental *load” OR ”cognitive modeling” OR ”cognitive walkthrough” OR ”heuristic evaluation” OR ”task analysis” OR ”cognitive ergonomics” OR ”situation awareness” OR ”human error”)
(”scientific data” OR ”scientific tool” OR ”scientific software” OR ”escience”) AND (”data sourcing” OR ”data acquisition” OR ”data cleaning” OR ”data wrangling” OR (”data” AND ”labeling”) OR ”data* curat*” OR ”data* management” OR ”data* maintenance”) OR ”parallel processing” OR ”cloud computing” OR ”high-performance computing” OR ”infrastructure” OR ”system design” OR ”system specification” OR ”requirement* engineering” OR ”modeling” OR ”scientific programming” OR ”scientific computing” OR ”scientific visualization”) AND (”usability” OR ”human-computer interaction” OR ”human-data interaction” OR ”human factors” OR ”ergonomics” OR ”computer-supported cooperative work”)

Table 2: Summary of papers selected in the literature review process according to scientific domain and type of study. ”Experiment” refers to papers presenting experimental results assessing usability; ”best practices” are papers that list guidelines and recommendations for usability based on literature and case studies; finally, ”proposal” are papers that present solutions for a better usability of scientific data analysis tools, (although not testing these experimentally).

	type			
domain	experiment	best practices	proposal	*total*
bioinformatics	3	4	3	10
citizen science	3	0	0	3
environmental science	3	0	2	5
public health	2	0	0	2
general	5	7	5	17
total	16	11	10	37

and implemented modules) in existing or recently developed activity, system, sequence, and data flows”. This also means that, sometimes, the specificity and even transitoriness of the features being sought does not afford for a complete software engineering process. After a comparative analysis, the authors thus also conclude that no present software development process is ”especially proposed towards the scientific software solution development in academia”, and propose a new one specifically directed to that, called the butterfly model.

In order to deal with this, the software has to make available a whole lot of flexible and customizable features, which generally increase its complexity and need of training in order to use it (Queiroz et al., 2017). A common solution is the use of scientific scripting languages such as MATLAB or R, command line interfaces and domain-specific languages (DSLs) (Ahmed and Zeeshan, 2014). Hossain et al. (2020) notices that this amounts to a flexibility-usability tradeoff, which can be approached by identifying and working with an adequate *extensibility model*; i.e., offering extensible features for experienced users but without sacrificing usability for beginner users (Lacroix and Critchlow, 2003; Queiroz et al., 2017).

In the specific case of data science, the uncertainty regarding its uses and users may also contribute to a challenge in developing machine learning models and visualization strategies (Hossain et al., 2020; Wald et al., 2016). Efforts toward visualization of data workflow models are thus constantly being investigated and tested (Liu et al., 2015).

3.2.2 Expertise-dependent Requirements

The previous challenge also leads to the question of dealing with diverse developer and user backgrounds and expertises in scientific data analysis tools. The software requirements are very much dependent on expert knowledge and judgement, which can be difficult to transfer to the developers in a satisfactory way. Both work by Machado Paixão-Cortes et al. (2018) and Ramakrishnan and Gunter (2017) remarked this difficulty during a scientific process and pointed to the crucial role of multidisciplinary work in the development of a web server for bioinformatics research. Similarly, Overmyer (2019) has worked with ”structured learning phases” in which users, developers and experts come together to exchange expertise and perspectives on their data related tasks. Michener et al. (2012) has employed a participatory approach capable of engaging all the key stakeholders in the process of defining the tasks of collecting, managing, preserving, analysing and sharing biological and environmental data — and, in this case, has dealt not only with the expertise of scientific practitioners, but perspectives from many other communities such as from government, industry, non-profit and community. Douglas et al. (2011) also concluded that integrating this diversity of viewpoints is crucial for the development of bioinformatics databases.

With these examples, it can also be pointed that we are in need of models and directives for development involving diverse expertises and backgrounds. Harry Collins and Robert Evans' "periodic table" of expertises and his reflections on trading zones and interactional expertise (Collins et al., 2007) can be an interesting starting point for better organizing and understanding the multidisciplinary groups involved in the development of data analysis tools. Data may have multiple sources, meanings and are be collected for multiple objectives, which have to be all recognized (Kogan et al., 2020).

3.2.3 Expertise-dependent Operation

Another challenge related to expertise refers to development of tools which respond and adapt to expert users. This means that not always it will be possible to have a general user to try to "befriend" to. The "friendly-user interface" will not necessarily the most intuitive one, from the point of view of someone who does not possess a basic expertise in the scientific domain. And this requires that usability studies go beyond analyses on perception and action of human beings "in general" and now turns itself to studying the workings of *skilled perceptions and actions* (Ribeiro, 2014). For example, the study by Kalakoski et al. (2019) has reflected on what should be the optimal position of presented information in a data-based judging task and concluded that this depends very much on what we can identify as the "disciplined gaze" of the expert. Which leads to the question: how to work out an interaction that does not go against or disregarding this skilled perception, but rather works with it? Swaid et al. (2017), for example, notices that a common heuristic evaluation "may not be completely representative of expert user interaction with the tools".

As indicated earlier, an extensibility model capable of switching between a beginner and an expert user can be an interesting direction (Lacroix and Critchlow, 2003). Multi-layer interfaces are also proposed as a way to tackle these problems (Hwang and Yu, 2011). And even beyond that, Queiroz et al. (2017) argues for the advantage of have multiple input modes in the interface, which aim for different objectives (for example, one for accuracy, and other for speed), and which can be used differently by different users.

However, when dealing specifically with data analysis tools, the most pressing issue related to expertise refers to problems of misunderstanding and ambiguity of the information presented. This may happen either in the type of language used in the interface (which is commonly tackled by the area known as UX writing) — for example, Baca (2009) has noted issues on terminology regarding job status in a scientific workflow software — or also in the very task of interpreting data: Michener et al. (2012), for example, concluded that "data heterogeneity and interoperability issues are the single greatest obstacles to addressing many scientific grand challenges requiring generalizable data synthesis solutions". Another example was given by Macaulay et al. (2009), noticing that "ou users considered the use of the word 'scope' in the Omero search interface—that is, the extent of data on which the search will be conducted—confusing, as 'scope' is the scientist's abbreviation for the instrument they use to capture images". Néron et al. (2009) also noticed difficulties for the users understanding some terms and ambiguities in data storage. Thus, as Collins et al. (2007) put it, scientific domains are formed by different *language games* in which it may not be easy to construct bridges, and data analysis tasks are really prone to these kinds of problem — as put by Lin et al. (2016), we have "challenges of managing, standardising, and integrating different epistemic cultures, especially when amateurs meet experts". Solutions such as constructing and relying on *boundary objects* are commonly employed to tackle these problems, however, to date, there are few known solutions being applied to scientific software (Fremont et al., 2018).

3.2.4 Engagement and Retention

Many of the papers also investigated factors of engagement and retention of users in the data analysis tools. Most of theses covered studies on citizen science, however were not limited to it. Hossain et al. (2020) remarked the challenge to identify the socio-cultural characteristics of user communities in order to develop engaging interfaces, and Newman (2010) employed a set of engagement metrics. Wald et al. (2016) concluded that although the usual heuristic evaluation procedure performed well in their software, engagement and retention of users was less consistent, pointing to the necessity of further studies on motivation and attention of users. Lin et al. (2016) has put five requirements to be considered in this task: considering 1) local, personal and tacit knowledge of users, 2) socialisation of users, 3) embodiment of users ("physical, emotional and cognitive activities involved in recording, observing, transcribing and editing"), 4) attitudes towards professional stan-

Table 3: Description of the methodology used in the 16 papers selected. Three main types of methods were identified: usability tests, heuristic evaluations and user-centered design procedures. These are then detailed according to their use in each paper.

paper	domain	usability test	heuristic evaluation	user-centered design
(Scotch et al., 2007)	public health	think-aloud, task completion time, success rates, problem spaces, cognitive walkthrough, closed questionnaires		
(Macaulay et al., 2009)	bioinformatics	user feeback, closed questionnaires	Nielsen's heuristics	design ethnography
(Oliver et al., 2009)	public health	open and closed questionnaires		user advisory panels, groups and workshops
(Baca, 2009)	general	cognitive walkthrough,closed questionnaires		user advisory panels, groups and workshops
(Néron et al., 2009)	bioinformatics			think aloud, user advisory panels, groups and workshops
(Newman, 2010)	citizen science	task completion time, open questionnaires		
(Michener et al., 2012)	environmental science	task completion time, open questionnaires		design ethnography, personas and user scenarios
(Lin et al., 2016)	citizen science			observation, interviews
(Wald et al., 2016)	citizen science		Nielsen's heuristics, engagement and retention heuristics	
(Volentine et al., 2017)	environmental science	think-aloud, task completion time, success rates, closed questionnaires		
(Swaid et al., 2017)	general		Nielsen's heuristics + 2 custom heuristics	
(Machado Paixão-Cortes et al., 2018)	bioinformatics	closed questionnaires	Nielsen's heuristics, inspectors with different backgrounds	
(Iqbal Chunpir et al., 2018)	environmental science			interviews, action-research
(Zhang, 2018)	general	cognitive walkthough		interviews
(Overmyer, 2019)	general			ethnography, interviews
(Hossain et al., 2020)	general	closed questionnaires (SUS), cognitive load assessment		

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

dards and data quality and 5) trust between the parties involved. For this, a careful planning and better understanding of the scientific practices and communities is needed in order to achieve a successful reception of new solutions and avoid technological lock-in (Douglas et al., 2011; Poole, 2015). Furthermore, both Wald et al. (2016) and Lin et al. (2016) notice that motivation also stems from the perception of the value of time in performing the activities. Users may implicitily ask, for example, if it's really worthy to spend time recording or preparing data. Depending on the answer, the very overall task may be compromised, not only in its success, but also ethical and social implications (Kogan et al., 2020).

3.2.5 Reusability and Reproducibility

Issues of reproducibility are constantly being discussed and tackled in science (Berg, 2018), and specifically with data-intensive science also comes the related issue of data reusability (Hardwicke et al., 2018). Terms like open science abound in the literature (Levin and Leonelli, 2017), and it is expected to find this also in discussions of usability of data analysis tools, since an interface that is difficult to approach also disencourages their availability of reproduction (Sanchez Reyes and McTavish, 2022).

The FAIR principle — findable, accessible, interoperable and reusable — is constantly being sought and respected in the design of scientific databases (Vogt, 2021). Hossain et al. (2020), Douglas et al. (2011) and Hunter-Zinck et al. (2021) present guidelines for portability of databases, which include needs of curation (Poole, 2015) and data formats and standards (Parsons and Duerr, 2005). Work by Cisar et al. (2016) has demonstrated an application of a strategy of standardizing repositories and using back-tracking in order to increase reproducibility, and Samourkasidis and Athanasiadis (2020) has employed ontologies and templates in order to speed and facilitate curation of databases in the domain of environment science. Other types of work also dedicate attention to accessibility (and, thus, also usability) of scientific computing services such as in the frameworks of event-based computing and "function as a service" (Chard and Foster, 2019).

Furthermore, encouraging practices of logging and user accounting are also mentioned, such as by Queiroz et al. (2017) and Néron et al. (2009), and this can, in some cases, be obtained automatically by the software — Chin Jr and Lansing (2004), for example, has devised mechanisms of visualizing the software execution history as a way to increase reproducibility. Callahan et al. (2006), by their turn, proposed a way to trace pipelines for generated scientific visualizations. These kinds of strategy touch on one of the most common practices today in data science, the use of story-like or literate programming (Sanchez Reyes and McTavish, 2022), such as Jupyter notebooks (Perez and Granger, 2015). Sanchez Reyes and McTavish (2022) indicate that this not only helps reproducibility and reusability, but also supports memory and understanding of the practicioner. The experiment done by Kalakoski et al. (2019) pointed out the need to articulate coherent sequences of presentation of information in order to reduce the cognitive load of data analysis tasks, and this may be understood as a search for a good usability of these interactive computing notebooks, which remain to be explored. The work by Zhang (2018) is a good example of this, by doing an usability evaluation of a visualization enhancement in the Jupyter Lab interface.

3.2.6 Policy, Rights and Privacy

Finally, a much discussed issue around data science revolves around policy, rights and privacy, and usability studies can offer an important contribution to these by designing interactions that make all the stakeholders better aware of these questions (Mortier et al., 2013). Such is what is sought in fields such as HDI, however, as stated previously, most studies are focused in end-user applications and less on scientific data practices. Poole (2015) makes an important discussion around policy procedures for data curation in the sciences. The authors review an extensive literature to notice that current policies focus more on access of data than preservation, and that researchers are often "suspicious of policies", since the perception is that they sometimes can hinder work.

In fact, security measures in software such access control and privacy statements can be badly designed so as to present an annoyance to users, as noticed in the usability evaluation by Iqbal Chunpir et al. (2018), which noticed difficulties related to user login, data download, data search and user registration. However, on the other side, good usability design on these issues can be crucial both as to increase trust among the parties so as to raise awareness of ethical, social and legal implications in the users (Lin et al., 2016). Work by Machado Paixão-Cortes et al. (2018), for example, has explicitly included in their usability evaluation the need for users to have a clear view on copyrighted materials and data. Good usability can and should be directed to encourage a reflective practice, such as argued by Aragon et al. (2022) in their proposal of a human-centered data science.

3.3 Directions for Usability

3.3.1 Visualization Strategies

Many of the papers investigated propose and evaluate new visualization strategies in order to increase the usability of the data analysis tools. Overmyer (2019), for example, proposes an adaptation of UX methods for obtaining usable scientific visualizations. Baca (2009), by means of user focus groups and usability walkthroughs, obtained useful results for an interface for job monitoring in high performance computing. Hossain et al. (2020) devise a visual scripting framework for scientific data analysis. And as already cited, Callahan et al. (2006) propose visualizations of the pipelines used to generate scientific visualizations. Zhang (2018) used personas and usability evaluations to propose an improvement on Jupyter Lab that deals with the difficulties of users in understanding new datasets, programming libraries and exporting data. Finally, Stepanyan (2021) make important remarks on the cognitive ergonomics of visualizing long nucleotide sequences, as the strategy the authors proposed help observing geometrical patterns in data — and indeed, it has been long noticed that studies on cognition and pattern recognition skills in humans can greatly contribute to these efforts (Patterson et al., 2014).

3.3.2 User Support and Documentation

The usability assessments considered in this review many times conclude that investing more in user support and documentation can greatly reduce the usability problems (Macaulay et al., 2009). Ahmed and Zeeshan (2014), for example, noticed that "deployment and configuration procedures are very complex and there is no proper documentation which can help scientists in easily deploying the system". This touches, after all, in the problem of getting with just "forgivable" interfaces, without considerations for their further use by other scientists and communities. As indicated by Levin and Leonelli (2017), however, it is important to foster a culture of incentives and rewards for constructing solid and documentable scientific tools and databases.

Relating to that, some solutions also rely on development of better troubleshooting features, which are able to show users' actual and even potential errors during their use of the software (Sanchez Reyes and McTavish, 2022). The already common use of live programming environments, such as iPython (Pérez and Granger, 2007), in scientific workflows, can be an important point of entry to this, as shown by Ayres et al. (2019).

3.3.3 Automation and Human-in-the-loop Usability

Lastly, many of the studies develop ways to automate part of the work of scientists dealing with data, and propose that this can greatly increase the usability and even motivation of users by removing the need of dealing with tedious, tiresome or even too complex tasks (Iqbal Chunpir et al., 2018; Cid-Fuentes et al., 2021). Ayres et al. (2019), for example, develops an automatic tool for resource selection in high performance computing, and concludes that this can greately increase usability. Serverless approaches such as advocated by Chard and Foster (2019) also promise to automate and outsource many common tasks faced by data analysis and visualization today. Kim et al. (2016) also proposes a framework for managing computing resources in scientific domains in automated ways. Finally, Cid-Fuentes et al. (2021) propose even a data structure format to support reuse and automation of scientific tasks (although it should be noted that the article is an arXiv preprint).

In all of these efforts, studies can be greatly enriched with insights from human-in-the-loop usability studies, which focus on ways to better integrate tasks delegated to computers (Wu et al., 2022). Dragut et al. (2021), for example, point to the need of studies and solutions dealing with the cooperation between human and computers in preparing, analysis and representing data. One of the pressing questions in this regard is how to make this cooperation in a way that avoids biases or other maleficial effects that would make hard an attribution of responsibility (Cornelissen et al., 2022).

4 CONCLUSIONS

Our literature review has retrieved some common themes, challenges and proposals in diverse studies assessing and reflecting on usability of data analysis tools in scientific research. By categorizing and observing them in tables such as Table 2, it can be possible to visualize domains of science which are still not much covered by usability studies. As an example, domains such as citizen science, environmental science and public health science still lack discussions of best practices for assuring usability of their data analysis tools. Table 3, by its turn, can highlight methods which are still not employed or discussed in some domains — for example, our literature review could not find a heuristic evaluation of tools in the areas of environmental and public health science. Also, even when methodological contributions are made, these are only applied to the paper that proposes them and still are not reutilized in other studies (such as, for example, the custom heuristics proposed in Swaid et al. (2017)).

Among the challenges encountered in the literature, we have encountered the problem of designing interaction in software that is constantly changing requisites and that are very specific to certain areas. We discussed the problem of designing a human-data interaction for expert users, and in software with requirements that depend of specific scientific expertise. Issues of engagement and retention of users in these tools were also a target of assessment and reflection in the literature, and is becoming more and more important in tasks related to data. The issues of reusability and reproducibility of the data analysis and visualization tasks are also being constantly discussed, such as under the FAIR principle, and by improvements in practices of live and story-like programming. Finally, we presented literature discussing how usability studies can deal with issues pertaining to policy, rights and privacy of databases and tools. Next, we laid out some of directions of solutions being proposed to increase the overall usability of these systems. We pointed to studies presenting new visualization strategies for dealing with scientific data and the need to study them in relationship to cognitive ergonomics. Then, we discussed the need to include better user support and documentation in scientific tools, and finished with a discussion on the need for human-in-the-loop usability studies in the cases of automation and human-computer cooperation. In the end, with this study, we hope to point directions for research in usability of data analysis tools and show their importance in face of the growing digitalization of science.

ACKNOWLEDGEMENTS

This research was sponsored by the German Federal Ministry of Education and Research (BMBF) at the Käte Hamburger Kolleg Cultures des Forschens/Cultures of Research.

REFERENCES

Ahmed, Z. and Zeeshan, S. (2014). Cultivating software solutions development in the scientific academia. *Recent Patents on Computer Science*, 7(1):54–66.

Amer-Yahia, S. (2018). Human factors in data science. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pages 1–12. IEEE.

Aragon, C., Guha, S., Kogan, M., Muller, M., and Neff, G. (2022). *Human-Centered Data Science: An Introduction*. MIT Press.

Ayres, D. L., Cummings, M. P., Baele, G., Darling, A. E., Lewis, P. O., Swofford, D. L., Huelsenbeck, J. P., Lemey, P., Rambaut, A., and Suchard, M. A. (2019). Beagle 3: improved performance, scaling, and usability for a high-performance computing library for statistical phylogenetics. *Systematic biology*, 68(6):1052–1061.

Baca, J. (2009). Designing for usability in a high performance computing application: A case study. In *Proceedings of the Cybernetics and Information Technologies, Systems and Applications Conference: CITSA 2009*.

Berg, J. (2018). Progress on reproducibility. *Science*, 359(6371):9–9.

Bødker, S. (2015). Third-wave hci, 10 years later—participation and sharing. *interactions*, 22(5):24–31.

Callahan, S. P., Freire, J., Santos, E., Scheidegger, C. E., Silva, C. T., and Vo, H. T. (2006). Vistrails: visualization meets data management. In *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*, pages 745–747.

Chard, K. and Foster, I. (2019). Serverless science for simple, scalable, and shareable scholarship. In *2019 15th International Conference on eScience (eScience)*, pages 432–438. IEEE.

Chin Jr, G. and Lansing, C. S. (2004). Capturing and supporting contexts for scientific data sharing via the biological sciences collaboratory. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 409–418.

Cid-Fuentes, J. Á., Álvarez, P., Solà, S., Ishii, K., Morizawa, R. K., and Badia, R. M. (2021). ds-array: A distributed data structure for large scale machine learning. *arXiv preprint arXiv:2104.10106*.

Cisar, P., Soloviov, D., Barta, A., Urban, J., and Stys, D. (2016). Biowes-from design of experiment, through protocol to repository, control, standardization and back-tracking. *BioMedical Engineering OnLine*, 15(1):129–147.

Clemmensen, T., Kaptelinin, V., and Nardi, B. (2016). Making hci theory work: an analysis of the use of activity theory in hci research. *Behaviour & Information Technology*, 35(8):608–627.

Collins, H., Evans, R., and Gorman, M. (2007). Trading zones and interactional expertise. *Studies in History and Philosophy of Science Part A*, 38(4):657–666.

Cornelissen, N., van Eerdt, R., Schraffenberger, H., and Haselager, W. F. (2022). Reflection machines: increasing meaningful human control over decision support systems. *Ethics and Information Technology*, 24(2):1–15.

Douglas, C., Goulding, R., Farris, L., and Atkinson-Grosjean, J. (2011). Socio-Cultural characteristics of usability of bioinformatics databases and tools. *Interdisciplinary Science Reviews*, 36(1):55–71.

Dragut, E., Li, Y., Popa, L., and Vucetic, S. (2021). Data science with human in the loop. In *Proceedings of the*

27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, pages 4123–4124.

Fremont, V. H. J., Frick, J. E., Åge, L.-J., and Osarenkhoe, A. (2018). Interaction through boundary objects: controversy and friction within digitalization. *Marketing Intelligence & Planning*.

Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., Hofelich Mohr, A., Clayton, E., Yoon, E. J., Henry Tessler, M., et al. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal cognition. *Royal Society open science*, 5(8):180448.

Hey, A. J., Tansley, S., Tolle, K. M., et al. (2009). *The fourth paradigm: data-intensive scientific discovery*, volume 1. Microsoft research Redmond, WA.

Hossain, M. M., Roy, B., Roy, C. K., and Schneider, K. A. (2020). Vizsciflow: A visually guided scripting framework for supporting complex scientific data analysis. *Proceedings of the ACM on Human-Computer Interaction*, 4(EICS):1–37.

Hunter-Zinck, H., de Siqueira, A. F., Vásquez, V. N., Barnes, R., and Martinez, C. C. (2021). Ten simple rules on writing clean and reliable open-source scientific software. *PLoS computational biology*, 17(11):e1009481–e1009481.

Hwang, T. and Yu, H.-Y. (2011). Accommodating both expert users and novice users in one interface by utilizing multi-layer interface in complex function products. In *International Conference on Internationalization, Design and Global Development*, pages 159–165. Springer.

Iqbal Chunpir, H., Rathmann, T., and Zaina, L. M. (2018). An empirical evidence of barriers in a big data infrastructure. *Interacting with Computers*, 30(6):507–523.

Kalakoski, V., Henelius, A., Oikarinen, E., Ukkonen, A., and Puolamäki, K. (2019). Cognitive ergonomics for data analysis. Experimental study of cognitive limitations in a data-based judgement task. *Behaviour & Information Technology*, 38(10):1038–1047.

Kim, C.-W., Yoon, H., Jin, D., and Park, S. O. (2016). Integrated management system for a large computing resources in a scientific data center. *The Journal of Supercomputing*, 72(9):3511–3521.

Kogan, M., Halfaker, A., Guha, S., Aragon, C., Muller, M., and Geiger, S. (2020). Mapping out human-centered data science: Methods, approaches, and best practices. In *Companion of the 2020 ACM International Conference on Supporting Group Work*, pages 151–156.

Lacroix, Z. and Critchlow, T. (2003). Compared evaluation of scientific data management systems. In *Bioinformatics*, pages 371–391. Elsevier.

Levin, N. and Leonelli, S. (2017). How does one "open" science? questions of value in biological research. *Science, Technology, & Human Values*, 42(2):280–305.

Lin, Y.-W., Bates, J., and Goodale, P. (2016). Co-observing the weather, co-predicting the climate: Human factors in building infrastructures for crowdsourced data. *Science and Technology Studies*, 29(3):10–27.

Liu, J., Pacitti, E., Valduriez, P., and Mattoso, M. (2015). A survey of data-intensive scientific workflow management. *Journal of Grid Computing*, 13(4):457–493.

Macaulay, C., Sloan, D., Jiang, X., Forbes, P., Loynton, S., Swedlow, J. R., and Gregor, P. (2009). Usability and user-centered design in scientific software development. *IEEE software*, 26(1):96.

Machado Paixão-Cortes, V. S., da Silva Tanus, M. d. S., Paixão-Cortes, W. R., de Souza, O. N., de Borba Campos, M., and Silveira, M. S. (2018). Usability as the key factor to the design of a web server for the cref protein structure predictor: The wcref. *Information*, 9(1):20.

Michener, W. K., Allard, S., Budden, A., Cook, R. B., Douglass, K., Frame, M., Kelling, S., Koskela, R., Tenopir, C., and Vieglais, D. A. (2012). Participatory design of dataone—enabling cyberinfrastructure for the biological and environmental sciences. *Ecological Informatics*, 11:5–15.

Mortier, R., Haddadi, H., Henderson, T., McAuley, D., and Crowcroft, J. (2013). Challenges & opportunities in human-data interaction. *University of Cambridge, Computer Laboratory*.

Néron, B., Ménager, H., Maufrais, C., Joly, N., Maupetit, J., Letort, S., Carrere, S., Tuffery, P., and Letondal, C. (2009). Mobyle: a new full web bioinformatics framework. *Bioinformatics*, 25(22):3005–3011.

Newman, G. J. (2010). *Designing and evaluating participatory cyber-infrastructure systems for multi-scale citizen science*. PhD thesis, Colorado State University.

Oliver, H., Diallo, G., De Quincey, E., Alexopoulou, D., Habermann, B., Kostkova, P., Schroeder, M., Jupp, S., Khelif, K., Stevens, R., et al. (2009). A user-centred evaluation framework for the sealife semantic web browsers. *BMC bioinformatics*, 10(10):1–15.

Overmyer, T. (2019). UX methods in the data lab: arguing for validity. In *Proceedings of the 37th ACM International Conference on the design of communication*, SIGDOC '19, pages 1–6. ACM.

Parsons, M. and Duerr, R. (2005). Designating user communities for scientific data: challenges and solutions. *Data Science Journal*, 4:31–38.

Patterson, R. E., Blaha, L. M., Grinstein, G. G., Liggett, K. K., Kaveney, D. E., Sheldon, K. C., Havig, P. R., and Moore, J. A. (2014). A human cognition framework for information visualization. *Computers & Graphics*, 42:42–58.

Pérez, F. and Granger, B. E. (2007). Ipython: a system for interactive scientific computing. *Computing in science & engineering*, 9(3):21–29.

Perez, F. and Granger, B. E. (2015). Project jupyter: Computational narratives as the engine of collaborative data science. *Retrieved September*, 11(207):108.

Perry, M. (2003). Distributed cognition. *HCI models, theories, and frameworks: Toward a multidisciplinary science*, pages 193–223.

Poole, A. H. (2015). How has your science data grown?

digital curation and the human factor: a critical literature review. *Archival Science*, 15(2):101–139.

Queiroz, F., Silva, R., Miller, J., Brockhauser, S., and Fangohr, H. (2017). Good Usability Practices in Scientific Software Development. *arXiv:1709.00111 [cs]*, page 376271 Bytes.

Ramakrishnan, L. and Gunter, D. (2017). Ten principles for creating usable software for science. In *2017 IEEE 13th International Conference on e-Science (e-Science)*, pages 210–218. IEEE.

Ribeiro, R. (2014). The role of experience in perception. *Human Studies*, 37(4):559–581.

Samourkasidis, A. and Athanasiadis, I. N. (2020). A semantic approach for timeseries data fusion. *Computers and Electronics in Agriculture*, 169:105171.

Sanchez Reyes, L. L. and McTavish, E. J. (2022). Approachable case studies support learning and reproducibility in data science: An example from evolutionary biology. *Journal of Statistics and Data Science Education*, pages 1–20.

Scotch, M., Parmanto, B., and Monaco, V. (2007). Usability evaluation of the spatial olap visualization and analysis tool (sovat). *Journal of Usability Studies*, 2(2):76.

Stepanyan, I. (2021). Cognitive ergonomics of dna-algorithms. In *IOP Conference Series: Materials Science and Engineering*, volume 1129, pages 12–48. IOP Publishing.

Swaid, S., Maat, M., Krishnan, H., Ghoshal, D., and Ramakrishnan, L. (2017). Usability heuristic evaluation of scientific data analysis and visualization tools. In *International Conference on Applied Human Factors and Ergonomics*, pages 471–482. Springer.

Vogt, L. (2021). Fair data representation in times of escience: a comparison of instance-based and class-based semantic representations of empirical data using phenotype descriptions as example. *Journal of Biomedical Semantics*, 12(1):1–25.

Volentine, R., Owens, A., Tenopir, C., and Frame, M. (2017). Usability testing to improve research data services. *Qualitative and Quantitative Methods in Libraries*, 4(1):59–68.

Wald, D. M., Longo, J., and Dobell, A. R. (2016). Design principles for engaging and retaining virtual citizen scientists. *Conservation biology*, 30(3):562–570.

Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., and He, L. (2022). A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems*.

Xu, W. (2019). Toward human-centered ai: a perspective from human-computer interaction. *interactions*, 26(4):42–46.

Zhang, B. Y. and Chignell, M. (2021). Reshaping human factors education in times of big data: Practitioner perspectives. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 65, pages 1574–1578. SAGE Publications Sage CA: Los Angeles, CA.

Zhang, J. (2018). *JupyterLab_Voyager: A Data Visualization Enhancement in JupyterLab*. PhD thesis, California Polytechnic State University, San Luis Obispo, California.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS