

## Open Science in Health Research: Lessons Learned from the CIELO Project

Philip R.O. Payne, PhD<sup>1</sup>; Omkar Lele, MS, MBA<sup>2</sup>; Beth Johnson, MPH<sup>3</sup>;  
Erin Holve, PhD, MPH, MPP<sup>4</sup>

<sup>1</sup>*Washington University in St. Louis School of Medicine, Institute for Informatics, St. Louis, MO*

<sup>2</sup>*The Ohio State University, Department of Biomedical Informatics, Columbus, OH*

<sup>3</sup>*AcademyHealth, Washington, DC*

<sup>4</sup>*Department of Health Care Finance, Government of the District of Columbia*

### Abstract

**Introduction:** There is an emergent and intensive dialogue in the United States concerned with 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 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 resulting in greater productivity and a reduction in redundant research investments. **Methods:** AcademyHealth's Electronic Data Methods Forum (EDMF) implemented a proof-of-concept open science platform for health research, known as CIELO (the Collaborative Informatics Environment for Learning on Health Outcomes). EDMF conducted a user-centered design process in order 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, EDMF 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. **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 which capabilities 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.

**Keywords:** Healthcare Research, Information Dissemination, Open Access to Information, Social Networking, Reproducibility of Results

## Introduction

There is an emergent and intensive national dialogue concerned with 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 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 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, 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 EDM Forum, called 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 in order to achieve a compelling vision of open science in healthcare research and all of its concomitant benefits.

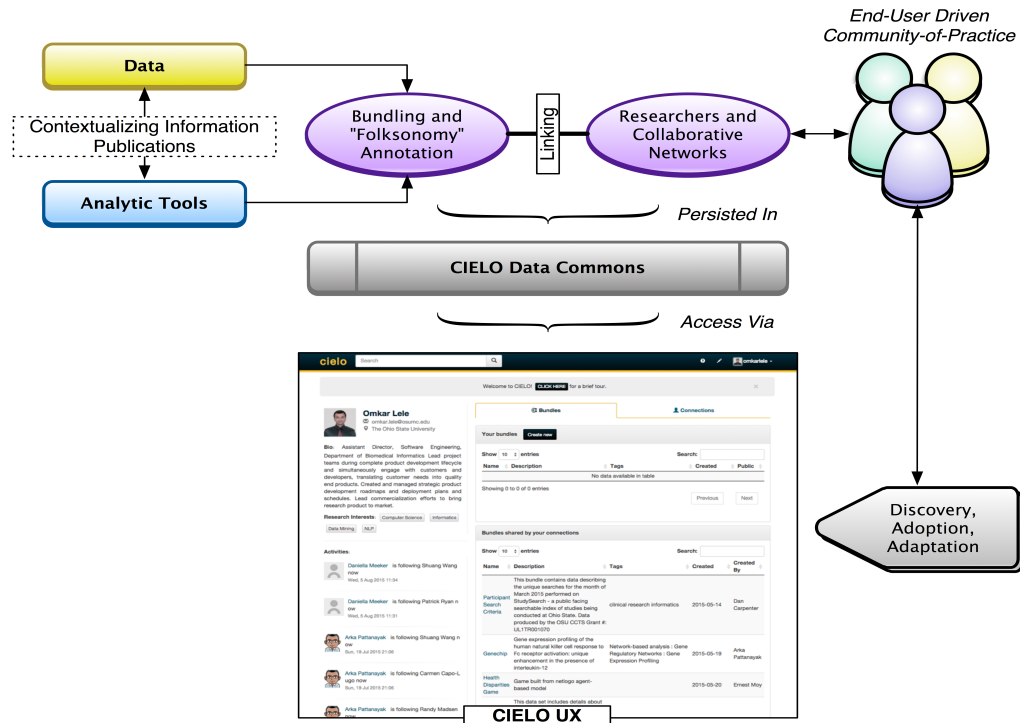
## Methods

As mentioned above, we have implemented a proof-of-concept open science platform for health research, known as CIELO (the Collaborative Informatics Environment for Learning on Health Outcomes). [11] Our primary goal in developing CIELO has been 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 has been 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, notably:

- *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.
- *The “app store” paradigm reflects a user experience (UX) paradigm that is comfortable and desirable for both technical and non-technical 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; and
- *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 in order to provide members of the healthcare 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 three key features: 1) a content and version management systems (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 have used an agile and user-centered design and evaluation process, with a specific emphasis on end-user engagement in all project phases.



**Figure 1:** Overview of CIELO architecture and workflow, emphasizing the bundling of data and analytic tools, and the provisions of social search capabilities.

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

## Results

By implementing CIELO and engaging a variety of potential users in its public beta testing, we have been 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, including:

- It is important to *allow users to bundle data and code in variable ways* (e.g., 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; and
- *Support for provisioning of durable resource identifiers, such as digital object identifiers (DOIs) can increase uptake and impact.* DOIs enable attribution of work and creates a value proposition for both the contribution and subsequent re-use, adaptation, and re-contribution 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 which capabilities 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 multi-disciplinary collaborators, we believe CIELO project provides a successful technical prototype to facilitate collaboration in health research. We have also surfaced a number of important questions that will require further community dialogue and agreement, including:

- 1) 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)?
- 2) 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?
- 3) 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 that reflects the technical and socio-cultural needs of our intended end-user community. Encouraged by the robust capabilities of the platform and early user experiences, we will continue to explore CIELO’s potential by: 1) identifying opportunities to deliver reference data sets 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 surface additional functional requirements representative of end-user needs and expectations. Ultimately, we believe that projects like 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.

### Acknowledgements

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

### Abbreviations

EDMF: AcademyHealth Electronic Data Methods Forum

CIELO: Collaborative Informatics Environment for Learning on Health Outcomes

API: Application Programming Interface

UX: User Experience

DOI: Digital Object identifiers

OSF: Open Science Framework

## References

1. Goodman SN, Fanelli D, Ioannidis JP. What does research reproducibility mean? *Science translational medicine*. 2016;8(341):341ps12-ps12. PMID: 27252173.
2. Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JP. Reproducible research practices and transparency across the biomedical literature. *PLoS Biol*. 2016;14(1): e1002333. PMID: 26726926.
3. Nosek BA, Alter G, Banks G, Borsboom D, Bowman S, Breckler SJ, ... Yarkoni, T. Promoting an open research culture. *Science*. 2015;348(6242):1422-5. PMCID: PMC4550299.
4. Warren E. Strengthening Research through Data Sharing. *New England Journal of Medicine*. 2016;375(5):401-3.
5. Holve E. Open Science and eGEMs: Our Role in Supporting a Culture of Collaboration in Learning Health Systems. *eGEMs*. 2016;4(1). PMID: 27429993.
6. McKiernan EC, Bourne PE, Brown CT, Buck S, Kenall A, Lin J, ... Yarkoni, T. How open science helps researchers succeed. *Elife*. 2016;5:e16800.
7. Moher D, Glasziou P, Chalmers I, Nasser M, Bossuyt PM, Korevaar DA, ... Boutron, I. Increasing value and reducing waste in biomedical research: who's listening? *The Lancet*. 2016;387(10027):1573-86.
8. FOSTER. Open Science Taxonomy: FOSTER (Facilitate Open Science Training for European Research) 2016 [cited 2016 September 14]. Available from: <https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>.
9. Longo DL, Drazen JM. Data sharing. *New England Journal of Medicine*. 2016;374(3):276-7. PMID: 26789876.
10. AcademyHealth. (2015). Electronic Data Methods Forum. Retrieved from <http://www.edm-forum.org>. Archived by WebCite® at <http://www.webcitation.org/6nJoJtv1b>.
11. AcademyHealth. (2016). CIELO: an open science environment for health analytics. Retrieved from <http://www.edm-forum.org/cielo>. Archived by WebCite® at <http://www.webcitation.org/6nJohD4Am>.
12. Center for Open Science. 2011. Open Science Framework. Retrieved from <https://osf.io>. Archived by WebCite® at <http://www.webcitation.org/6nJp1k5Vr>.
13. Moed HF. The effect of open access on citation impact: An analysis of ArXiv's condensed matter section. *Journal of the Association for Information Science and Technology*. 2007; 58(13):2047-54.
14. Priem J. Altmetrics (Chapter from Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact). arXiv preprint arXiv:150701328. 2015.