Working Group Report on Innovation and Partnerships
In Preparation for the Seattle DB Meeting, October 9-10, 2018

As a community we aim to encourage and facilitate highly innovative research. While easily stated, this goal is challenging for a variety of reasons. For one thing, innovation is difficult to judge in the moment. Ideas that appear innovative in one context, may have already been discovered in a different context, with similarities obscured by differences in terminology and emphasis. Alternatively, in an active research area such as databases, it's quite possible that an idea was not previously published, not because no one had ever thought of it, but rather, because it turns out to be not a very good idea. Finally, some of the most innovative ideas are too far ahead of their time, making it difficult for the community to evaluate them.

Despite these challenges, it is imperative for our community to foster and demand innovation. To do this successfully, we must first acknowledge and embrace the idea that innovation comes in many forms, and is difficult to measure. As a community we often find ourselves falling back on various proxy metrics: paper counts, dollars raised, h-indices, revenue, number of users, popular press mentions, etc. These proxies are at best inaccurate and at worst misleading. Furthermore they evolve over time: papers get ignored and rediscovered, hot products become outdated, etc. On the other hand, some proxies are more useful than others, and the metrics we choose to emphasize say a lot about who we are as a community.

We believe that the healthiest approach that our community can take towards innovation is to have an expansive view, to foster and encourage pushing boundaries of all kinds, and to let time sort out the true innovations from the rest.

In terms of partnerships, there is a widespread agreement that our community stands to benefit greatly and can have tremendous impact by encouraging many sorts of partnerships. We need to foster partnerships between academia and industry, with other areas in computer science such as ML, HCI, Security, Systems, etc., and with the many domains in the sciences and humanities that are becoming or need to become more data-driven. There has never been a better time to find collaborators with real data problems who are willing to engage. As an academic, you can simply walk down to the next department and find real data problem with real challenges that the "customer" is willing to invest in/collaborate.

Below we address some of the key things that our community must do in terms of innovation and partnerships.

Innovation: Encouraging and Supporting Systems Work and Engagement

There is little incentive under the existing metrics for rewarding academic researchers in doing much more than pursuing least-publication-units. The academic reward system does not typically value actual product impact. Many academics also don't want to take the time to really understand the problems that end-customers face or the industry faces, as this is a time-consuming process and often does not lead to immediate publication.

In addition, there is a large tendency by program committees (made of both academic and industry researchers) to dismiss and reject papers that are focused on end-to-end-systems designs as these are considered as "engineering" projects and thus not "novel". Perversely, this attitude is also sharply present in funding reviews at agencies like NSF (which tends to have similar reviewers as PCs). Thus, junior faculty are quickly trained to not chase projects involving building actual systems. As a contrast, consider an example from outside the field: The construction of the Hoover dam. This was a massive project that was a "Test of Engineers' Theories" (cf. C. G. Sims, "Hoover Dam was Test of Engineers' Theories", New York Times, Oct 15, 1985). A parallel idea of building a large system to test end-to-end complex set of existing methods often will simply be rejected by our top conferences and review panels under the existing mindset.

We simply need to expand our notion of what Innovation means not just in furthering the research in our community (e.g., how we review papers), but more importantly to stand out as a community that has "Impact" in the broader field of Computer Science and society in general.

To this effect, an actionable item is to have a sustained 10 year funding marked primarily for the development of data related tools that are develop in open-source. Just like the Hoover dam project, there is higher value to be delivered and much to be learnt from hardening research ideas into actual systems and tools that are usable by others outside the community. While there are small amounts of funding directed to such tasks (e.g. the NSF CSSI program that funds 3-year projects) the timeline for building usable data-related software tools requires a longer-term effort, and more importantly a mechanism to continue funding tools that has demonstrable tractions beyond the initial funding.

Another key form of innovation is taking early ideas and spinning that off into companies with venture capital financing. The database community has been quite successful with this, but the know-how about how to do this is only disseminated informally. Conferences could do more to more formally encourage such dialogs between aspiring entrepreneurs and veteran entrepreneurs by holding panels, hosting entrepreneurial experiences talks, and small group meetings.

Partnerships: Academia and Industry

Since its earliest days, the database community has benefited from close interaction and intellectual collaboration between academia and industry. Being a practical discipline, data management research problems and agendas have often been driven by real-world challenges. Our students go on to both industry and faculty jobs, and individual researchers move between research and academia. There is, however, a mounting concern that the pendulum is swinging a bit towards industry careers these days. **In the long run, a brain drain from academia to industry is likely to harm both.** Academia trains the industrial innovators of the future, not just to solve today's problems but to learn how to create and assimilate new knowledge as the technology and societal landscape evolves. The timescales, values and rewards of academia and industry differ in significant ways, which provides homes for a variety of types of innovation

and impact. Thus, we benefit from ensuring that both are vibrant and healthy and we can benefit even more by encouraging collaboration, idea and people flow between them.

Innovation and Partnerships: Grand Challenges

As a community, we lack a small number of grand challenges where a number of researchers can rally around and advance the state of the art. We should identify a set of Grand Challenges that collectively help achieve a number of goals: More work on useful and usable end-to-end systems, development of useful Open Source systems, collaboration among large groups of database researchers, collaboration with other computer scientists, collaboration with researchers across the sciences, humanities, policy and other areas, and more meaningful interaction between industry and academia. The definition of such a set of challenges would be a major output from this meeting.

One example of a grand challenge is to address the end of the "free ride" that Moore's law with Dennard's scaling has offered. Since data growth continues exponentially in many domains and exponential speedup in single thread performance has ended (and not expected to come back), a grand challenge is to run a benchmark (e.g. TPC-DS) with 25% lower energy consumed per unit of data (e.g. at 1PB) per year, while keeping performance constant. Can we do this for two decades using as reference the faster open-source that is available today?

A second potential grand challenge is the creation of an accurate, real-time "fact-checker" that could assess the validity of claims made in political speeches and social media. This challenge would involve collaboration with researchers in NLP and other areas, and could have both automated and human-in-the-loop variants.

Summary

History has shown that real innovation is nearly alway only clear in hindsight. Thus academic researchers, industry, and funding agencies should take a broader view of research. In addition, the practical and lasting value of research ideas only crystalize when they are put to use in actual end-to-end working systems. Far more emphasis, in both evaluation of academic performance and funding agency priorities, must be put in end-to-end systems and tools building, especially in open-source. Funding agencies should also be encouraged to continue funding open-source tools that show demonstrable tractions beyond the initial funding that seed the effort.

Academics are encouraged to make as early connection as possible to actual problems. Given the ubiquity of data problems, it is trivially easy for academics to access real problems and forge synergies with cross-domain collaborators. In many cases, the academic does not even have to leave campus to find interesting data problems to solve with "real customers." In addition, with the vast amount of open-source software there has never been a better time to leverage suitable building blocks from open-source to build new systems and tools. There is also a powerful learning element associated with such exercises, which has obvious pedagogical value.

Finally, conferences are encouraged to find mechanisms to allow aspiring entrepreneurs to learn from veterans, as often a powerful way to take ideas to actual practice is via startups.

Action items for the community

What academia and academics can do

- Be more open minded about what good work is
- Encourage systems work and avoid systems reviewing pitfalls
- Encourage big ideas but understand that "delta papers" are a large part of how science progresses
- Reward faculty and students for building real systems and prototypes, open source
- Recognize that data people are in high demand and compensate them appropriately, as is currently done by business and economics faculty, among others
- Provide flexible arrangements with industry engagement such as split appointments between a university and a company or lab and create policies that encourage rather than penalize engagement with the outside world, such as starting companies.
- Encourage interaction and joint work among multiple academic database research groups - there are surprisingly few examples of this happening
- Encourage interaction with industry, with other domains, and with groups across institutions
- Provide mentoring by successful innovators to young faculty and students who want to have real-world impact

What funding agencies and promotion and tenure committees can do

- Instruct review panels and tenure committees to have an expansive view of innovation and impact including creating real systems, starting companies, etc.
- Develop policies that enable and encourage faculty and students to pursue these alternative paths for innovation and impact
- Work with companies to build new funding programs
- Provide access to computing infrastructure, data resources etc of a magnitude that is competitive with what industry is able to provide

What companies, industry researchers and practitioners can do

- Provide academics with access to problems, data, logs, workloads, etc
- Avoid restrictive employment agreements that prevent company people from collaborating outside the company or on Open Source projects
- Support academic research through membership in research consortia
- Support students and faculty through fellowships, internships, research visits
- Reward employees for participation in the greater research community
- Enable employees to do "sabbaticals" in universities for teaching and research
- Help remind funding agencies of the crucial role that academic research plays in the data ecosystem

Innovative Models

- Company collaborative sponsorships such as: AMP/RISE/DAWN/DSAIL research collaborations
- NSF Funding Calls that include Corporate funds (e.g., Intel and VMWare)
- Split appointments: Faculty and Industry
- Enabling PhD students do their thesis work at companies, and allowing qualified company practitioners to serve on their committees
- Industry outreach programs such as Faculty forums (e.g., MSR, Google, Facebook)
- Placing lablets near universities (e.g., old Intel lablets, various Google labs, etc.)
- Grand Challenges that engage many parts of the community

Breaking down research silos

- DB and Systems support for ML and vice versa
- We should play a key role in Data Science
- Working with data- and compute- intensive scientists and rewards for doing so
- Social good and policy impact

Questions to consider at the meeting

1. What grand challenges can we propose?
2. What can we do to foster innovation that is being stifled today at various levels, including funding agencies and program committees?
3. What models of collaboration between industry and academia work? Which ones don't work and why?