Open Regional Science

Sergio J. Rey[1,2]

February 18, 2014

PRESIDENTIAL ADDRESS

55th Annual Meeting of Western Regional Science Association
San Diego, California

[1] GeoDa Center for Geospatial Computation and Analysis, School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ. email: srey@asu.edu

1 Introduction

The title of this talk could have multiple meanings. If one were to use *Open* as an adjective, what would follow would be a discussion of the many ways in which our world of regional science has embraced open science, an important offspring of the open source revolution. An alternative meaning would have *Open* as a verb in which case the talk that follows would be a call to arms for regional scientists to engage with open science and open source.

As should be clear in what follows, I hold the second inteprertation to be correct, and this for two reasons. First, I have personally seen the impact that open source has had on scientific software development, primarily through my experiences with the packages STARS (Rey & Janikas, 2006) and PySAL (Rey & Anselin, 2010). While software was my point of entry into this open world, the impacts have spread beyond better software development models to influence many aspects of my day-to-day existence as an academic regional scientist. These impacts, which I outline in more detail below, have been profoundly positive. Second, I feel that we, the regional science community, have been largely disengaged from developments in open science. I am excited about the possibilities that a deeper engagement could have on the future of regional science and want to do all that I can to hasten its development.

To do so I will contrast two worlds of science. The first, and the one we regional scientists currently find ourselves embedded within, is what I will call *captured science*. While this is our status quo, it is not generally what holds everywhere in the broader scientific community where a second and new type of science is operative. This is what I will call *open science*. My purpose in this talk is to situate regional science within the paradigm of open science by arguing that our future should be linked to open data, open modeling, open software, open collaboration, and open publication.

I will begin by outlining what exactly I mean by captured science, highlighting its key characteristics and pointing to some of the challenges and constraints it poses. Some of these issues may be familiar to many of us, but I think we, as a community, have not had a healthy open discussion about their implications. I then move on to consider the alternative model of open science, what its components are, and how regional science might benefit from migrating to this model and away from the current status quo. I close the talk with some thoughts about the likely future for open regional science.

2 Science: Riches and Challenges

I, and I suspect many of you, take the *science* part of regional science seriously. We see science as one of the best constructions of the human kind, running close to beer. We hold it in such high regard as science has delivered on the noble goals of:

- uncovering new knowledge
- deepening our understanding of the world
- improving our realities for the betterment of human kind

Many historians point to Descartes's *Discourse on Methods* written in 1637 as the birth of science. From the beginning the concept of reproducibility was at the core of science. As Schroeder (2013) notes, the motto of the Royal Society founded in 1671 was

Nullius in Verba[1]

A key development in the advancement of early science was the decision by the Royal Society to publish results in the form of letters as a way to hasten dissemination relative to book publication. Both the format and process of scientific publishing had self-correction baked into them from the very beginning. Indeed, the use of three referees as a model for the review process dates back to this period. At the risk of a gross simplification, this model has been successful on the whole as the amount of knowledge produced by the scientific process has grown enormously. One consequence of this success is that the age of the generalist has long been eclipsed and it is simply impossible for any one scientist to be conversant in the ocean of specializations in which we now swim.

As a metaphor, consider a map of scientific domains on a sphere where each domain is represented as a simple polygon. The polygons would likely be nonuniform, but they would represent an exhaustive and mutually exclusive partitioning of the body of scientific knowledge. Take the centroid of each domain polygon as a point representing the core of that domain. We can measure the distance between each of the domains using, say, great circle distance. We could adopt a spatial interaction model that incorporates this distance to measure the level of interaction between two domains i and j at time t. We adopt a time index to reflect the evolution of the associated parameters and distance metrics.

The impact of increasing distance on inter-core interactions as knowledge domains expand will of course also be shaped by the evolution of $\alpha(t)$, our distance decay exponent. Generally speaking, this will reflect the costs of overcoming the distance between different bodies of knowledge. My thesis is that, in general terms, we have seen changes from a world in which the institutions of science that originally worked to dampen this effect to one where increasing access costs have been exacerbating this effect, and this transformation has everything to do with a walling off of science to capture economic rents.

2.1 Captured Science

There are multiple channels through which science has become increasingly commercialized since the heady days of Descartes. Perhaps the most widely debated of these is the commercialization of scientific publishing. The costs of scientific publication has attracted much criticism from members of the academy. A well known lament is that the scientific community carries out the peer review process largely gratis yet the commercial journal publishers charge billions of dollars per year for access to the final scientific product.

Although the true costs of access to journals is often obscured by non-disclosure agreements that university libraries must sign to access journals, some estimates put the annual revenue of the science-publishing industry at \$9.4 billion in 2011 with an average revenue per article of roughly \$5,000, average per article costs around \$3,500-4,000, and profit margins at 20-30% (Van Noorden, 2013). These are Apple-like profit margins. Publishers own the copyright of the paper and have enclosed the theory

---

[1] "Take nobody's word for it."

within the confines of the article. Software and data are generally not to be found, even in cases where the paper access costs can be borne. Captured science is the result.

Commercial publishers counter these claims by pointing to the role that journal branding can play through the peer review process. Rejection by high prestige journals serves to filter manuscripts to the most appropriate outlets and lowers the search costs for future researchers.

2.2   Open Access

A recent development in scientific publishing has been the move to the so called open access publication model. Although it can take on different forms, the basic premise is that once a scientific paper has been published, and requisite fees paid, it is made freely available to interested readers. The model has gained much momentum; as of 2011 estimates are that some 50% of all scientific papers published are in some form of open access. It is useful to take a look at the pressures that have driven its evolution.

Three general forces have been responsible for open access coming into being. First, and most prominently, is the widespread disenchantment by the academic community with the traditional commercial publishing model. This is reflected in such documents as *The Cost of Knowledge*, signed by some 14,000 scientists who pledged to no longer participate in commercial scientific publishing, from refusing to submit their articles to journals under the traditional model, to declining referee requests or serving in any aspect of the editorial process. This was a key component of the boycott of Elsevier.

A second pressure reflects movements by governments to begin requiring that research supported by federal funding make the findings and data publicly available. This is reflected in the US National Science Foundation requiring a data management component in all research proposals with accessibility as a central concern.

The third pressure flowing from academia is reflected in the University of California's (UC) recent adoption of its Open Access Policy (University of California, 2013) in which faculty grant a license to UC prior to any contractual agreement with publishers. This permits the UC to archive the research in its eScholarship system, thus providing access to the public at no charge.

The open access model has enormous potential for fueling wider dissemination of and access to scientific research, and it is often held in high regard by the scholarly community for this very reason. However, in large part the open access movement can be viewed as a reaction by commercial publishers to the three pressures mentioned above. Moreover, the model is not without its problems, some of which are threatening the basic integrity of scientific publishing.

There have been a number of high profile cases where the poor quality control of open access journals have been exposed. Particularly striking is the example outlined in Bohannon (2013) where a single author submitted 304 versions of the same spoofed article to open access journals. The key theme in these papers addressed the role of an extract from lichen, and its anti-cancer properties. Author names for each version of the paper were randomly created, and the dimensions of the arguments surrounding combinations of different types of molecules, species, and cancer cells were exploited to provide some differentiation of the alternative versions of the paper.

Several profound results stem from this experience. First, more than half of the papers were accepted. Second, both the key anti-cancer agents and the author of these manuscripts did not exist. Thirdly, among the 147 journals that accepted the article were those owned by Elsevier and Sage, as well as prestigious academic institutions such as Kobe University. Combined, these results raise doubts about the legitimacy of the open access journal as a repository for the state of scientific knowledge as, in this case, acceptance of patently false findings was far from the exception.[2]

In addition to the problem of quality control in open access journals, other major challenges that arise in captured science are the roadblocks to reproducibility and the weakening of science's self-correction mechanism. Prominent cases of plagiarism, data cooking, and fraud are to be found in the social sciences. Frank Fischer, a political scientist at Rutgers, was accused by a graduate student and Alan Sokal[3] of plagiarism. Similar to the case of Doris Kearns Goodwin, the accused claimed that it was a simply sloppiness on their part rather than outright plagiarism.[4]

Closer to home for regional scientists is the Rogoff & Reinhart (2010) paper purporting that debt-to-G.D.P. ratios above 90% dampened growth. The paper was widely cited in policy circles in support of austerity programs. However, examination of the study by others revealed a mixture of spreadsheet errors, omission of available data, favorable weighting and transcription errors. Follow-up studies have shown the magnitude of the effect is reduced when these changes are incorporated, but not the sign (Herndon, Ash, & Pollin, 2013). One ray of shining light is that the heroes of these stories often tend to be graduate students who uncovered the fraud or questionable practices. They were able to do so because they had access to the data and methods underlying the original studies in question.

2.3 Data Hoarding

In order for science's error-correction mechanisms to kick in, data underlying research projects must be made accessible to the wider research community. Unfortunately, current institutional constraints and individual practices are standing in the way of realizing this.

2.4 Privacy Concerns

There are important concerns regarding the protection of personal information on the one hand, and the rich set of empirical analyses that micro data support on the other. A number of strategies have been explored to strike this balance. Anonymization of public records attempt to minimize the risk of revealing information about individuals. However, this has its limits as a number of high profile failures have demonstrated.

---

[2]Questions of selectivity bias in Bohannon (2013) have been raised since the sample included only open access journals and not journals published under the traditional model.

[3]The namesake of the *Sokal affair* in which the author submitted a completely nonsensical manuscript entitled "Transgressing the Boundaries: Towards a Transformative Hermeneutics of Quantum Gravity" to the journal *Social Text*. After the paper was published Sokal revealed the hoax.

[4]For a running commentary of plagiarism in the social sciences see the blog of Andrew Gelman at http://andrewgelman.com.

One example was the case of the State of Massachusetts Group Insurance Commission (GIC), the agency responsible for purchasing health insurance for state employees (Sweeney, 2005). As part of that effort GIC anonymized data by removing names, addresses, and Social Security numbers before releasing to researchers. What remained in the released data were the ZIP code, birth date and gender of each person along with diagnoses and prescription information. For a mere $20 a researcher was able to purchase a voter registration list for Cambridge that contained the name, address, ZIP code, birth date and gender of each voter. Linking this with the GIC data, the researcher was able to determine that only six people had the same birth date, of these three were men, and only one of these had the same ZIP code. That person was William Weld, the state's governor.

2.5   Data Silos

Clearly there are problems with anonymization procedures, and a very active research agenda is developing around privatization (Reiter, 2012). An alternative is the use of safe havens as secure sites for data containing sensitive personal information, with access being granted to authorized researchers. In the US, Census Research Centers (CRC) play this role and offer remarkable opportunities for regional scientists to have access to micro data subject to a number of restrictions designed to ensure confidentiality.

Although the CRCs are successful in protecting privacy concerns, it important to keep in mind that they also place limits on the network effect and science's self-correcting mechanism. This is because replication of studies that come out of CRC research is difficult, as any researcher seeking to do so requires access to the same data used in the original study. That access is only granted by the CRC which faces difficult choices in determining what proposals get approved for access. Given the choice between proposals for new novel studies versus studies that seek to replicate previous studies, it is conceivable that the former may be viewed more favorably.

Data hoarding is not limited to institutions, as the problem can be found at the individual level as well. Our existing tenure and reward systems stress the number of publications produced, and for researchers who have invested time and resources in constructing or acquiring unique data sets it is rational to seek a return on that investment by maximizing their exclusive use of the data. However, the individual scholar model doesn't really scale well, and one must consider the opportunity costs from the scientist restricting access to the data by the wider scientific community. After all, if the data are really so wonderful, just imagine what might flow from releasing it to the field!

This does not have to be a zero sum game where the private gains are sacrificed for social goods - we can tweak the reward structure and attribution norms to make data provision a first class contribution to the scientific process. In other words, rather than the single scholar producing a series of 5 papers with the exclusive use of the data, she could produce one article and releases the data. This in turn empowers a larger group of scholars to generate vastly more than 5 papers using the same data, with each of these papers citing the single paper produced by the data generating/contributing scholar. The citations to her original article would grow exponentially in this world rather than linearly in the hoarding model.

2.6 Modeling Islands

Since the early days of regional science, Walter Isard envisioned that integration would be a hallmark of our discipline (Hewings, Nazara, & Dridi, 2004). It is sobering to contrast that grand vision with today's state of our modeling science. To a very real extent, integration of different modeling efforts has fallen far short of this vision. Rather than a rich ecosystem of interconnected modeling components, the silo business model appears to have become dominant. In part this reflects the economics of the regional modeling business where the development, continued enhancement, and support of modeling frameworks requires stable and constant financial support.

At the same time, we as a community of regional modelers have paid scant attention to model interoperability. A search of the leading proprietary regional models (e.g., REMI, IMPLAN) failed to turn up any references to application programing interfaces (API) which could be used to couple different modeling frameworks together. The finger should not just be pointed at proprietary modeling systems, as the academic community has also largely ignored interoperability concerns. Indeed, developers of alternative modeling frameworks are often viewed as competitors rather than as potential collaborators.

The lack of interoperability has hindered progress in the area of integrated modeling as much of the research effort has focused on the challenges of fusing existing modeling frameworks using different integration strategies (Rey, 2000). If model designers had paid more attention to interoperability, modularity and basic object oriented practices (Jackson, 1994), less of the research effort would have been spent on refactoring integration strategies and more on enhancing and applying integrated models to pressing regional economic issues. This lack of interoperability is particularly worrisome, given the growing recognition of the importance of research on coupled natural and human systems and the need for integrated analytical frameworks to support inquiry.

3 Open Science

While the previous sections of this address have painted a less than rosy portrait of the current state of science, the brush is not intended to be overly broad. At the same time, I am also optimistic that rich opportunities lie before us, and if we grasp them we can affect a stronger, more open science going forward. Here I outline what I see as the key pillars of open science and the roles that regional scientists might play in their realization.

3.1 Open Source

Much attention has been paid to the open source revolution and its impacts on many aspects of the modern world (Rey, 2009). Remarkably, these widespread and deep impacts had their origins in the seemingly obscure question of how software teams organized themselves. As chronicled by Raymond (1999), the traditional model adopted by proprietary software houses was to bring together wizards or high priests working in small groups isolated in towers walled off from users and markets. The dominant model since the early days of the software industry, Raymond contrasted this cathedral model

with a new upstart represented by early work on the Linux kernel. Here thousands of seemingly unorganized and decentralized developers were working to build the kernel. The chaotic nature of this form of organization struck Raymond as much more of a bazaar than a cathedral.

That the bazaar model resulted in software that became critical infrastructure underlying many of the internet services we rely on today seemed surprising. Its impacts on science are less well recognized, but we don't have to look far to see concrete evidence. For example, the beauty contest that is the annual ranking of the worlds fastest super computers is massively dominated by Linux as the Top500 supercomputer list reported that 476 of the 500 fastest machines ran Linux (Vaughan-Nichols, 2013).

Perhaps more profound, though subtle, are the soft innovations that open source may bring to the practice of science. For science to be truly open, two components need to be operative. *Open data* constitutes available, intelligible, accessible and usable data. *Open access* to scientific publications and knowledge allows the realization of the building on shoulders of giants.

3.2   Open Data

The increasing availability of open data is playing a pivotal role in the evolution of the so called *fourth paradigm of science*. The classic pairing of experiment and theory (first and second paradigms) were married to the third paradigm of large-scale computational simulation in the mid-20th century. In this triad, data have provided observations about phenomena and were either collected to test particular theories or generated as output in process based simulations about those phenomena. In the fourth paradigm, data assumes a more leading role. Applications of exploratory and data mining technologies to massive and heterogeneous datasets are increasingly being used to generate, rather than test, new hypotheses. Indeed the central role of data in this context is reflected in an alternative name for this fourth paradigm: *data-intensive science* (Tolle, Tansley, & Hey, 2011).

Regional science has been slow to engage with this new paradigm, and at times has been hostile to the exploratory nature of much of this work. I think that hostility reflects traditional paradigms dominating economics in which theory-informed specifications are viewed in high regard, while exploratory data analysis is often dismissed as measurement without theory. Yet, in an increasingly interdependent and complex world, one which is also generating unprecedented and overwhelming amounts of empirical data, the gap between what extant theory can shed light upon and what remains unknown will only widen. If we don't change our ways, we regional scientists will incite the criticism of practicing theory without measurement. In a world of big data, over-reliance on theory will limit what we can contribute to science.

There are positive developments, however. The rise of exploratory spatial data analysis, geocomputation, agent based models, microsimulation, data mining, and related new computationally based methods are enabling an expansion of the scope of regional science. Very often the outcome of these lines of investigation are new types of questions arising from newly discovered empirical patterns, and this stands in stark contrast to the tradition approach of hypotheses based investigation. These do not, however, have to be mutually exclusive and I would argue they can be complementary approaches.

3.3 Open Modeling

Interoperability and open modeling are crucial for moving regional science into the high performance computing era. Much of our regional science modeling toolkit consists of frameworks that were developed and designed for the single desktop era. In order to tap into the power offered by multiprocessing, cluster and grid computing architectures, a significant refactoring of this code base will be required (Anselin & Rey, 2012).

While earlier I remarked that the current state of the science in regional modeling consists of individual modeling efforts that are largely isolated from one another, there are some important exceptions that point the way forward towards supporting a genuine form of model integration. The UrbanSim project (Waddell, 2002) is fully open source and released under the GPL. Designed for use in operational planning it has enjoyed wide application across the U.S., Europe, Asia and Africa. The open source aspect has fueled both the application and development of UrbanSim, as researchers seeking to apply the modeling framework in their own projects are free to do so having full access to the source code.

A second set of exceptions are spatial analytical libraries the are freely available to regional scientists doing spatial econometrics and exploratory spatial data analysis. Although not strictly open source in a licensing sense, Jim LeSage's Spatial Econometric Toolbox was one of the first freely available (as in price and full source code), libraries for spatial econometrics. A full open source stack of spatial data analysis models has been available through the spdep and related packages in the R environment. Finally, PySAL is an open source library for exploratory spatial data analysis and spatial econometrics for the Python language. The key aspect of these libraries is that they provide detailed APIs for their components, which allows end users to combine these components in flexible ways

3.4 Open Collaboration: Release Early and Release Often

Data analysis is not easy, and honest mistakes are made. Uncovering those mistakes is vital to sciences self-correction mechanism. Yet, uncovering those mistakes is also not easy, especially in our current publishing system. Publication pressures exacerbate this by leaving scholars precious little time to fully document the research process that goes into the final publication of a given manuscript. The lack of replication infrastructure is a major impediment to identifying errors of both nefarious and honest species.

One possible antidote for this problem is to adopt more open forms of collaboration. These would tap into Linus' Law (Himanen, 2001):

> Given enough eyeballs all bugs are shallow.

A fascinating example of open collaboration is the Polymath Project developed by Gowers & Nielsen (2009). Inspired by the open source practices seen in the Linux project and Wikipedia, Gowers began the project on his blog with a description of a research problem: namely to develop a proof of the Hales-Jewett Theorem in the realm of combinatorics. He also provided links for background materials and rules of engagement designed to encourage peoples' and collaboration.

The project began on February 1 of 2009, and had its first contribution 7 hours latter from a UBC mathematician. Comments quickly followed from a diverse array of individuals spanning the spectrum from an Arizona high school math teacher to a winner of a Fields Medal. Within a few weeks of its inception, the project had accumulated over 800 contributions comprising 170,000 words and by early March the collective effort had generated an elementary proof. This initial success of the Polymath project has been extended to other math problems, and similar open source collaborative models have been employed in other fields including biology, physics and computer science. More broadly, one can point to the popularity of sites such as Stack Exchange, a community-powered question and answer forum programmers frequent in seeking help on problems, as evidence that this approach scales well. These models provide new pathways for pushing the limits of our problem solving abilities.

3.5   Open Publishing

Advances in cyberinfrastructure are having impacts not just on how we *do* science but on how we will *report* science. The traditional vessel of reporting findings has been the regular journal article, which has served us well but, as pointed out above, is showing its age. Tapping the possibilities of electronic publication opens up new ways to explore the scientific literature.

The open publication model also provides an entry point to access the data and methods that underlie an article. This lowers the barriers to reproducing reported work by other scientists. It is exciting to contemplate the impact that this could have on research in regional science. Take the case of the literature on regional convergence where a number of meta-analyses have attempted a synthesis of what we know about the processes of regional growth (de Groot, Abreu, & Florax, 2005). These entail an enormous amount of traditional literature review and careful extraction of estimation results from previous studies, the latter then being used as inputs in meta-regressions to quantify the relationship between say speed of convergence and aspects of research design employed in the individual studies.

But, consider a meta-analysis on steroids where what is available to the meta-researcher is not just data in the form of the final estimation results of previous papers, but rather the original data, estimation code and software, and ancillary materials used to generate the reported estimation results.That is an entirely different ballgame.

Because of our reliance on hard copy journals, too much attention has been placed on documents as the only research artifact - data and software are not part of the scientific corpus. Meta-analysis of research areas becomes prohibitively expensive, and this severely constrains the synthesis of knowledge. All of this leads to major problems of irreproducibility. Essentially what is being compromised is:

> Science's capacity for self-correction comes from its openness to scrutiny and challenges (Boulton, 2012)

If openness is in question, science is in question.

Complexity in doing science, computational burden, and related technological developments are challenging the traditional scientific publishing model. No longer does a regular length paper in a hardbound journal adequately capture the scientific process

underlying a research effort. As a result, reproducibility is almost invariably impossible in the current model. The open science model offers a way out of this morass.

There are signs that alternative open models for scientific publishing are beginning to gain traction. A recent example that I had experience with was the publication of the conference proceedings for the 2013 Scientific Computing with Python (SciPy2013) conference. The entire process was produced using open source tools, including GitHub, for file submission, reviewing and ultimately publishing.

Several aspects of this process are worth noting. The process relied on technologies that were already familiar to this community of scholars who use code repositories for collaboration on the development of scientific software on a regular basis. In addition to the technologies, the traditional roles of author, reviewer, and editor were mapped into those of participants in an open source software development model. In this model developers (i.e., authors) wishing to contribute a new software feature (i.e., article) to a field, issue a pull request (i.e., submission) to the project (i.e., journal). That pull request is then reviewed by the community (editor and reviewers) and bug reports (referee reports) are submitted. The developer then incorporates the feedback from the bug report (reviews) into the manuscript and updates the pull request.

At the end of the process the project maintainer (editor), has to make a final decision. The pull request would be merged in the case where the paper is accepted, or simply closed but not merged if the paper were rejected. In the former case the paper would appear in the final published proceedings. In contrast to the traditional publication model, however, papers that were rejected in the process actually remain in the repository as the trail of pull requests, bug reports and publication decisions are available for all to see.

Perhaps more importantly, this design enhances the collaborative nature of the enterprise, as the reviewers took on roles of allies in helping to improve the papers. This stood in stark contrast to the traditional review process in which papers can be shredded by reviewers. At the same time, the open source model was highly efficient, as the review process started with an initial pull request deadline of May 19 and final publication of the proceedings just over two months later. The organization of the review process akin to an open source software development model tapped into the power of community and is a clear reflection that science publication is not a solitary endeavor but rather is done by groups of scholars.

It takes a village to publish an article.

4   Conclusion

Notwithstanding my sincere hope that regional scientists fully embrace open science, more realistically the way forward will likely be forged by some combination of open and proprietary endeavors. In other domains, this mixed model has been highly successful (von Hippel, 2004). This hybrid model will, however, represent a rebalancing of many components of regional science as it engages with open science.

Our choice is whether that engagement takes an active or passive form. If it is passive, I fear we will have missed an opportunity for reinvention. Most of us here work in academia, which has been recognized as the second most conservative institution every

invented, the first being classic opera (Sui, 2014). In other words, we cannot wait for the institutions to change on their own; we must take a bottom-up, grass-roots approach to change those institutions.

Each of us in our own way can be a change agent in this regard. This address represents my attempt towards a contribution, and I hope that this starts a discussion.

Finally, I will leave you with a recommendation from his Holiness the Dalai Lama that captures the spirit of open science:

> Share your knowledge, it is a way to achieve immortality.

5 References

Anselin, L., & Rey, S. J. (2012). Spatial econometrics in an age of cyberGIScience. *International Journal of Geographic Information Science*, *26*, 2211–2226.

Bohannon, J. (2013). Who's afraid of peer review? *Science*, *342*(6154), 60–65. doi:10.1126/science.342.6154.60

Boulton, R. (2012). *Science as an open enterprise* (p. 104pp). London: Royal Society.

de Groot, H. L. F., Abreu, M., & Florax, R. J. (2005). A meta-analysis of beta-convergence: the legendary 2%. *Journal of Economic Surveys*, *19*(3), 389–420.

Gowers, T., & Nielsen, M. (2009). Massively collaborative mathematics. *Nature*, *461*(7266), 879–881.

Herndon, T., Ash, M., & Pollin, R. (2013). *Does high public debt consistently stifle economic growth?: A critique of Reinhart and Rogoff*. Political Economy Research Institute.

Hewings, G. J., Nazara, S., & Dridi, C. (2004). Channels of synthesis forty years on: integrated analysis of spatial economic systems. *Journal of Geographical Systems*, *6*(1), 7–25.

Himanen, P. (2001). *The hacker ethic and the spirit of the information age*. Secker; Warburg.

Jackson, R. W. (1994). Object-oriented modeling in regional science: An advocacy view. *Papers in Regional Science*, *73*(4), 347–367.

Raymond, E. C. (1999). *The cathedral & the bazaar: musings on Linux and open source by an accidental revolutionary*. Sebastopol: O'Reilly.

Reiter, J. P. (2012). Statistical approaches to protecting confidentiality for microdata and their effects on the quality of statistical inferences. *Public Opinion Quarterly*, *76*(1), 163–181.

Rey, S. J. (2000). Integrated regional econometric+input-output modeling: Issues and opportunities. *Papers in Regional Science*, *79*, 271–292.

Rey, S. J. (2009). Show me the code: Spatial analysis and open source. *Journal of Geographical Systems*, *11*, 191–207.

Rey, S. J., & Anselin, L. (2010). PySAL: A Python library of spatial analytical methods. In M. M. Fischer & A. Getis (eds.), *Handbook of applied spatial analysis* (pp. 175–193). Berlin: Springer.

Rey, S. J., & Janikas, M. V. (2006). STARS: Space-Time Analysis of Regional Systems. *Geographical Analysis*, *38*(1), 67–86.

Rogoff, K., & Reinhart, C. (2010). Growth in a time of debt. *American Economic Review*, *100*(2), 573–78.

Schroeder, W. (2013). The new scientific publishers. In *Keynote: Scientific Computing with Python*. Austin.

Sui, D. (2014). Opportunities and impediments for Open GIS. *Transactions in GIS*, *18*(1), 1–24.

Sweeney, L. (2005, October). Recommendations to identify and combat privacy problems in the Commonwealth. http://dataprivacylab.org/dataprivacy/talks/Flick-05-10.html#testimony.

Tolle, K. M., Tansley, D., & Hey, A. J. (2011). The fourth paradigm: Data-intensive scientific discovery [point of view]. *Proceedings of the IEEE*, *99*(8), 1334–1337.

University of California. (2013). UC Open Access Policy. Retrieved from http://osc.universityofcalifornia.edu/open-access-policy/

Van Noorden, R. (2013). Open access: The true cost of science publishing. *Nature*, *495*(7442), 426–429.

Vaughan-Nichols, S. J. (2013). Linux continues to rule supercomputers. Retrieved from http://www.zdnet.com/linux-continues-to-rule-supercomputers-7000016968/

von Hippel, E. (2004). *Democratizing innovation*. Cambridge: MIT Press.

Waddell, P. (2002). UrbanSim: Modeling urban development for land use, transportation, and environmental planning. *Journal of the American Planning Association*, *68*(3), 297–314.