

**INVITING THE PUBLIC: THE IMPACT ON INFORMATICS ARISING FROM
EMERGING GLOBAL HEALTH RESEARCH PARADIGMS**

RICHARD GAYLE

SpreadingScience

Woodinville, WA 98072

Email: gayler@spreadingscience.com

MARK MINIE

University of Washington

Seattle, WA

Email: mark@meminie.com

ERIK NILSSON

inSilicos

Seattle, WA

Email: erik.nilsson@insilicos.com

This workshop will focus on disruptive processes impacting research arising from the increasing ability of individuals to create, curate and share data with scientists. Encompassing processes from funding research to providing samples to creating algorithms, including the public will require new approaches even as it opens up new possibilities. We will hear from a few researchers at the forefront of these disruptive processes, followed by a moderated discussion with the audience about these topics.

1. Why invite the public?

1.1. *Two trends*

Two accelerating trends in biomedical informatics are producing disruptive effects on research projects. They are also creating new opportunities to apply novel approaches in the analysis of data and the creation of knowledge.

Due to the increasing computational power of smartphones, medical apps and powerful handheld medical devices, we are seeing a radical change in how data are generated and who controls access. The onrushing onslaught of medical data that will arise from the quantified-self^a and the personalized medicine^b revolutions holds the promise of generating enormous storehouses of data, waiting to be examined using novel informatics approaches. However, this also can create barriers for analysis of the data as access is often dispersed to patients.

Perhaps, the second trend may provide a novel approach for solving the difficulties of the first. The formerly separate, once distinct boundaries between patient, researcher, funder, student and entrepreneur are overlapping. In some instances, they are totally disappearing. This allows collaborations and data sharing to happen in unique ways. It also permits the creation of new research and analytical approaches for informatics – if we are creative enough to use them.

Tremendous stores of data are being produced. Yet, knowledge can only be created by examining information in a social setting. We need to understand and utilize new approaches for generating, examining or analyzing the data – by expanding the idea of who collaborators are.

1.2. *Some Problems and Some Questions*

Research into human health is becoming more global every day. Health records are increasingly digitized, decentralized and personalized. This process can disrupt the ability of researchers to ask new questions and investigate new solutions. It is also impacting the lay population in ways to add greater obstacles as well as opportunities.

Electronic health records (EHR) are rapidly being adopted in developed countries. The number of US hospitals that adopted EHR^c tripled between 2010 and 2012, with over 40% of all hospitals using such systems. This rapid adoption is also apparent in many developing countries. For example, in Malawi, touchscreen digital devices^d are being used for the healthcare needs of over 40,000 HIV patients. Can researchers get ethical access to such widely dispersed or incompatible data? What new informatics approaches will be required to acquire and analyze this data?

While digitization of medical records is being driven from the top-down, the quantified-self and personalized medicine movements are being driven from the bottom-up, with individuals controlling their own health data and creating large personal databases of increasingly granular information. Can community-building approaches help gain access to this data? How will this affect analysis?

^a http://en.wikipedia.org/wiki/Quantified_Self

^b http://en.wikipedia.org/wiki/Personalized_medicine

^c <http://health.usnews.com/health-news/news/articles/4013/07/08/us-hospitals-triple-use-of-electronic-health-records-report>

^d <http://www.plosmedicine.org/article/info:doi/10.1371/journal.pmed.1000319%23s5>

Companies such as Theranos^e are working to fragment the collection of human medical data further, putting devices into pharmacies for customer use that can produce large amounts of personal health data for a few dollars. The rapidly dropping costs for testing of more “-omes” means that the data are becoming not only decentralized, under the control of individuals, but also potentially accessible to a much larger groups, including researchers. What approaches for dealing with the torrents of data that will satisfy all participants will need to be developed?

The continuing digitization of worldwide health data will provide important information simply unattainable before. Although this presents opportunities for examination, it also raises novel barriers for accessing the data as control of the data becomes dispersed amongst individuals. How will the increasing lack of boundaries between individuals controlling the data and those wishing to examine it affect the analysis of the data?

2. The Solutions?

Highly collaborative, open access research networks often destroy the barriers between doctors and patients, between the researcher and the entrepreneur, between the lab bench and the clinic. While disruptive to some processes, these efforts can make it easier for the researcher to gain access to an individual’s data, simply because in many cases the people collecting the data on themselves become a part of the research collaboration.

The rapidly dropping costs for many informatics investigations remove many obstacles formerly seen, allowing the public to fund research they find useful, without asking or waiting for permission^f. For example, the people of Puerto Rico used bake sales and art auctions to raise money to produce the genomic sequence of the Puerto Rican parrot. The resulting data were then assembled and annotated by the local college.^{1,2} This full circle connection between community and academia bootstrapped a much larger effort to examine tropical parrot genomes, produced multiple informatics tools and created the Caribbean Genome Center^g. Research continues to appear derived from this example of public-academia cooperation.³

The Pacific Symposium on Biocomputing has always been at the forefront of new technologies, algorithms and research processes. One paradigm already highlighted has been increasingly collaborative approaches between different researchers to leverage network effects while examining large amounts of data. These collaborations are just one subset of the "public", which can extend from a handful of scientists to millions of people.

The tremendous pressure for reducing the cost/time of drug development is driving new approaches. The overlap of what had been separate roles along the drug development pipeline (i.e. nonprofit biomedical research centers now working on drug development and manufacturing) is just one example. The very idea of who is a collaborator, as well as who is the public, is changing in ways that can actually enhance the ability to examine and analyze the torrent of health data beginning to appear. In all cases, scalable informatics approaches need to be created to deal with the workflow and to help produce results in novel ways.

^e http://news.walgreens.com/article_display.cfm?article_id=5820

^f <http://blogs.biomedcentral.com/gigablog/2014/09/30/community-genomes-from-the-peoples-parrot-to-crowdfunding/>

^g <https://www.facebook.com/Caribbean.Genome.Center>

Informatics are dealing with the collision of these formerly separate spheres in ways that hold enormous potential, but only if we can become more comfortable with the new paradigms these endeavors require. Coupled with the emerging and exponentially increasing ability of individuals to collect or control huge amounts of personal medical data, informatics requires appreciation of the unique approaches needed to take advantage of this important emerging resource of raw data.

There are many efforts attempting to leverage the two emerging trends of widespread digital health records and the disintegrating boundaries between stakeholders along the drug development pipeline. Inviting the public to fund research has spawned a dedicated site – Experiment^h. New databases are being created by connecting with the publicⁱ. Researchers have mimicked fantasy football^j to help invite the public into the lab. Organizations such as Ingenuity are asking people to become involved in informatics with the chance of winning a prize^k.

Deeper entanglements of research with the public are also happening. For example, several projects have used open challenges that permit the public, along with researchers, to examine data in order to crowdsource better algorithms. These have involved breast cancer models^l, cancer survival rates^l, Alzheimer's^m and Parkinson's diseaseⁿ. Open competitions for predicting drug responses to arthritis drugs by examining genetic data have also been announced.⁵

In many cases, the creation and support of open science communities may permit greater collaborations, as well as decreased barriers between the public and academia. Sites such as 23andme^o or patientslikeme^p are forcing new approaches for communicating science as well as analyzing data. Understanding how communities deal with information and innovation will be critical for many future research projects.

We hope to hear from researchers who are embracing new approaches for creating informatics databases, for crowd sourcing the analysis of the databases, for disintermediating the examination of data and for leveraging low cost, low barrier-to-entry, open research approaches to more rapidly understand the complex systems being studied today in health investigations.

3. For Discussion

This workshop has invited researchers working to discover the best practices for combining the dual trends of increasing digital health data with increasingly overlapping roles of scientific collaboration between researchers and interested communities that may control the data.

Richard Gayle, President of SpreadingScience, will be moderating this workshop. Using his commercial experience (as a researcher at a biotechnology company and as Vice-President of a

^h <https://experiment.com>

ⁱ <http://www.northeastern.edu/pollastri/collaborate/>

^j <http://www.kplu.org/post/seattle-scientists-look-make-drug-research-more-fantasy-football>

^k <http://www.ingenuity.com/blog/customer-stories/causal-variant-challenge-update-our-ipad-winner-and-next-challenge>

^l <http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1003047>

^m <http://www.ceoalzheimersinitiative.org/global-ceo-initiative-alzheimers-disease-announces-big-data-challenge-find-new-predictors-cogniti-0>

ⁿ <http://news.patientslikeme.com/press-release/patientslikeme-and-sage-bionetworks-launch-open-science-study-people-parkinsons-diseas>

^o <https://www.23andme.com>

^p <http://www.patientslikeme.com>

small startup), he has recently been examining how information transits communities. Enlarging the communities of practice that can examine data and produce knowledge is a key aspect. He has helped organize successful crowdfunding projects for research, including the Arkyd project^q, which raised over \$1.5 million for a satellite for the people. He has also raised money for his own research using these approaches^r.

Daniel McDonald is currently at the University of Colorado at Boulder, working in the lab of Rob Knight. He is involved with the American Gut Project,^s examining the microbiomes found in humans. This project not only provides a service to the public but also collects data from the same public for research, requiring new informatics tools to be developed. It has so far raised almost \$600,000 through crowdfunding approaches, with over 6000 people submitting the samples used in the study^t. Some preliminary data are already openly available^u. He has been directly involved in designing and implementing bioinformatics tools for reducing data bottlenecks when dealing with a variety of “-omics.”⁶

Jonathan Eisen, at the University of California, Davis, has long been focused on open science approaches for engaging the public. These have also involved gamifying^v the research. He has extended high throughput sequencing approaches from the human gut to the built environment^w. He has taken a lead on examining the communication barriers that these processes can produce^x.

Following short presentations from the three speakers, we will have a moderated discussion. We hope to create a vibrant conversation including all participants regarding current and future approaches for including a wider public in research processes.

The entire healthcare field has finally found Moore’s Law, seeing enormous and rapid change as patient records become fully digitized. Quickly understanding the processes involved to effectively manage successful human health research projects using these unique datasets will be as important as developing new analytical tools to examine the data.

References

1. S. J. O’Brien, *GigaScience*, **1**, 13 (2012)
2. T. Oleksyk, *GigaScience*, **1** 14 (2012)
3. Y. Afanador, *Conservation Genetics Resources*, doi 10.1007/s12686-014-0232-6 (2014)
4. A. Margolin et al., *Sci Transl Med*, **5**, 381 (2013)
5. R. Plenge et al., *Nature Genetics*, **45**, 468 (2013)
6. D. McDonald et al., *Gigascience*. **1**, 7 (2012).

^q <https://www.kickstarter.com/projects/arkydforeveryone/arkyd-a-space-telescope-for-everyone-0>

^r <http://www.rockethub.com/projects/28741-consider-the-facts-moving-people-to-deliberative-thinking>

^s <http://americangut.org>

^t <https://fundrazr.com/campaigns/4Tqx5>

^u http://americangut.org/wordpress/wp-content/uploads/2013/09/module1_Sept_16_small.pdf

^v <http://phylogenomics.wordpress.com/gut-check-the-microbiome-game/>

^w <http://microbe.net>

^x <http://icis.ucdavis.edu>