DELSA Global for "Big Data" and the Bioeconomy: Catalyzing Collective Innovation

Eugene Kolker,[1,2,3,4] Elizabeth Stewart,[1,4] and Vural Özdemir[4,5,6]

[1]Bioinformatics & High-Throughput Analysis Lab and High-Throughput Analysis Core, Seattle Children's Research Institute, [2]Predictive Analytics, Seattle Children's Hospital, and [3]Departments of Biomedical Informatics & Medical Education and Pediatrics, University of Washington School of Medicine, Seattle, WA
[4]Data-Enabled Life Sciences Alliance International (DELSA Global), Seattle, WA
[5]Group on Complex Collaboration, Faculty of Management, and [6]Centre of Genomics and Policy, Department of Human Genetics, Faculty of Medicine, McGill University, Montreal, Canada

Introduction

Big data—on the order of zeta-bytes—was once the most daunting problem in addressing the complex challenges we face in fields as diverse as pharmaceuticals and healthcare, genomics, watershed and ecosystem safety, meteorology, oceanography, astronomy and physics, but now it is being sought as part of the solution. Data-intensive scientific research and development is an approach whose time has come, and scientists are not the only ones jumping on the big-data bandwagon. Direct-to-consumer availability of products that rely on big data (e.g., whole genome sequencing for prediction of individual health risks) has brought consumers and scientists into closer proximity. Citizen scientists also contribute to big data R&D; for example, in collecting high-granularity observations for ecosystem epidemiology where geographically distributed real-time science is crucial. Moreover, social media analysis generates volumes of data that can reveal insights into consumer behaviors, computer hacker networking traffic, or the spread of infectious diseases, to name a few.

From Big Data to Innovation

While this data deluge brings about hope to realize the post-genomics bioeconomy and a systems approach to life sciences, progress in translating big data into innovative products has been a pedestrian stroll rather than a quick march to the future.[1,2] Chief among the barriers is the recent breach of Moore's Law, the prediction that the number of transistors (and by extension, computation and data storage and analysis capacity) on a computer chip doubles every 24 months, allowing "chip scale" to be reduced proportionately. Absent Moore's Law, we have no choice but to exploit parallelism and program multiple processors on multicore chips to cope with the current data deluge.

Yet, merely enhancing computation capacity is not sufficient. The entire big-data-to-product innovation trajectory needs to be examined as a whole and in real-time to enable seamless linking of data production with data access and sharing, and most importantly, data analysis. It used to be that scientists would handle such complex innovation trajectory through years of tacit professional expertise. But this is no longer tenable in the age of big data and geographically distributed forms of data-enabled global science.

Hopes for knowledge-based innovative products and the bioeconomy are shared on both sides of the Atlantic. Indeed, the Europe 2020 economic reform and growth agenda in the European Union (EU) is defined to a large extent by science and technology. One of the five EU-wide targets is R&D and innovation.[3] But for big data to have a compelling impact on the recent investments in bioeconomies such as the US, EU, and beyond, both people and technology need to be connected in order to achieve a form of "collective innovation" that is more effective than any individual or singular group of people and computers.[3–5] This both elevates and challenges the old notions of innovation that tend to rely on "single scientist, single computer, single project" models.

To be sure, science for too long rested on an "individual entrepreneurship" model. The arrival of big data firmly challenges this long-held (and often unexamined) assumption in translating science to innovative products. By virtue of its reliance on transdisciplinary expertise, not to mention consumers and citizen scientists, a new kind of *hybrid* entrepreneurship model is emerging in the age of big data—one that recognizes and rewards both individuals' and groups' collective contributions so that large-scale datasets can be accessed, shared, and analyzed on the critical path to innovative products.

Finally, big data calls for new modes of doing business whereby the R&D actors, ranging from scientists, industry, regulators, and policymakers, utilize the assistance of independent and standalone observatories—a type of "innovation narrator"–that measure the performance of the big data innovation ecosystem from a credible analytical distance while simultaneously being firmly engaged with it.

DELSA: Creating an Innovation Ecosystem for Big Data R&D

To facilitate the progress from data to knowledge to action, leading members of the life sciences community and interested stakeholders

formed DELSA Global (Data-Enabled Life Sciences Alliance International) in November 2011. The founding members of DELSA Global came together through a series of National Science Foundation (NSF)-sponsored workshops, as part of a nationwide search for big data solutions captured in part in a set of reports published by the NSF Advisory Task Force for Cyberinfrastructure in 2011.[6] It was evident, through the workshops and reports, that the innovation process as it had existed in the past was no longer adequate for the present or future.

However, the push towards innovation and the need to solve our society's problems expeditiously still exist. DELSA's mission and vision were crafted with this in mind. DELSA's mission is to accelerate the impact of data-enabled life sciences research on the pressing needs of our global society, while its vision is, through interdisciplinary research and transdisciplinary engagement, that the life sciences community will move from a "single scientist-single project" model to collective innovation.

The first Alliance workshop was held in conjunction with Super-Computing-11 in November 2011 in Seattle, WA. The event's morning session attracted an overflow crowd. Topics tackled in the afternoon included DELSA Recruitment and Structure, Data Access and Sharing, DELSA Identification of Grand Challenges, and DELSA Benefits and "Added-Value." The second workshop, in May 2012, was held in Washington, DC. The by-invitation-only workshop, "Supporting Ecosystem for the Life Sciences in the Beginning of the 21st Century," hosted more than 80 participants of diverse expertise from nine countries—Belgium, Canada, Germany, India, Israel, People's Republic of China, Russian Federation, United Kingdom, and United States. The participants discussed a range of topics, including new transformative ways of doing data-enabled, data-intensive, data-driven discovery in life sciences; accelerators, successful projects, and criteria to evaluate them; and DELSA Global as a worldwide community development alliance.

A major outcome of the second workshop was the identification of high-potential and community-endorsed projects for the DELSA mission. Eight of these projects were subsequently selected by the Board of Directors as DELSA-Endorsed Projects (*Table 1*). This set of

Table 1. List of DELSA–Endorsed Projects*
1. Social networking platform/matchmaking community portal
2. Data set accessibility project
3. Training data scientists
4. Global protein atlas
5. Internet2 application
6. Virtual 3D cell
7. Pregnancy atlas use case
8. ParaMEDIC use case
*www.DELSAglobal.org

projects is the focus of continued discussion by the community. DELSA is facilitating the establishment of project groups and will aid these projects in a number of crucial ways, including through collaborative connections and assistance in identifying funding resources.

DELSA Global is neither a funding agency nor a consortium, but an alliance of individuals and organizations from around the world with a common commitment to the future of data-enabled life sciences. Each of these eight opportunities is open to further development and enhancement, and all members of the community are invited to indicate interest and submit comments or suggestions. We expect varying levels of success for the projects, that is the reality of innovation—but each one could contribute solutions to the challenges we face.

As a way forward for sustained big data community engagement, DELSA Global has adopted a semi-annual workshop approach. Each year we will hold a two-day workshop in the spring to develop and review DELSA-Endorsed Projects, and identify new avenues for DELSA Global impact. In the fall we will host a 1-day workshop, in conjunction with another organization when possible, to further Alliance partnerships, exchange information, and update communities on DELSA efforts. This year we will be hosting a one-day workshop on October 9th in Chicago as part of the 8th IEEE Annual International Conference on eScience (www.ci.uchicago.edu/escience2012/). This workshop will serve as a means to introduce DELSA to the eScience community and to update the DELSA community on the progress of the DELSA-Endorsed Projects.

Catalyzing Innovation with Big Data: Where to from Here?

DELSA's inception comes during a critical time, when the US federal government is undeniably recognizing the need for accelerated development of the bioeconomy and new ways of conceptualizing knowledge-based innovation. On April 26, 2012, the White House unveiled its National Bioeconomy Blueprint, "a comprehensive approach to harnessing innovation in biological research."[3] The Blueprint identifies five strategic objectives: (1) support R&D investments that will provide the foundation for the future US bioeconomy; (2) facilitate the transition of bioinventions from research lab to market, including an increased focus on translational and regulatory sciences; (3) develop and reform regulations to reduce barriers, increase the speed and predictability of regulatory processes, and reduce costs while protecting human and environmental health; (4) update training programs and align academic institution incentives with student training for national workforce needs; and (5) identify and support opportunities for the development of public-private partnerships and precompetitive collaborations where competitors pool resources, knowledge, and expertise to learn from successes and failures.

All of these points are being discussed by DELSA Global members as efforts continue to build relationships and infrastructure that will help life scientists tackle 21st century challenges. The DELSA-Endorsed Projects support these objectives, and the Alliance

New Peer–Reviewed Journal on Big Data

In a world where we create over two quintillion bytes of data every day, global leaders in academia, industry, and government are grappling with the problem of how to organize, store, evaluate, share, and protect this vast amount of information. To address the questions surrounding this powerful and growing field of data discovery, **Mary Ann Liebert, Inc., publishers** announces the launch of *Big Data*, a highly innovative, peer-reviewed journal that will provide a unique forum for world-class research exploring the challenges and opportunities in collecting, analyzing, and disseminating vast amounts of data, including big data analytics. The Journal will launch in 2013.

A multidisciplinary editorial team of opinion leaders is gathering to build this new forum for the big data community, including Editor-in-Chief **Edd Dumbill**, Principal Analyst & Program Chair, O'Reilly Media; Executive Editor **Eugene Kolker, PhD**, Chief Data Officer, Seattle Children's, Co-founder & President, DELSA Global, Seattle, WA; **Geoffrey Charles Fox, PhD**, Associate Dean for Graduate Studies & Research, Professor of Computer Science and Informatics, Indiana University, Bloomington; **Sorin Istrail, PhD**, Julie Nguyen Brown Professor of Computational and Mathematical Sciences, Professor of Computer Science, Brown University, Providence, RI; **Folker Meyer**, Computational Biologist, Argonne National Laboratory, Argonne, IL, Senior Fellow at the Computation Institute at the University of Chicago, Chicago, IL, and Associate Division Director of the Institute of Genomics and Systems Biology, Chicago, IL; **Claudia Perlich, PhD,** Chief Scientist, Media6Degrees, New York, NY; **Gregory Piatetsky–Shapiro, PhD,** President, KDNuggets, Boston, MA, Founder of Knowledge Discovery in Database (KDD), and Co-founder of ACM SIGKD; **Anne Wright,** Director of Operations, BodyTrack Project, CREATE Lab, Carnegie Mellon University, Pittsburgh, PA; and **Rick Stevens, PhD,** Associate Laboratory Director for Computing, Environment, and Life Sciences at Argonne National Laboratory, Chicago, IL, Professor of Computer Science, University of Chicago, Senior Fellow of the University of Chicago & Argonne National Laboratory Computation Institute (CI), and co-Director of the Argonne Futures Lab.

continues to grow and support national and international transdisciplinary efforts while examining the novel paths to knowledge-based innovation and bioeconomy. The most fruitful areas of innovation are frequently where fields of expertise combine and perspectives widen. With DELSA Global, the progression of data to knowledge to action can be accelerated, transforming and realizing the development of innovative solutions.

Contact Eugene Kolker, Ph.D. Chief Data Officer, Bioinformatics & High-Throughput Analysis Laboratory, Seattle Children's Research Institute, C9S-9, 1900 9th Avenue, Seattle, WA 98101. Phone: 206-884-7170. Fax: 206-987-7175. Email: eugene.kolker@seattlechildrens.org or eugene.kolker@delsaglobal.org

REFERENCES

1. Kolker E, Stewart E. Data to Knowledge to Action. *The Scientist* (2012) Available at: http://the-scientist.com/2012/04/18/opinion-data-to-knowledge-to-action (Last accessed July 2012).

2. Kolker E. Greater than the sum of its parts: DELSA Global and collective innovation. *Seattle Trade Alliance Newsletter.* Summer 2012.

3. European Commission. Europe 2020, EU-Wide Targets. (2010) Available at: http://ec.europa.eu/europe2020/targets/eu-targets/index_en.htm (Last accessed June 2012).

4. Office of Science and Technology Policy Press Release. National Economy Blueprint, Executive Summary and Fact Sheet (2012) www.whitehouse.gov/administration/eop/ostp/library/bioeconomy (Last accessed June 2012).

4. European Commission. Europe 2020, EU-Wide Targets. (2010) Available at: http://ec.europa.eu/europe2020/targets/eu-targets/index_en.htm (Last accessed June 2012).

5. Ozdemir V, Muljono DH, Pang T, et al. Asia-Pacific Health 2020 and Genomics without Borders: Co-Production of Knowledge by Science and Society Partnership for Global Personalized Medicine. *Curr Pharmacogenomics Person Med* **2011**;9(1):1–5.

6. National Science Board. Advisory Committee for Cyberinfrastructure (ACCI) Task Force Reports. (2009) www.nsf.gov/od/oci/taskforces (Last accessed June 2012).