

# Open Source Research, A Quiet Revolution

OWEN DENSMORE  
Complexity Workshop

## Abstract

The Open Source revolution is not limited to operating systems, file sharing and web servers. It is flourishing in other areas as well. This paper introduces the reader to a quiet revolution occurring in the research community, where open source research includes not only algorithms and shared techniques. It also includes hardware design and open community collaboration as well. One research community in particular, the complex adaptive systems researchers, is introduced here in detail.

## Introduction

Imagine a world wide project to land on an asteroid. Imagine this done with no barriers and total participation by interested parties. All information about the “rocket science” is freely available with no proprietary or secret technologies. The details of the pilot's console are online, and have even been turned into this year's hottest game. Every fastener used within the system is standard, many available at high end hobbyist supply stores or at Silicon Valley's famous Fry's Electronics. The “buckyball,” nanotube straps used for securing pairs of asteroids in mining operations were created by Cal Tech's undergraduate advanced studies group and are available in kit form. The entire project has been made the topic of MIT's<sup>1</sup> graduate engineering school, and all course materials are freely available on their website.

As odd and unapproachable as this may appear to us today, it is slowly becoming fact. Researchers world wide are lowering the barriers to collaboration and communication. Not only is software being “open sourced” but so is hardware and scientific research itself. Many researchers are going beyond passively being open; they are actively seeking participation and creating communities to refine the research.

This discussion will not attempt an exhaustive coverage of open source research as the topic is simply too huge. Instead I'll give several in-depth examples from research, primarily from the field of complexity science. For now, we'll define complexity science as the exploration of systems consisting of individual interacting components acting with only local knowledge in a spatial (2D or 3D) environment over time. The purpose of complexity science is to understand the emergent properties of such a system as it evolves. Examples of these systems include economies, societies, biological systems, the Web and Internet, peer network systems, and even autonomous robots. They typically do not yield to closed form mathematical solutions.

## Evolution of Cooperation

One of the early examples of open source research came from the social sciences. In the early 1980s, Robert Axelrod was puzzled by a seeming anomaly in a non-zero sum game seen in a game theory problem demonstrated in the Prisoners Dilemma. This game's name derives from an imagined scenario in which two criminals are arrested without sufficient evidence against them. The two players can choose between two moves—“cooperate” or “defect”—in an attempt to minimize the length of their prison sentences. Each player gains when both cooperate. However, if only one of them cooperates, the player that does not cooperate will serve a shorter sentence. If both players choose not to cooperate and thus “rat” on each other, they will both serve the longest sentence. This outcome is what game theory grimly predicts will happen.

Because the dilemma stems from not knowing the other player's plan, Axelrod decided to resolve this by “iterating” the game: having it played many times, thus introducing memory and learning. He held two

famous tournaments where participants (both experts and hobbyists) were invited to submit simple programs playing against each other. The results were surprising. The simplest program won in both tournaments. It was called TIT-FOR-TAT, which called for initial cooperation and then always repeated the opponent's prior move: defecting if the opponent had defected, cooperating if the opponent had cooperated.

This is an extraordinary result! For example, it can be proven that TIT-FOR-TAT never beats its opponent, but only ties at best. Yet because it promotes high scores for itself and its opponent, it comes out on top during prolonged play with many different strategies. This discovery became one of the cornerstones of current Computational Social Science, the study of the Evolution of Cooperation.

From the beginning, Axelrod and his collaborators decided to not only make the tournament public, but also to promote open source research by publishing the algorithms and extensive details of the actual tournament play. This was done initially in book form,<sup>2</sup> and later on FTP and websites.<sup>3</sup> From this effort arose dozens of sites and papers using these freely available details to further refine the initial work.

## Papers

Scientific papers, generally published in professional or academic journals, are the keystone for research. In the 1980s, many papers started to be made available in electronic form, generally formatted as text (ASCII), PostScript, or later PDF (Portable Document Format) and Word (doc). Although some of these formats were initially proprietary, all have since become standard and available from multiple vendors. This may be the single most important component of open source research—the easy and free availability of research papers in standard formats.

Initially these papers were exchanged among colleagues via email and FTP. Later, as the Web emerged, individuals and then institutions began creating sites, collecting their papers in standard formats. Generally these are made freely available to the public, although a few do still persist in requiring either subscriptions or institutional affiliation. However, even these are beginning to become available via public library interlibrary loan, or at worst, making individual papers available for a fee.

Perhaps the most impressive collection today is the CiteSeer Scientific Literature Digital Library index.<sup>4</sup> This is a very large collection of scientific literature with complete citation linkage among the papers. This linkage is managed by Autonomous Citation Indexing (ACI), which also automatically generates the paper in several formats: PostScript and PDF being the most used. Well over 100,000 documents are included in the index. Steve Lawrence also provides convincing evidence of the increasing access to well-indexed, freely available research papers in his book *Online or Invisible*.<sup>5</sup>

A similarly ambitious index is the arXiv archive,<sup>6</sup> which evolved from the former [xxx.lanl.gov](http://xxx.lanl.gov) site. It is a fully automated electronic archive for research papers within physics, mathematics, nonlinear sciences, computational linguistics, and neuroscience.

The Directory of Open Access Journals<sup>7</sup> is more directly focused on freely available journals. There are now 1,149 journals in the Directory, with 314 journals searchable at the article level. Currently 55,343 articles are included in the service. This is a broader index than CiteSeer, covering many nonscientific disciplines.

An interesting related phenomenon is the emergence of online journals. An example in the complexity science field is the *Journal of Artificial Societies and Social Simulation*.<sup>8</sup> This is a quarterly, Web-only, peer-reviewed journal in its seventh year focused on articles discussing Agent Based Modeling.

Sadly, some of the most important journals still make their articles available only through subscription. The most egregious example is the Association for Computing Machinery (ACM),<sup>9</sup> one of the pioneers in

computer science, which makes its articles available to non-subscribers for \$10.00 USD. The worst example of this is ACM's Turing Award series,<sup>10</sup> in which the recipient of the award delivers a lecture that is then published in the journal only available to subscribers. These are cornerstone computer science discussions that are simply not available to most interested readers.

Some journals have a more reasonable subscription access limitation: the articles are made generally available after a brief period. For example, the Proceedings of the National Academy of Sciences<sup>11</sup> makes its articles freely available six months after publication.

The difficulty with expensive and restricted access to research journals is being widely discussed amongst respected practitioners. John Baez,<sup>12</sup> the researcher who has written or taught about such subjects as loop quantum gravity, quantum geometry and spin foams, is not to be taken lightly! His recent detailed discussion entitled "What We Can Do About Science Journals"<sup>13</sup> is a great read from the viewpoint of the mathematics researcher.

## Docking

One of the more interesting outcomes of freely available research papers is "docking" experimental results. Docking means performing similar experiments in different environments, thereby verifying and expanding the results of the initial work.

This has led to an openness not usually associated with academic environments. A fascinating example is from a pair of researchers at Brookings Institute, Robert Axtell<sup>14</sup> and Joshua Epstein,<sup>15</sup> authors of a seminal study on artificial societies in the book *Growing Artificial Societies: Social Science from the Bottom Up*<sup>16</sup>. In this book, they presented SugarScape, an artificial world in which inhabitants are easily programmed to explore a wide variety of behaviors. Just as Axelrod's books and papers published the Prisoner's Dilemma algorithms in detail, Axtell and Epstein openly provided the details for their SugarScape world.<sup>17</sup> Indeed, a companion CD-ROM to the book included the entire simulation environment so that readers could repeat the experiments as they read the book.

Not unexpectedly, Axtell, Epstein, and Axelrod joined forces to write a paper on the docking process, "Aligning Simulation Models: A Case Study and Results."<sup>18</sup> This study not only discussed the difficulties and value of docking experiments, it also quantified the costs and made recommendations on creating a general practice of doing so. The paper includes complete work logs detailing to the minute the effort involved in the study. It was discussed more generally later in book form in an article entitled "Advancing the Art of Simulation in the Social Sciences."<sup>19</sup>

Other docking efforts have taken place, such the works of North and Macal,<sup>20</sup> and Xu, Gao, and Madey.<sup>21</sup> Perhaps the most detailed is Replication of Eight "Social Science" Simulation Models<sup>22</sup> found at the Center for the Study of Complex Systems website. These efforts have led to the first "Model to Model" workshop<sup>23</sup> in which modelers met to discuss collaborating to strengthen the validity of experimental results.

## Books

A surprising number of books are beginning to be freely available in electronic form. All of Stephen Wolfram's *A New Kind of Science*<sup>24</sup> has recently become available online,<sup>25</sup> although requiring free registration. The book is already an amazingly encyclopedic work, but the online version actually augments each page with additional material. It also enhances the book by providing a full search facility. Finally, by making the thousands of images within the book available, researchers can exchange notes on specific parts of the books along with supporting images.

A very popular undergraduate text, *Introduction to Probability* by Grinstead and Snell, has similarly

become available online.<sup>26</sup> They also provide considerable additional material and a complete set of programs are made available in four formats: Mathematica, Maple, Basic, and Java Applets. In addition, the answers to the odd-numbered problems are available, and instructors can get the even ones with an email to the authors. It's striking to read their discussion about how the GNU version of the book will improve over the existing paper version, and about their gratitude to the American Mathematical Society, publisher of the paper version.

Perhaps the most respected introduction to computer science is Abelson and Sussman's *Structure and Interpretation of Computer Programs*.<sup>27</sup> It also is freely available in electronic form. In this case, the electronic version evolves the book into an entire course. For self study, the book's website<sup>28</sup> includes sample programming assignments, all the source code used in the book, and assistance in obtaining Scheme, the language used by the book.

It appears that making books available in electronic form is likely to increase rather than decrease the sales of the paper version. Personally, I always obtain both forms, using the digital version to expand on and enrich the paper version.

As an example of using the electronic form to enhance the paper book, consider another amazing instance, David MacKay's *Information Theory, Inference and Learning Algorithms*,<sup>29</sup> which is becoming a classic in the complexity community. A person in our reading group presented the "Efficient Monte Carlo Methods" chapter, which stimulated discussions on other sections and motivated many of the group's members to order the book after witnessing its effectiveness. In addition, I've begun translating the book's examples into the J programming system. The electronic version of the book naturally includes enhancements to the paper version. For instance, many of the examples are available as Octave source code, downloadable from the website. Also, additional future chapters are first made available electronically.

A similar, book-like offering occurs with subject reviews made by respected researchers. These are typically candidates for sections in books, but are published initially in reviewed journals. Two examples within complexity science are Mark Newman's overview of graph theory in "The Structure and Function of Complex Networks,"<sup>30</sup> and Cosma Shalizi's overview of complexity science in "Methods and Techniques of Complex Systems Science."<sup>31</sup> These large (58 pages and 48 pages respectively) reviews are invaluable to new researchers in the field.

The idea of freely available books may seem unimportant to the casual observer. But there are many ways their availability accelerates research and enhances collaboration. For example, my ability to send a couple of pages of MacKay's book to a European collaborator saved a month of copying and mailing material. And the ability to read the book while the paper version is in transit speeds up projects considerably. In the MacKay example, I was able to cut and paste an image used to study noisy communication channels, writing my own versions of the redundancy algorithms using the actual example data in the book! Obviously, the time savings are huge.

## Courseware

With papers and books being freely available on the Net, a natural next step is for courses to use them. One of the earliest examples is the Open University<sup>32</sup> in England, started in 1969. This amazing amalgam of home study, multimedia learning (tapes, web, radio and TV broadcasts), local tutors, mailed course packets, and even telephone conferences is Britain's largest university, with 200,000 current students and over two million participants in its lifetime. At this university, a graduate typically earns a degree in six years while working, at a total cost of roughly £4,000. In addition, the university's structure and openness makes higher education much more available to those with disabilities (nearly 8,000 such students). And, despite its unique approach to higher learning, the Open University even has support for research and funding for research projects.

Perhaps the best publicized example in the United States is MIT's OpenCourseWare.<sup>33</sup> While not a degree-granting facility, it is an ambitious project to make all of MIT's course materials available freely online. For example, Abelson and Sussman's introduction to computer science referred to previously has a section within the OpenCourseWare site.<sup>34</sup>

## Tools

A powerful tool used by the complexity science community is Agent Based Modeling (ABM). As mentioned above, complexity science studies the (generally nonlinear) interactions of many objects (agents). They interact both spatially (neighbors and with motion) and repetitively, iterated over time. One of the first Agent Based Modeling systems built was Swarm,<sup>35</sup> developed by Chris Langton and colleagues at the Santa Fe Institute. Several are now in existence, many written in Java so as to be portable across different types of computers. Two prominent examples today are RePast<sup>36</sup> and NetLogo,<sup>37</sup> both freely available, Java-based, platform independent systems. They are easy to install and have very active developer communities sharing ideas via mail lists. Naturally there are numerous tools in addition to ABM that are freely available and invaluable to research.<sup>38</sup>

## Hardware

People often assume that freely available implementations of research projects are limited to software. Amazingly, this is not the case. Many digital electronic projects are following the open source trend. And, because the digital formats are standard for creating digital boards, one can often receive the digital layout documents, send them to an inexpensive web fabrication facility, and obtain the hardware within a few days! One of the more well known examples is from Berkeley's TinyOS and related hardware project TOS.<sup>39</sup> Downloading a package brings detailed information in the readme file, including a pointer to a file which literally can be sent to a web-fab house that will return the board within a week. A similar project, RoboMote,<sup>40</sup> is being built and deployed by USC's robotics lab.<sup>41</sup> Their website includes board layout schematics and bill of materials.

## Summary and Futures

This paper has presented a very narrow review of open source research activities, limited to recent experiences in complexity science. We have shown that a growing number of research papers, broad topical reviews, book chapters, and even entire books are becoming freely available in standard digital formats on the Web. We also explored active collaboration efforts such as the docking of experiments and making considerable supplementary material openly available, including the initial papers and descriptions of experiments.. Some of these efforts become community efforts involving a group of experts honing initial experimental findings such as the iterated Prisoner's Dilemma and the SugarScape artificial societies. Beyond this, entire courses and degree programs are being made openly available, and even plans for hardware devices are being exchanged with fabrication-ready machining files.

This article could have mentioned many more examples and certainly practitioners in every field are striving to make their systems more and more open and accessible over time. Our mythic asteroid landing is not far away!

There are still many barriers, however. Business plans for journal publishing should evolve to use resources differently, much as software manufacturers are doing. For example, they may give away their resources but sell their document research expertise, bundling groups of papers into sector reports and evaluations. An intermediate strategy would be for such businesses to join together in a brokering arrangement, providing access to broad sets of subscriptions at a much reduced price. The Creative Commons organization<sup>42</sup> is succeeding in its effort to find the middle way. As it states on its website:

“Thus, a single goal unites Creative Commons' current and future projects: to build a layer of

reasonable, flexible copyright in the face of increasingly restrictive default rules.”

Significantly, one of their projects is the Creative Science Commons.<sup>43</sup> This will help unify the various individual efforts mentioned here, and give them the legal structure to manage the tension between control and free access.

Innovative approaches and new business models such as the ones described in this chapter are needed to facilitate open research so that we can build upon, rather than recreate information. Faced with declining subscriptions for its traditional publication models, publishers and other intellectual property holders would benefit from embracing the principles of standardization in which competing companies cooperate and share information to create better products and larger markets. This type of business model migration is normal for the expansion phase of rapidly developing technologies such as computing and communications and is needed for the publishing industry to flourish under this new economy. It is a migration from Intellectual Property to Intellectual Capital, from secrets to know-how.

## Notes

---

<sup>1</sup> MIT: Massachusetts Institute of Technology.

<sup>2</sup> Axelrod, Robert. *The Evolution of Cooperation* (New York: Basic Books), 1984.

<sup>3</sup> Examples include: Complexity of Cooperation.

<http://www.cscs.umich.edu/Software/ComplexCoop.html> and Robert Axelrod's research papers. [http://www-personal.umich.edu/~axe/research\\_papers.html](http://www-personal.umich.edu/~axe/research_papers.html). Note "supporting material" sections which include additional material used within the paper.

<sup>4</sup> CiteSeer Scientific Literature Digital Library index. <http://citeseer.ist.psu.edu/>.

<sup>5</sup> Lawrence, Steve. "Online or Invisible," *Nature* 411, no. 6837 (2001): 521.

<sup>6</sup> Cornell University arXiv e-Print archive. <http://arxiv.org/>.

<sup>7</sup> Directory of Open Access Journals. <http://www.doaj.org/>.

<sup>8</sup> *Journal of Artificial Societies and Social Simulation*. <http://jasss.soc.surrey.ac.uk/JASSS.html>.

<sup>9</sup> Association for Computing Machinery. <http://www.acm.org/>.

<sup>10</sup> Association for Computing Machinery A. M. Turing Awards of the Association for Computing Machinery. <http://www.acm.org/awards/waward.html>.

<sup>11</sup> National Academy of Sciences. *Proceedings*. <http://pnas.org/>.

<sup>12</sup> John Baez's official Web site. <http://www.math.ucr.edu/home/baez/>.

<sup>13</sup> "What We Can Do About Science Journals." <http://www.math.ucr.edu/home/baez/journals.html>.

<sup>14</sup> Robert Axtell's official Web site. <http://www.brook.edu/scholars/raxtell.htm>.

<sup>15</sup> Joshua Epstein's official Web site. <http://www.brook.edu/scholars/jepstein.htm>.

<sup>16</sup> Epstein, Joshua and Robert Axtell. *Growing Artificial Societies: Social Science From the Bottom Up* (Washington D.C.: Brookings Institution Press, 1996).

<sup>17</sup> Growing Artificial Societies. <http://www.brook.edu/press/books/ARTIFSOC.htm>

<sup>18</sup> Axtell, Robert, et al., "Aligning Simulation Models: A Case Study and Results," *Computational and Mathematical Organization Theory* 1 (1996): 123-141.

<sup>19</sup> Axelrod, Robert. "Advancing the Art of Simulation in the Social Sciences," in Conte, Rosario, Rainer Hegselmann and Pietro Terna (eds.). *Stimulating Social Phenomena* (Berlin: Springer, 1997).

<sup>20</sup> North, Michael and Charles M. Macal. "The Beer Dock: Aligning Three and a Half Implementations of the Beer Game," *SwarmFest* (Seattle, WA: Swarm Development Group, 2002).

<sup>21</sup> Xu, Jim, Yongqin Gao and Gregory Madey. "A Docking Experiment: Swarm and Repast for Social Network Modeling," (paper presented at the Seventh Annual Swarm Researchers, Notre Dame, IN., April 2003) *Swarm* 2003.

<sup>22</sup> Replication of eight "Social Science" simulation models. <http://www.pscs.umich.edu/Software/CAR-replications.html>.

- 
- <sup>23</sup> Model to Model Workshop. <http://cfpm.org/m2m>.
- <sup>24</sup> Wolfram, Stephen. *A New Kind of Science* (Champaign, IL: Wolfram Media, 2002).
- <sup>25</sup> A New Kind of Science. <http://www.wolframscience.com/nksonline/toc.html>.
- <sup>26</sup> Grimstead, Charles and J. Laurie Snell. *Introduction to Probability*, 2<sup>nd</sup>. rev. ed. (Providence, RI: American Mathematical Society, 1997); [http://www.dartmouth.edu/~chance/teaching\\_aids/articles.html](http://www.dartmouth.edu/~chance/teaching_aids/articles.html).
- <sup>27</sup> Abelson, Harold, Gerald Jay Sussman, and Julie Sussman. *Structure and Interpretation of Computer Programs*, 2<sup>nd</sup> ed. (Cambridge, Mass.: MIT Press, 1996).
- <sup>28</sup> Structure and Interpretation of Computer Programs. <http://mitpress.mit.edu/sicp/>.
- <sup>29</sup> McKay, David J. C. *Information Theory, Inference and Learning Algorithms* (Cambridge, U.K.: Cambridge University Press, 2003).
- <sup>30</sup> Newman, Mark E. J. "The Structure and Function of Complex Networks," *SIAM Review* 45 (2003): 167-256.
- <sup>31</sup> Shalizi, Cosma Rohilla. "Methods and Techniques of Complex Systems Science: An Overview" in Deisbroeck, Thomas, J. Yasha Kresh and Thomas B. Kepler (eds.). *Complex Systems Science In Biomedicine* (New York: Kluwer, forthcoming).
- <sup>32</sup> Open University. <http://www.open.ac.uk/>.
- <sup>33</sup> MIT OpenCourseWare. <http://ocw.mit.edu/index.html>.
- <sup>34</sup> Abelson, Sussman and Sussman. Structure and Interpretation of Computer Structures course site. <http://ocw.mit.edu/OcwWeb/Electrical-Engineering-and-Computer-Science/6-001Structure-and-Interpretation-of-Computer-ProgramsFall2002/CourseHome/index.htm>.
- <sup>35</sup> Swarm. [http://wiki.swarm.org/wiki/Main\\_Page](http://wiki.swarm.org/wiki/Main_Page).
- <sup>36</sup> University of Chicago's Social Science Research RePast. <http://repast.sourceforge.net/>.
- <sup>37</sup> NetLogo. <http://ccl.northwestern.edu/netlogo>.
- <sup>38</sup> A list of the freely available tools used by myself and colleagues in the last 3 months include:
- GnuPlot: the standard scientific plotting system, available on all platforms, <http://www.gnuplot.info/>.
  - Octave: enhances GnuPlot by adding a sophisticated Matrix algebra system using a language compatible with the commercial MatLab system, <http://www.octave.org>.
  - R: is a statistical computing and graphics language largely compatible with S, its commercial counterpart, <http://www.r-project.org>.
  - J: a general and complete mathematical environment descending from APL, <http://www.jsoftware.com/index.htm>.
  - GraphViz: a graph (nodes and edges) drawing system, <http://www.research.att.com/sw/tools/graphviz>.
  - VisAD: a Java Component Library for interactive analysis and visualization of numerical data, <http://www.ssec.wisc.edu/~billh/visad.html>.
  - GeoTools: , Geo Tools, <http://www.geotools.org/>. OpenMap, <http://openmap.bbn.com/> and Jump: Java-based Graphical Information Systems (GIS), <http://www.vividsolutions.com/jump/>.
  - ImageJ: a Java image processing and analysis application and library, <http://rsb.info.nih.gov/ij/>.
- <sup>39</sup> TinyOS and Hardware. <http://www.tinyos.net/scoop/special/hardware>.
- <sup>40</sup> Dantu, Karthik, et al., *Robomote: Enabling Mobility in Sensor Networks* Technical Report CRES-04-006 (Los Angeles: Center for Robotics and Embedded Systems, Viterbi School of Engineering, University of Southern California, 2004).
- <sup>41</sup> RoboMote. <http://www-robotics.usc.edu/projects/robomote/>.
- <sup>42</sup> Creative Commons. <http://creativecommons.org/>.
- <sup>43</sup> Creative Science Commons Site. <http://creativecommons.org/projects/science>.