Purdue University
Purdue e-Pubs

Proceedings of the IATUL Conferences | 2022 IATUL Proceedings

Jun 15th, 9:00 AM - 10:15 AM

Leveraging socio-technical collaborations to support researchers at the University of Florida

Plato Smith
University of Florida, plato.smith@ufl.edu

Plato Smith, "Leveraging socio-technical collaborations to support researchers at the University of Florida." *Proceedings of the IATUL Conferences.* Paper 9.
https://docs.lib.purdue.edu/iatul/2022/clr/9

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Introduction

A U.S. Department of Health and Human Services IR-ORI-22-001grant proposal was submitted on March 14, 2022 to conduct participatory action research using a mixed methods approach including but not limited to hybrid meetings, previously IRB approved surveys, and training workshops including to improve transparency in the conduct of research, reporting of research, and development of tools and resources to ensure research integrity and compliance with title 42 Code of Federal Regulations 93.100 General Policy (b)[1]. 42 CFR 93.100(b) states:

> 'The U.S. Department of Health and Human Services (HHS) and institutions that apply for or receive Public Health Service (PHS) support for biomedical or behavioural research, biomedical or behaviora research training, or activities related to that research or research training share responsibility for the integrity of the research process. HHS has ultimate oversight authority for PHS supported research, and for taking other actions as appropriate or necessary, including the right to assess allegations and perform inquiries or investigations at any time. Institutions and institutional members have an affirmative duty to protect PHS funds from misuse by ensuring the integrity of all PHS supported work, and primary responsibility for responding to and reporting allegations of research misconduct, as provided in this part."

The project proposed implementing a general data repository pilot (Figshare), electronic research notebook support (RSpace), Open Science Framework (OSF), and protocols.io resources through an innovative socio-technical approach to distribute responsibility for the integrity of the research process. Through the selection of electronic research solutions that offer multiple integrations such as RSpace (See: Fig. 7), researchers will have the option of developing research notebooks that integrate with existing productivity products in use by researchers to enhance description, aggregation, representation, and dissemination of data for transparency and reproducibility. Vendor training coupled with related Collaborative Institutional Training Initiative (CITI Program) courses will support research, development, and demonstration activities to advance the evolving field of research integrity through the development of good laboratory practices and electronic research notebook (ELN) tools with integrations to support components of the research ecosystem. Electronic research notebook with audit trails, data repository that integrates with open researcher and contributor ID (ORCID), and protocols that integrate with data repositories such as Open Science Framework (OSF) and protoocls.io to support effective communication between authors and/or collaborators which can resolve authorship/collaborator disputes and/or issues related to research integrity and patents. Select CITI Program courses can provide foundational training necessary to better understand the significance of use of the proposed resources such as ELNs, Figshare, Open Science Framework (OSF), and protocols.io for funded research. Institutional site licenses versus free/personal versions not only enable the aggregation, organization, and representation of current users of ELNs, Figshare, OSF, and protocols.io but also provide enhanced features such as single sign-on for authentication, metadata, security, and increase public and private sharing of detailed research methods.

University-wide collaborations require the identification, aggregation, and collaboration of key stakeholders such as the Office of Research, Research Compliance Office, Information Technology Department, researchers, academic units, and libraries (Erway, 2013) for institutional resources to better support researchers throughout the research ecosystem . This paper reflects efforts to budget institutional support for 1.) Figshare site license, 2.) Open Science Framework Institutional Membership, 3.) protocols.io site license, 4.) electronic research notebooks, and 5.) relevant CITI Program courses from an unsuccessful grant proposal.

[1] https://www.ecfr.gov/current/title-42/chapter-I/subchapter-H/part-93/subpart-A/section-93.100.

Literature review

What is reproducibility?

The Turing Way (2021) provides two definitions for reproducibility (and replication). One definition of reproducibility used in computational science is by Claerbout and Karrenbach (1992). Claerbout and Karrenbach (1992) define reproducibility as "authors provide all the necessary data and computer codes to run the analysis again, re-creating the results"(Turing Way, 2021). Another definition of reproducibility used in microbiology, immunology and computer science is by Association for Computing Machinery (ACM) introduced in 2013 (Turing Way, 2021). ACM defines reproducibility as the following:

> "(Different teams, different experimental setup.) The measurement can be obtained with stated precision by a different team, a different measuring system, in a different location on multiple trials. For computational experiments, this means that an independent group can obtain the same result using artifacts which they develop completely independently" (Turing Way, 2021; Ivie and Thain, 2018).

Reproducibility challenges

"Reproducing computational research poses challenges in many environments" (Schwab et al., 2000, p. 61). Reproducibility challenges result from lack of protocols, budget, documentation, infrastructure, resources, and tools which impact teams and experimental setup. A study by Stagge et al. (2019) assessing data availability (e.g. annotation, data, models, code, directions for use) and research reproducibility reflects replicability issues in hydrology and water resources revealed out of 360 of the 1,989 articles published by six hydrology and water resources journals in 2017, only 4 articles were fully reproducible (See: Fig. 1).

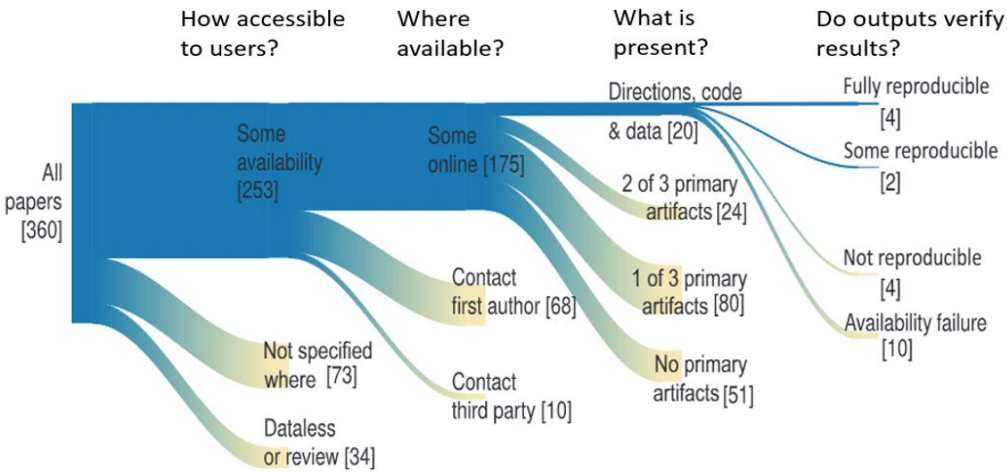


Fig. 1 Number of papers progressing through a reproducibility survey (Stagge et al., 2019).

What is an electronic lab notebook (ELN)/electronic research notebook (ERN)?

Electronic lab notebook (ELN) originated from electronic notebook research that began at the Pacific Northwest National Laboratory (PNNL) in 1994 (Myers, 2003). "Laboratory notebooks (LNs) are vital documents of laboratory work in all fields of experimental research. The LN is used to document experimental plans, procedures, results, and considerations based on these outcomes" (Vaas et a., 2016, p. 1). "An electronic research notebook (ELN) describes a software resource which records how research resulting in the production of data is conducted"

(Smith II b, 2022). An electronic lab notebook, also referred as electronic research notebook (ERN), can serve as "a well-established routine for documentation [that] discourages data falsification by ensuring the integrity of entries in terms of time, authorship, and content" (Vaas et al., 2016, p 1; Myers, 2003). "Many ERN packages offer features enabling data management, collaboration, integration with other software, laboratory information management (LIMS) and many more" (University of Cambridge, 2022). ERN represents use beyond wet lab research.

Criteria and rationale for selecting an institution-wide ERN

An official University Committee comprised of management, PIs, research, academic units, scientists at all levels and the institute's IT department and the Libraries is best qualified for formal assessment and selection of an institution-wide ELN/ERN solution (Smith II a, 2022). The criteria for selection based on ELN/ERN features include but are not limited to:

- Local hosting and storage
- Recovery Options (Deleted files can be located and restored at any time (follows 21 CRF requirements)
- Import/Data Entry/Data Migration from Other Systems (incoming migration)
- Export/Exit Strategies (outgoing migration)
- Able to search for documents within the platform
- Multiple platform support
- Integration with other systems & instruments
- Powerful and open API
- Security: GLP (Good Laboratory Practice) & FDA 21 CFR Part 11 (Supports digital signatures and witnessing)
- Audit Trail
- Electronic signature
- Creates stable URLs or persistent identifiers for entries
- Metadata creation prompts
- Sample and inventory system integrated for entries
- Integrated Apps
- Support to training
- Guest access
- Protocol Integration (Özkan, 2022)

The rationale for selection an institution-wide ERN include but are not limited to:

- Scope to go beyond traditional wet lab research
- Good documentation and help guides; Easy integration with institutional structures (e.g. storage) and research services (e.g. DMPTool)
- Ability to export data in multiple formats
- Avoiding vendor lock-in
- Good application programming interface (API)
- Working in partnership with shared support (Wilson, 2022)

Prior work

A survey titled "Investigating Data Assets, Management, and Planning at UF" was conducted in March 2017. One hundred and fifty-six respondents participated in the institutional review board (IRB) approved (IRB#20160203) 26 questions survey. The respondents represented 81 faculty, 34 graduate/professional students, 26 staff, 6 postdoctoral fellows, 2 residents, 2 undergraduates, and 2 others (Prof. Emeritus; no research role). Some respondents did not answer all questions. The participants were asked in question 15 of 26, "What are some

barriers for you with regard to managing and storing your research?" (See: Fig. 2). The survey illustrated some Research Data Management barriers. The top barriers to managing and sharing data include budget/funding 39.7% (62/156), data management plan (DMP) expertise/experience 38.4% (60/156), storage/technology 36.5% (57/156), and infrastructure/resources 33.1% (52/156) followed by other 8.9% (14/156) and stakeholders 8.3% (13/156). These barriers can create a research culture that negatively impacts the quality of research resulting in quality, replicability, and reproducibility issues (See: Fig. 2).

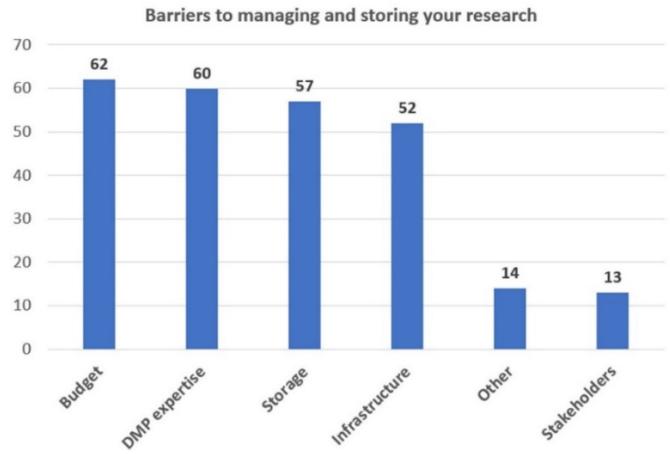


Fig. 2 What are some barriers for you with regards to managing and storing your research data? (Q15, IRB#20160203 2017 Survey)

Project Description

Good laboratory practice (GLP) and responsible research require structural, relational, and transformative change. Structural change includes policies, practices, and resource flows. Relational change includes relationships and connections and power dynamics. Transformative change includes mental models (Kania, Kramer, & Senge, 2018, p. 4). Structural, relational, and transformative changes require a socio-technical systems approach (See: Fig. 4) which includes the complex interdependencies of social and technical systems that can impact people and their understanding of research integrity, research misconduct, and protection of science, public health, and public health service (PHS) funds. Thus, academic libraries as a broker of socio-technical resources can improve practices related to transparency in the conduct/reporting of research through development of resources. The potential impacts include but are not limited to improvements in the documentation, organization, representation, and dissemination of research. Introduction to good laboratory practices and standards such as 21CFR11 inherent in electronic lab notebook enhanced by protocols.io and CITI training courses are important for researchers, research administrators, and stakeholders. The proposed project has potential to support institutional efforts concerning the digital persistent identifiers, information sharing, and international collaboration components of the White House Office of Science and Technology Policy (OSTP) National Security Presidential Memorandum (NSPM) – 33 announced by the Consortium of Social Science Associations (COSSA) on January 11, 2022, and supported by the ORCID US Community Call: NSPM-33 & ORCID on March 15, 2022.

Innovations in technology and computational availability have allowed researchers to generate an unprecedented volume of health-related data. Thus, documenting research through all stages of the research lifecycle is important to ensure integrity, reproducibility, and transparency. Researchers that create research data must add context and meaning (metadata), undertake early research data management (file naming, data organization, location), and follow applicable standards for data creation, description, and representation. The promotion of research integrity through deterrence and education (Titus et al., 2008) requires shared

responsibility among researchers, staff, and students. The CITI Program courses proposed for this project which were vetted and confirmed by a scientist in the University of Florida Department of Molecular Genetics & Microbiology include .1) Research Study Design, 2.) Good Laboratory Practice (GLP), 3.) Data Management and Security for Student Researchers: An Overview, 4.) Responsible Conduct for Research (RCR), and 5.) False Claims Act: A Primer and guide for Research Organizations. These CITI Program courses were selected to promote "a notion of integrity not just in science but in the science-society relationship" (Mitcham, 2003, p. 273) as an approach to advance 'co-responsibility' for scientific integrity through education as a deterrent to issues related to research and climate that negatively impact the reporting of research. Thus, it is the opportunity of the academic libraries to provide resources such as the aforementioned CITI Program training courses to share in the responsibility of the integrity of the research process.

Rather than focus on individual traits and local (laboratory and departmental) related to misconduct, this project aims to promote general scientific integrity and 'responsible conduct of research' (Martinson, et al., 2005, p. 738) with education, resources, and tools. The roles of academic libraries can transition from stakeholder to partner to resources broker throughout the *biomedical data lifecycle* (Harvard University Medical School, 2022) and research data lifecycle across multiple communities of practice.

Artificial intelligence (AI) requires cores, graphics processing unit (GPU), power, and storage to facilitate data-intensive research. Academics libraries must collaborate with IT research computing, institutes, and external partners to support AI. "Reproducibility requires multiple, progressive components such as all data, models, code, directions, and other digital artifacts used in the research are available to others to reuse" (Stagg et al., para. 2, 2019). Reproducibility is "obtaining consistent results using the same input data, computational steps, methods, and conditions of analysis; it is synonymous with computational reproducibility" (NASEM, 2019, #1). Replicability is "obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data" (NASEM, 2019, #1). "Academic institutions, journals, conference organizers, funding organizations, and policy makers can all play a role in improving the reproducibility and replicability of research" (NASEM, 2019, #10) while also improving the supporting infrastructure (Wilkinson et al, 2016).

Researchers need research support capacity, infrastructure, and resources to comply with funding requirements such as the improvement of practices enabling transparency in the conduct of research, reporting of research, and effective communication between authors and collaborators to resolve authorship/collaborator issues related to research integrity, including but not limited to patent disputes. The proposed project includes a development of tools and resources with features and complimentary integration capabilities (See: Fig. 3) that aim to further researchers' efforts to improve practices related to transparency in the conduct of research, reproducibility, and replicability of research. Researchers with different teams and different experimental setups must provide all the necessary data and computer code to run the analysis again, re-creating the results for reproducibility (Claerbout and Karrenbach, 1992; The Turing Way, 2021). Researchers with different teams and same experimental setups must produce a study that arrives at the same scientific findings as another study, collecting new data (possibly with different methods) and completing new analyses for replicability (Claerbout and Karrenbach, 1992; The Turing Way, 2021). Protocols.io and ELN are resources to aid researchers' efforts.

The proposal built upon two prior UF Institutional Review Board (IRB) approved surveys "Exploring electronic lab notebooks (ELN) at UF" in 2020 (#IRB202001407) and "Investigating Data Assets, Management and Planning at UF" in 2017 (#IRB201602303). Both surveys asked participants to provide their name and email address if they were interested in 1.) contributing to data management use cases, 2.) partnering with the Libraries and Research Computing, and 3.) participating in more in-depth data management assessments, education, interviews to improve research data management at UF. The 2017 survey provided the names and email addresses of 36 participants that completed the #IRB2016023030 survey. The 2020

survey provided the 38 names and 37 email addresses of participants that completed the #IRB202001407 survey. The participants from both surveys will be contacted and informed of the new proposed project resources to stimulate initial grassroot support. The participants will then be asked to share proposed resources across their communities of practice. Quantitative and qualitative survey data will be collected using Qualtrics survey for assessment. Quantitative data will be analyzed using a statistical package. Qualitative data will be analyzed using Atlas.ti. Basic referential statistics for quantitative data using Qualtrics and categorization of themes for qualitative data using Atlas.ti will be used to measure lessons learned.

The proposed activities will help achieve the project goals through enhanced research support services, upgrade to enterprise licenses of SaaS tools and introduction of CITI training courses on good laboratory practice, responsible conduct for research, and False Claims Act.

The Proposed IR-ORI-22-001 Project and its ecosystem (See: Fig. 3) can be readily adapted for specific research settings and/or disciplines to improve practices at the level of stakeholders (i.e. individual researchers, research groups, research administrators, or research institutions). Since many UF researchers are currently using Figshare, OSF, protocols.io, and electronic lab notebooks, the proposed project to pilot institutional licenses will aggregate currently distributed UF researchers under respective institutional licenses and increase the practice of persistent identifiers, information sharing, and international collaboration.

The clear connection between the focus area of developing practices to improve the transparency in the conduct and reporting of research is evident by preliminary investigations of the proposed activities. Current UF researchers using OSF, Figshare, and protocols.io do so with free versions or paid versions with limited features. Upgrading to enterprise versions of these products increases features and institutional support for these products that will improve how researchers conduct research. The CITI Program courses proposed for this project will expand the current CITI courses available for researchers to broaden understanding of select core competencies of research for downstream benefits in behavior that enable responsible conduct for research.

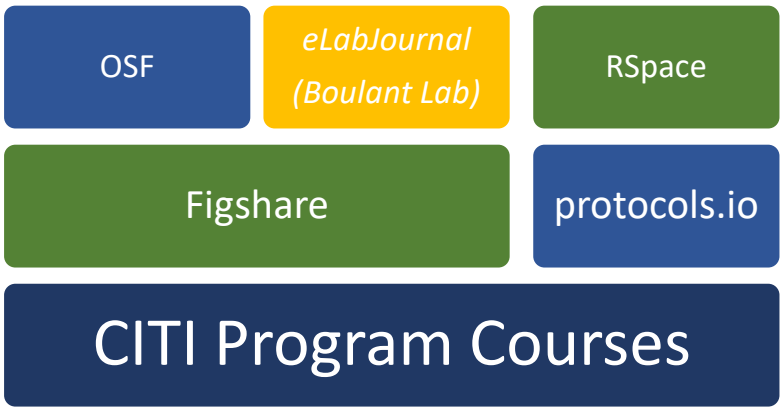


Fig. 3 Proposed IR-ORI-22-001 Project and its ecosystem supporting transparency.

Socio-technical systems framework

A socio-technical systems theory approach (See: Fig. 4) includes focus on critical subsystems of "a social system in which the members form relationships through activities, and a technical system where they perform a series of tasks related to specific goals" (Baek et al, 2018). Organizational change programs objectives met by joint optimization of the social and technical aspects (Cherns, 1976) of a system embedded within an external environment (Davis, et al, 2014) can help mitigate potential social system problems related to alters (people) and egos in the system. Parsons (1951) asserts the problem of access to facilities (alters as 'supplier' of facilities), disposed problem (alters as 'consumer'), problem of cooperation (alters as 'partners') and renumeration problem (alters as 'income sources') are interdependently related

to technical instrumental goal-orientation of ego (Parson, 1951, p. 73) which can impact mental models that affect structural changes such as policies, practices, and resource flows.

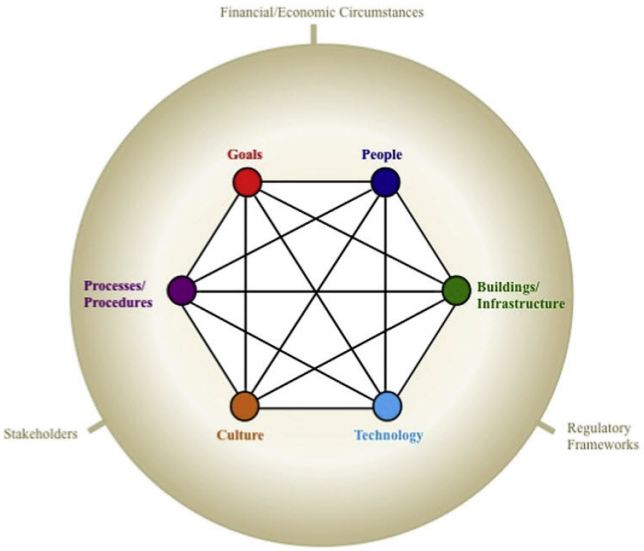


Fig. 4 Socio-technical system, illustrating the interrelated nature of an organizational system, embedded within an external environment (Davis et al., 2014).

The socio-technical system framework attempts to address some of these problems in an organizational system embedded within an external environment of financial/economic circumstances, regulatory frameworks, and stakeholders which impact suppliers, consumers, income sources, and partners. “Organizational change programs often fail because they are too focused on one aspect of the system, commonly technology [or people - research integrity, research misconduct, transparency in the conduct of research], and fail to analyze and understand the complex interdependencies that exist” (University of Leeds, 2022) such as stakeholders, financial/economic circumstances, and regulatory frameworks. Organizational culture perspectives such as integration, differentiation, or fragmentation (Martin, 1992) impact research integrity, research conduct, and reproducibility.

The Corey Lab at Harvard University Medical School is a good example of electronic lab notebook use to support researchers in better documenting, organizing, and sharing research that promote transparency in the conduct and reporting of research. ELNs have audit trails for effective communication and delineation between authors and/or collaborators.

> “[The Corey Lab] have been searching for an ELN to manage the diverse storage and security requirements of data collected in the lab and our extensive inventory of equipment and reagents, and to aid the efficient management of multi-collaborator projects in our group. After attending the ELN seminar run by Julie Goldman and meeting with Joanna Loveluck, it was evident that the time was right to proceed. The valuable support from Joanna and the Research Computing (RC) team turned what initially seemed like a daunting task, within the limited time allowed from research requirements, to an efficient evaluation platform for a suitable product. With the constant exchange of ideas between our lab, especially my colleague Andy Ward, and the RC team, we successfully arrived to the selection of the ELN that will best serve our lab's needs and, we hope, the needs of many HMS groups.” - Domenica Karavitaki, Research Associate, Department of Neurobiology, Corey Lab (Loveluck, 2020)

Protocols.io (section developed in collaboration with protocols.io)

The protocols.io platform is for developing and sharing research methods; its versatility ensures that it is an ideal tool for researchers in all disciplines that use a step-by-step methodology, from life sciences, engineering, chemistry, physics, data science, computer science, all the way through to the social sciences. And as an online platform that can be accessed by many users simultaneously, protocols.io is also an incredibly powerful and adaptable teaching resource. protocols.io enables researchers to structure their research for developing and sharing reproducible methods. The editor feature enables creation of protocols, documents, or collections; advanced components for precision and details; concurrent editing for teams; workspace commenting functionality; and detailed history with ability to roll back changes. The workspaces feature enables shared files and reagent library; shared protocol editing options; secure file sharing; individual and workspace commenting features; and fully integrated task manager. The file manager feature enables support any file type; granular permissions; archiving, auditing, exporting; connect to Dropbox, Google Drive, OSF; and enterprise grade security and backup. Protocols.io enhances public sharing, improves reproducibility, and gives researchers more credit for their method-development work. Many UF researchers share research via protocols.io. As of February 23, 2022, there are 179 accounts associated with ufl.edu domain. The protocols.io CEO presented a brief demo presentation for UF Libraries on March 4,2022.

On August 9, 2022, the Data Management Librarian reached out to protocols.io to request an update on the number of users with University of Florida domain (ulf.edu). The number of protocols.io users with ufl.edu domain though not comprehensive (i.e. users registered with private email addresses and ORCID ID) have increased since February 2022. As of August 10, 2022, there are 190 users of protocols.io with ufl.edu as reported by protocols.io Sales Director.

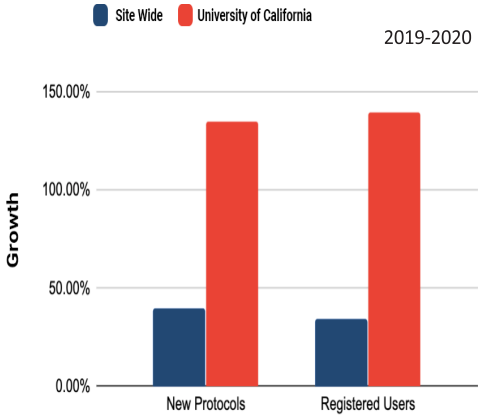
On August 11, 2022 the protocols.io CEO provided recommendations to improve this article, particularly the protocols.io section, and provided a link to protocols.io brand usage[2]. The protocols.io logos/infographics in Appendix B represent Commenting, Editor, File manager, Groups, and Run features of protocols.io.

protocols.io Institutional Benefits case study - University of California

The University of California protocols.io case study articulates institutional benefits. A protocols.io Institutional Plan increases productivity, facilitates teaching, improves collaboration and recordkeeping, and accelerates progress across most research disciplines. University of California (UC) is a system of public research universities in the US routinely ranked among the world's best. Founded in 1869, UC has established a reputation for pioneering research, innovation, and discovery. UC signed up for a three-year institutional plan of protocols.io in June 2019. The protocols.io academic contract applicable to the proposed project is available online via https://www.protocols.io/view/protocols-io-academic-and-non-profit-contract-b33iqqke.

The University of California site wide new protocols and user growth after adoption of academic/non-profit license is evidence of potential of increase of growth from a peer institution.

[2] protocols.io brand usage. https://www.protocols.io/brand.



School	User Growth	New Protocol Growth
UC Merced	575%	300%
UC Santa Cruz	94%	114.43%
UC Santa Barbara	109.09%	280%
UC San Francisco	136.21%	230.48%
UC San Diego	89.72%	155.11%
UC Riverside	106.25%	125%
UC Los Angeles	122.41%	152.09%
UC Irvine	54.35%	179.17%
UC Davis	75.70%	213.16%
UC Berkeley	34.89%	28.90%
Site Wide	34.43%	39.56%

Fig. 5 University of California site wide new protocols.io and user growth.

Utilizing protocols.io provides all researchers with a means to record all of the details of methods they devise or adjust. Regarding teachers, protocols.io provides a means to place a method online for all students to follow, either for a practical or theoretical lesson. Equally, students can be tasked with submitting their approach on the platform for evaluation. Ultimately, protocols.io provides a solution to some of the many challenges faced by universities today: to ensure reproducibility, open science, and support researchers and teachers with adaptable tools.

The proposed project includes foundation CITI Program courses available to the entire institution that teaches different competencies for responsible research. Protocols.io enables tracking, commenting, and sharing of active research processes. Figshare enables researchers to share research outputs. RSpace enables researchers to document their research and provides audit trails for effectively allowing authors and/or collaborators to communicate with each other throughout the research process. The project seeks to aggregate distributed OSF, Figshare, and protocol.io users under respective institutional licenses to further transparency. “Adoption of protocols.io for internal organization and tracking of methods subsequently will lead to better reporting and optionally to more reproducible publications” (protocols.io, 2022). Some of the reproducibility capabilities and functionalities features are illustrated in Appendix B.

CITI Program Training Courses (purposively selected for this project)

The Collaborative Institutional Training Initiative (CITI Program) offers training needs to universities, technology and research organizations, and government agencies to foster integrity and professional advancement of their learners. After searching CITI for integrity and data management and cross-checking against current CITI training courses available to the University of Florida, several CITI training courses were identified as relevant for this project. The identified CITI courses were then shared with Dr. Steeve Boulant from UF Department of Molecular Genetics & Microbiology who confirmed the following as relevant to this project related to focus areas: 1.) transparency in the conduct or reporting of research and 2.) effective communication between authors/collaborators to avoid issues related to integrity. Proposed CITI training includes:

1. Research Study Design - https://tinyurl.com/2yv8ar65.
2. Good Laboratory Practice (GLP) - https://tinyurl.com/yb8a4yqq.
3. Data Management and Security for Student Researchers: An Overview - https://tinyurl.com/y874n3ct.
4. RCR series: https://about.citiprogram.org/series/responsible-conduct-of-research-rcr/.

5. False Claims Act: A Primer and Guide for Research Organizations - https://tinyurl.com/yac7eq89.

Figshare generalist data repository (consulted with Figshare)

Figshare is a generalist data repository that enable researchers to share research outputs. On January 26,2022, the National Institutes of Health (NIH) Office of Data Science Strategy announced new initiative to improve access to NIH-funded research. The Generalist Repository Ecosystem Initiative (GREI) will make it easier to find and reuse NIH-funded research. The GREI includes six established generalist repositories. Open Science Framework (OSF) and Figshare are two of the six generalist repositories. The GREI will work together to establish consistent metadata, develop use cases for data sharing, train and educate researchers on FAIR data and the importance of data sharing, and more (HHS, 2022a). Figshare and Open Science Framework (OSF) are listed among the generalist repositories for data guidance according to scientific data. Figshare is a general data repository the enables researchers to store, share, and discover research via DOI. Figshare provides researcher 100GB free per Scientific Data manuscript, 1 TB per dataset, and upload via Scientific Data (scientific data, 2022). As of March 1, 2022, there are 102 accounts using ufl.edu on Figshare. UF researchers are currently sharing 327 records comprising 1,088 files, and totalling 64.3 GB on Figshare.

The number of records shared using by UF researchers using Figshare is illustrated in Fig. 6. Figshare was selected for this project due to listing as generalist repository by the NIH National Library of Medicine Trans-NIH Biomedical Informatics Coordinating Committee (BMIC) https://www.nlm.nih.gov/NIHbmic/generalist_repositories.html. Figshare has an established termination policy related to user accounts, profiles, portals, DOIs, and content.

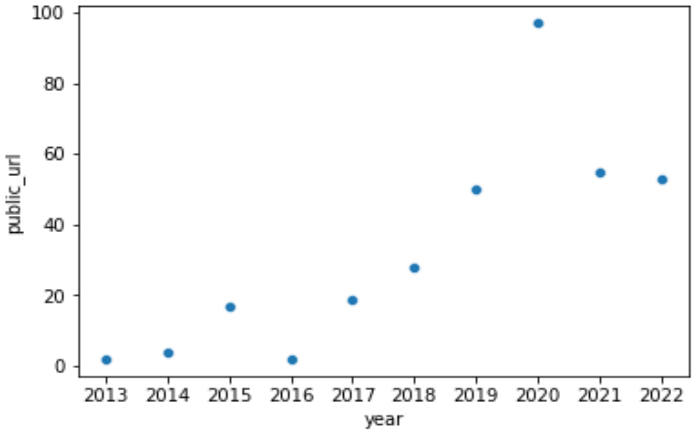


Fig. 6 Number of records shared by UF researchers using Figshare.com (March 2022).

On August 9, 2022, the Data Management Librarian reached out to Figshare to request an update on the number of users with University of Florida domain (ufl.edu). The number of Figshare users with ufl.edu domain though not comprehensive (i.e. third-party email to register Figshare, published through another repository, or publish by external collaborators) have increased since March 1, 2022. As of August 9, 2022, a Figshare.com use by University of Florida August 2022 report generated by Figshare documented that there are 124 figshare.com accounts using a University of Florida email, 348 records, 1213 files, total file size of 74 GB, and 76 citations on Figshare.com (Figshare, 2022). The Links to Funding Sources section in the report stated that 60% of records did not list funding source in the metadata. 6% of records are linked to a dimensions.ai grant record. 34% of the records have free-text grant name. An institutional Figshare implementation would enhance outputs, ensure metadata quality, future data stewardship, and enable integration with ORCID. UF Research enabled ORCID API

integration with UF's Integrated Research Support Tool (UFIRST). UFIRST is the proposal and award management system of the University of Florida.

RSpace (section developed in collaboration with RSpace)

The RSpace ELN is able to function as more of an ELN/ERN. RSpace is a digital data hub, enabling collection and coordination of data from disparate sources (See Fig. 7). RSpace can operate as a valuable conduit between the planning and archiving and re-use phases of the research ecosystem. ELNs are digital analogues of the traditional paper lab notebook. Current ELNs/ERNs technologies facilitate enhanced transparency of research, communication between authors/collaborators, and research integrity. Some benefits of using RSpace include:

1. Controlled sharing of data within and between labs.
2. Complete record of the contributions of all participants in research projects for higher quality attribution of involvement in research thereby enhancing transparency and providing a platform for sustained reproducibility.
3. Capture of more and higher quality data relating to experiments, resulting in improved research integrity and rigor.
4. Integration with Figshare, DMPTool, ORCID, Protocols.io… See: integrations.

Universities in the United States that have completed RSpace pilots are Harvard University, Harvard Medical School, and the University of Nebraska. Universities in Europe that have completed RSpace pilots are UiT Arctic University of Norway, University of Edinburgh, Delft University of Technology (TU Delft), École polytechnique fédérale de Lausanne (EPFL), and University of Liege.

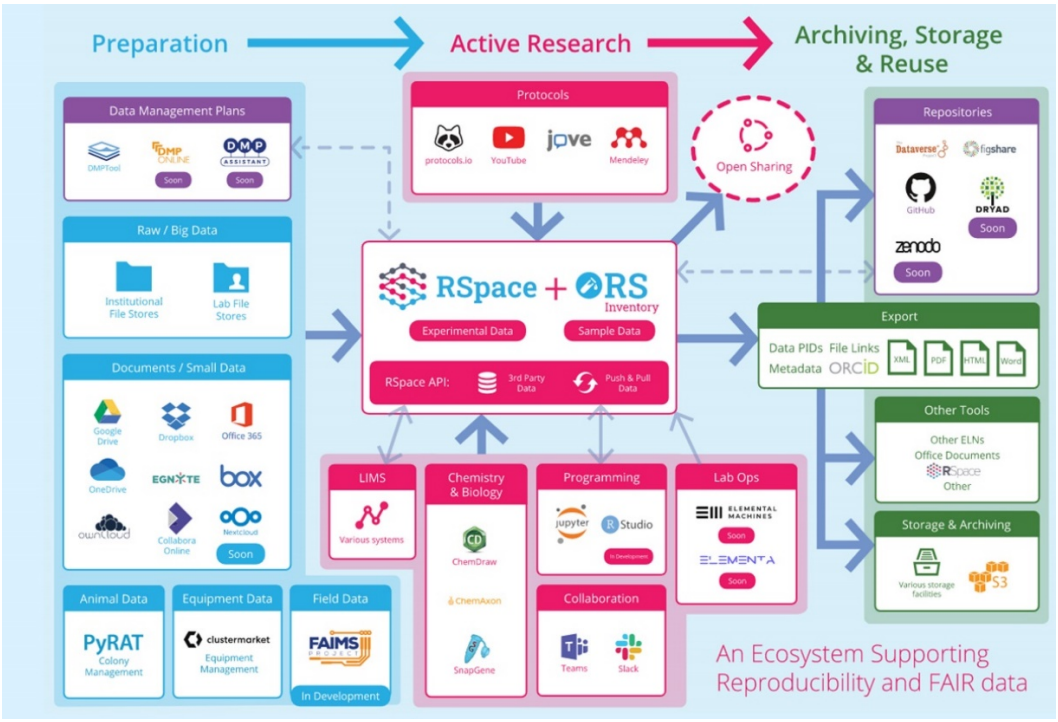


Fig. 7 RSpace and its ecosystem supporting reproducibility and FAIR data.

Select an ELN with at least Title 21 C.F.R. Part 11 standard

ELNs with standards such as Title 21 C.F.R. Part 11 are necessary for quality research. Title 21 CFR Part 11 (often shortened to "21CFR11") is the part of Title 21 of the Code of Federal

Regulations that establishes the United States Food and Drug Administration (FDA) guidelines for electronic records and electronic signatures (ERES). Part 11 defines the criteria under which electronic records and electronic signatures are considered trustworthy, reliable, and equivalent to paper records. Any scientist who goes into commercial research will probably need to know about these regulations. Additionally, because universities are becoming more focused on licensing research discoveries or intellectual property (IP) as a source of revenue, academic labs will certainly need to become more familiar with these regulations in the coming years. 21CFR11 supports promotion of research integrity and prevention of research misconduct. Developed in the US during the 1990s to ensure that records in a digital system can be reasonably assumed to be a true and accurate representation of the work performed, 21CFR11 helps to ensure that the records were created by the author(s) who claims they did the work, and that they performed the work at the date and time that they say they did it. 21CFR11 helps reviewers to understand who has had access to, or has contributed to the work. Digital records, particularly PHS-funded research should be collected and stored in a 21CFR11 compliant system if you want to eventually use your records as supporting evidence in any type of formal, legal or regulatory proceeding, or if you want the records to be used in patent hearings, or other "due diligence" / intellectual property verification events. Ownership and accuracy disputes related to scientific research are surprisingly common and expensive in both industry and academia. Many disputes can be avoided if records are kept in a 21CFR11 compliant system.

Initially 21CFR11 was only considered an absolute requirement for US-based, FDA regulated organizations (i.e. research related to human health and safety), however over the last two decades, 21CFR11 has gradually been accepted as the international "gold standard" for scientific digital data management and is becoming common in many research sectors, or indeed, anywhere that accurate, reliable, legally defensible or patent-related record keeping is required. 21CFR11 record keeping should be considered mandatory for any modern, efficient, paperless lab. Additionally, 21CFR11 compliance is probably the main reason why the University of Florida should use a dedicated, professional ELN system vs. home-grown systems, or ad-hoc, non-compliant data management tools like Dropbox, MS Onenote, MS Sharepoint, Evernote, shared file servers, which have some features and safeguards that seem appealing.

Open Science Framework (OSF)

The OSF Institutional Open Science Membership includes: access to quarterly webinars; public resources on https://help.osf.io/hc/en-us; OSFI project (https://osf.io/ba3dy/); publicly listed on COS supporter page; single sign on (SSO) login access and sign up; affiliate projects and registrations (public and private); aggregated page for all public affiliated projects and registrations; publicly listed on university partner page (https://osf.io/institutions); dashboard for internal access to SSO user list; public and private project counts; department for SSO users (if configured by institution to share metadata attribute); publicly listed on Open Science member page; 25% off training fees; support priority; monthly newsletters sharing use cases, examples, training notes, tips and tricks for workflow efficiency, and OSF new and future features. As of March 1, 2022, there are 705 active accounts associated with ufl.edu domain. On August 9, 2022, the author requested an update on the number of OSF users with .ufl.edu domains. On August 15, 2022, OSF responded that there are over 750 users with ufl.edu associated accounts.

Potential problems include delay in implementation of Figshare, OSF, and protocol.io. Thus, the principle investigator (PI) suggests soft and firm deadlines for implementation to allow buffer of 5-7 days for unexpected delay. Potential problem with RSpace is selecting the most collaborative users/lab that are willing to interact regularly with utilizing and developing use cases from use of the products. The potential problem with CITI Training Program is working with the UF CITI Program rep to facilitate rapid acquisition of the new proposed courses. Strategies include incentivizing participation of researchers through the promotion of their research via repositories and tools.

eLabJournal (section developed in collaboration with eLabJournal)

In preparation for submission of the HHS grant, the author contacted eLabJournal to request feedback on the targeted focus areas of 1.) improving practices related to transparency in the conduct or reporting of research and 2.) improving practices related to "effective communication between authors and/or collaborators to avoid, mitigate, and resolve authors, collaborators disputes and/or issues related to research integrity" (U.S. HHS OASH, 2022). A scientist in the UF Department of Molecular Genetics & Microbiology recruited from Heidelberg University who previously used eLabJournal for his lab at Heidelberg University was recommended by the University Libraries to submit a UF Integrated Risk Management (IRM) Fast Path Solutions (FPS) risk assessment for eLabJournal. eLabJournal was approved by UFIRM FPS for open data. Risk assessments of ELN resources by researches is recommended.

An ELN can enable significant benefits to laboratories in improving organization in the physical and digital lab spaces, streamlining collaboration amongst several authors, and accelerating life science research discoveries. A larger issue is whether the electronic lab notebook enables its users to practice research integrity while eliminating potential for research misconduct.

eLabNext includes integrated high-security standards expressed in daily practice evident via:

1. **Datacenter and network security**: regular monitoring and encryption.
2. **Application and product security features**: proper training, quality assurance checks, security measures, and utilizing security features.
3. **Certification and compliance**: eLabNext is ISO27001 and GDPR certified and 21 CFR part 11 compliant.

> "The EU General Data Protection Regulation (GDPR) requires organizations to adopt appropriate technical and organizational measures – including policies, procedures and processes – to protect the personal data they possess. ISO 27001, the international standard for an ISMS (information security management system), provides an excellent starting point for achieving the technical and operational requirements necessary to reduce the risk of breach" (IT Governance, 2022).

Fig. 8 eLabNext dashboard with articulated key ELN features supporting FAIR.

Conclusion

The proposed project includes leveraging existing good working relationships with UF IT Research Computing, UF Informatics Institute, UF Clinical and Translational Science – IT, UF Research, UF Department of Molecular Genetics & Microbiology, UF Department of Biomedical Engineering, and Figshare, RSpace, OSF, protocols.io, and eLabNext (eLabJournal) representatives. The reason why these products were selected for the proposed project is that all of the resources have been approved by UF Integrated Risk Management (IRM) Fast Path Solutions (FPS) for open data, excluding OSF (*in final process*). Fast Path Solutions (FPS) is a comprehensive list of pre-assessed software and computing environments (https://irm.ufl.edu/). Thus, the libraries must consult and collaborate with information technology research computing on capacity for risk assessments of resources for data (i.e. open, sensitive, restricted).

Horizon issues

Collectively, the proposed resources for this project contribute to support of the United States White House Office of Science and Technology Policy (OSTP) National Security Presidential Memorandum (NSPM)-33, particularly use of digital persistent identifiers (DPIs) such as digital object identifiers (DOIs) in ELNs and data repository enhanced by Open Researcher and Contributor ID (ORCID) of authors and collaborators connecting research to researchers. The Consortium of Social Science Association (COSSA) provide a good primer on NSPM-33. “Successful data sharing ultimately depends on cultural and social infrastructures as much as on technical infrastructures” (Ruediger et al., 2022).

The development of socio-technical collaborations both internal and external coupled with costs for institutional resources integrated into grant budgets are crucial to advance research support services and the role of Libraries in the research data ecosystem and support NSPM-33.

Acknowledgements

The author acknowledges Lenny Teytelman and Karen Wilson (protocols.io), Olivia Lin and James St. Pierre (eLabNext), Eric Olson (Open Science Framework), Rory Macneil (RSpace), Patrick Reakes, Benjamin Walker, Valrie Minson, Dr Laurie Taylor, and support from the Agriculture and Food Research Initiative Competitive Grant no. 2018–67016-27578 awarded as a Center of Excellence from the USDA National Institute of Food and Agriculture.

References

Baek, J. S., Kim, S., Pahk, Y., & Manzini, E. (2018). A sociotechnical framework for the design of collaborative services. Design Studies, 55, pp. 54–78. https://doi.org/10.1016/j.destud.2017.01.001.

Cherns, A. (1976). The Principles of Sociotechnical Design. Human Relations, 29(8), pp. 783–792. https://doi.org/10.1177/001872677602900806.

Claerbout, J. F. and Karrenbach, M. (1992). Electronic documents give reproducible research a new meaning. URL: https://library.seg.org/doi/abs/10.1190/1.1822162, doi: https://doi.org/10.1190/1.1822162.

COSSA. (2022). OSTP Releases Guidance for NSM-33, Long Awaited Research Security Roadmap. https://cossa.org/2022/01/11/ostp-releases-guidance-for-nspm-33-long-awaited-research-security-roadmap/.

COGR. (2022). Council On Government Relations. COGR Releases Summaries of NSPM-33 Disclosure Requirements & Other Provisions. https://www.cogr.edu/cogr-releases-summaries-nspm-33-disclosure-requirements-other-provisions.

Erway, R. (2013). Starting the Conversation: University-wide Research Data Management Policy. Why IT Matters to Higher Education. Educause Review. Retrieved from https://tinyurl.com/tjzlrk8.

Figshare. (2022). Figshare.com use by University of Florida August 2022. Excerpt from an internal report developed and shared by Figshare. (unpublished).

Harvard University Medical School. (2022). Biomedical Data Lifecycle. What is the Data Lifecycle? https://datamanagement.hms.harvard.edu/about/what-research-data-management/biomedical-data-lifecycle.

Ivie, P. and Thain, D. (2018). Reproducibility in Scientific Computing. ACM Comput. Surv., 51(3):1–36. doi:10.1145/3186266.

IT Governance. (2022). ISO 27001 and the GDPR. https://www.itgovernance.co.uk/gdpr-and-iso-27001.

Kania, J., Kramer, M., & Senge, P. (2018). The Water of Systems Change. http://efc.issuelab.org/resources/30855/30855.pdf.

Loveluck, J. (2020). Harvard Medical School. Finding the Right Electronic Lab Notebook with the Corey Lab. https://datamanagement.hms.harvard.edu/news/finding-right-electronic-lab-notebook-corey-lab.

Martin, J. (1992). Cultures in Organizations: Three Perspectives. Oxford University Press.

Martinson, B. C., Anderson, M. S., & De Vries, R. (2005). Scientists behaving badly. Nature, 435(9), pp. 737–738.

Mitcham, C. (2003). Co-responsibility for research integrity. Science and Engineering Ethics, 9(2), pp. 273–290. https://doi.org/10.1007/s11948-003-0014-0.

Myers, J. D. (2003). Collaborative electronic notebooks as electronic records: Design issues for the secure electronic laboratory notebook (eln). (July). Retrieved from http://collaboratory.emsl.pnl.gov/resources/publications/papers/seceln(final1)1-22Nov.pdf.

National Academies of Sciences, Engineering, and Medicine (NASEM). (2019). 10 things to Know About Reproducibility and Replicability, https://tinyurl.com/yc4qposf.

NSTC. (2022). National Science and Technology Council. Guidance for Implementing National Security Presidential Memorandum 33 (NSPM-33) on National Security Strategy for United States Government-Supported Research and Development. https://www.whitehouse.gov/wp-content/uploads/2022/01/010422-NSPM-33-Implementation-Guidance.pdf.

Özkan, Ö. (2022). ELN (RSpace) @ MDC. Research Notebooks Community Discussion. Jisc. 06/16/2022. Zoom. (unpublished).

Parson, T. (1951). The social system. American Book-Stratford Press, Inc., New York.

protocols.io. (n.d.). Case Study, The Benefits of Institutional Plans. University of California.

protocols.io (2022). What are the key benefits of Institutional Plans? https://www.protocols.io/help/institutional-plans.

Ruediger, D., MacDougall, R., Cooper, D., Carlson, J., Herndon, J., & Johnston, L. (2022, August 9). Leveraging Data Communities to Advance Open Science: Findings from an Incubation Workshop Series. https://doi.org/10.18665/sr.317145.

Schwab, M., Karrenbach, N., & Claerbout, J. (2000). Making scientific computations reproducible. In Computing in Science & Engineering (Vol. 2). https://doi.org/10.1109/5992.881708.

Smith II a, P. L. (2022, June 27). Exploring Electronic Research Notebooks (ERNs) at a R1 institution in the southeast USA. Zenodo. https://doi.org/10.5281/zenodo.6761977.

Smith II b, P.L. (2022), "Exploring electronic lab notebooks (ELNs) at a R1 institution in the Southeast USA", Digital Library Perspectives, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/DLP-02-2022-0013.

Stagge, J. H., Rosenberg, D. E., Abdallah, A. M., Akbar, H., Attalah. N. A., James., R. (2019). Assessing data availability and research reproducibility in hydrology and water resources. Scientific Data. Retrieved from https://www.nature.com/articles/sdata201930.

Titus, S. L., Wells, J. A., & Rhoades, L. J. (2008). Repairing research integrity. Nature, 453(19).

Teytelman, L. (2022). Introduction to protocols.io for The University of Florida. https://tinyurl.com/ycnm9u3j.

The Turing Way. (2021). Guide for Reproducible Research: Overview of Reproducible Research – Definitions. https://the-turing-way.netlify.app/reproducible-research/overview/overview-definitions.html#rr-overview-definitions.

United States Department of Health & Human Services (HHS). (2022). Office of Research Integrity. Notice of Funding Opportunity: Ensuring Research Integrity – Research, Development, and Demonstration, Opportunity Number: IR-ORI-22-001. https://www.grants.gov/web/grants/view-opportunity.html?oppId=334952.

University of Cambridge. (2022). Electronic Research Notebooks. The Basics. https://www.data.cam.ac.uk/data-management-guide/electronic-research-notebooks.

University of Leeds Business School (2022). Socio-technical systems theory. https://business.leeds.ac.uk/research-stc/doc/socio-technical-systems-theory.

Vaas, L. A. I., Witt, G., Windshügel, B., Bosin, A., Serra, G., Bruengger, A., … Kohler, M. (2016). Electronic laboratory notebooks in a public-private partnership. PeerJ Computer Science, 2016(9), 1–22. https://doi.org/10.7717/peerj-cs.83.

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18.

Wilson, J. A.J. (2022). UCL Electronic Research Notebook – the story so far. Research Notebooks Community Discussion. Jisc. 06/16/2022. Zoom. (unpublished).

Appendix A

The following Budget Narrative Content from a submitted yet unfunded 2022 U.S. Department Health & Human Services IR-ORI-22-001 grant proposal was to support institutional resources.

Overall Cost of the Project Justification

The primary goal of this project is to acquire five new Collaborative Institutional Training Initiative (CITI Program) courses, institutional licenses for Open Science Framework (OSF), protocols.io, five licenses for RSpace, five licenses for eLabJournal, and an institutional general data repository solution to enable researchers to improve the practice of transparency in the conduct and reporting of research university-wide. The project teams seeks funding for the following activities, excluding staff time to execute the project: research to acquire institutional general data repository Figshare instance (pilot) for ($16,057); five CITI Program courses for (GLP $675 per year; Research Study Design $675 per year; Data Management and Security for Student Researchers: An Overview $300 per year; Responsible Conduct for Research (RCR) Basic $750 per year; False Claims Act: A Primer and Guide for Research Organizations $675 per year) ($6,150); Open Science Framework (OSF) Institutional Open Science Member ($9,998); protocols.io ($30,000); RSpace ($120 per year * 10 users + one time setup costs/training $1800) ($4,200).

Necessity, reasonableness, and allocation of the proposed costs:

It is necessary for a Research 1 to have a university-wide institutional data repository solution to enable researchers to comply with funders data management and sharing requirements such as the new NIH Data Management & Sharing Policy[3] effective January 25, 2023. Currently, the University of Florida does not have a university-wide general data repository implementation that support researchers in the stewardship of research. It is necessary to support a general data repository implementation that ensures dissemination of research findings acquired using Public Health Service (PHS) funds. The negotiated reduced costs for a university-wide Figshare data repository institutional pilot (up to 1 Terabyte) of data storage implementation in Year 1 for $8,135 and in Year 2 for $7,922 are reasonable. The purpose of the CITI Program courses is to broaden the educational opportunity for researchers and student researchers to learn the basics of responsible conduct for research, good laboratory practice, research designs, and the False Claims Act. The costs of the CITI Program courses are (GLP $675 per year; Research Study Design $675 per year; Data Management and Security for Student Researchers: An Overview $300 per year; Responsible Conduct for Research (RCR) Basic $750 per year; False Claims Act: A Primer and Guide for Research Organizations $675 per year) which totals in Year 1 $3,075 and $3,075 in Year 2. Researchers at a R1 institution need enterprise licenses for select products to maximize research productivity. While some open-source and free-version are feasible for some research activities, researchers should not have to rely on potentially insecure systems for PHS-funded research. Thus, institutional licenses for select products researchers are currently and which have recently been approved by UF Integrated Risk Management (IRM) Fast Path Solutions (FPS) for open data is prosed. OSF allows researchers to connect to select data storages and share data. Protocols.io enable researchers to develop reproducible methods to promote the practice of reproducibility. RSpace is a 21 CFR Part 11 complaint electronic lab notebook that enables secure collaboration, sharing, and audit trail between authors/collaborators to help mitigate issues involving disputes, patents, or authorship. The Open Science Framework (OSF) Institutional Open Science Membership in Year 1 is $4,999 and $4,999 in Year 2. The academic license for protocols.io is $15,000 in Year 1 and $15,000 in Year 2. RSpace is $3,000 ($120 per user times 10 users plus a

[3] https://sharing.nih.gov/data-management-and-sharing-policy/about-data-management-and-sharing-policy/data-management-and-sharing-policy-overview

one-time setup/training costs of $1800) in Year 1 and $1,200 ($120 per user times 10 users) in Year 2.

Year 1 Budget Table

Object Class	Federal Funds Requested	Non-federal Resources	Total Budget
Personnel	$	$0	$
Figshare	$8,135	$0	$
CITI Program	$3,075	$0	$
RSpace	$3000	$0	$
Open Science Framework (OSF)	$4,998	$0	$
Protocols.io	$15,000	$0	$

Year 2 Budget Table

Object Class	Federal Funds Requested	Non-federal Resources	Total Budget
Personnel	$	$0	$
Figshare	$7,922	$0	$
CITI Program	$3,075	$0	$
RSpace	$1,200	$0	$
Open Science Framework (OSF)	$4,998	$0	$
Protocols.io	$15,000	$0	$

Appendix B

The following select screenshots of protocols.io include Commenting, Editor, File Manager, Groups, and Run features.

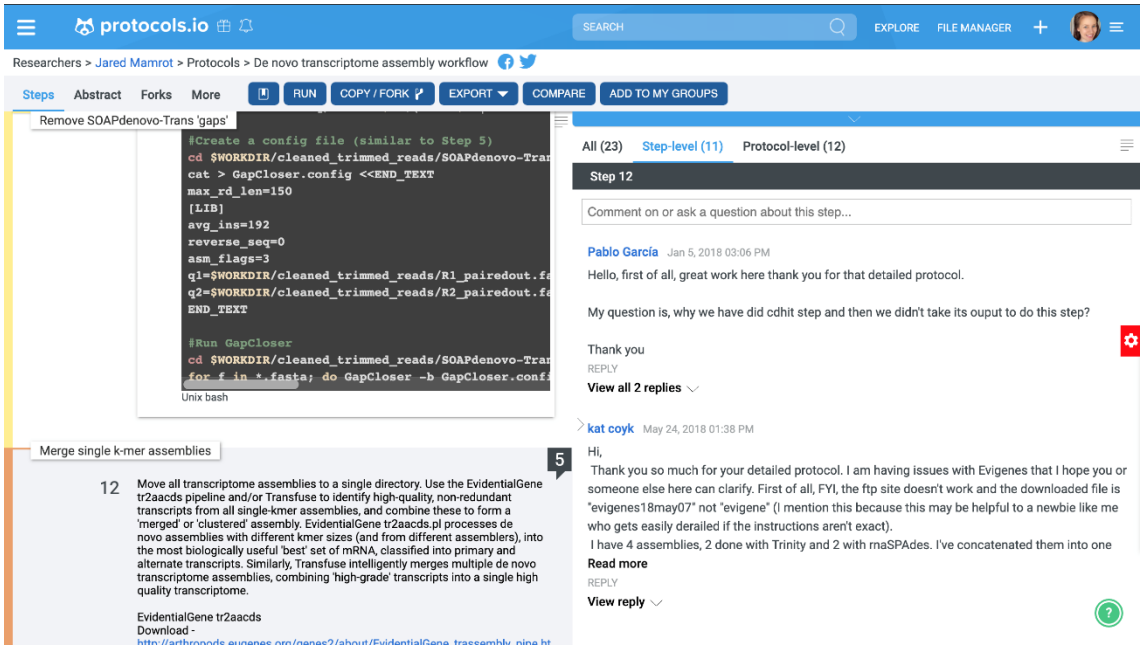


Fig. 9 protocols.io - Commenting

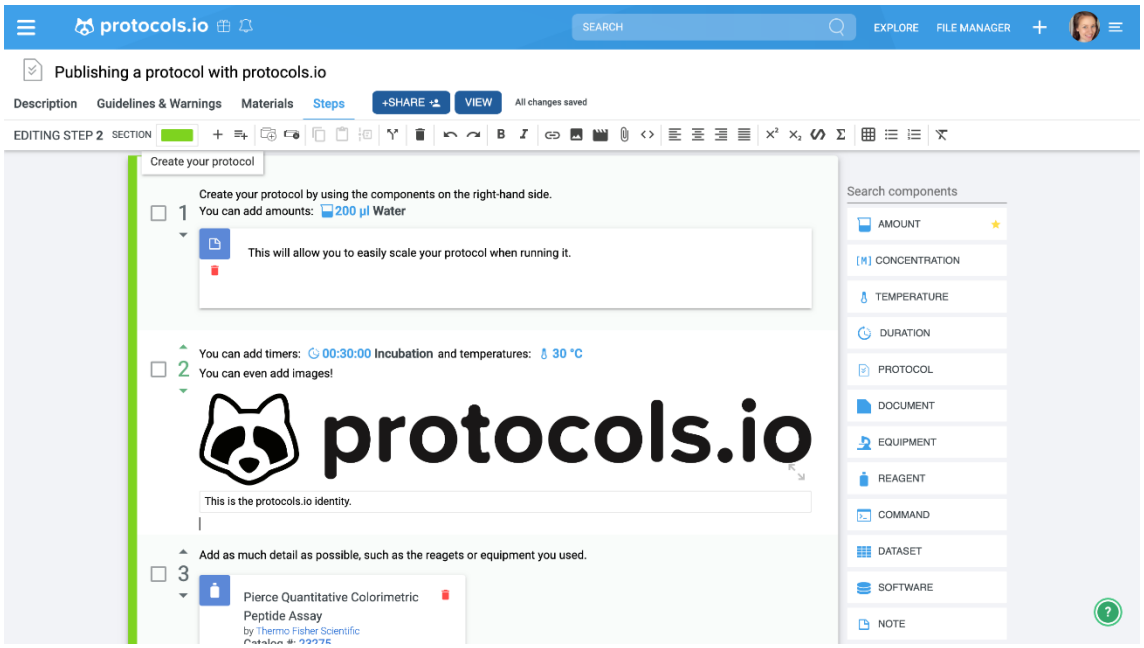


Fig. 10 protocols.io - Editor

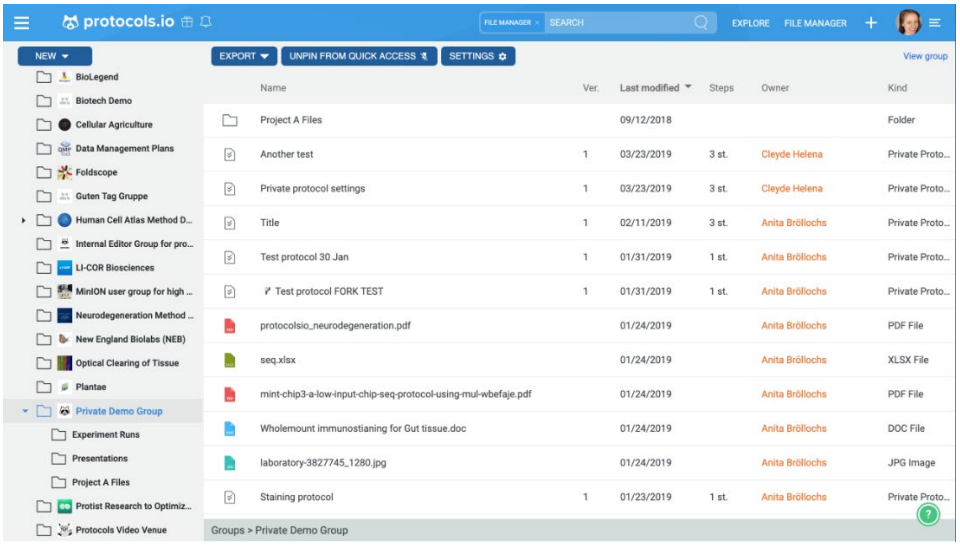


Fig. 11 protocols.io – File Manager

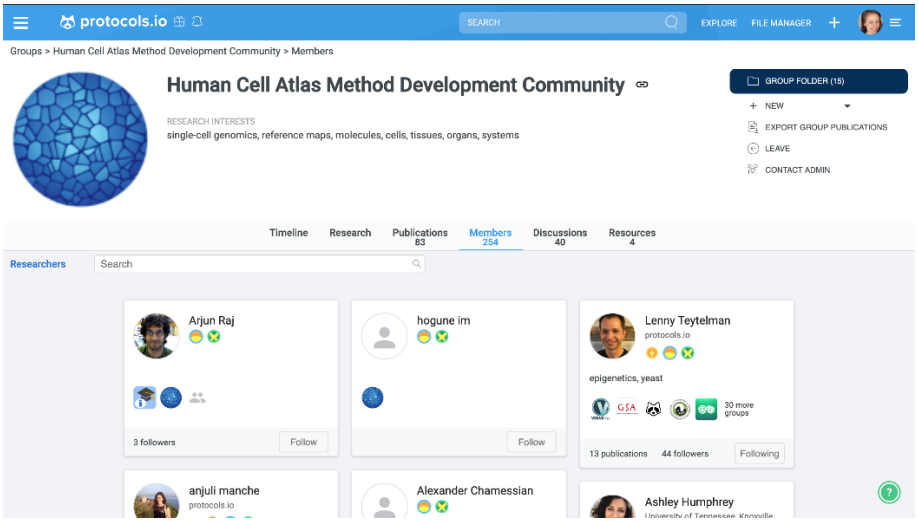


Fig. 12 protocols.io – Groups

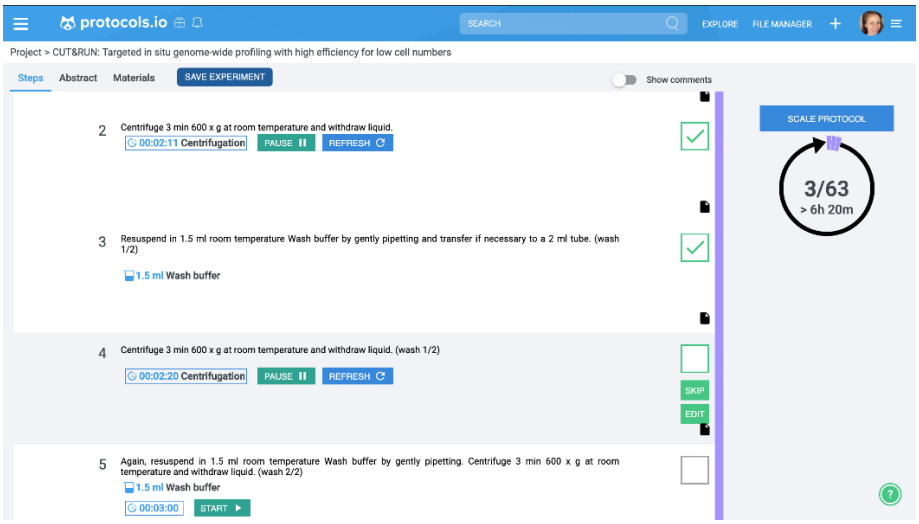


Fig. 13 protocols.io - Run