

Open Science and Crowd Science: Selected Sites and Resources

Diane (DeDe) Dawson
Natural Sciences Liaison Librarian
University of Saskatchewan
Saskatoon, Saskatchewan, Canada
diane.dawson@usask.ca

Table of Contents

Introduction

 Open Science

 Crowd Science

Methods and Scope

Open Science – Definitions and Principles

Open Science – Open Lab Notebooks of Individuals and Lab Groups

Open Science – Blogs

Crowd Science – Projects for Individuals or Small Teams

Crowd Science – Volunteer Distributed Computing Projects

 The Main Software

 Organizations

 Selected Projects

 Further Sources for Projects

Selected Examples of Collaborative Science Sites for Specialists

Main Software & Online Tools for Open Science

Open Science Conferences and Community

 Conferences

Further Reading/Viewing

 Videos

 Reports and White Papers

 Open e-Books

 Selected Essays, Articles, and Interviews

References

Introduction

“To take full advantage of modern tools for the production of knowledge, we need to create an open scientific culture where as much information as possible is moved out of people’s heads and laboratories, and onto the network” (Nielsen 2011, p.183).

New Internet technologies are radically enhancing the speed and ease of scholarly communications, and are providing opportunities for conducting and sharing research in new ways. This webliography explores the emerging “open science” and “crowd science” movements which are making use of these new opportunities to increase collaboration and openness in scientific research.

The collaboration of many researchers on a project can enhance the rate of data-collection and analysis, and ignite new ideas. In addition, since there are more eyes to spot any inaccuracies or errors, collaborative research is likely to produce better quality results. Openness early in the research process alerts others to the work resulting in less duplication of efforts. Later on in the process, openness can amplify the visibility and impact of the research results and create more opportunities for future collaborations. An increase in both openness and collaboration has the potential to significantly accelerate the progress of science.

The Internet makes these trends possible and allows discussion across space and across disciplines. Indeed, it facilitates connections between scientists and the general public. Although citizen science is not a new phenomenon, the Internet is enabling more science enthusiasts to participate in the discourse than was previously feasible and more scientists are beginning to recognize the valuable contributions collaborations of this kind can make.

Open Science

Taking inspiration from the open source software and open access movements, some scientists are now sharing their lab notebooks and raw experimental data openly online. Open science is a broad concept that includes these closely related areas of open notebook science and open data. Advocates of open science believe that there should be no insider information, and all protocols and results – even those of failed experiments – should be made visible and open to reuse as soon as possible in open lab notebooks and data repositories. Additional definitions of open science are listed in the first section of this webliography.

The primary concern expressed by many researchers when first confronted with this method is the fear of being ‘scooped.’ However, Williams (2010) argues that the web is essentially an improved printing press: once something is posted online, it can be considered published. So, practicing open science is actually a means to establish priority. Other concerns may relate to research that involves private medical records or proprietary information. A researcher committed to open science methods might be able to accommodate these concerns, but there also might simply be some areas of science unsuitable for this approach.

Besides establishing priority, there are many benefits to practicing open science. Science done in the open increases the potential for collaborations to occur. In the past, researchers might have worked in parallel for years on the same topic and only discovered their shared interest when one published the work. Today, if these same researchers practice open science, they can find each other quickly online and perhaps form a mutually beneficial partnership. This would eliminate the duplication of efforts and potentially speed the progress of their research. Additionally, once others in the same field are aware of an open scientist’s work, they can

follow the results produced and alert the investigator to any anomalies that might have been missed. By clearly and thoroughly posting protocols and data online, open scientists are also allowing others to replicate the experiment and reproduce the results adding to the robustness of the conclusions. And finally, unsuccessful experiments published in an open lab notebook can save the time of other researchers who may be considering performing similar experiments. All of these factors can result in increasing the speed and quality of scientific discoveries.

An added benefit of practicing open science may be in building better relationships between scientists and the public. Certainly, open scientific notebooks and data files may not be entirely comprehensible to the average citizen, but the simple fact that this information is openly posted will increase the transparency of science. This transparency can help enhance trust between scientists and the tax-paying citizens who likely ultimately fund much of the research. In fact, one of the main criticisms arising from the “climategate” controversy of 2009 was that scientists refused to share data. This created an impression that they had something to hide (Hayes 2010).

Although the community of open science practitioners is growing, it is still a relatively small network. While the online technologies facilitating openness have developed quickly, the culture of science and its incentive systems are much slower to catch up. Unfortunately, practicing open science involves a time commitment that is not rewarded in most tenure processes at the moment, and there is currently no method to track the impact of open science practices.

Crowd Science

Modern tools have rapidly increased our capacity for producing massive quantities of scientific data and scientists at all levels must collaborate to manage this data deluge (Wilbanks 2009). Enabling the average citizen to participate in the collection and management of this data can be one solution to this deluge. Although crowd science is not a widely used term, I use it here to refer to the phenomenon of innovative online “crowdsourcing” science projects, in contrast to more traditional and smaller scale offline citizen science activities. However, an individual involved in a project will continue to be referred to as a “citizen scientist” herein.

Opening up the research process to interested non-academics has the potential to increase understanding of how science functions by engaging and educating the public. Enthusiastic amateurs may have a lot to contribute to scientific progress. They may have valuable insights from unique perspectives and the spare time to commit to data collection and analysis. Indeed, citizen scientists often help with data and specimen collection. The difference today is that the web allows a new range of opportunities for such individuals to participate in science. According to Clay Shirky (2010), the world’s educated population has well over a trillion hours of free time each year, which he refers to as a “cognitive surplus.” This is a huge social asset to be harnessed for the benefit of large community projects enabled by the recent invention and spread of Web 2.0 applications. This is the basis of crowd science.

Methods and Scope

I conducted the research for this webliography over the period of one year (2011), primarily by following the discussion of open scientists on various social networks and listservs, as well as following the blog postings of the main proponents of open science.

A number of closely related topics, large enough to merit their own treatment, are beyond the scope of this project. These include open access and open source software as well as the growing e-Science phenomenon, which creates large amounts of research data.

This webliography is not intended to be exhaustive, but instead attempts to introduce science librarians to the significant proponents of, and sites describing, the open science and crowd science movements.

Open Science – Definitions and Principles

There have been several attempts at defining “open science”, and below I have listed the most commonly referred to examples. Additional definitions appear in the sources found in the “Further Reading/Viewing” section. In addition to definitions of this movement, this list includes links to sets of principles formulated to guide scientists and administrators towards openness.

Open Notebook Science

<http://drexel-coas-elearning.blogspot.com/2006/09/open-notebook-science.html>

This is the 2006 blog post by Jean-Claude Bradley, a Drexel University chemist, in which he coined the term “open notebook science.” At the time, some proponents referred to this concept as “open source science,” causing some confusion with the open source software movement. In this definition Bradley refers specifically to maintaining an online lab notebook freely visible to anyone.

What, Exactly, is Open Science?

<http://www.openscience.org/blog/?p=269>

In this 2009 blog posting on *The OpenScience Project* Dan Gezelter attempts to define the concept of open science. He boils it down to greater transparency in four fundamental areas: methodology, data, communication, and collaboration. This post initiated a series of animated responses. Gezelter is a chemist at the University of Notre Dame and director of the Open Science Project, a group of researchers that develop open source scientific software.

Definitions of Open Science?

<http://lists.okfn.org/pipermail/open-science/2011-July/subject.html>

This discussion of a definition for open science took place on the open science listserv in July 2011. The link above is to the subject archive for that month. Look for the thread initiated by Jo Walsh and follow the resulting discussion. Cameron Neylon contends that it is easier to articulate shared aims than to define this movement. However, Michael

Nielson offers his informal definition: "Open science is the idea that scientific knowledge of all kinds should be openly shared as early as is practical in the discovery process" (<http://lists.okfn.org/pipermail/open-science/2011-July/000907.html>).

Principles for Open Science

<http://sciencecommons.org/resources/readingroom/principles-for-open-science/>

This is a set of principles drafted by Science Commons in 2008. The principles encourage the development of an open cyberinfrastructure to support the flow of research information and open, barrier-free access to: research literature, research tools needed to replicate that research, and research data and protocols.

Panton Principles

<http://pantonprinciples.org/>

The Panton Principles define and promote one important facet of the open science movement: open data. The proponents of open data contend that all data related to published science should be placed in the public domain. The Panton Principles are a set of recommendations that deal with how best to make scientific data available for re-use. Peter Murray-Rust, Cameron Neylon, Rufus Pollock, and John Wilbanks developed these Principles in 2009 at the Panton Arms pub in Cambridge, U.K.

Open Definition

<http://opendefinition.org/>

The Open (Knowledge) Definition (OD) is one of the projects of the Open Knowledge Foundation. It is a broad set of principles attempting to define the "open" in "open knowledge" as it relates to all kinds of content and data, not just scientific. Some open science practitioners prefer the term "open knowledge" since it is more inclusive of other disciplines.

Open Science – Open Lab Notebooks of Individuals and Lab Groups

This section presents a selection of the open lab notebooks of the most active practitioners and proponents of open science in a variety of disciplines, as well as others that follow this practice for their own purposes. In ideal terms, open lab notebooks should expose all research protocols and results (including failed experiments) in as close to real time as possible. However, researchers vary widely in their actual practice. Additionally, open notebook science practitioners often use a combination of wiki and/or blog platforms, so there is a range of platform approaches among the notebooks as well.

With no widespread uptake as yet, active practitioners of open science are a small group. Also, the Internet reveals many open lab notebooks that are not maintained. It is a reasonable assumption that supervisors or principle investigators may introduce their graduate students to the philosophy and methods of open science, and encourage them to adopt the practice.

However, the open lab notebooks of students are transitory. Therefore, most of the following links point to the supervisor's lab group rather than the notebooks of individual students.

Jean-Claude Bradley – UsefulChem

<http://usefulchem.wikispaces.com/>

<http://usefulchem.blogspot.com/>

Bradley is the Drexel University organic chemist, mentioned earlier, who coined the term "open notebook science;" he is one of the best known proponents of this movement. He leads the UsefulChem project, posting all research work in this open wiki and blog. Researchers can also link to his students' open notebooks from this wiki. Bradley's research focuses on the synthesis of new anti-malarial compounds.

Steve Koch – KochLab & Research Blog

http://www.openwetware.org/wiki/Koch_Lab:Notebooks

<http://stevekochresearch.blogspot.com/>

Steve Koch is an experimental biophysicist at the University of New Mexico. His research involves developing new molecular DNA techniques, and he has a very active open lab group. The first link leads to the wiki-based notebooks of lab members. Anthony Salvagno (<http://research.iheartanthony.com/>) is perhaps one of Koch's most productive students in open science and regularly posts blog updates of his research. And Andy Maloney (http://openwetware.org/wiki/User:Andy_Maloney) is a former student who is also a very strong advocate for open science, even writing his dissertation online. The second link is to Koch's research blog that includes further discussion of the lab's research and grant funding.

Cameron Neylon – Cameron's LaBlog

http://biolab.isis.rl.ac.uk/camersons_labblog

Cameron Neylon is a biophysicist and Senior Scientist in Biomolecular Sciences at the ISIS Neutron Scattering facility of the Science and Technology Facilities Council (STFC) in the U.K. As of July 2012 he will be taking up a new position as Advocacy Director at the Public Library of Science. His open lab notebook details his research in structural biology. Neylon is also a very outspoken proponent of open research.

Carl Boettiger's Open Notebook

<http://www.carlboettiger.info/research/lab-notebook>

Carl Boettiger is a PhD student in theoretical ecology and evolution at University of California, Davis and employs an "integrated notebook" approach. He uses several open web tools to track his various research activities: a Wordpress blog, a Mendeley literature database, a Github code database, and a Flickr image database. See this entry for his explanation for this approach: <http://www.carlboettiger.info/archives/211>.

Garrett Lisi – Differential Geometry

<http://www.differentialgeometry.org/>

Garrett Lisi is a theoretical physicist whose research involves the application of differential geometry in this field. He received his PhD from the University of California, San Diego and currently conducts independent research funded by private sources. He describes his open lab notebook as a “choose your own adventure book in theoretical physics.” Lisi has maintained this notebook since 2006.

Martin Johnson – Open Notebook

<http://www.uea.ac.uk/~e356/>

Martin Johnson is an atmospheric chemist at the University of East Anglia in the U.K. This open lab notebook, started in 2010, shares selected content from his research. Johnson studies air-sea gas exchange and marine microbial geochemistry.

Rosie Redfield – RRRResearch

<http://rrresearch.fieldofscience.com/>

Rosemary Redfield of the Department of Zoology, University of British Columbia, studies the evolution of genetic exchange systems in bacteria. She has been maintaining her open lab notebook in blog format since 2006, and has more than 100 entries for most years. Also see some students’ open notebooks under the “what we’re doing” link.

Dror Bar-Natan – Academic Pensieve

<http://katlas.math.toronto.edu/drorbn/AcademicPensieve/>

Dror Bar-Natan, a mathematician at the University of Toronto, keeps an open online archive of all of his research notes (mostly handwritten on a tablet), and photos of his office blackboard scribbles. He has been consistently archiving his notes since 2008, with a few files dating back to 2000. Bar-Natan is not an active proponent of open notebook science, maintaining an open notebook online simply fits best into his workflow.

Greg Lang – Notebooks

<http://www.genomics.princeton.edu/glang/notebooks.htm>

Greg Lang is currently a post-doctoral researcher at Princeton University and studies the molecular basis for evolution with yeast as the model system. His approach to open notebook science is to scan his handwritten lab notebooks into electronic format and post the PDFs online. The files are arranged by research topic, and only those associated with a publication are available.

Open Notebook Science – Wikipedia Entry

http://en.wikipedia.org/wiki/Open_notebook_science

The Wikipedia entry for open notebook science includes a list of active and archived open notebooks and divides them by experimental, theoretical, and “partial/pseudo” (not all experimental results are shared, or there is a delay in the sharing). Members of the open science community periodically update the list, but it is still somewhat outdated.

Open Science – Blogs

Here is a list of blogs by well-known members of the open science community and others who frequently post entries of interest to this community. Many of these individuals also are active in various other online forums such as Twitter, FriendFeed, LinkedIn, and Mendeley.

Science in the Open – Cameron Neylon

<http://cameronneylon.net/>

This is not just a blog, but the “online home of Cameron Neylon,” and includes links to his presentations, publications, and other Web 2.0 activities. Neylon is one of the most active and ardent advocates of this movement. This is the blog to read if you only track one.

Michael Nielsen

<http://michaelnielsen.org/blog/>

Michael Nielsen is a physicist and pioneer in the field of quantum computing. He was a senior faculty member at the Perimeter Institute for Theoretical Physics, but resigned in order to devote all of his time to writing a book promoting open science. This book, *Reinventing Discovery*, was released in October 2011 (see Nielsen 2011). You will see most of his thoughts on this topic in this blog and under the essays tab.

Steve Koch Science

<http://stevekochscience.blogspot.com/>

Steve Koch, mentioned earlier, is a passionate proponent of openness in science. He has a strong online presence in various social media, including several blogs. In this blog he posts entries of a professional nature but not directly related to his teaching and research.

A Scientist and the Web – Peter Murray-Rust

<http://blogs.ch.cam.ac.uk/pmr/>

Peter Murray-Rust is a chemist at the University of Cambridge with research interests in crystallography and informatics; he mainly blogs about open science/knowledge topics, and more recently the semantic web. He is an active advocate of open data in particular and is deeply involved with the promotion of the Panton Principles.

Intermolecular – Mat Todd

<http://intermolecular.wordpress.com/>

Mat Todd is an organic chemist at the University of Sydney with a research interest in synthesizing drugs for neglected tropical diseases. He is active in The Synaptic Leap, an open online community of researchers collaborating to develop such drugs. This community is discussed in the “Selected Examples of Collaborative Science Sites for Specialists” section.

Circle of Complexity – Pawel Szczesny

<http://www.pawelszczesny.org/>

Pawel Szczesny is a Polish biologist with a diverse professional profile, and a strong interest in using new Internet technologies to improve the process of science. This blog is maintained as a series of notebooks on categories of research interest to Szczesny. The notebook entitled Science 2.0 contains his entries on open science.

A Blog Around the Clock – Bora Zivkovic

<http://blogs.scientificamerican.com/a-blog-around-the-clock/>

Bora Zivkovic, a well-known and very active science blogger, is also the organizer of the annual ScienceOnline conference (a popular meeting for open science practitioners). He frequently posts blog updates regarding the conference and other open science topics. Zivkovic has a research interest in circadian rhythms and photoperiodism, so he often blogs about these subject areas.

It is not Junk – Michael Eisen

<http://www.michaeleisen.org/blog/>

This is the blog of Michael Eisen, the University of California at Berkeley evolutionary biologist and co-founder of the Public Library of Science (PLOS). Eisen is a strong proponent of open science and open access, and posts substantial and thoughtful blog entries on these topics as well as genetics, evolution, and baseball.

Research Remix – Heather Piwowar

<http://researchremix.wordpress.com/>

Heather Piwowar is a post-doctoral research associate working with the Dryad team at the National Evolutionary Synthesis Center (NESCent). She studies data sharing and reuse behavior, and her blog focuses mainly on the open data side of the open science movement. She writes blog posts on open data and scholarly publishing behavior.

Crowd Science – Projects for Individuals or Small Teams

New Internet technologies greatly facilitate the collaboration of scientists with the general public to collect and process data. This section lists the most popular, active, online, crowd science projects in a range of disciplines. These are projects that require the active participation of the individual citizen scientist, as opposed to distributed computing projects that only require processing time on a computer while it is idle.

Zooniverse

<https://www.zooniverse.org/>

Zooniverse claims to be home to the “...largest, most popular and most successful citizen science projects” on the Internet. Galaxy Zoo, launched in 2007, helps astronomers classify the millions of galaxy images taken by a robotic telescope. The human brain is much more reliable than computers in such pattern-recognition tasks,

but the amount of data being collected by telescopes is too much for the professional astronomy community to process on their own. The first two Galaxy Zoo projects are now complete; three more are currently underway, along with a growing number of other projects. Participants create one Zooniverse account to take part in any of the projects offered.

Citizen Science Alliance (CSA)

<http://www.citizensciencealliance.org/index.html>

In order to launch a project through Zooniverse teams must submit proposals via the Citizen Science Alliance, a group of five different universities and museums. The philosophy of CSA is to involve the public in academic research and share in the excitement of discovery.

Selected Zooniverse Projects:

Zooniverse – Space Projects

<https://www.zooniverse.org/projects#space>

There are currently eight different astronomical projects available for citizen scientists. Projects range from classification of galaxy images taken by the Hubble Telescope to a detailed exploration of the surface of the Moon.

Zooniverse – Old Weather

<https://www.zooniverse.org/project/oldweather>

The Old Weather project asks participants to transcribe the weather observations handwritten in logbooks by crews of Royal Navy ships in the early 20th century. The information transcribed helps scientists improve climate model predictions and may also inform the work of historians researching this time period.

Zooniverse – Ancient Lives

<https://www.zooniverse.org/project/ancientlives>

This is the first Zooniverse project in the humanities. Participants virtually examine fragments of papyrus manuscripts from a 4th Century BC Egyptian city. They can measure, identify, and mark the characters observed using several virtual tools. These texts could include the lost manuscripts of great authors or the ephemera of everyday events – all of which could be of great value to classics scholars.

Zooniverse – Whale FM

<https://www.zooniverse.org/project/whalefm>

The most recently added project (November 29, 2011) in the Zooniverse suite invites citizen scientists to categorize the sounds made by killer whales in order to help researchers better understand their communication patterns.

Foldit

<http://fold.it/portal/>

Foldit is a game devised by a team at the University of Washington to appeal to our innate puzzle-solving capabilities and competitive tendencies. Players compete in teams to design the best protein structures based on optimal folding of amino acid chains. Knowing the 3-D structure of a protein helps scientists understand its role in the body, and how to target it if it is involved in a disease. Some gamers recently solved a long-standing scientific problem related to the structure of an enzyme from the family of retroviruses that includes HIV (see Khatib et al. 2011).

EteRNA

<http://eterna.cmu.edu/content/EteRNA>

Similar to Foldit, EteRNA is a puzzle-solving game where players design new molecules, in this case RNA. The growing library of synthetic RNA designs could one day contribute to the development of new ways to control living cells. Stanford University researchers select the best design each week to synthesize.

eBird

<http://ebird.org/content/ebird>

The wealth of knowledge and fervor in the recreational birdwatching community is a highly valuable resource for professional ornithologists to tap. For example, “bird counts” have long been used to gather bird distribution data. eBird is an online checklist program, launched in 2002 by the Cornell Lab of Ornithology and the National Audubon Society, to harness such data in real time. Volunteers fill out a checklist of the birds they see or hear on a particular outing, as well as time and location information. These data add to millions of other observations, building a database of ornithological biodiversity and distribution information.

The Open Dinosaur Project (ODP)

<http://opendino.wordpress.com/>

Three vertebrate paleontologists initiated the Open Dinosaur Project in 2009. This project seeks to understand the evolution of quadrupedality in ornithischian dinosaurs and requires the analysis of thousands of measurements of limb bones from hundreds of fossil specimens. These data have already been published in hundreds of disparate scientific papers. With the help of citizen scientists, the ODP assembles these measurements into one database for analysis. Besides doing good science, the goals of the researchers are also to do science in the most open way possible, and allow anyone to participate (Taylor et al. 2010).

Herbaria@home

<http://herbariaunited.org/atHome/>

This is a project coordinated by the Botanical Society of the British Isles. Volunteers help decipher hand-written plant specimen labels from photographed herbarium sheets. The preserved specimens originate from various botanical archives in the U.K. participating in a digitization project. Often the hand-written information is difficult to read, and it is

a time-consuming task for small, under-funded archives to transcribe this data. The project, started in 2006, has seen more than 90 000 specimens documented so far.

Stardust@Home

<http://stardustathome.ssl.berkeley.edu/index.php>

The Stardust@Home project is aimed at the study of tiny interstellar dust particles formed in distant stars. An aerogel collection medium holds the first such samples of these particles, collected in space and returned to Earth in 2006 aboard the Stardust spacecraft. The search for them is time-consuming because of their extremely small size (a few microns at the most), and their rarity. Participants in this project view digital movies taken through optical microscopes of small sections of the aerogel and watch for any potential dust particles. Stardust@Home is a project of The Planetary Society, a public space organization.

Phylo

<http://phylo.cs.mcgill.ca/>

This game, developed in 2010 by researchers at the Centre for Bioinformatics at McGill University in Montreal, appeals to a broader gaming community, due to its design as an abstract puzzle. Phylo seeks to solve multiple sequence alignment problems, arranging sequences of DNA or RNA to identify regions of similarity that could indicate functional, structural, or evolutionary relationships between the two sequences. All of the alignments used contain human DNA sections that are suspected of being involved in various genetic diseases.

SciStarter (formerly Science for Citizens)

<http://scistarter.com/>

SciStarter is an index of citizen science projects collated by volunteer contributors. The SciStarter team reviews each project before it is approved for posting. Scientists can also contribute their citizen science projects directly to this website. Currently, there are more than 400 projects listed.

Crowd Science – Volunteer Distributed Computing Projects

Volunteer distributed computing projects make use of the idle time on many individual home computers to process data in support of large initiatives. Often the user will download a platform, and then join the project of interest. When the volunteer's computer is inactive for a certain length of time, it retrieves a packet of data to process and returns the packet to the project when complete, with no other contribution usually required.

The Main Software:

Berkley Open Infrastructure for Network Computing (BOINC)

<http://boinc.berkeley.edu/>

BOINC is an open source software platform developed by a team at the University of California, Berkley for volunteer distributed computing. Originally designed for the SETI@home project in 2002, it now hosts many projects in a variety of scientific disciplines. BOINC is the largest platform of its kind. Users download the BOINC software then choose the project they prefer to join; alternatively, they can go directly to the project website of interest and download the software from there. Since the software is open-source, anyone may use it for public or private projects. For this reason, BOINC developers cannot guarantee that it is safe to download each application. Most of the projects listed below run on BOINC software.

Organizations:

Citizen Cyberscience Centre (CCC)

<http://www.citizencyberscience.net/>

The Citizen Cyberscience Centre, established in 2009, helps scientists in developing countries set up Internet-based volunteer-computing projects. Scientists attend workshops that introduce them to the concept and help them formulate proposals for projects. The CCC then evaluates each proposal and matches the projects to appropriate developers and computer resources. The CCC partners are CERN, the UN Institute for Training and Research, and the University of Geneva.

World Community Grid

<http://www.worldcommunitygrid.org/index.jsp>

World Community Grid is devoted to advancing humanitarian research that might not otherwise be completed due to the high cost of computer processing infrastructure. IBM donates hardware, software, technical service, hosting, and maintenance to this site. Public and non-profit organizations submit proposals for research projects. Currently there are nine completed projects and ten active projects.

Selected Projects:

SETI@home

<http://setiathome.berkeley.edu/>

This is the first and best known, distributed computing project out there. SETI (Search for Extraterrestrial Intelligence) is an area of research aimed at detecting intelligent life in the universe. Radio telescopes identify signals from space that might indicate the presence of extraterrestrial technologies. These data are analyzed digitally, which requires massive amounts of computing resources. SETI@home was originally launched in 1999 on purpose-written software.

Einstein@home

<http://www.einsteinathome.org/>

Albert Einstein proposed that the universe is full of gravitational waves created by the movements of heavy objects like black holes and pulsars. Participants' computers process the data from gravitational wave detectors to search for evidence of these waves.

Climateprediction.net

<http://climateprediction.net/>

Current state-of-the-art models of climate change include approximations. By running the models thousands of times, with slight changes in these approximations, scientists can gain a better understanding of climate change predictions. This improves confidence in the projections under different scenarios. Participants' computers run these models.

LHC@home

<http://lhathome.web.cern.ch/LHCathome/>

The CERN Large Hadron Collider (LHC) offers two distributed computing projects: SixTrack and Test4Theory. These projects run simulations of particles travelling through the accelerator at CERN. Essentially they turn home computers into virtual versions of the LHC. Computer simulations provide theoretical references to compare to the actual measurements taken at CERN. Any discrepancies between the simulations and the actual data could lead to discovery of new phenomena.

MalariaControl.net

<http://www.malariacontrol.net/>

This project runs large-scale models testing diverse sets of social and biological parameters to determine optimal strategies for combating the spread of malaria in Africa. By running thousands of simulations, researchers will be better able to predict and control this deadly disease. It is the first and only project of AFRICA@home (<http://africa-at-home.web.cern.ch/africa-at-home/index.html>), a website of distributed computing applications focused on solving humanitarian causes in Africa. The AFRICA@home website does not seem to have been updated recently, so the status of this initiative is unclear.

Rosetta@home

<http://boinc.bakerlab.org/rosetta/>

Similar to Foldit (described above), Rosetta@home is a project attempting to discern and predict the 3-dimensional shapes of proteins. The shape of a protein plays a large role in its function and how it interacts with other molecules. Determining how a particular disease-causing protein folds, for example, can lead to development of new drugs to target it. Scientists need massive amounts of computing resources to help accurately predict and design protein structures.

Folding@home

<http://folding.stanford.edu/>

Similar to Rosetta@home, Folding@home focuses on protein-folding simulations. While Rosetta@home aims at predicting protein structure, Folding@home is directed at understanding how the proteins fold. These are complimentary, not competing, projects. Folding@home, developed by the Pande Lab at Stanford University, is one of the few distributed computing projects that doesn't run on BOINC software; it functions on graphics processing units, multi-core processors, and PlayStation 3s.

Further Sources for Projects:

BOINC Project List

http://boinc.berkeley.edu/wiki/Project_list

This page is a list of active BOINC-based projects. Since BOINC is open source software, anyone can use it without the permission or knowledge of the BOINC developers. Therefore, this list is not comprehensive.

List of Distributed Computing Projects (Wikipedia)

http://en.wikipedia.org/wiki/List_of_distributed_computing_projects

This Wikipedia article gives quite a comprehensive list of distributed computing projects arranged by discipline. Each entry has a very brief description and link to the project.

Selected Examples of Collaborative Science Sites for Specialists

A number of online sites exist for open collaborative research, or for the contributions of specialists towards larger goals. These sites facilitate the sharing of data and results among individuals with similar research interests, or they may be projects that require expertise in order to make a contribution (not for the average citizen scientist).

The Synaptic Leap

<http://www.thesynapticleap.org/>

The Synaptic Leap was launched in 2005 as an open online community of biomedical researchers collaborating to develop drugs to treat tropical diseases. Patenting drugs for these diseases that mostly afflict the poor in developing countries promises little profit, so there is less incentive for secrecy compared to other pharmaceutical research. The organizers believe that promoting openness and collaboration will generate ideas more quickly and reduce redundancy in research. In turn, they hope drugs will be developed more quickly and efficiently. The site is organized into four communities focused on malaria, schistosomiasis, toxoplasmosis, and tuberculosis. Woelfle et al. (2011) describe how this approach was successful in producing a drug for schistosomiasis.

Open Source Drug Discovery (OSDD)

<http://www.osdd.net/>

This is an initiative led by the Council of Scientific and Industrial Research (CSIR), a large, publicly funded research organization in India. OSDD was launched in 2008 to provide a global platform where researchers can collaborate and share information to hasten the discovery and development of drugs directed at neglected tropical diseases. OSDD hopes to encourage participation by providing incentives for researchers' contributions in the form of "credit points" for solving particular problems. The first target of this project is tuberculosis.

Sage Bionetworks

<http://www.sagebase.org/>

In 2009 Stephen Friend left his high-powered job as senior vice president of cancer research at Merck in order to start up this non-profit organization. Sage Bionetworks is still in development, but the idea is that it will provide a platform for open collaboration within the pharmaceutical research community. This will expedite the "...pathway to knowledge, treatment, and prevention of disease." Friend envisions the sharing of methodological tools and analytical results, as well as an online open access journal hosted from this site.

The Polymath Projects

<http://polymathprojects.org/>

[http://michaelnielsen.org/polymath1/index.php?title=Main Page](http://michaelnielsen.org/polymath1/index.php?title=Main_Page)

The Polymath Projects is a well-known example of the potential effectiveness of open collaboration among specialists. Timothy Gowers, a mathematician at the University of Cambridge, initiated this project on his blog in January 2009 by posing a difficult math problem with no solution to his readers. Within six weeks the group had collaboratively solved the problem on the blog comment threads. The project continues with a group blog and wiki devoted to solving more math problems in the same manner. Michael Nielsen is one of the blog and wiki administrators and frequently refers to this project in many of his writings and talks.

Open Notebook Science Challenge

<http://onschallenge.wikispaces.com/>

In 2008 well-known open science practitioners, Jean-Claude Bradley and Cameron Neylon, initiated this project as an opportunity for the chemistry community to openly collaborate on the production of a reference source for solubility data. The Challenge invites participants to measure the solubility of various common solvents and contribute all of their data to an open spreadsheet. This has become a good project for students to practice their lab skills and win cash prizes from sponsors for their contributions. Besides the openly available spreadsheet, a freely downloadable book, now in its third edition, is also available.

The Spectral Game

<http://www.spectralgame.com/>

Jean-Claude Bradley and his collaborators Andrew Lang, Antony Williams, and Robert Lancashire developed this game as a fun way for undergraduate chemistry students to learn how to match molecules to their associated spectra. Files uploaded by researchers into the open chemistry database, ChemSpider, are used for the problem sets. The spectra files that are missed most often by the players are flagged for the ChemSpider curators to assess; it is likely that these files are low quality or incorrect. In this way the game also assists in improving the quality of the ChemSpider database.

Nutrient Network (NutNet)

<http://www.nutnet.umn.edu/home>

NutNet was established in 2005 to answer fundamental ecological questions by coordinating consistent data-collection methods among a network of collaborators and sites worldwide. Data are available to all members without restriction, and are publicly available on a three-year moving window. The project operates on a small NSF fund and relies on the volunteer input of researchers. This separates it from similar networks, e.g. Long Term Ecological Research Network & National Ecological Observatory Network, that are mainly based in the U.S. and require more substantial funds to operate (Stokstad 2011).

The Blue Obelisk

<http://blueobelisk.org/>

This group formed as an unfunded grassroots organization at the American Chemical Society meeting in 2005. The founding members are concerned about the lack of open data, open standards, and open source innovations in chemistry. They have collaborated over the years to develop many cheminformatics tools freely available for researchers. See O'Boyle et al. (2011) for a comprehensive discussion of the resources produced through this collaboration.

The Human Genome Project

http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml

Although this project ended in 2003, it is included in this list as a major success story for open science. This publicly funded international effort aimed to identify and sequence the entire human genome. Because of technological advances and widespread collaboration, the project was completed earlier than anticipated, beating out a similar, private project that intended to patent and restrict access to their data. As a result the gene sequences reside in open databases (e.g. NIH's GenBank <http://www.ncbi.nlm.nih.gov/genbank/>) and are therefore publicly available for anyone to use in research.

Main Software & Online Tools for Open Science

Open science practitioners use a variety of online tools to support their activities. There are also a number of proprietary software applications for maintaining lab notebooks (a general

Internet search on “Electronic Laboratory Notebooks” will offer a sampling). Selected specialized tools, popular among open scientists, are listed below.

Open Notebook Science Claims and Logos (ONS Claims)

<http://onsclaims.wikispaces.com/>

Logos developed by Jean-Claude Bradley and Andrew Lang and posted on this site are freely available for use on open lab notebooks. There are several versions of the logos to indicate the timing of data release and level of openness of a notebook, e.g., I = Immediate release of information/data; D = Delayed release; AC = All Content released; SC = Selected Content.

OpenWetWare (OWW)

http://openwetware.org/wiki/Main_Page

OpenWetWare is a wiki-based community primarily for biologists and biological engineers. Individuals or lab groups can create and maintain their own open lab notebooks on this site.

FigShare

<http://figshare.com/>

FigShare is an increasingly popular tool for sharing research data and figures. Creative Commons licenses cover all data uploaded to permit reuse with attribution, and each figure or dataset receives a persistent identifier. Users must register for a free account to upload data, but visitors may browse the files without logging in.

BioTorrents

<http://www.biotorrents.net/browse.php>

BioTorrents is a file-sharing site developed specifically for scientists to share large datasets with collaborators. This site uses the well-known BitTorrent peer-to-peer file sharing technology and is hosted by Jonathan Eisen’s lab at the UC Davis Genome Center. Illegal file-sharing is not permitted and all uploaded files are open access.

myExperiment

<http://www.myexperiment.org/>

Launched in 2007, myExperiment is a wiki-based collaborative environment and social hub where scientists can plan their experiments and share workflows, methodologies, and other digital research objects.

GitHub

<https://github.com/>

Many open scientists also write some computer code for their research. GitHub is a popular repository for sharing this code with others, but also includes tools to encourage collaborative projects.

Mendeley – Future of Science Group

<http://www.mendeley.com/groups/530031/future-of-science/>

The Future of Science group in Mendeley currently has nearly 700 members who contribute and tag relevant citations for articles. The group is open, so anyone can join and view the articles collected under the broad topics of the future of science, peer review, open access, and science 2.0/3.0.

Open Science Conferences and Community

Open science is a movement rooted in collaboration and community. As such, many opportunities for discussing this topic have developed online and in the form of “unconferences.”

FriendFeed and LinkedIn Groups

<http://friendfeed.com/science-2-0>

<http://friendfeed.com/opensci-info>

<http://www.linkedin.com/groups/Open-Science-Supporters-2622476>

Unsurprisingly, members of the open science community are very active in social networking forums. The two FriendFeed groups, Science 2.0 and Open Science Info, are a valuable way to listen in or join the conversation in this area. Recently, the LinkedIn group, Open Science Supporters, was changed to an open group to allow anyone to join. However the archived discussions from the previous group remain private.

Open Knowledge Foundation

<http://okfn.org/>

The Open Knowledge Foundation (OKF) is a non-profit organization established in 2004 in the U.K. to promote open knowledge in all of its forms. It has since grown into an international network of active communities that develop tools, applications, and guidelines to encourage the adoption and spread of open data practices.

Open Science Working Group

<http://science.okfn.org/About/>

<http://lists.okfn.org/mailman/listinfo/open-science>

The Open Science Working Group, established in 2009, is an active community within the Open Knowledge Foundation. Members of this group developed the Panton Principles to encourage scientists to publish their data openly. Their discussion listserv, “open-science”, is an excellent source of information about this movement.

Twitter Hashtag: #openscience

<https://twitter.com/#!/search/realtime/%23openscience>

Another good way to follow discussions on this topic online is to watch the Twitter hashtag #openscience. This link will take you directly to a search of the recent tweets using this hashtag.

Conferences:

ScienceOnline

<http://scienceonline2012.com/>

This is a popular unconference where the participants build the program collaboratively on a wiki and design sessions that will foster discussion. Started in 2007, this annual three-day gathering is usually held in North Carolina. It attracts interdisciplinary delegates with a common interest in the way science is carried out, taught, and communicated online.

Science Online London

<http://www.scienceonlinelondon.org/>

Started in 2008, this is an annual two-day conference held in London that brings together an international range of participants from many disciplines to explore the ways in which the Internet has transformed scientific research and collaboration.

Open Science Summit

<http://opensciencesummit.com/>

This two-day annual conference, begun in 2010, brings together researchers and others interested in discussing the future of collaborative science and innovation. This conference has a focus on medicine and the life sciences but many sessions are also of a broader appeal.

Open Knowledge Conference (OKCon)

<http://okfn.org/okcon/>

<http://okfestival.org/>

This is a one-day interdisciplinary conference of presentations and workshops hosted by the Open Knowledge Foundation. Started in 2007, this annual conference has so far been held in London and Berlin. In 2012, OKCon will be joining with the Open Government Data Camp for a weeklong Open Knowledge Festival.

Further Reading/Viewing

Videos:

Open Science Now!

http://www.ted.com/talks/michael_nielsen_open_science_now.html

This 2011 TEDxWaterloo Talk by Michael Nielsen is an excellent introduction to the topic of open science. Nielsen's recently released book, *Reinventing Discovery*, expands in detail upon the themes presented in this talk.

How Cognitive Surplus will Change the World

http://www.ted.com/talks/clay_shirky_how_cognitive_surplus_will_change_the_world.html

This is a 2010 TedTalk by Clay Shirky explaining his idea of “cognitive surplus,” the basis of his book of the same title. New Internet technologies enable people to be creative, and to collaborate in online community projects in their spare time instead of being idle consumers. Shirky contends that there is huge potential in harnessing this surplus time for civically-oriented projects.

Declarations, Reports and White Papers:

Open Science for the 21st Century

<http://www.alllea.org/Pages/ALL/31/397.bGFuZz1FTkc.html>

This declaration in support of open science was made by ALLEA (ALL European Academies) at their General Assembly in April 2012.

Open Science at the Web-Scale: Optimising Participation and Predictive Potential

<http://www.jisc.ac.uk/publications/reports/2009/opensciencerept.aspx>

This consultative report, produced for the British public body JISC (Joint Information Systems Committee), describes several emerging areas of science including open science and citizen science. This document, written by Liz Lyon and released November 2009, is intended as a discussion piece and raises many questions and challenges for various stakeholder groups.

Open Science Project: Final Report

http://crc.nottingham.ac.uk/projects/rcs/OpenScience_Report-Sarah_Currier.pdf

In response to the JISC report listed above, the U.K. Centre for Research Communications (CRC) conducted interviews of seven UK-based open science and citizen science practitioners and advocates. This report, written by Sarah Currier and released June 2011, communicates the results of these interviews including discussions on definitions, benefits and risks, and strategic recommendations.

Open to All? Case Studies of Openness in Research

<http://www.rin.ac.uk/our-work/data-management-and-curation/open-science-case-studies>

The Research Information Network (RIN) and the National Endowment for Science, Technology and the Arts (NESTA) in the U.K. produced this 2010 report. It communicates the results of interviews with 18 researchers from six U.K. research institutions. The individuals and groups selected represent a range of scientific disciplines and levels of openness at different stages in the research process. The authors present a comprehensive discussion of the benefits and barriers of open science, as well as recommendations aimed at policy-makers on how best to support openness.

To Share or not to Share: Publication and Quality Assurance of Data Research Outputs

<http://www.rin.ac.uk/our-work/data-management-and-curation/share-or-not-share-research-data-outputs>

This 2008 report was also produced by RIN and communicates the results of over 100 detailed interviews of researchers across eight disciplines. The study investigated whether or not researchers made their data available to others for re-use, and any issues they may have encountered in the process. The report includes numerous conclusions and recommendations.

Open Source for Neglected Diseases: Magic Bullet or Mirage?

<http://healthresearchpolicy.org/assessments/open-source-neglected-diseases-magic-bullet-or-mirage>

Consultants Rachelle Harris and Hassan Masum produced this 2011 assessment report for the non-profit organization Results for Development (R4D) Institute. R4D's mission is to accelerate social and economic progress in developing nations. The assessment gives an overview of the existing initiatives that use an open approach in neglected disease research and drug development. The consultants discuss successes and challenges, as well as suggestions for moving forward and supporting this approach.

Science as a Public Enterprise

<http://royalsociety.org/policy/projects/science-public-enterprise/>

This is a major study initiated by the Royal Society in 2011 to "...identify the principles, opportunities and problems of sharing and disclosing scientific information." It was not yet complete at the time of submission of this webliography, but those interested should watch the "Reports & Publications" section of the Royal Society's website for the final report and recommendations.

Open e-Books:

Digitize Me, Visualize Me, Search Me – Open Science and its Discontents

[http://www.livingbooksaboutlife.org/books/Digitize Me, Visualize Me, Search Me](http://www.livingbooksaboutlife.org/books/Digitize_Me,_Visualize_Me,_Search_Me)

This is a *Living Books About Life* title, a series of open access e-books released in 2011 that attempt to bridge the disciplines of the sciences and the humanities. The e-books repackage existing, openly available content and are therefore similar to webliographies, although without annotations for every source. This e-book brings together links for articles, websites, and videos all related in some way to a theme of openness in research. Although several links overlap with this webliography, there is considerable divergence in coverage overall.

I Have Seen the Paradigm Shift, and It Is Us

http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_part4_wilbanks.pdf

This is a chapter in the 2009 book, *The Fourth Paradigm*, published by Microsoft Research (the entire book is freely available online). The production of scientific data is accelerating and accumulating, rapidly in what has been widely referred

to as the “data deluge.” Software incompatibilities and copyright restrictions often inhibit the ability of researchers to collaborate in the use and reuse of these data, thereby impeding the progress of science. In this chapter, John Wilbanks argues that in order to keep up with the data deluge scientists need to make their data open to foster more online collaboration in processing the data.

Selected Essays, Articles, and Interviews:

The Future of Science

<http://michaelnielsen.org/blog/the-future-of-science-2/>

Michael Nielsen’s 2008 essay is often referred to and much discussed. Nielsen’s main theme is that science is currently undergoing a period of rapid change brought about by the Internet and new online collaborative tools. He advocates a kind of “extreme openness” that makes many types of content freely available online for creative reuse.

The Impact of Open Notebook Science

<http://www.infotoday.com/it/sep10/Poynder.shtml>

Richard Poynder conducted this interview of Jean-Claude Bradley in September 2010. Bradley elaborates on the definition of “open notebook science,” and discusses his motivations for doing open research.

The Open Knowledge Foundation: Open Data Means Better Science

<http://www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.1001195>

This is a 2011 Community Page article in PLoS Biology by Jenny Molloy, the Coordinator of the Open Data in Science Working Group of the Open Knowledge Foundation. The article discusses the importance of opening up data and describes some of the tools developed by the working group to promote the open sharing of scientific data.

Who Shares? Who Doesn't? Factors Associated with Openly Archiving Raw Research Data

<http://dx.plos.org/10.1371/journal.pone.0018657>

This study by Heather Piwowar was published in PLoS ONE in 2011. Using bibliometric methods she investigated whether patterns exist in the frequency with which some genetics researchers openly archived their raw data between 2000 and 2009. Her results show that sharing of research data is still minimal, especially in areas such as cancer research where the data could have the most impact.

Crowd Science Reaches New Heights

<http://chronicle.com/article/The-Rise-of-Crowd-Science/65707/>

This 2010 Chronicle of Higher Education article by Jeffery Young is among the first to coin the term “crowd science” for describing the growing trend of large

distributed computing and citizen science projects developing online. It also gives a personal account of the development of the popular Galaxy Zoo project.

Scientists Embrace Openness

http://sciencecareers.sciencemag.org/career_magazine/previous_issues/articles/2010_04_09/credit.a1000036

In this April 2010 article in *Science* Chelsea Wald interviews several proponents of open science: Jonathan Eisen, Steve Koch, Carl Boettiger, and Jean-Claude Bradley.

Earn a Nobel Prize in your Lunch-Break! The Best “Citizen Science” Games Reviewed!

<http://realdocstorstu.com/2011/02/15/earn-a-nobel-prize-in-your-lunch-break-the-best-citizen-science-games-reviewed/>

Doctor Stu’s Science Blog rates and compares five crowd science online games. He gives scores out of ten for playability, fun factor, and value to humanity. Dr Stuart Farrimond is a former medical doctor and teacher in the U.K.

References

Hayes, J. 2010. MPs call for greater transparency after “climategate” scandal. *British Medical Journal* [Internet]. [Cited 2011 Dec 19]; 340:c1855. Available from: <http://www.bmj.com/content/340/bmj.c1855>

Khatib, F., DiMaio, F., Foldit Contenders Group, Foldit Void Crushers Group, Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., Zabranska, H., Pichova, I., Thompson, J., Popovic, Z., Jaskolski, M., Baker, D. 2011. Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature Structural & Molecular Biology* [Internet]. [Cited 2011 Oct 28]; 18:1175-1177. Available from <http://www.nature.com/doifinder/10.1038/nsmb.2119>

Nielsen, M. 2011. *Reinventing discovery: The new era of networked science*. Princeton (NJ): Princeton University Press.

O'Boyle, N.M., Guha, R., Willighagen, E.L., Adams, S.E., Alvarsson, J., Bradley, J., Filippov, I.V., Hanson, R.M., Hanwell, M.D., Hutchison, G.R., James, C.A., Jeliakova, N., Lang, A.S., Langner, K.M., Lonie, D.C., Lowe, D.M., Pansanel, J., Pavlov, D., Spjuth, O., Steinbeck, C., Tenderholt, A.L., Theisen, K.J., Murray-Rust, P. 2011. Open data, open source and open standards in chemistry: The Blue Obelisk five years on. *Journal of Cheminformatics* [Internet]. [Cited 2011 Dec 19]; 3(1):37. Available from: <http://www.jcheminf.com/content/3/1/37>

Shirky, C. 2010. *Cognitive surplus: Creativity and generosity in a connected age*. New York (NY): The Penguin Press.

Stokstad, E. 2011. Open-source ecology takes root across the world. *Science* [Internet]. [Cited 2011 Dec 9]; 334 (6054): 308-309. Available from: <http://www.sciencemag.org/content/334/6054/308.full>

Taylor, M.P., Farke, A., Wedel, M.J. 2010. The Open Dinosaur Project. *The Palaeontological Association Newsletter* [Internet]. [Cited 2011 Oct 28]; 73:59-63. Available from: <http://newsletter.palass-pubs.org/pdf/News73.pdf>

Wilbanks, J. 2009. I have seen the paradigm shift, and it is us. In: Hey, T. et al., editors. *The fourth paradigm: Data-intensive scientific discovery*. Redmond (WA): Microsoft Research. p. 209-214. Available from: http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_part4_wilbanks.pdf

Williams, B.S. 2010. Sceptical chymists online: How the practice, teaching, and learning of science will be affected by Web 2.0. In: Belford, R. E. et al., editors. *ACS Symposium Series 1060: Enhancing learning with online resources, social networking, and digital libraries*. Washington (DC): American Chemical Society. p. 95-114.

Woelfle, M., Oliaro, P., Todd, M.H. 2011. Open science is a research accelerator. *Nature Chemistry* 3:745-748.