A Review of Scholarly Communication Research in Open Science

Authors: Liping Gu

Date: 2016-06-08T00:00:00+00:00

Abstract

This article provides an overview of the connotation, structure, and driving factors of open science, explains that the demands for new-generation scientific and technological information exchange constitute the primary driving force for open science development, analyzes the development conditions for open science computational engineering, and proposes several ideas for supporting services for academic exchange.

Full Text

Academic Communication in Open Science: A Research Review

Author: Associate Researcher and Deputy Director, Science and Technology Information Research and Consulting Center, National Science Library, Chinese Academy of Sciences, gulp@mail.las.ac.cn

Abstract: This paper briefly introduces the meaning, structure, and driving forces of open science, explains that new demands in scientific information exchange constitute the primary motivation for open science development, analyzes the development conditions for open science computational engineering, and proposes several ideas for supporting academic communication services.

Keywords: grid, cloud computing, social networks, public joint resources, academic communication needs, science and technology information policy, Web2.0, eScience

Funding: This work was supported by the Directed Talent Research Fund of the National Science Library, Chinese Academy of Sciences, for the project "Science and Technology Information Policy Research and Consulting" (Project No. Guan 1203).

Received: July 11, 2012
Published: February 28, 2013

Open science represents a novel form of scientific communication. As science and technology advance at an unprecedented pace and grow increasingly complex, both large-scale global collaborative research projects (in the "big science" sense) and independent creative work (in the "small science" sense) must understand the implications of open science in this rapidly changing era.

1.1 What Is Open Science?

"Open science" refers to a new research paradigm where scientists share every element of their research process—not just final products—to foster collaboration among researchers and enable spontaneous new forms of virtual cooperation [1]. For instance, according to institutional repository deposit policies, research data accumulated during scientific investigations and final revised manuscripts of submitted papers are considered materials from the research process beyond final products (such as published journal articles or synthesized chemical compounds). Providing such materials constitutes one way to participate in open science, which encompasses a broader knowledge management mechanism.

Open science functions as a dynamic system for knowledge production. From a technological foresight perspective, the progress of future research depends on how current knowledge production, storage, utilization, and dissemination (in terms of openness and readability) shape and constrain the design of future scientific activities. From the economic principles of equity and efficiency, open science serves as a balancing mechanism between information disclosure and secrecy, optimizing efficiency through equilibrium. The benefits of information disclosure depend on subsequent researchers' use of portions of that disclosed information, making the costs of proprietary knowledge and individuals' relative disadvantages in knowledge discovery key factors in open science' s comparative advantages [2]. From a long-term strategic perspective in the sociology of knowledge, maintaining scientific openness provides higher-level social welfare. In certain public policy domains, open science offers benefits such as gaining public support, appropriately planning intellectual property rights, and strengthening scientific norms and institutional roles.

From a computer engineering perspective, open science applications include four key elements: (1) principles for publishing scientific data and metadata; (2) open-source and web-based tools for use, verification, and exploration of research; (3) cloud computing and distributed computing efficiency; and (4) infrastructure sharing [3]. These four foundational elements represent the current common scope and content of open science in practice, whose specific meanings and development conditions are discussed below.

1.2 Driving Forces and Structure of Open Science

The driving forces behind open science lie in how scientists decide whether to share information with colleagues, followed by the infrastructure and computational engineering of open science, with its development structure constrained by specific science and technology information policies.

Research indicates that social capital (i.e., anticipated reciprocity) correlates with the degree to which scientific communities adhere to open science information-sharing norms, while competitive interests within the scientific community exert a moderate influence, depending on which social networks and industry competition systems scientists belong to [4]. In other words, the institutional structures of different disciplines, scientific communities, and national regions also affect open science progress in those fields. For example, a survey of Danish life science researchers on the significance of patents for academic research and their views on open science revealed that scientists engaged in basic research (particularly those with relatively low productivity) are skeptical about patents' academic significance, while research grant recipients, scientists closely related to industry, and full professors with high productivity show less concern for open science [5]. While the concept of open science is sound, its implementation faces several issues worth further investigation. Particularly as a knowledge production dynamic system, open science' s driving force rests on the principle of anticipated reciprocity within the scientific community.

Another structural factor is market design, whose three fundamental principles are market thickness, congestion reduction, and market safety. Good market design may foster healthy development of innovative societies, while violating these principles may inhibit resource allocation efficiency. Some have questioned whether the normative value of disclosure in open science might undermine authors' ability to receive market feedback for their ideas [6]. Although crises in open science development have been noted, why is open science still necessary?

New Demands in Scientific Information Exchange

Academic exchange has always been a crucial component of scientific development, with a robust system promoting technological innovation and vice versa. Current and foreseeable future technological innovation will increasingly rely on academic information exchange as a foundational element. New demands in scientific information exchange have four characteristics: the need for large-scale computing resources, the need for global collaborative research, data-driven research, and the need for reuse of research results. These four demands drive scientific information exchange toward open science.

2.1 The Need for Large-Scale Computing Resources

Although personal computing tools continue to proliferate (e.g., fast cache memory, CPUs, high-capacity memory in laptops), scientific research data are growing at an even more massive scale. Trends indicate that managing large-scale complex data will become a major infrastructure issue for future scientific development, as such data may saturate users' computing resources, leaving them in difficult situations and stifling collaboration and sharing [7]. From a big science perspective, computing resources have even become part of national strategic resources, as important as strategic reserves of oil or grain.

Recent advances in computational science involve using petascale resources to increase the granularity of scientific observation and expand the scale of scientific achievements. These resources can address complex scientific problems such as materials science research, HIV species studies, and biomedical applications in cerebrovascular hemodynamics [8]. Put simply, such research requires not only verifiable paradigms but also the "good luck" that comes from continuous experimentation. Researchers must leverage large-scale computing resources provided by open science to increase the number of "dice rolls" and gain even the slightest opportunity for "good luck" (see *Does God Play Dice? A History of Quantum Physics* for further reading).

2.2 The Need for Global Collaborative Research

In marine science literature, academic resource aggregator EBSCO provides the scientific community with an open science guide—a directory of scientific journals offered free to institutions and scientists in developing countries [9]. However, this merely accommodates rather than truly meets the needs of global collaborative research projects. What scientists truly need from open science models is the ability to analyze data through advanced statistical methods and tools such as classification, association rule mining, and cluster analysis, to simplify technology for tagging output to other storage through the stability and relatively simple patterns of event metadata records, and to utilize existing data mining components from other communities [10].

This demand arises from research that must be conducted through global cooperation. For example, sound is a crucial factor in the survival activities of many marine organisms, and the impact of noise on marine life represents a major unknown variable in current marine science. Substantial evidence indicates that human-generated noise has become the primary component of ocean noise over the past few decades, particularly in increasingly industrialized regions. Achieving global ocean observation and predictive capabilities for important marine biological functions and relationships requires collection, observation, and modeling research, making scientific data exchange and its support mechanisms critical issues [11]. Similar to marine noise research, numerous other fields exist, but the key point is that open science is essential for achieving data sharing and collaborative computing.

2.3 The Need for Data-Driven Research

Data-driven research implies a series of needs and problems regarding the preservation, circulation, management, and curation of research data. In structural biology, for example, early development stages lacked systems for maintaining experimental data. However, as research questions become increasingly complex, the structural biology community increasingly believes that these inadequately preserved data hold important reference value for each macromolecular structure study [12]. Archiving and collecting data and releasing it at appropriate times are important measures in open science. Additionally, the storage, access, and management of complex data from widely used shared research facilities such as synchrotron beamlines represent new topics derived from open science [12]. Currently, people recognize that the best practice for data-driven research at the practical level is open science.

2.4 The Need for Full Utilization of Research Results

Research results are buried for many reasons, most regrettably when good research findings are buried by growing literature noise because they are not immediately transformed into commercial products. For over a decade, cheminformatics has made outstanding contributions to analytical chemistry, biochemical pharmacology, and drug discovery. However, its challenge lies in how to develop better products more quickly: open science models help make these judgments, and some open science tools can transform projects in the worst situations (lacking funding and utilization) into best-case scenarios (full integration from systems biology to virtual physiology) [13]. The key lies in chemical product development, which often requires choosing between mainstream (i.e., accepted, predictable projects) and novelty. Therefore, the full utilization of research results, or the reuse of research data, has become a strong driving force pulling open science forward.

New Tools Emerging from New Demands

Under these new demands in scientific information exchange, new tools have emerged. Open science grid technology, security and authorization mechanisms, cloud computing and smart grid applications, and data exchange packages with social networking tools represent various new research tools in this wave of scientific transformation, ranging from large-scale to small-scale solutions.

3.1 Open Science Grid Technology

The Open Science Grid (OSG) supports new science, new scientists, and new research methods, generating new community activities at various scales through online education, collaborative participation, and task distribution [14]. The open science grid is a large-scale distributed computing infrastructure that provides increasingly diverse services and software development for the scientific community, promotes teamwork among practitioners and researchers, and realizes the initial ideals of open science [15].

Grids consist of resources owned by dispersed institutions, necessitating policy-based scheduling technology for grid computing environments to handle heterogeneous resources for next-generation multimedia applications: (1) it must support scheduling with resource usage constraints; (2) it must provide optimized solutions for grid resource allocation problems; (3) it must account for the geographical distribution and heterogeneity of resources; and (4) it must be able to track resources and adjust workloads [16].

Through the open science grid, hundreds of grid interfaces can provide access to computing and storage resources via standards [17]. Consequently, countries have advanced network infrastructure construction, which from a research perspective can support virtual organizations in creating and utilizing knowledge more effectively [18].

3.2 Security and Authorization Mechanisms for Open Networks

To build complete grid infrastructure, resource providers must publish information about their potential resources and services, enabling users and virtual organizations to more easily judge resource selection and scheduling. However, such websites, which provide access including sensitive data, can easily expose users' private information and reveal important entity data. Due to these potential security vulnerabilities, data collected by the open science grid—including resource selection, monitoring, auditing, fault diagnosis, logging, and site verification—require a risk assessment and potential threat warning model to manage and protect sensitive data [19].

In open grids among thousands of users and hundreds of institutions, data storage locations are vulnerable to hacker attacks, making network security an institutional issue. Particularly, the management of large-scale scientific experiments and collaboration rules for scientists' social networks require jointly discussed and adhered-to systems. Through mixed-integer programming (MIP) models and cooperative rules (e.g., suspending cooperation or shutting down sites when necessary), threat levels can be minimized [20].

Security mechanisms are inseparable from authorization mechanisms. Initiated in 2006, interoperability authorization was designed to promote interoperability between the Open Science Grid (OSG) and the Enabling Grids for E-sciencE (EGEE) project grids, integrating and socially converging different software products through a common authorization protocol [21]. As a result, OSG and EGEE have a common security model based on public key infrastructure, granting grid resource access to users with a virtual organization (VO) identity rather than personal identity. To this end, people seek a joint project for authorization interoperability to define numerous definitions including a common communication protocol, authorization and attribute identities, and infrastructure and data integration mechanisms that can implement important grid middleware [22], thereby meeting open science' s data circulation needs while ensuring open network security.

3.3 Cloud Computing and Smart Grid Applications

Large-scale open science grids provide researchers with substantial computing resources that can restrict access based on requirements but may affect users' workflows. Therefore, in many cases, Infrastructure as a Service (IaaS) cloud computing reduces restrictions for individual users and communities. Due to flexibility in resource allocation, remote access, and environment configuration that can combine local resources with one or more IaaS providers based on demand, grid infrastructure management (such as Network Virtual Workstations (WOW)) becomes easier to grasp [23].

For example, the U.S. Deformable Elastic Network (DEN) integrates into the SBGrid scientific portal through a simple and intuitive web interface, providing access to smart grid services that can complete in hours what previously required thousands of hours of standard computing [24].

Smart grids achieve unified resource access through standard interfaces for resource gateways, granting permissions based on users' virtual organization (VO) membership in the Open Science Grid (OSG) [25]. These measures facilitate mobilizing computing resources to achieve optimal computational efficiency.

3.4 Data Exchange and Work Collaboration Supporting Open Science

As previously mentioned, scientific research and development have needs for reusing, re-mining, and reorganizing research data. Therefore, how to facilitate open data exchange and communication becomes a key issue at a smaller scale. Take the Talkoot toolkit as an example: this small program can systematically collect, tag, and share earth science data, becoming a web-based academic information exchange tool for sharing data, work processes, and research notes [1]. Although not involving massive funding and equipment investment like smart grids or cloud computing, it is an essential new tool in open science. On the other hand, beyond exchanging research results and experimental data, self-organizing research projects online are becoming a new form of scientific research and development. For instance, on the myExperiment social website, people can collaboratively discover, share, and curate scientific workflows and experimental plans, representing a model that supports the socialization of research workflows for researchers and communities [26]. This new attempt enables a new generation of researchers to recognize that open science makes research content dissemination more widespread and may shorten knowledge circulation cycles and discovery time.

New Problems Arising from New Tools

The key to technological innovation often lies not in scientific theory or engineering practice but in the previously mentioned changes in demand, community communication patterns, production conditions, and, most crucially, the metaphysical factors often overlooked and discussed below.

4.1 Ethical Issues in Open Science

Respect for fundamental bioethical principles (such as autonomy, privacy, beneficence, justice, and research integrity) must be balanced with technical tools for open scientific data sharing to effectively develop open science. Excessive concern for ethical safeguards at the expense of scientific development needs is unrealistic, yet emphasizing technical tools while ignoring scientific ethics will not gain social group support. Therefore, science and technology information policy must mediate and serve as a lever balancing both sides. For example, in stem cell research, publicly accessible SNP databases have become controversial due to new methods and tools enabling individual identification. Current technical safeguards for scientific ethics cannot fully protect donors' identity information. In response, the International Stem Cell Forum (ISCF) published a "Policy Statement on Publishing SNP Genotypes of Human Embryonic Stem Cell Lines," emphasizing ethics in related research [27]. This demonstrates that "open"science also requires a certain degree of self-regulation and common norms to develop robustly.

4.2 Policy Research Issues Supporting Open Science

The scientific community often has different understandings of open science, with ambiguous meanings regarding reproducibility, computing resources, scientific credit, data sharing, results sharing, peer review, and impact evaluation. This necessitates extensive opinion and motivation surveys and various forms of outreach to build consensus and reach agreements. These efforts rely on professional science and technology information policy research and consulting working groups to enable researchers to more easily access convenient academic information exchange services and better benefit from them [28]. This perspective views open science policy decision support from a pure knowledge services standpoint.

From a structural and institutional perspective, the contradiction between open science mechanisms and patent incentive systems requires mediation by science and technology policy. For example, within the biomedical research field' s institutional system, intellectual property (IP) systems extend patent protection to new areas, creating two risks: (1) high-cost research tools and few profitable results weaken knowledge dissemination; and (2) the destruction of open norms traditionally associated with scientific and technological progress [29]. A survey of French university professors showed that technology transfer from universities to industry delays publication of research results, thereby hindering scientific knowledge dissemination, particularly the commodification of knowledge in life sciences and medicine [30]. It can be said that publicly funded researchers currently operate within a hybrid system that forces them to oscillate between the sharing conventions of "open science" and the market rules of the "patent system." Coupled with the close interaction between latest IP practices and growing technology products, this profoundly affects research funding policies, making the situation more complex and difficult to stabilize. These are current policy issues constraining open science development.

Toward New Library and Information Services for the Future

Libraries are important organizations for describing, collecting, preserving, managing, and accessing scientific research outputs in innovative societies.

The scale of shared research resources of various types is increasing, making the practical models, obstacles and facilitators, and innovation drivers of new forms of academic communication (such as Web2.0 or other informal communication modes) worthy of further study [31]. For example, we can view emerging "social networks" as collective creation by millions of people on a collective knowledge system (including communities, blogs, forums, mailing lists, collaborative portals, and other online systems). These technologies supplement traditional knowledge-sharing mechanisms such as conference presentations, paper publications, and book publishing, and at lower cost. Increasing numbers of scientists and researchers are using these "open resource tools" as transformative methods for collective discussion and scientific output release, leveraging colleagues' expertise to accelerate scientific knowledge discovery [32]. In this environment, what challenges do library and information services face?

Open innovation and open resource tools represent new ideas emerging to support collaboration among different people, providing new opportunities for scientific research and education. Open Science 2.0 includes citizen science papers, public lectures, and collective intelligence creation, which not only increases individual research visibility but also encourages exchange between academic theory and practical work [33]. This traditionally informal information exchange mode is becoming mainstream, raising questions about how library and information institutions will face this transformation. A survey of UK e-Science projects showed that although policies require open resource sharing and a broad spirit of sharing exists in daily life, many uncertainties and unresolved issues remain, such as publication forms and methods and their rights management [34]. In the future, science and technology information policy will increasingly demonstrate its importance in open science development. How to deeply understand technology, how to use technology to solve policy problems, and how to use policy to guide the healthy development of open science are all new challenges of this era. Librarians must reconsider how to innovate practical work in line with the spirit of open science on the basis of existing services.

References

[1] Ramachandran R, Maskey M, Kulkarni A, et al. Talkoot: Software tool to create collaboratories for earth science[J]. Earth Science Informatics, 2012, 5(1):33-41.

[2] Mukherjee A, Stern S. Disclosure or secrecy? The dynamics of open science[J]. International Journal of Industrial Organization, 2009, 27(3):449-462.

[3] Kauppinen T, de Espindola G M. Linked open science-communicating, sharing and evaluating data, methods and results for executable papers[J]. Procedia Computer Science, 2011(4):726-731.

[4] Haeussler C. Information-sharing in academia and the industry: A comparative study[J]. Research Policy, 2011, 40(1):105-122.

[5] Davis L, Larsen M T, Lotz P. Scientists' perspectives concerning the effects of university patenting on the conduct of academic research in the life sciences[J]. Journal of Technology Transfer, 2011, 36(1):14-37.

[6] Gans J S, Stern S. Is there a market for ideas?[J]. Industrial and Corporate Change, 2010, 19(3):805-837.

[7] Miller S D, Herwig K W, Ren S, et al. Data management and its role in delivering science at DOE BES user facilities—past, present, and future[EB/OL].[2012-11-04].http://iopscience.iop.org/1742-6596/180/1/012049/pdf/1742-6596_180_1_012049.pdf.

[8] Saksena R S, Boghosian B, Fazendeiro L, et al. Real science at the petascale[EB/OL].[2012-10-27]. http://rsta.royalsocietypublishing.org/content/367/1897/2557.full.

[9] Stokes-Rees I, Levesque I, Murphy F V, et al. Adapting federated cyber infrastructure for shared data collection facilities in structural biology[EB/OL].[2012-09-22].http://www.ncbi.nlm.nih.gov/pubmed/22514186.

[10] van Gemmeren P, Malon D. Event metadata records as a testbed for scalable data mining[EB/OL].[2012-10-22]. http://iopscience.iop.org/1742-6596/219/4/042057/pdf/1742-6596_219_4_042057.pdf.

[11] Boyd I L, Frisk G, Urban E, et al. An International Quiet Ocean Experiment[J]. Oceanography, 2011, 24(2):174-181.

[12] Goovaerts M. Oceandocs and open science directory: Two facets of the information policy of UNESCO/IOC-IODE[EB/OL].[2012-11-04].https://darchive.mblwhoilibrary.org/bitstream/handle/1912/3777/Goovaerts_iamslic2009.pdf?sequence=1

[13] Oprea T I, Taboureau O, Bologa C G. Of possible cheminformatics futures[J]. Journal of Computer-Aided Molecular Design, 2012, 26(1):107-112.

[14] Pordes R, Altunay M, Avery P, et al. New science on the open science grid[EB/OL].[2012-10-27]. http://osg-docdb.opensciencegrid.org/0007/000761/001/NewScienceontheOpenScie

[15] Altunay M, Avery P, Blackburn K, et al. A science driven production cyber infrastructure—The open science grid[J]. Journal of Grid Computing, 2011, 9(2):201-218.

[16] In J U, Lee S, Rho S, et al. Policy-based scheduling and resource allocation for multimedia communication on grid computing environment[J]. IEEE Systems Journal, 2011, 5(4):451-459.

[17] Garzoglio G, Levsbina T, Mhashilkar P, et al. ReSS: A resource selection service for the open science grid[EB/OL].[2012-08-16].http://lss.fnal.gov/archive/2008/pub/fermilab-pub-08-016-cd.pdf.

[18] Zou Guangyu, Yilmaz L. Dynamics of knowledge creation in global participatory science communities: Open innovation communities from a network perspective[J]. Computational and Mathematical Organization Theory, 2011, 17(1):35-58.

[19] Cholia S, Porter R J. Publication and protection of sensitive information in a grid infrastructure[EB/OL].[2012-10-27]. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=453427

[20] Altunay M, Leyffer S, Linderoth J T, et al. Optimal response to attacks on the open science grid[J]. Computer Networks, 2011, 55(1):61-73.

[21] Garzoglio G, Bester J, Chadwick K, et al. Adoption of a SAML-XACML profile for authorization interoperability across grid middleware in OSG and EGEE[EB/OL].[2012-08-16].http://iopscience.iop.org/1742-6596/331/6/062011/pdf/1742-6596_331_6_062011.pdf.

[22] Garzoglio G, Alderman I, Altunay M, et al. XACML profile and implementation for authorization interoperability between OSG and EGEE[EB/OL].[2012-10-27].http://iopscience.iop.org/1742-6596/219/6/062014/pdf/1742-6596_219_6_062014.pdf.

[23] Wolinsky D I, Figueiredo, R. Experiences with self-organizing, decentralized grids using the grid appliance[C]//HPDC 11: Proceedings of the 20th International Symposium on High Performance Distributed Computing, 2011:183-194.

[24] O' Donovan, D J, Stokes-Rees, I, Nam, Y, Blacklow, S C, Schroder, G F, Brunger, A T, Sliz, P. A grid-enabled web service for low-resolution crystal structure refinement[EB/OL].[2012-11-04].http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3282622.

[25] Garzoglio G, Wang N B, Sfiligoi I, et al. SVOPME: A scalable virtual organization privileges management environment[EB/OL].[2012-09-22].http://www.fnal.gov/docs/products/voprivilege/documents/SVOPME-iSGC09-Paper.pdf.

[26] De Roure D, Goble C, Aleksejevs S, et al. Towards open science: The myExperiment approach[J]. Concurrency and Computation—Practice and Experience, 2010, 22(17): 2335-2353.

[27] Knoppers B M, Isasi R, Benvenisty N, et al. Publishing SNP genotypes of human embryonic stem cell lines: Policy statement of the international stem cell

forum ethics working party[J]. Stem Cell Reviews and Reports, 2011, 7(3):482-484.

[28] Grubb A M, Easterbrook S M. On the lack of consensus over the meaning of openness: An empirical study[J]. PLOS ONE, 2011, (8):e23420.doi:10.1371/journal.pone.0023420.

[29] Jonjic T. Juggling between open science and the market: Public science responses to the patentability of biomedical research tools[J]. Periodicum Biologorum, 2010, 112(4):381-390.

[30] Penin J. On the consequences of patenting university research: Lessons from a survey of French academic inventors[J]. Industry and Innovation, 2010, 17(5):445-468.

[31] Procter R, Williams R, Stewart J, et al. Adoption and use of Web 2.0 in scholarly communications[J]. Philosophical Transactions of the Royal Society A-Mathematical Physical and Engineering Sciences, 2010, 368(1926):4039.

[32] Ramachandran R, Movva S, Conover H, et al. Talkoot software appliance for collaborative science[EB/OL].[2012-09-22]. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=54176

[33] Tacke O. Open science 2.0: How research and education can benefit from open innovation and Web 2.0[EB/OL].[2012-11-07].http://www.olivertacke.de/wp-content/uploads/2011/02/Tacke-2010-Open_Science_2.0-preview-100613-OLT.pdf.

[34] Fry J, Schroeder R, den Besten M. Open science in e-science: Contingency or policy?[J]. Journal of Documentation, 2009, 65(1):6-32.

Note: Figure translations are in progress. See original paper for figures.