

6-27-2016

# Open Science and eGEMs: Our Role in Supporting a Culture of Collaboration in Learning Health Systems

Erin Holve

*AcademyHealth*, erin.holve@academyhealth.org

Follow this and additional works at: <http://repository.edm-forum.org/egems>

---

## Recommended Citation

Holve, Erin (2016) "Open Science and eGEMs: Our Role in Supporting a Culture of Collaboration in Learning Health Systems,"  
*eGEMs (Generating Evidence & Methods to improve patient outcomes)*: Vol. 4: Iss. 1, Article 12.

DOI: <http://dx.doi.org/10.13063/2327-9214.1271>

Available at: <http://repository.edm-forum.org/egems/vol4/iss1/12>

This Learning Health System Commentary/Editorial is brought to you for free and open access by the the Publish at EDM Forum Community. It has been peer-reviewed and accepted for publication in eGEMs (Generating Evidence & Methods to improve patient outcomes).

The Electronic Data Methods (EDM) Forum is supported by the Agency for Healthcare Research and Quality (AHRQ), Grant 1U18HS022789-01. eGEMs publications do not reflect the official views of AHRQ or the United States Department of Health and Human Services.

---

# Open Science and eGEMs: Our Role in Supporting a Culture of Collaboration in Learning Health Systems

## Abstract

“Open science” includes a variety of approaches to facilitate greater access to data and the information produced by processes of scientific inquiry. Recently, the health sciences community has been grappling with the issue of potential pathways and models to achieve the goals of open science—namely, to create and rapidly share reproducible health research. *eGEMs*’ continued dedication to and milestones regarding the publication of innovative, useful, and timely research to help contribute to the push towards open science is discussed, as well as the EDM Forum’s new data sharing platform, CIELO. Although strides have been made, there is still more work to be done to help health sciences community truly embrace open science.

## Acknowledgements

My great thanks to Samantha Weiss for her research and editorial assistance, and Puneet Kishor for his comments and extensive knowledge of open science and all things open.

## Keywords

learning health system, open science, research

## Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](https://creativecommons.org/licenses/by-nc-nd/3.0/).



# Open Science and eGEMs: Our Role in Supporting a Culture of Collaboration in Learning Health Systems

Erin Holve, PhD, MPH, MPP<sup>i</sup>

## ABSTRACT

“Open science” includes a variety of approaches to facilitate greater access to data and the information produced by processes of scientific inquiry. Recently, the health sciences community has been grappling with the issue of potential pathways and models to achieve the goals of open science—namely, to create and rapidly share reproducible health research. eGEMs’ continued dedication to and milestones regarding the publication of innovative, useful, and timely research to help contribute to the push towards open science is discussed, as well as the EDM Forum’s new data sharing platform, CIELO. Although strides have been made, there is still more work to be done to help health sciences community truly embrace open science.

<sup>i</sup>AcademyHealth

## A New Model for Open Health Science

Over the past few months, the health sciences community has engaged in a fierce debate regarding pathways to achieve the goals of open science.<sup>1,2,3</sup> There is a significant divide between the potential for the rapid, reproducible health research that is critical to advancing health science, on the one hand, and the current infrastructure and incentives in place that are designed to keep information private and unavailable to others who may benefit from using health data, on the other. All the communities we work with as part of the Electronic Data Methods (EDM) Forum—health services research, clinical research informatics, medicine, behavioral health, and numerous others—are grappling with these issues to better assess the feasibility, cost, and ultimate benefits of a substantial culture change toward greater “openness.” Indeed, as the President’s Cancer Moonshot demonstrates,<sup>4</sup> there is a new imperative to accelerate collaboration in implementation science, learning health systems, and precision medicine.<sup>5</sup>

Writ large, “open science” includes a suite of approaches to facilitate greater access to both data and the information produced by a process of scientific inquiry (Fig. 1).<sup>6,7</sup> Open access (OA) publishing is the most well-known of these concepts. As currently defined, OA publishing is designed to enable the free access, use, modification, and sharing of published research for any purpose.<sup>8</sup> Increasingly in physics and astronomy,<sup>9</sup> and more recently in biological health, open data and open source code<sup>10,11</sup> accompany OA manuscripts to accelerate the scientific process by contributing research materials and findings in a computable format.

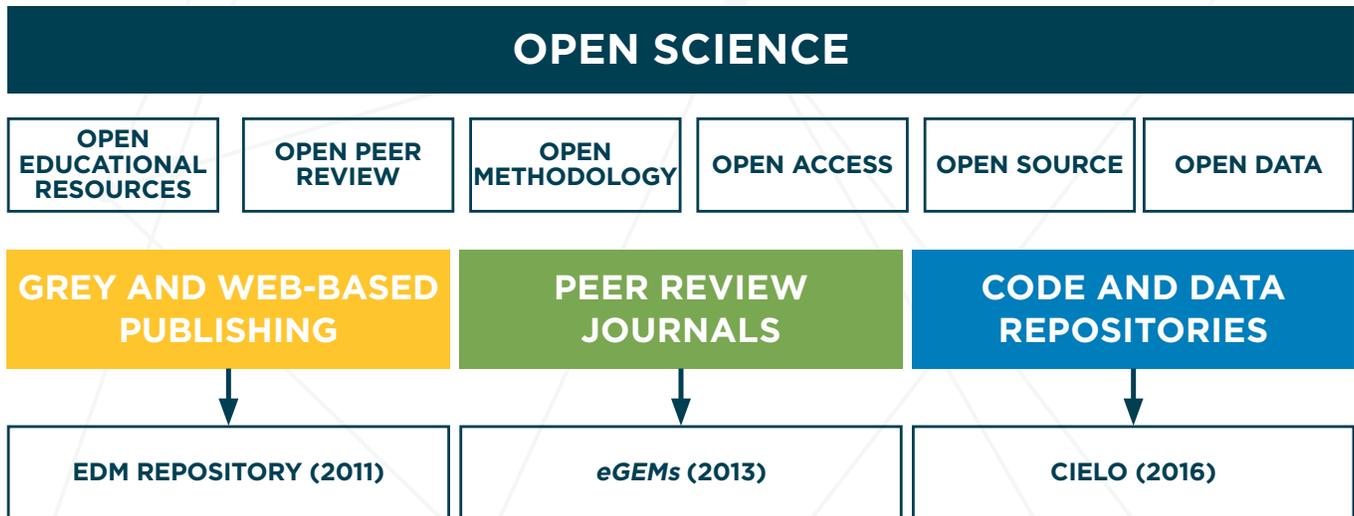
The goal of these open scientific efforts is to facilitate more rapid sharing and, therefore, learning across the field, and to promote reproducibility of programs committed to research, discovery,

and continuous improvement. Certainly, there are legitimate concerns about the need to ensure adequate security and privacy of individual, protected health information (PHI) in an era of open science. However, there are also substantial prevailing interests arguing against open sharing that do not necessarily represent the public’s interest. These include the potential value of intellectual property to investigators who withhold data sets in order to preserve the ability to conduct new analyses over time, and even concerns that more reproducible research enables competing researchers to identify flaws in the original work if the original data is reanalyzed. These issues can be heightened in the clinical health domain given its close relationship with a trillion dollar marketplace for health services that is increasingly dependent on data and analytic tools. Both examples highlight incentives *not* to share and *not* to collaborate that, if not entirely wholesome, are nonetheless entirely human.

There are many reasons to believe OA and open science are crucial to enabling a national learning health system. As demonstrated by arXiv in the physics, mathematics, engineering, and astrophysics communities,<sup>12</sup> a key value proposition for open science is the ability to spread promising scientific approaches rapidly in a more open and transparent environment—both to accelerate innovation and to reduce redundancy. Note, however, that as we move beyond the current period in which unprecedented support has been provided for health IT infrastructure and research, it is imperative that we consider ways to provide open source tools that can reduce disparities between wealthier systems and regions on the one hand and those that may be more resource constrained, including safety net and public health systems, on the other. Open science offers tools to minimize the digital and analytic divide between disparate systems while promoting continuous learning and improvement.



Figure 1. The Open Science Continuum and EDM Forum Resources



### The EDM Forum's Journey from Open Access (OA) to Open Science

The result of these debates over OA can be extremely polarizing because open science is, at its core, a philosophy. While formal requirements to instantiate open science principles as part of the scientific workflow may be on their way, for now we're relying on the goodwill of data liberators<sup>13</sup> and individual actors to demonstrate the benefits of openness. Those in health care who more freely share data and code believe that the public good resulting from sharing their work is of greater value than the "rent" they may be able to derive from their work at a later date—and this can feel like a big leap of faith. In applied health research, findings and approaches to improving health care delivery are sufficiently proximal to commercial health care in that they are perceived to have more commercial value. This has historically led to a phenomenon wherein the performance of new predictive algorithms and other tools are published, while the full code or a reference data set needed to replicate the findings are not.

With support from the Agency for Healthcare Research and Quality, the EDM Forum's natural experiment to facilitate a culture of open science started in 2010, with the creation of a repository to facilitate contributions from the community, including an opportunity to crowdsource comments and review. Fairly quickly however, we perceived a need for an independent OA journal that would ensure good peer review for cutting edge work pertaining to electronic health data. As a result, the EDM Forum launched Generating Evidence & Methods to improve patient outcomes (*eGEMs*) in 2013. In an early editorial, I commented on our philosophy for the journal,<sup>14</sup> recognizing that *eGEMs* was an experiment in understanding the culture of OA publishing and the willingness to embrace manuscripts more focused on methods, data, and the implementation issues affecting our ability to build learning health systems.<sup>15,16</sup>

*eGEMs* has kept to this commitment. Our inaugural efforts published articles discussing the funding from the American Recovery and Reinvestment Act (ARRA)<sup>17</sup> and the Health Information Technology for Economic and Clinical Health (HITECH) Act.

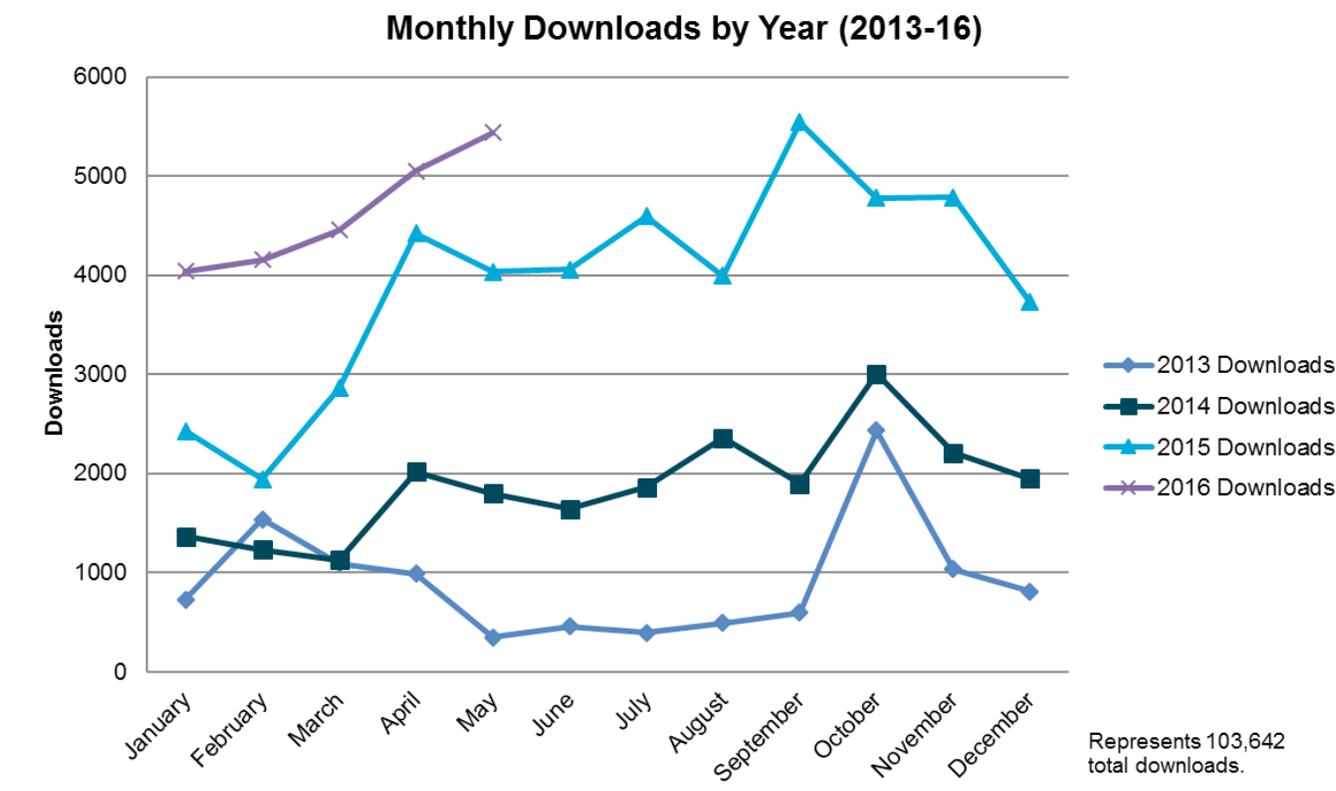
Subsequently, an infusion of support from the Affordable Care Act for patient-centered outcomes research supported by the Patient-Centered Outcome Research Institute (PCORI)<sup>18</sup> and the Center for Medicare & Medicaid Innovation (CMMI) has led to new insights into data, methods, and implementation science. New opportunities to reinvest in paying for quality care via the Medicare Access and CHIP Reauthorization Act (MACRA)<sup>19</sup> and the new regulation implementing payment reform, the Merit-Based Incentive Payment System (MIPS),<sup>20</sup> will no doubt generate further opportunities for learning in practice.

I am pleased to report that as of June 1, 2016, *eGEMs* has published its 139<sup>th</sup> article, and has passed a personal milestone for readership of 100,000 paper downloads. A great boon to reaching this goal

has been the journal's inclusion in PubMedCentral (PMC), which ensures *eGEMs* is fully searchable via keywords in PubMed<sup>21</sup> and makes the full text of all articles available in HTML. The number of citations of *eGEMs* papers is also growing rapidly in step with increasing readership.

While Figure 2 demonstrates the rapid rate of increase in downloads year after year, what is not observable is an important change we refer to as the "PubMed effect." Since March of 2015, when *eGEMs* papers were made available in PubMed, access to our manuscripts has undergone a rapid shift from accessing *eGEMs* papers on our own website ([www.egems.org](http://www.egems.org)) to downloads in PMC—as of May 2016 access to manuscripts via PMC now comprises 67 percent of all *eGEMs* downloads.

Figure 2. *eGEMs* Downloads by Month, 2013–2016





This transition is informative for our team and the field because it demonstrates the important public good provided by the National Library of Medicine's catalog as an international marketplace for health research that includes credible OA journals. It also demonstrates the extraordinary power of semantic search to ensure that new ideas are found within the massive catalog available through PMC. Interestingly, many of the most highly downloaded papers from PMC were different from the most downloaded papers on the *eGEMs* website, highlighting that the two sources are accessed for different purposes and perhaps by different types of researchers. An ongoing effort for our team is to understand differences in search and citation patterns due to the portals used to access *eGEMs* papers—bibliometric and sociometric research of interest on its own.

Our global readership, according to statistics for our website (excluding downloads from PubMed), shows significant international readership (Figure 3). Since the journal's launch, roughly 60 percent of downloads can be traced to countries outside of North America, demonstrating a strong international readership.

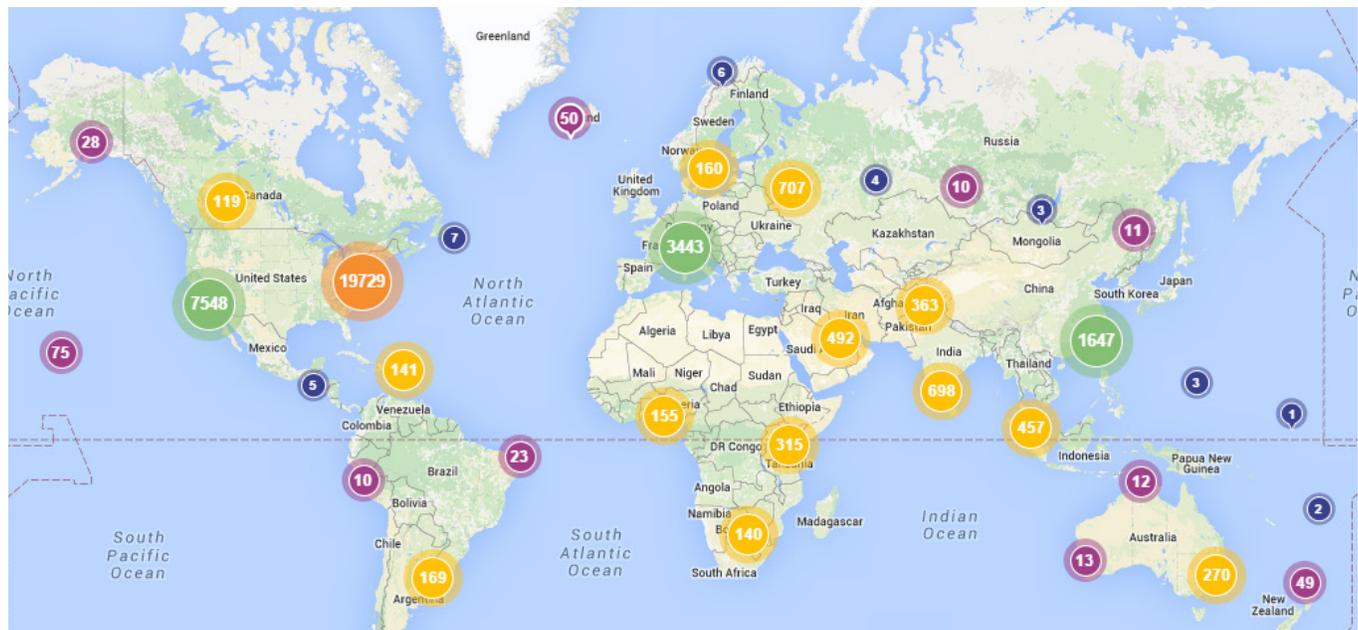
At the same time, we have struggled with many of the same issues as other journals, such as swift recruitment of peer reviewers and increasing the rapidity of the peer review process, and are continuously improving our timeliness while maintaining the quality of peer review. A key strategic decision arrived at with leadership from the *eGEMs* editorial board was to dedicate time to making sure we get things right. At this early stage in the development of science with electronic health data (EHD), it is most important to publish the most useful stories about the journey toward timely and relevant science at the nexus of precision medicine, delivery system science, and implementation science. And it isn't always a simple or quick process to tease out the most salient lessons from ongoing

efforts in this space. In what may be the world's greatest understatement—this evolving scientific paradigm is complex!

One critical element of developing a culture of greater experimentation is to promote transparency of the work in an open commons of data, code, and scientific processes. Notable examples of communities that have demonstrated effective use of platforms to show and share data include GitHub for software engineering and an increasing number of disciplines, and the Open Science Framework, which supports many projects in biomedicine and other fields. To achieve a similar success in our collaborations with learning health systems interested in improving health and health care, our team has been working with colleagues at the Ohio State University to build the third pillar of open resources needed to support a culture of open science—in a commons for sharing data and code in applied health science.

Enter CIELO ([www.edm-forum.org/cielo](http://www.edm-forum.org/cielo)). CIELO formally stands for “Collaborative Informatics Environment for Learning about Outcomes.” However, the most salient concept is the translation of the acronym as the Spanish word for “sky,” since the type of sharing we envision is a blue sky concept. CIELO is designed to include the version control framework, Git, developed by Linus Torvalds.<sup>22</sup> However, unlike Github, which does not enable efficient semantic search for work in key areas of health science, CIELO enables the community to develop new tags (a “folksonomy”) that will evolve to create the most salient tags or taxonomy. The goal is for CIELO, over time, to develop a robust knowledge-management structure that has strong links to *eGEMs* keywords and PubMed Medical Subject Headings (MeSH) so that items can be found easily and unambiguously cited using digital object identifiers (DOIs). If these goals can be

Figure 3. eGEMs' International Readership, 2013–2016



achieved we'll be pleased to contribute this test bed to the community to facilitate a culture of more rapid adoption and adaptation of new tools techniques.

### Significant Progress—and Remaining Sociotechnical Hurdles

If health care is to join the ranks of other fields that have embraced open science, we must address several issues that we already know are critical to preserving the public's trust in scientific publishing:

- *The ability to unambiguously cite all published scientific material.*  
Ideally, this will occur at the levels of both the object—such as DOIs, as well as that for individual researchers—such as Open Researcher and Contributor ID (ORCID).<sup>23</sup>
- *Infrastructure and standards to support metadata (data about the data).*  
This is to ensure unambiguous interpretation at the level of individual data elements, which is key to executing programs in new environments.

- *Seamless links between formal peer-reviewed publications and the materials that support the research.*

Executable papers are perhaps the most advanced concept in this arena<sup>24</sup> and propose a promising direction for our work to “show not tell.”

- *Rigorous quality control of peer review.*  
Increasingly, standards for reviewers are being developed,<sup>25</sup> in addition to changes in the academic marketplace that promote and reward higher quality reviews.
- *Review for plagiarism.*  
This is to ensure that standards are met, with respect both to the uncredited use of others' work, and to self-plagiarism.
- *Credentialing and strong management of journals to ensure predatory practices do not prevail.*

Resources such as Retraction Watch<sup>26</sup> and Beall's list of predatory journals for 2016<sup>27</sup> are examples of efforts to promote credibility and best business practices among scientific publishers. eGEMs has



readily participated in a review of our editorial practices to demonstrate and certify that we manage our journal with the highest standards of scientific credibility and openness, even as these standards evolve to be increasingly rigorous.<sup>28</sup>

Having enumerated this substantial wish list for open science in health, I should acknowledge that the person-hours to achieve these aims are significant and will require strong community engagement and support. We must define a strong value proposition for researchers and other analysts who produce ideas and evidence as their livelihood. Likewise, the costs of producing and maintaining repositories of code or data are not well appreciated, and must be understood in order to achieve a rational and sustainable approach. Even with CIELO, we acknowledge it will be an ongoing effort to achieve our idealized components for a health repository, or “repo,”<sup>29</sup> yet I am optimistic we will resolve these issues by working closely with the scientific communities engaged in the EDM Forum.

Health researchers have adapted to great change in the past five years, and we are just now beginning to get our bearings on ways to make the most effective use of the enabling health IT infrastructure at our disposal. I anticipate tremendous work from *eGEMs* authors and readers in the years to come and am pleased to provide resources to help promote the journey toward open science in health research.

## Acknowledgements

My great thanks to Samantha Weiss for her research and editorial assistance, and Puneet Kishor for his comments and extensive knowledge of open science and all things open.

## References

1. Longo, D and Jeffrey M. Drazen, M.D., “Data Sharing,” Dan L. Longo, M.D., and Jeffrey M. Drazen, M.D. *N Engl J Med* 2016; 374:276-277. January 21, 2016 DOI: 10.1056/NEJMe1516564
2. Drazen, J, “Data Sharing and the *Journal*” *N Engl J Med* 2016; 374:e24 May 12, 2016 DOI: 10.1056/NEJMe1601087
3. Comments on ICMJE Proposals for Sharing Clinical Trial Data. <https://forms.acponline.org/icmje-comments> Accessed June 15, 2016.
4. Cancer Moonshot Fact Sheet <https://www.whitehouse.gov/the-press-office/2016/02/01/fact-sheet-investing-national-cancer-moonshot>. Accessed June 15, 2016
5. *Chambers* DA, WG Feero WG, MJ Khoury. “Convergence of Implementation Science, *Precision Medicine*, and the Learning Health Care System: A New Model for Biomedical Research. *JAMA*. 2016 May 10;315(18):1941-2. doi: 10.1001/jama.2016.3867.
6. Amsen, E. “What is open science?” *F1000 Research*. 11 Nov 2014. <http://blog.f1000research.com/2014/11/11/what-is-open-science/>. Accessed 9 Jun 2015.
7. Gezelter, D. “What, exactly, is open science?” *The Open Science Project*. 28 Jul 2009. <http://www.openscience.org/blog/?p=269>. Accessed 9 Jun 2015.
8. The Open Definition. <http://opendefinition.org/od/2.1/en/>. Accessed June 15, 2016
9. Djorgovski, SG, Mahabal AA, Crichton DJ, Chaudhry B (2015). From Stars to Patients: Lessons from Space Science for the Health Care Informatics. Poster presented at IEEE International Conference on Big Data, Santa Clara, CA. [http://repository.edm-forum.org/edm\\_briefs/19](http://repository.edm-forum.org/edm_briefs/19)
10. Sage BioNetworks. <http://sagebase.org/>
11. The Sage Bionetworks Governance project. [www.sagebase.org/governance](http://www.sagebase.org/governance). Accessed June 15, 2016
12. arXiv. <https://arxiv.org/> Accessed June 15, 2016.
13. Hactivist Fred Trotter receives fourth annual Health Data Liberator award. [http://www.eurekalert.org/pub\\_releases/2016-05/a-hft051016.php](http://www.eurekalert.org/pub_releases/2016-05/a-hft051016.php) Accessed June 17, 2016
14. Holve, Erin (2013) “*eGEMs*: Pragmatic Publishing to Build a Learning Health System,” *eGEMs (Generating Evidence & Methods to improve patient outcomes)*: Vol. 1: Iss. 1, Article 2. DOI: <http://dx.doi.org/10.13063/2327-9214.1001>
15. Simpson L. Learning How to Learn: How AcademyHealth is Supporting Evidence Generation in a Transforming World. *eGEMs (Generating Evidence & Methods to improve patient outcomes)* 2013 Jan 17;1(1). DOI: 10.13063/2327-9214.1002.
16. Wallace P. *eGEMs*: An Opportunity for Better Science. *eGEMs (Generating Evidence & Methods to improve patient outcomes)* 2013 Jan 17;1(1). <http://repository.academyhealth.org/egems/vol1/iss1/6/>
17. American Recovery and Reinvestment Act (ARRA) of 2009, Pub. L. No. 111-5, 123 Stat. 115, 516 (Feb. 19, 2009).

18. Patient Protection and Affordable Care Act (PPACA), Pub. L. No. 111-148 (Mar. 2010).
19. Medicare Access and CHIP Reauthorization Act of 2015. Public Law No: 114-10 (04/16/2015)
20. Medicare Program; Merit-Based Incentive Payment System (MIPS) and Alternative Payment Model (APM) Incentive under the Physician Fee Schedule, and Criteria for Physician-Focused Payment Models (CMS-5517-P). <https://federalregister.gov/a/2016-10032>. Accessed June 15, 2016.
21. National Center for Biomedical Informatics, the National Institutes of Health National Library of Medicine, <http://www.ncbi.nlm.nih.gov/pmc/journals/2686/> Accessed June 15, 2016
22. Scopatz, Anthony; Huff, Kathryn D. (2015). *Effective Computation in Physics*. O'Reilly Media, Inc. p. 351. ISBN 9781491901595.
23. ORCID. <http://orcid.org/> Accessed June 15, 2016
24. SHARE: A Web Portal for Creating and Sharing Executable Research. Elsevier. <https://www.elsevier.com/physical-sciences/computer-science/share-a-web-portal-for-creating-and-sharing-executable-research>. Accessed June 15, 2016.
25. Suggested Guidelines for HSR Methods Reviews. April 2010 <http://www.hsrmethods.org/About/-/media/Files/PDFs/HSRMetSugGuide.ashx>. Accessed June 15, 2016.
26. Retraction Watch. <http://retractionwatch.com/> Accessed June 17, 2016.
27. Beall's List of Predatory Open Access Publishers. <https://scholarlyoa.com/individual-journals/> Accessed June 15, 2016.
28. op.cit.
29. Perfect Health Reop Desiderata. <http://punkish.org/Perfect-Health-Repo-Desiderata>. Accessed June 15, 2016.