

Viewpoint

Enabling Open Science for Health Research: Collaborative Informatics Environment for Learning on Health Outcomes (CIELO)

Philip Payne¹, PhD; Omkar Lele², MS, MBA; Beth Johnson³, MPH; Erin Holve⁴, MPH, MPP, PhD

¹Institute for Informatics, School of Medicine, Washington University in St. Louis, St. Louis, MO, United States

²Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, OH, United States

³AcademyHealth, Washington, DC, United States

⁴Department of Health Care Finance, Government of the District of Columbia, Washington, DC, United States

Corresponding Author:

Philip Payne, PhD

Institute for Informatics

School of Medicine

Washington University in St. Louis

660 South Euclid Avenue

MS 8102-13-610

St. Louis, MO, 63110

United States

Phone: 1 314 747 6119

Fax: 1 314 747 6119

Email: prpayne@wustl.edu

Abstract

Background: There is an emergent and intensive dialogue in the United States with regard to the accessibility, reproducibility, and rigor of health research. This discussion is also closely aligned with the need to identify sustainable ways to expand the national research enterprise and to generate actionable results that can be applied to improve the nation's health. The principles and practices of *Open Science* offer a promising path to address both goals by facilitating (1) increased transparency of data and methods, which promotes research reproducibility and rigor; and (2) cumulative efficiencies wherein research tools and the output of research are combined to accelerate the delivery of new knowledge in proximal domains, thereby resulting in greater productivity and a reduction in redundant research investments.

Objectives: AcademyHealth's Electronic Data Methods (EDM) Forum implemented a proof-of-concept open science platform for health research called the Collaborative Informatics Environment for Learning on Health Outcomes (CIELO).

Methods: The EDM Forum conducted a user-centered design process to elucidate important and high-level requirements for creating and sustaining an open science paradigm.

Results: By implementing CIELO and engaging a variety of potential users in its public beta testing, the EDM Forum has been able to elucidate a broad range of stakeholder needs and requirements related to the use of an open science platform focused on health research in a variety of "real world" settings.

Conclusions: Our initial design and development experience over the course of the CIELO project has provided the basis for a vigorous dialogue between stakeholder community members regarding the capabilities that will add the greatest value to an open science platform for the health research community. A number of important questions around user incentives, sustainability, and scalability will require further community dialogue and agreement.

(*J Med Internet Res* 2017;19(7):e276) doi: [10.2196/jmir.6937](https://doi.org/10.2196/jmir.6937)

KEYWORDS

healthcare research; information dissemination; open access to information; social networking; reproducibility of results

Introduction

There is an emergent and intensive national dialogue regarding the accessibility, reproducibility, and rigor of health research. This discussion is also closely aligned with the need to identify sustainable ways to expand the national research enterprise and to generate actionable results that can be applied to improve the nation's health. The principles and practices of *Open Science* offer a promising path to address both goals by facilitating (1) increased transparency of data and methods, which promotes research reproducibility and rigor [1-4]; and (2) cumulative efficiencies wherein research tools and the output of research are combined to accelerate the delivery of new knowledge in proximal domains, thereby resulting in greater productivity and a reduction in redundant research investments [5-7]. For the purposes of the remainder of this viewpoint, we provide the following working definition for Open Science: "Open Science is the practice of science in such a way that others can collaborate and contribute and where research data, lab notes, and other research processes are freely available under terms that enable reuse, redistribution, and reproduction of the research and its underlying data and methods" [8].

Unfortunately, contradictory and sometimes conflicting positions on open science—and the way the open science paradigm might best be operationalized—demonstrate the need for greater community engagement to test the theory that open science in the health sciences can indeed improve the rigor and efficiency of research. This challenge is exemplified by the recent controversy regarding research "parasites," [9] and the vigorous debate that ensued as a result. In response to these important and timely issues, in this viewpoint, we describe a set of lessons learned and future directions associated with an open science initiative conducted by AcademyHealth's Electronic Data Methods (EDM) Forum, called the Collaborative Informatics Environment for Learning on Health Outcomes (CIELO) [10], targeting the broad health research community. We also highlight policy, social, cultural, and implementation-level issues, setting the stage for a vigorous community-wide dialogue concerning future activities as are needed to achieve a compelling vision of open science in health care research and all of its concomitant benefits.

Methods

As mentioned above, we implemented a proof-of-concept open science platform for health research called CIELO [11]. Our primary goal in developing CIELO was to explore real-world information needs and end-user expectations for health research, a domain in which data provenance, privacy, security, and stewardship are of utmost importance. In pursuit of these goals, CIELO was designed and implemented based upon a set of conceptual models and functional requirements informed by

systematic and rigorous user needs assessments involving representatives from the academic, private, and public sectors.

During the course of the aforementioned user-centered design process, we elucidated a number of important and high-level requirements for a research data and analytics commons. The essential requirements are as follows:

First, *a successful data and analytics commons must be able to interoperate with and leverage a variety of technologies and approaches*. There are an increasing number of technologies that can be used to enable open science, such as content management systems and standard data-centric APIs (application programming interfaces). To be successful, a commons must be able to interoperate with such technologies in a scalable and user-friendly manner.

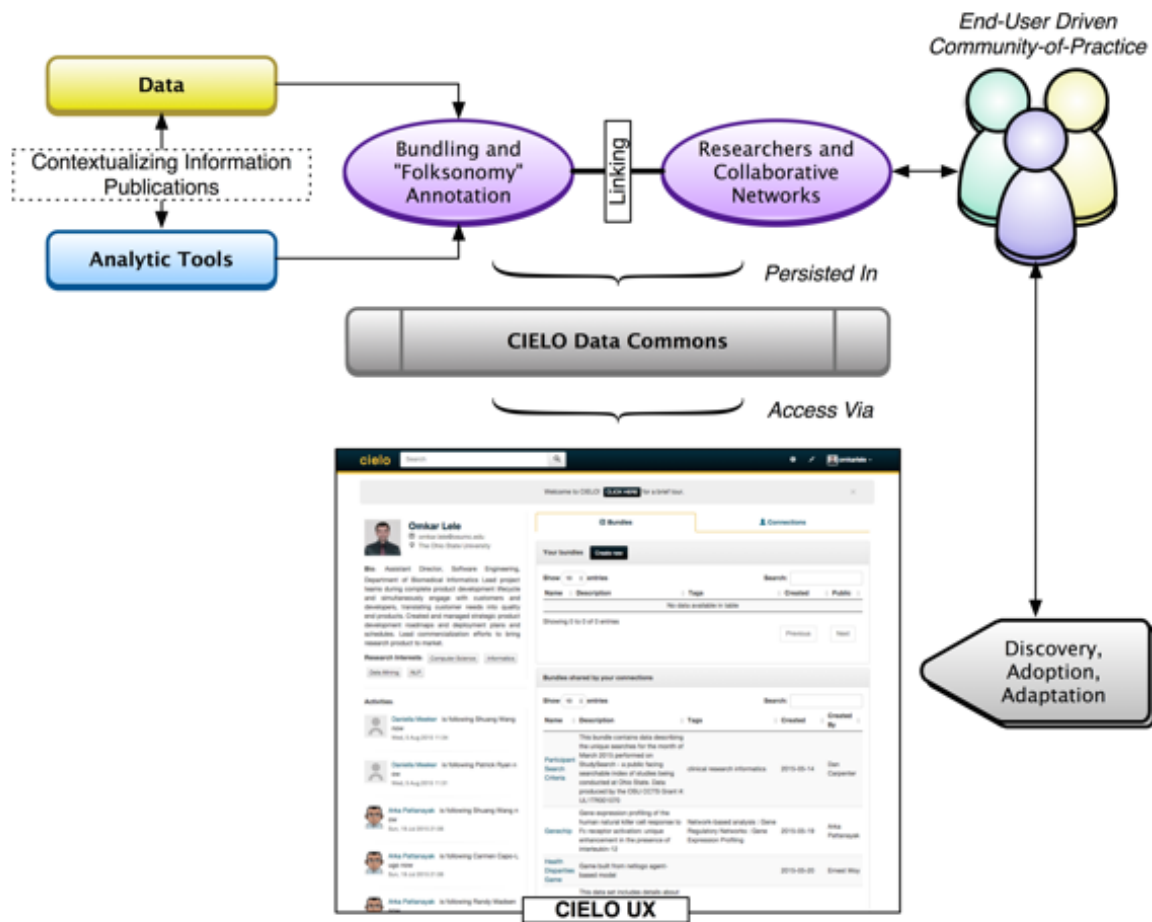
Second, *the "app store" paradigm reflects a user experience (UX) paradigm that is comfortable and desirable for both technical and nontechnical users and can create an effective marketplace for sharing ideas across disciplines*. There exists a similarly promising body of "app store" constructs for the user-friendly submission, quality assurance, distribution, and community-wide documentation of technical artifacts, all of which can be leveraged to build an effective exchange.

Third, *social search and discovery is a critical feature to promote interaction with data and analytical tools in a data and analytics commons*. The need for social search and discovery is reflective of the primary foci of many potential commons users who seek to engage in collaborative data and analytics projects with a group of trusted and known colleagues.

In response to these preliminary user needs, CIELO was developed to provide the members of the health care research communities with *a fully functional platform and dynamic community-of-practice designed to collectively reduce time and cost of research while enhancing the reproducibility, transparency, and rigor of health research*. To achieve these aims, we implemented CIELO using a combination of the following three key features: (1) a content and version management system (such as GitHub); (2) a "folksonomy"-driven annotation and search mechanism; and (3) a simplified user experience leveraging prevailing Web application technologies. All software design and implementation activities associated with CIELO used an agile and user-centered design and evaluation process, with a specific emphasis on end-user engagement in all project phases.

The resulting platform enabled the users to create analytic "bundles" (comprising both data and analytical code) to show and share their work (see Figure 1). As an early proof-of-concept to provide user feedback and demonstrate the potential impact of CIELO, we undertook a public beta release program. At the time of this submission, nearly 90 registered users from more than 20 different institutions had used CIELO.

Figure 1. Overview of the Collaborative Informatics Environment for Learning on Health Outcomes (CIELO) architecture and workflow, emphasizing the bundling of data and analytic tools and the provisions of social search capabilities.



Results

By implementing CIELO and engaging a variety of potential users in its public beta testing, we were able to elucidate a number of additional information needs and requirements based on using an open science platform focused on health in a “real world” setting, which are as follows:

- It is important to *allow users to bundle data and code in variable ways* (eg, mapping multiple versions of code to multiple versions of data, as opposed to a one-to-one mapping of such artifacts).
- There is a need to *support multi-level sharing permissions that can evolve gracefully over the lifecycle of a project or bundle* (from private collaborative or enclaves to fully open releases of data and code).
- *Flexible and dynamic metadata management functionality* can assist in responding to the ongoing evolution of standards and requirements.
- *Cross-linkage to external data and code resources where contribution to a centralized repository is not possible*, due diverse data and code stewardship, ownership, and technical requirements, is highly desirable.
- *Support for provisioning of durable resource identifiers, such as digital object identifiers (DOIs), can increase uptake and impact.* DOIs enable attribution of work and

create a value proposition for both the contribution and subsequent reuse, adaptation, and recontribution of data and analytics bundles, particularly for scholars.

Discussion

Our initial design and development experience over the course of the CIELO project has provided the basis for a vigorous dialogue between stakeholder community members regarding the capabilities that will add the greatest value to an open science platform for the health research community. Particularly because CIELO is designed to address the needs of multidisciplinary collaborators, we believe that CIELO project provides a successful technical prototype to facilitate collaboration in health research. We have also raised a number of important questions that will require further community dialogue and agreement, as follows: How do we incentivize and sustain *participation* in these types of platforms and sharing frameworks (for example, current funding and career advancement models and metrics of scholarly success may serve as a barrier to participation)?

How do we create a sustainable *fiscal* strategy that aligns with the evolving needs of a high performing healthcare research community and the ways in which it may utilize such a commons platform?

How can we make such a platform elastic and scalable from a technical standpoint so that it can evolve gracefully over time and not become obsolete? For example, in parallel to the development of CIELO, a robust community has also arisen around the Open Science Framework (OSF) [12], which we envision as providing a complementary platform for shared data analytics workflow management and sharing of such workflows and their products. It will be important for environments such as CIELO to interoperate with those like OSF in order to create a broad-based open system “ecosystem.”

Despite these open questions, we see CIELO as a proof-of-concept for what is required to establish a functional data and analytics commons reflecting the technical and sociocultural needs of our intended end-user community. Encouraged by the robust capabilities of the platform and early user experiences, we will continue to explore the potential of CIELO by (1) identifying opportunities to deliver reference

datasets within the environment that will make it even easier for individuals to share their analytics tools when source data sharing is infeasible; (2) creating incentive models to encourage the adoption and use of CIELO by a variety of stakeholders; (3) investigating novel methods to address diverse and challenging privacy and data-sharing constraints incumbent to health data in a systematic manner; and (4) continuing rigorous user-centered design processes to highlight additional functional requirements representative of end-user needs and expectations. Ultimately, we believe that projects such as CIELO represent an important effort to enable the health research community to achieve greater parity with other scientific communities, such as the natural and physical sciences, that have adopted open science paradigms and seen concomitant and exponential increases in research productivity as impact [5,6,13,14], setting the stage for achieving a compelling vision of open science and all of its concomitant benefits in the health research domain.

Acknowledgments

CIELO is a collaborative project of AcademyHealth’s EDM Forum, which is funded through a cooperative agreement from the Agency for Healthcare Research and Quality (Grant #U18 HS022789).

Conflicts of Interest

None declared.

References

1. Goodman SN, Fanelli D, Ioannidis JP. What does research reproducibility mean? *Sci Transl Med* 2016 Jun 01;8(341):341ps12. [doi: [10.1126/scitranslmed.aaf5027](https://doi.org/10.1126/scitranslmed.aaf5027)] [Medline: [27252173](https://pubmed.ncbi.nlm.nih.gov/27252173/)]
2. Iqbal SA, Wallach JD, Houry MJ, Schully SD, Ioannidis JP. Reproducible Research Practices and Transparency across the Biomedical Literature. *PLoS Biol* 2016 Jan;14(1):e1002333 [FREE Full text] [doi: [10.1371/journal.pbio.1002333](https://doi.org/10.1371/journal.pbio.1002333)] [Medline: [26726926](https://pubmed.ncbi.nlm.nih.gov/26726926/)]
3. Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, et al. Scientific standards. Promoting an open research culture. *Science* 2015 Jun 26;348(6242):1422-1425 [FREE Full text] [doi: [10.1126/science.aab2374](https://doi.org/10.1126/science.aab2374)] [Medline: [26113702](https://pubmed.ncbi.nlm.nih.gov/26113702/)]
4. Warren E. Strengthening research through data sharing. *N Engl J Med* 2016 Aug 04;375(5):401-403. [doi: [10.1056/NEJMp1607282](https://doi.org/10.1056/NEJMp1607282)] [Medline: [27518656](https://pubmed.ncbi.nlm.nih.gov/27518656/)]
5. Holve E. Open science and eGEMS: our role in supporting a culture of collaboration in learning health systems. *EGEMS (Wash DC)* 2016;4(1):1271 [FREE Full text] [doi: [10.13063/2327-9214.1271](https://doi.org/10.13063/2327-9214.1271)] [Medline: [27429993](https://pubmed.ncbi.nlm.nih.gov/27429993/)]
6. McKiernan EC, Bourne PE, Brown CT, Buck S, Kenall A, Lin J, et al. How open science helps researchers succeed. *Elife* 2016 Jul 07;5:16800 [FREE Full text] [doi: [10.7554/eLife.16800](https://doi.org/10.7554/eLife.16800)] [Medline: [27387362](https://pubmed.ncbi.nlm.nih.gov/27387362/)]
7. Moher D, Glasziou P, Chalmers I, Nasser M, Bossuyt PM, Korevaar DA, et al. Increasing value and reducing waste in biomedical research: who's listening? *Lancet* 2016 Apr 09;387(10027):1573-1586. [doi: [10.1016/S0140-6736\(15\)00307-4](https://doi.org/10.1016/S0140-6736(15)00307-4)] [Medline: [26423180](https://pubmed.ncbi.nlm.nih.gov/26423180/)]
8. Fosteropenscience. Open science taxonomy: FOSTER (facilitate open science training for European research) 2016 URL: <https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition> [accessed 2017-07-14] [WebCite Cache ID [6rxJcVaY7](https://www.webcitation.org/6rxJcVaY7)]
9. Longo DL, Drazen JM. Data sharing. *N Engl J Med* 2016 Jan 21;374(3):276-277. [doi: [10.1056/NEJMe1516564](https://doi.org/10.1056/NEJMe1516564)] [Medline: [26789876](https://pubmed.ncbi.nlm.nih.gov/26789876/)]
10. AcademyHealth. Edm-forum. 2015. Electronic data methods forum URL: <http://www.edm-forum.org/home> [accessed 2017-01-06] [WebCite Cache ID [6nJoJtv1b](https://www.webcitation.org/6nJoJtv1b)]
11. AcademyHealth. Edm-forum. 2016. CIELO: an open science environment for health analytics URL: <http://cielo.edm-forum.org/login/auth> [accessed 2017-01-06] [WebCite Cache ID [6nJohD4Am](https://www.webcitation.org/6nJohD4Am)]
12. OSF. 2011. Open science framework URL: <https://osf.io> [accessed 2017-01-06] [WebCite Cache ID [6nJp1k5Vr](https://www.webcitation.org/6nJp1k5Vr)]
13. Moed H. Arxiv. 2007. The effect of open access on citation impact: an analysis of ArXiv's condensed matter section URL: <https://arxiv.org/ftp/cs/papers/0611/0611060.pdf> [accessed 2017-07-20] [WebCite Cache ID [6s5tRWsI2](https://www.webcitation.org/6s5tRWsI2)]
14. Priem J. Arxiv. 2015. Altmetrics URL: <https://arxiv.org/ftp/arxiv/papers/1507/1507.01328.pdf> [accessed 2017-07-20] [WebCite Cache ID [6s5tWEvMX](https://www.webcitation.org/6s5tWEvMX)]

Abbreviations

API: application programming interface
CIELO: Collaborative Informatics Environment for Learning on Health Outcomes
DOI: digital object identifiers
EDM: Electronic Data Methods
OSF: Open Science Framework
UX: user experience

Edited by G Eysenbach; submitted 02.11.16; peer-reviewed by S Pais, J Till, J Apolinário-Hagen; comments to author 01.12.16; revised version received 24.05.17; accepted 10.06.17; published 31.07.17

Please cite as:

Payne P, Lele O, Johnson B, Holve E

Enabling Open Science for Health Research: Collaborative Informatics Environment for Learning on Health Outcomes (CIELO)

J Med Internet Res 2017;19(7):e276

URL: <http://www.jmir.org/2017/7/e276/>

doi: [10.2196/jmir.6937](https://doi.org/10.2196/jmir.6937)

PMID: [28760728](https://pubmed.ncbi.nlm.nih.gov/28760728/)

©Philip Payne, Omkar Lele, Beth Johnson, Erin Holve. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 31.07.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.