Conclusions

TONY HEY, STEWART TANSLEY, AND KRISTIN TOLLE | Microsoft Research

By the mid-1990s, Jim Gray had recognized that the next "big data" challenges for database technology would come from science and not from commerce. He also identified the technical challenges that such data-intensive science would pose for scientists and the key role that IT and computer science could play in enabling future scientific discoveries. The term "eScience" was coined in the year 2000 by John Taylor, when he was director general of the UK Research Councils. Taylor had recognized the increasingly important role that IT must play in the collaborative, multidisciplinary, and data-intensive scientific research of the 21st century and used the term eScience to encompass the collection of tools and technologies needed to support such research. In recognition of the UK eScience initiative, Jim Gray called his research group at Microsoft Research the eScience Group, and he set about working with scientists to understand their problems and learn what tools they needed.

In his talk to the Computer Science and Telecommunications Board of the U.S. National Research Council in 2007, Jim expanded on his vision of data-intensive science and enumerated seven key areas for action by the funding agencies:

1. Foster both the development of software tools and support for these tools.
2. Invest in tools at all levels of the funding pyramid.
3. Foster the development of generic Laboratory Information Management Systems (LIMS).
4. Foster research into scientific data management, data analysis, data visualization, and new algorithms and tools.

5. Establish digital libraries that support other sciences in the same way the National Library of Medicine supports the bio-sciences.
6. Foster the development of new document authoring tools and publication models.
7. Foster the development of digital data libraries that contain scientific data (not just the metadata) and support integration with published literature.

We believe that these challenges to the funding agencies are just as important today. This is why we have introduced this collection of essays, along with a version of Jim's talk to the NRC-CSTB constructed from the transcript of his lecture and his presentation slides. It is also educational to see the continuing momentum and progress of the eScience community since the report "Towards 2020 Science" published by our colleagues at Microsoft Research, Cambridge, UK.[1] That was based on a workshop in July 2005, attended by some of the authors in this new book, and subsequently inspired *Nature's* "2020 Computing" special issue in March 2006.[2]

At the heart of scientific computing in this age of the Fourth Paradigm is a need for scientists and computer scientists to work collaboratively—not in a superior/subordinate relationship, but as equals—with both communities fueling, enabling, and enriching our ability to make discoveries that can bring about productive and positive changes in our world. In this book, we have highlighted healthcare and the environment, just two areas in which humanity faces some of its biggest challenges. To make significant progress, the research community must be supported by an adequate cyberinfrastructure comprising not only the hardware of computing resources, datacenters, and high-speed networks but also software tools and middleware. Jim also envisaged the emergence of a global digital research library containing both the research literature and the research data. Not only are we seeing the maturing of data-intensive science, but we are also in the midst of a revolution in scholarly communication. This is driven not only by technologies such as the Internet, Web 2.0, and semantic annotations but also by the worldwide movement toward open access and open science.

This book is really a labor of love. It started with Jim's desire to enable scientific research through the technologies of computer science—cutting across the disciplines highlighted herein and beyond. We see this book as a continuation of Jim's work with the science community. We deliberately asked our scientific contributors

[1] http://research.microsoft.com/en-us/um/cambridge/projects/towards2020science/background_overview.htm
[2] *Nature*, vol. 440, no. 7083, Mar. 23, 2006, pp. 383–580.

to move out of their professional comfort zones and share their visions for the future of their research fields on a 5-to-10-year horizon. We asked them to write their contributions not only in essay form, which is often a greater challenge than writing a purely technical research article, but often in collaboration with a computer scientist. We are grateful to all of our contributors for rising to this challenge, and we hope that they (and you!) will be pleased with the result.

Several decades ago, science was very discipline-centric. Today, as evidenced by the articles in this book, significant advances are being made as a result of multidisciplinary collaboration—and will continue to be made into the future. The essays in this book present a current snapshot of some of the leading thinking about the exciting partnership between science and computer science—a data revolution—which makes this information timely and potentially fleeting. However, it is our fervent hope and belief that the underlying message presented by the totality of these articles will be durable for many years.

Finally, we offer this book as a call to action for the entire research community, governments, funding agencies, and the public. We urge collaboration toward a common goal of a better life for all humanity. We find ourselves in a phase in which we need to use our scientific understanding to achieve specific goals for the sake of humanity's survival. It is clear that to achieve this aim, we very much need experts with deep scientific knowledge to work closely with those who have deep experience with technology.

This situation is somewhat analogous to the 1940s, when U.S. and European physicists answered an urgent call from governments to collaborate on the Manhattan Project. Today, scientists must collaborate globally to solve the major environmental and health problems facing humanity in a race that is perhaps even more urgent. And ironically, the nuclear physics developed in the Manhattan Project is likely to provide part of the answer in supplying the world with zero-carbon energy.

Tony Hey, Kristin Tolle, and Stewart Tansley

Microsoft External Research, http://research.microsoft.com/collaboration

NEXT STEPS

WE HOPE THIS BOOK WILL INSPIRE YOU to take action as well as embark on further study. We are "walking the talk" ourselves at Microsoft Research. For example, we have reformulated our academic partnership organization, External Research, to focus on the themes presented in this book.

These themes incorporate active research in dynamic fields, so it is hard to track and predict the future evolution of the ideas presented in this book. But here are some suggested ways to remain engaged and to join in the dialogue:

- If you're a scientist, talk to a computer scientist about your challenges, and vice versa.
- If you're a student, take classes in both science and computer science.
- If you're a teacher, mentor, or parent, encourage those in your care toward interdisciplinary study in addition to giving them the option to specialize.
- Engage with the editors and authors of this book through the normal scholarly channels.
- Keep up to date with our eScience research collaborations through our Web site: http://research.microsoft.com.
- Be active in the eScience community—at the Fourth Paradigm Web site below, we suggest helpful resources.

www.fourthparadigm.org