EAGER: CI PAOS: KnowLedger: An Open Digital Notebook for Research Data Management

Stuart Chalk, Department of Chemistry & Biochemistry, University of North Florida

Narrative

The current efforts of researchers in research data management (RDM) are fragmented in terms of the systems/strategies employed, and as a result much data is lost causing others to potentially re-run failed experiments and wasting significant research funding (estimated at over $200 billion and rising)[1]. This is compounded by a general lack of reproducibility/repeatability[2] in science, due in part to a hesitancy of researchers to share data and research workflows.

Science is expensive however, investment in better data curation and management efforts can improve the quality and increase the return on investment[3]. The research community is reluctant to spend research funding on relatively expensive electronic laboratory notebook (ELN) software that generally ends up not fitting their needs and thus results in poor adoption. Given this situation we can draw the following conclusions:

- Researchers need tailored solutions that fit their workflows and enable research progress rather than get in the way
- The cost of such solutions should be as cheap as possible and preferably free
- Any solution needs to ensure trust in the integrity of the stored data and make clear that a researcher (or the group PI) has complete control of how, when and if data is made available to the community
- The research community must be integral to the development of any such RDM system, so they are vested in its success

Thus, this project focuses on creation of an open ecosystem, KnowLedger, by developing a suite of components needed for an open digital research notebook (DRN) that will nominally serve all research communities, putting the control of the development in the hands of the research community (as far as possible), promoting data science/informatics as an important part of all research activities, and encouraging research data sharing and reuse. There is a deliberate intent to represent KnowLedger as a different system than just another electronic laboratory notebook (ELN) as ELN products have tended to be expensive (for what they are) and are only a computer-based simile of traditional paper laboratory notebooks[4]. KnowLedger is envisioned to be a complete rethink of what a research notebook should be able to do, enabling new ways to record data/take notes, connect with online resources, and flexible so it can be tailored to the needs of the science and the scientific workflow.

Giving researchers the ability within KnowLedger to develop resources they need for RDM in an open and free ecosystem should encourage contribution to/participation in the systems development and garner suggestions for configuration options and functionality broadly. However, this alone is not likely to move the needle on researcher compliance with even federally mandated data sharing. Many publications over the last few years[1, 5-9] have reported on the varied reasons why researchers are reluctant to share their data. Comments about proprietary data, being scooped in research, reuse of data outside of what is fit for purpose, for etc. are reasonable reflections on issues with sharing, however reading between the lines in many cases PIs are hesitant because they don't have a handle on the quality of the data and how well (or not) the research was done (or where the files are).

The general topic of digital record keeping (a 'ledger' of what was done, in what order, and by whom) is the main advantage of a comprehensive system that stores research data at time of collection, records how data workup is done, and does not allow editing of collected research data. It is hoped that KnowLedger will allow PIs to feel they have a tool that will let them see a dashboard view of research in their group, be able to drill down to specific experiments, review data analysis procedures, and ask individual researchers targeted questions (and answers) that will stay with the research. This picture is only that of the PI of this proposal and we hope the research community will have much grander plans for what such a system could become/enable.

Background

Scientists offer up many reasons to not share data, and even agree to do so, but then don't follow through. In a recent Nature article[8] almost 1800 papers stated that research data was available online or was available upon request. Sadly, when contacted, over 90% did not provide the data, either by not responding or if they responded the went back on what they said. It seems clear that many in the research community are still paying only lip service to the idea of sharing data. In many cases this is due to a misunderstanding about how you collect and store data in anticipation of sharing it, and a reluctance to invest in infrastructure to do so. To change this mindset, we need to give the research community a tool that gives them confidence they are in control of their data and can get the advantages of sharing data with the rest of the community.

Over the last few years, Dr. Samantha Pearman-Kanza at the University of Southampton has published a series papers studying the chemistry communities' interest in electronic laboratory notebooks[4, 10-12]. In her latest paper[4], the author concludes that; i) scientists need better digital tools, ii) scientists are currently using a wide array of disparate systems, iii) the metadata recorded with research data needs to be improved, and iv) digital data needs to be managed better. This work was based on surveys of the chemistry research community over the last few years and suggests that more efforts should be made to train/educate scientists in research data management and create tools more align with the research perspective.

The idea of creating a customizable replacement for the laboratory notebook, is not new. There are currently many tools on the market[13], and the PI has submitted proposals on this topic to NSF in the past. Looking back at one of those proposals[14], the data format of choice was the Extensible Markup Format[15] (XML), the protocols of choice were Simple Object Access Protocol (SOAP)[16], Open Archive Initiative – Protocol for Metadata Harvesting[17] (OAI-PMH), and Representational State Transfer (REST[18]). While the 'preferred' format may have changed (to JavaScript Object Notation – JSON[19]) and the protocol has standardized on REST, the idea in the proposal was very similar to this one, just not as broad in application. What are the same are the types of data (plus metadata) that would be captured in the system (reproduced below) – suggesting that there is a logical set of components for such a system, that is generally well understood. What is missing are connections to web resources and services that can enable a shared ecosystem, foster collaboration and incentivize creation of research workflows that can be reused (and improved) by others, akin to current developments in protocols[20], or study design[21].

- Annotation
- Calculation
- Chemical
- Citation
- Communication
- Customer
- Data
- Dataset
- Definition
- Element
- Equipment
- Experiment
- Instrument
- Observation
- Project
- Protocol
- Report
- Result
- Sample
- Solution
- Substance
- Task
- Timeline
- User
- Vendor

Datatypes to be captured in an digital research notebook, reproduced from [[14]].

Some communities are starting to advance the ideas behind open science by promoting the needs for standards for metadata collection and storage. In chemistry, a recent article[22] outlined what we need to think about to digitally annotate research data appropriately, that will standardize data from different laboratories, thus making data sharing easier. This paper, authored by the NFDI4Chem[23] project in Germany and the International Union of Pure and Applied Chemistry (IUPAC)[24], is a disciplinary example of what is needed to move a research community toward FAIR and standardized research data management.

In addition, there has been growing interest in the development of persistent identifiers for research, including e.g., instruments[25], samples (iSamples[26], IGSN[27]), organizations/vendors (ISNI[28]), general research resources (RRID[29]), in addition to common examples such researchers (ORCID[30]), publications (DOI[31]), and funders (ROR[32]). PID development is needed more broadly, and this will be highlighted to researchers that work with the KnowLedger system, and research communities more broadly.

The SciData Framework

The PI has worked for the last ten years on a project to create a flexible file format for the storage of research data, and it's associated metadata in one file. The SciData framework[33-35] uses the approach of defining three main categories of information relative to research data – methodology, system, and dataset (see figure below). Although the framework is file format agonistic, it has become clear in a recent NSF grant by this PI[36] that JSON-LD representation of data using this framework is the most suitable because of the inherent semantic representation of data. This means that context (metadata) of research data can not only be recorded and saved with research data, but the knowledge inherent in the file can be used for knowledge mining, inferencing, and semantic integration with other such data.

The SciData JSON-LD Layout: Insert - overall structure, left side – system metadata and datapoint

The SciData framework is also an atomistic example of the FAIR Digital Object (FDO) approach being proposed[37] as a practical way to find data online (though a finding aid), allowing user to know what can be done with the data (license), and understand what is available for download.

Philosophy behind the KnowLedger Concept

It has been ten years, since the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles were first outlined, an in that time many advances have been made to toward enabling data availability and

reuse, and policies around how research data should be managed and licensed. Unlike 2014, funding agencies are now requiring research grant awardees to make their data available, fulfilling promises made in data management plans submitted with research proposals. However, the proverbial stick may not be enough to tip the balance for the research community. It might be that they need a 'carrot' as well.

The carrot proposed here is a research data management system, KnowLedger, that is designed around the needs of research, that being:

- a place to capture research data digitally at birth (including a log of research activities)
- an interface that allows easy annotation (text, audio, video)
- a platform that provides data analysis and workflow tools (through integrated modules)
- a system that integrates with online research resources (visual API's)
- a resource that stores data semantically (for knowledge mining)
- a notebook that allows sharing of workflows, protocols, metadata standards
- an ecosystem that is focused on data management, trust, and enabling collaboration

This proposal, in of itself, cannot make the utopian system above a reality, it's just too much. Our goal is to outline an approach, setup several key resources, provide ideas of how the ecosystem could/should be setup, and do all we can to enable the research community to; understand the idea, see what the future of research data management could be, and dedicate resources to develop important components for their communities.

The approach to the KnowLedger concept is to:

- Develop the KnowLedger software (in Python) to be open platform upon which data can be collected, managed, processed, annotated, analyzed, published, and archived
- Use of a generic but flexible data container – the SciData framework[34] in JSON-LD – to store research data, sharable templates (experiments, resources, raw and processed data), adapters (to connect online resources), and converters (to import data from online systems)
- Modularize the ecosystem around this approach to encourage tool development, resource access, workflow monitoring/control by developing a JSON-LD frameworks for the data infrastructure
- Develop an open framework for persistent identifier (PID) assignment to data and resources in alignment with the recently published strategy for PID development[38] and current PID systems
- Provide open access to community-built data templates, resource modules, and workflow templates via a network of GitHub repositories that can be centrally linked though persistent identifiers[39]
- Provide documentation of all parts of the ecosystem and delineate common best practices for development of code in support of resources
- Engage with the International Unions, CODATA/ISC and scientific societies to pro
- Promote the FAIR principles[40], CARE[41], digital Data Management Plans (DMP's) and open science

Our goals, which are intended to focus the development of KnowLedger, are to:

- Remove barriers to the collection of rich metadata along with experimental data, analysis and results (e.g. help researchers get better metadata from instruments, sensors, surveys, etc.)
- Promote community development of minimal metadata standards for recording instrument, experiment, sample, workflow, analysis metadata, etc.
- Enable community development/adoption of best practices for workflows, analysis and data reporting
- Encourage data sharing, data reuse, and collaboration to identify interoperability challenges and develop approaches to tracking data provenance and versioning of datasets
- Define/integrate metrics for data quality, data reliability, data provenance, and data integrity
- Generally, adopt best practices, community standards, and open specifications where available and appropriate

The PI has invited researchers from many disciplines to be unfunded collaborators on this project (and will invite more should the grant be awarded). One of their activities is to help the PI convene a steering committee of technical and scientific advisors for this project as soon as, and if, it is awarded.

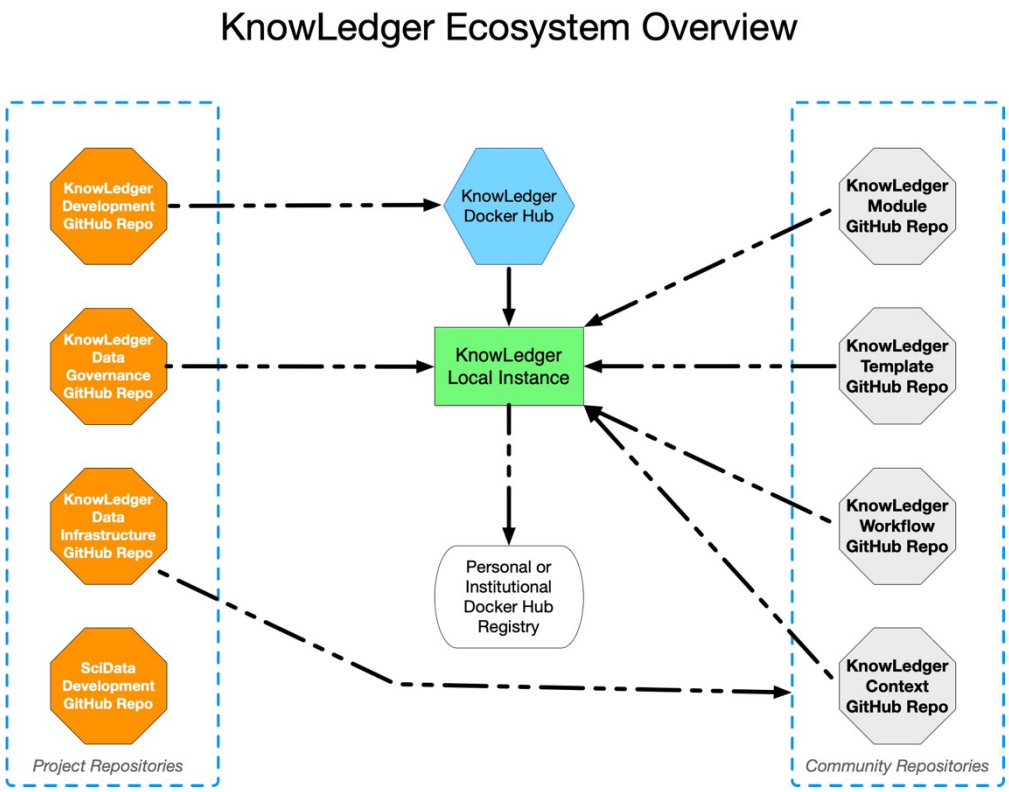
Technical Development/Implementation

Development of an ecosystem for such a broad approach to research data management is a challenge. To make priorities tractable in this project we will rely on those with time and enthusiasm to lead whatever priority aspect(s) is/are important to them. To shape and scope the ecosystem, a vision document will be developed by the PI and collaborators, as soon as the project has been funded (before the formal start date) so that by the formal start date we can start inviting the research community to get involved.

The detailed vision document will evolve during the project, but will start from the following high-level description and key pillars envisioned to be core to the project, as informed by the NIST Research Data Infrastructure (RDaF) 2.0[42]:

- KnowLedger ecosystem deployment/development
    - Inherently the ecosystem needs to be scalable through an open and free tool set. Our initial idea is to use open resources such as GitHub and Docker/Docker Hub as the basis for making available the resources outlined below (see diagram). Once the project is underway this approach will be refined based on input from the steering committee.
    - All components of the data infrastructure will be develop using JavaScript Object Notation for Linked Data (JSON-LD), an encoding format of the Resource Description Framework (RDF) for semantic data. This will mean that will be fully semantic in nature, describing both the data and the infrastructure at a granular level.
    - Wherever possible, the system will utilize best practices, e.g. for development of persistent identifiers[38], secure data transfer[43], or application programmable interfaces (APIs)[44].
- Data infrastructure
    - Each piece of the data infrastructure will be developed through project (invited collaborators) and community (open) GitHub repositories that will form the backbone of the KnowLedger ecosystem (see diagram).
        - KnowData - SciData framework for research data capture/data templates
        - KnowAll - framework for (local) activity tracking
        - KnowShow - framework for data governance
        - YouKnow - framework for user profiles
        - KnowModule - framework for module development
        - KnowFlow - framework for workflow/protocol development/implementation
        - WeKnow - framework for collaboration (internally/externally)
        - KnowHow - framework for education and training
        - KnowUs - framework for community engagement
- Technical development infrastructure
    - The core KnowLedger software will be developed in Python and made available as Docker container with separate KnowLedger and MongoDB[45] images (to deploy the system locally). After initial rollout of the individual/local deployment of KnowLedger an institutional deployment will be developed.
    - Data backup will be done initially using Docker both on the user's local intranet and via Docker hub[46] (with account)
- Data management/governance infrastructure
    - This part of the infrastructure will describe the functionality of the research dashboard in KnowLedger, the interface where availability of data locally, internal to a research group, externally via collaborative agreement, open after embargo, or generally open can be managed.

- Project communication infrastructure
    - Through the GitHub repository wiki's, communication about the evolution of parts of the ecosystem will be made available. In addition, a KnowLedger updates module will be developed to advertise these updates and link to details of the changes.
    - After the development of the module above, a second module will be developed that will allow users of KnowLedger to communicate to each other.
- Education/training infrastructure
    - The modular nature of the KnowLedger ecosystem will be leveraged to develop educational content and training modules for different research activities and make them available via a KnowLedger
- Community awareness infrastructure
    - An important activity in this project is engaging researchers and research communities with short informative communications about progress of the project and encouraging participation.
- Online semantic resources
    - As the system is underpinned with semantic data representation, various online resources such as Schema.org[47], Ontobee[48], Bioportal[49], Wikidata[50], etc. will be used to capture and iterate on the semantic representation of scientific knowledge.
    - The Python package SciContext[51] will be used to create and make available context files for the JSON-LD representation of metadata elements in metadata standards developed through the project.



Project Timeframe

- Year 0 (late 2024) (after award has been made but before formal start date)
  The PI will communicate with the unfunded collaborators announcing the award and asking to meet relative to creating the steering committee. The PI will review the project with this group and ask them for suggests of experts in their field that they see would be able to best contribute to this project. Advertisement/recruitment of Post-Doctoral Associate via discussion with collaborators.
- Year 1 (2025)
  - Q1: Hire postdoc; promote project; first meeting of the steering committee; recruit community members to development team; develop design criteria of KnowLedger software; design/implement JSON-LD data infrastructure.
  - Q2: Second meeting of the steering committee; review/revise technical infrastructure and advertise for the formation of teams for major components; develop/deploy KnowLedger ecosystem GitHub repositories; develop and document Docker install process; run and online town hall meeting discussing the project and seeking community input.
  - Q3: Third meeting of the steering committee; begin development of KnowLedger Software; develop initial GitHub wiki content targeted to develop community; run collaborator/community needs analysis survey, encourage formation of community groups; recruit first adopters of KnowLedger; test first version at UNF and some first adopters.
  - Q4: Fourth meeting of the steering committee; release of the beta version of the KnowLedger software, engage and support the community; encourage researcher sprints using KnowLedger relative to a common research activity; ask collaborators to identify conferences to prioritize for presentations on KnowLedger.
- Year 2 (2026)
  - Q1: Fifth meeting of the steering committee; release first full version of the Knowledger software, enable community groups to develop standards, train trainers in the development in the KnowLedger ecosystem, work with first adopter research groups to accelerate usage/development.
  - Q2: Sixth meeting of the steering committee; identify pain points in the ecosystem and triage solutions; hold online workshop for the KnowLedger community to gather feedback and seek input on functionality to be prioritized in the remaining six months of the project.
  - Q3: Seventh meeting of the steering committee; ask collaborators to organize workshops of KnowLedger in their communities; seek groups to develop research grants to continue funding for KnowLedger.
  - Q4: Eighth meeting of the steering committee; seek user reviews of the KnowLedger system; solicit feedback from the collaborators about successes/failures/lessons learned; write final report; write papers on the project to be published open access.

Additional funding

Additional funding will be requested via NSF 23-128 to enable collaborative development of KnowLedger in the UK's EPSRC PSDI[52] project, a research data infrastructure. While some of the funding through this mechanism will come to the PI (travel and summer support in the UK), the request will primarily be focused on engaging/enabling research groups in the UK (through seed funding) to deploy the KnowLedger system and use it to connect PSDI resources into groups research workflow. Subsequently, the PSDI team will use this research to fund more general KnowLedger deployment in the UK to encourage data sharing in addition to usage of PSDI resources.

How this project fits in with this NSF RDM DCL

This project is directly targeting the rapidly evolving data science cyberinfrastructure for code development, open standards for communication, security, etc. and its application to the important topic of research data management. This project expects to use the wealth of talent in cyberinfrastructure to address the needs of

RDM for the research community and link them to scientists in different disciplines so that together they can address the data grand challenges of the scientific community.

Appropriateness of this proposal for EAGER funding

The ideas in this proposal are based on the PI's experience on many different digital projects and some in-house development of KnowLedger. With this background the PI realized that the project can only be effectively done with a large group of collaborators, across many disciplines, which would create an unwieldy regular NSF proposal. As the nature of the project is to build an open ecosystem, where disciplinary 'ambassadors' are needed to promote and champion the project in their discipline, a regular NSF proposal is not feasible. This EAGER proposal will provide startup funding for this effort, with the idea that broadly engaging different research communities, and encouraging them to commit resources to development of the KnowLedger ecosystem, is the only avenue to get to a sustainable project. Once communities have developed resources and the idea and software get tracking, each community will be encouraged to seek additional funding for the development of domain specific tools, domain specific repository submission templates, and minimal metadata standards for domain specific knowledge representation.

Results from Prior NSF Support

In a recently concluded NSF grant, the PI developed the SciData framework approach to data representation across a variety of different datatypes, in several research communities. Software was developed in the grant to format SciData JSON-LD documents, migrate and organize SciData JSON-LD files in a graph database, and create JSON-LD context files needed for conversion to RDF triples. Several datasets are available and more will come online through the end of this year. One goal of the award that was not met was engaging different research communities to develop minimal metadata standards for discipline specific data. To the PI, this seems to be a critical roadblock that must be overcome to enable storage of data in any format, that allows any researcher to collect and report data of that type and thus significantly improve reproducibility and data quality. Thus, this proposal focuses on this activity, both to improve interoperability and encourage domain groups, at the society or International Union levels, to promote this approach and show how it will enable researchers to share and reuse data.

Intellectual Merit

This project is based on the adoption of the FAIR principles to develop an open system for RDM that can easily and cheaply be deployed in the laboratory to promote data capture at birth, provide high quality metadata about research experiments and enabling science through a customizable, yet interoperable data infrastructure. It is important in this project that we exemplify what FAIR means, and clearly communicate to researchers that FAIR does not mean open. This misconception is still a huge barrier to adoption, or even appreciation, of FAIR. In building KnowLedger, researchers will be given the tools to manage their data, whether they make data freely available, available behind a login and clear reuse license, available and for sale (e.g. large datasets available to industry) or locked until an embargo date is reached.

Broader Impacts

This concept is open to the community in terms of usage of the KnowLedger system, and contributions to the system by the community. All development will be as open as possible, as closed as necessary. Members from different communities will be encouraged to be champions for their community and organize/drive development/interest in the project, contributing tooling, minimum information standards, data converters etc. A research community based steering committee will be identified to oversee the development of the system. The PI has also identified several research groups to collaborate with on this project.

References

(1) Anderson, J. M.; Johnson, A.; Rauh, S.; Johnson, B.; Bouvette, M.; Pinero, I.; Beaman, J.; Vassar, M. Perceptions and Opinions Towards Data-Sharing: A Survey of Addiction Journal Editorial Board Members. *The Journal of Scientific Practice and Integrity* 2022. DOI: 10.35122/001c.35597.
(2) *Reproducibility and Replicability in Science*; 2019. DOI: 10.17226/25303.
(3) Steinhart, G.; Skinner, K. *The Cost and Price of Public Access to Research Data: A Synthesis.*; 2024. DOI: 10.5281/zenodo.10729575.
(4) Kanza, S.; Willoughby, C.; Knight, N. J.; Bird, C. L.; Frey, J. G.; Coles, S. J. Digital research environments: a requirements analysis. *Digital Discovery* 2023, *2* (3), 602-617. DOI: 10.1039/d2dd00121g.
(5) Stieglitz, S.; Wilms, K.; Mirbabaie, M.; Hofeditz, L.; Brenger, B.; Lopez, A.; Rehwald, S. When are researchers willing to share their data? - Impacts of values and uncertainty on open data in academia. *PLoS One* 2020, *15* (7), e0234172. DOI: 10.1371/journal.pone.0234172.
(6) Zuiderwijk, A.; Shinde, R.; Jeng, W. What drives and inhibits researchers to share and use open research data? A systematic literature review to analyze factors influencing open research data adoption. *PLoS One* 2020, *15* (9), e0239283. DOI: 10.1371/journal.pone.0239283.
(7) Gomes, D. G. E.; Pottier, P.; Crystal-Ornelas, R.; Hudgins, E. J.; Foroughirad, V.; Sánchez-Reyes, L. L.; Turba, R.; Martinez, P. A.; Moreau, D.; Bertram, M. G.; et al. Why don't we share data and code? Perceived barriers and benefits to public archiving practices. *Proceedings of the Royal Society B: Biological Sciences* 2022, *289* (1987). DOI: 10.1098/rspb.2022.1113.
(8) Watson, C. Many researchers say they'll share data — but don't. *Nature* 2022, *606* (7916), 853-853. DOI: 10.1038/d41586-022-01692-1.
(9) Kaiser, J.; Brainard, J. Ready, set, share: Researchers brace for new data-sharing rules. *Science* 2023, *379*. DOI: 10.1126/science.adg8470.
(10) Kanza, S. Understanding and Defining the Academic Chemical Laboratory's Requirements: Approach and Scope of Digitalization Needed. In *Digital Transformation of the Laboratory*, 2021; pp 179-189.
(11) Kanza, S.; Knight, N. J.; Willoughby, C.; Bird, C. L.; Frey, J. G.; Coles, S. J. Dataset supporting the PSDI pilot survey-case study 4-survey analysis & ELN data. 2022; https://eprints.soton.ac.uk/476413/5/PSDI_FullListofFeatures.pdf.
(12) Kanza, S.; Willoughby, C.; Bird, C. L.; Frey, J. G. eScience Infrastructures in Physical Chemistry. *Annual Review of Physical Chemistry* 2022, *73* (1), 97-116. DOI: 10.1146/annurev-physchem-082120-041521.
(13) Kanza, S.; Willoughby, C.; Gibbins, N.; Whitby, R.; Frey, J. G.; Erjavec, J.; Zupančič, K.; Hren, M.; Kovač, K. Electronic lab notebooks: can they replace paper? *Journal of Cheminformatics* 2017, *9* (1). DOI: 10.1186/s13321-017-0221-3.
(14) Chalk, S. J. *CIF21 DIBBs: Conceptualization of an ExptML Framework for the Chemical Sciences*; University of North Florida, 2012.
(15) W3C. *Extensible Markup Language (XML) 1.0 (Fifth Edition)*. 2008. https://www.w3.org/TR/xml/ (accessed.
(16) W3C. *Simple Object Access Protocol (SOAP) 1.2*. 2007. https://www.w3.org/TR/soap/ (accessed.
(17) OAI. *The Open Archives Initiative Protocol for Metadata Harvesting*. 2015. https://www.openarchives.org/OAI/openarchivesprotocol.html (accessed 6/30/24).
(18) Fredrich, T. *What Is REST?* 2016. https://www.restapitutorial.com/introduction/whatisrest (accessed 6/30/24).
(19) ECMA. *Introducing JSON*. 2022. https://www.json.org (accessed 12/28/22).
(20) Nature, S. *protocols.io*. 2024. https://www.protocols.io (accessed 6/30/24).

(21) Science, C. f. O. *The Open Science Framework (OSF)*. 2024. https://osf.io (accessed 6/30/24).
(22) Herres-Pawlis, S.; Bach, F.; Bruno, I. J.; Chalk, S. J.; Jung, N.; Liermann, J. C.; McEwen, L. R.; Neumann, S.; Steinbeck, C.; Razum, M.; et al. Minimum Information Standards in Chemistry: A Call for Better Research Data Management Practices. *Angewandte Chemie International Edition* 2022, *61* (51). DOI: 10.1002/anie.202203038.
(23) NFDI4Chem. *NFDI4Chemistry*. 2024. https://www.nfdi4chem.de (accessed 6/30/24).
(24) IUPAC. *The International Union of Pure and Applied Chemistry*. 2024. https://iupac.org/ (accessed 6/30/24).
(25) RDA. *Persistent Identification of Instruments WG*. 2024. https://www.rd-alliance.org/groups/persistent-identification-instruments-wg/ (accessed 6/30/24).
(26) Lehnert, K.; Walls, R.; Davies, N.; Vieglais, D. *Internet of Samples: iSamples*. 2024. https://isamplesorg.github.io/home/ (accessed 6/30/24).
(27) Organisation, I. *International Generic Sample Number (IGSN)*. 2024. https://ev.igsn.org (accessed 6/30/24).
(28) Ltd., I. I. A. *International Standard Name Identifier (ISNI)*. 2024. https://isni.org (accessed 6/30/24).
(29) Bandrowski, A.; Martone, M. E.; Grethe, J. S. *Research Resource Identication*. 2024. https://www.rrids.org (accessed 6/30/24).
(30) ORCID. *Open Researcher and Contributor ID (ORCID)*. 2024. https://orcid.org (accessed 6/30/24).
(31) Foundation, D. *Digital Object Identifier (DOI) System*. 2024. https://www.doi.org (accessed 6/30/24).
(32) Gould, M.; Chodacki, J.; Pentz, E.; Buys, M. *Research Organization Registry (ROR)*. 2024. https://ror.org (accessed 6/30/24).
(33) Chalk, S. J. *The SciData Data Portal*. 2022. https://scidata.unf.edu/ (accessed 12/28/22).
(34) Chalk, S. J. SciData: a data model and ontology for semantic representation of scientific data. *Journal of Cheminformatics* 2016, *8* (1). (acccessed 12/28/22).
(35) Chalk, S. J. *Scientific Data Model Ontology (SDO)*. 2016. http://stuchalk.github.io/scidata/ontology/scidata.owl (accessed.
(36) Chalk, S. J. RUI: Framework: Data - An Open Semantic Data Framework for Data-Driven Discovery. US National Science Foundation: Jacksonville, FL, 2018; https://www.nsf.gov/awardsearch/showAward?AWD_ID=1835643.
(37) Bonino da Silva Santos, L. O. *FAIR Digital Object Framework*. 2022. https://fairdigitalobjectframework.org/ (accessed 6/16/23).
(38) RDA; ORFG. *Developing a US National PID Strategy*; 2024. DOI: 10.5281/zenodo.10811008.
(39) W3ID. *Permanent Identifiers for the Web*. 2024. https://w3id.org/ (accessed 6/30/24).
(40) Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L. B.; Bourne, P. E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 2016, *3* (1). DOI: 10.1038/sdata.2016.18 (acccessed 12/28/22).
(41) Jennings, L.; Anderson, T.; Martinez, A.; Sterling, R.; Chavez, D. D.; Garba, I.; Hudson, M.; Garrison, N. A.; Carroll, S. R. Applying the 'CARE Principles for Indigenous Data Governance' to ecology and biodiversity research. *Nature Ecology & Evolution* 2023, *7* (10), 1547-1551. DOI: 10.1038/s41559-023-02161-2.
(42) NIST. *Research Data Framework (RDaF)*. 2024. https://www.nist.gov/programs-projects/research-data-framework-rdaf (accessed 6/30/24).
(43) NIST. *NIST Cybersecurity Framework 2.0*. 20204. https://www.nist.gov/cyberframework (accessed 6/30/24).
(44) SmartBear. *Swagger: Best Practices in API Design*. 2024. https://swagger.io/resources/articles/best-practices-in-api-design/ (accessed 6/30/24).
(45) Inc., M. *MongoDB*. 2024. https://www.mongodb.com (accessed 6/30/24).
(46) Inc., D. *How to back up and restore your Docker Desktop data*. 2024. https://docs.docker.com/desktop/backup-and-restore/ (accessed 6/30/34).

(47) W3C. *Schema.org (v27.01)*. 2024. https://schema.org/ (accessed 6/30/24).
(48) Ong, E.; Xiang, Z.; Zhao, B.; Liu, Y.; Lin, Y.; Zheng, J.; Mungall, C.; Courtot, M.; Ruttenberg, A.; He, Y. Ontobee: A linked ontology data server to support ontology term dereferencing, linkage, query and integration. *Nucleic Acids Research* 2017, *45* (D1), D347-D352. DOI: 10.1093/nar/gkw918.
(49) Whetzel, P. L.; Noy, N. F.; Shah, N. H.; Alexander, P. R.; Nyulas, C.; Tudorache, T.; Musen, M. A. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Res* 2011, *39* (Web Server issue), W541-545. DOI: 10.1093/nar/gkr469.
(50) Wikimedia. *Wikidata*. 2024. https://www.wikidata.org (accessed 6/30/24).
(51) Chalk, S. J. *SciContext: A Django app to create/publish JSON-LD context files*. 2024. https://github.com/chalklab/SciContext (accessed 7/3/24).
(52) Bicarregui, J.; Montanari, B.; Coles, S. J.; Knight, N. J.; Frey, J. G.; Bunakov, V.; Matthews, B. *The Physical Sciences Data Infrastructure (PSDI)*. 2024. https://www.psdi.ac.uk (accessed 6/30/24).