Foreword

The conventional image of scientific research includes two contrasting memes: the dogged/brilliant individual researcher making their breakthroughs in isolation or in a small laboratory and the industrial-sized team working on big problems, which may or may not be relevant to the lives of the masses in the short term. The reality encompasses both of these extremes and many points in between. What is not in dispute is the proposition that scientific progress involves the generation of shared knowledge, systematically expressed, perhaps systematically gathered or generated. What definitely varies is the degree of direct collaboration involved and the way that the networks that support it are constructed.

For some areas of observationally centered research, scale is a determining factor. Ships, satellites, radio telescopes, and particle accelerators require a collaborative approach, at least to the management and funding of the initiatives that use them. But the emergence of ubiquitous communication channels has encouraged an expectation that researchers will collaborate more generally. My own engagement with national and particularly international collaborators was only possible by the emergence of email, the Web, wikis, and other collaborative online workspaces, which support construction of virtual teams that can interact daily or better. A significant impact of this toolkit might be expected for informatics-based projects, but perhaps more impressive examples of large-scale collaborative knowledge development using online tools are found in a variety of natural sciences, most notably biomedicine and biodiversity. Meanwhile, e-research has emerged as something of a discipline in its own right, with an ecosystem of funding calls and conferences in support.

What are the ingredients for successful collaboration networks, and how can we evaluate them? Dipping into my experience, the development of rich data transfer standards through the International Union of Geological Sciences (IUGS) Commission for Geoscience Information (CGI) was grounded in a strong existing consensus on the basic scientific model (historical geology, exemplified in geological maps with their stratigraphic legends) supported by very standardized institutional arrangements (government-backed geological surveys). Each of these is possibly unique in the scientific realm in terms of their longevity and near-ubiquity, so the project was more about formalizing than developing knowledge. But while the semantic sophistication of the results is clear from a cursory inspection of the elaborate models and encodings, their impact is less clear. The primary delivery portal “OneGeology” now boasts over 100 national members, but fewer than one-fifth are delivering actual data (as opposed to map images), and many of these are limited to what is essentially a cartographic, rather than scientific, product. In response, CGI is now stripping the models back to a more implementable core, meeting the community where it is, rather than where we hoped it was. Another large-scale collaborative knowledge initiative I participate in—technology standards for spatial data infrastructures—has struggled with mandate and conflicting technical demands of different parts of the community. The requirements of experts from
the defence/intelligence sector and other industries dominated by large enterprises are largely alien to myriad developers of location-aware applications supported by GPS and Google maps. The latter have never heard of, or see little application for, “coordinate reference systems” or “interior boundaries,” for example. In this domain, the European experiment of using legal instruments to enforce compliance with formal standards by public sector agencies is also looking distinctly shaky and has little impact on the mass market anyway, which is often where the most dynamic applications are found.

Now we are developing some experience with self-consciously collaborative knowledge projects, a more systematic and perhaps formal approach and assessment is needed. That is what this book addresses. There are numerous concerns. The issue of frameworks, paradigms, and biases is tackled in the first group of chapters, which considers the degree to which collaboration is reinforced in certain networks and undermined when network participants are talking past each other. Back to the geosciences, the incomprehension, even disdain, between geologists and geophysicists is a conventional trope (storytellers both, but with different weightings to means and end), and I recently encountered a disjuncture between a pair of sub-disciplines whose names are identical, except that the qualifier and noun are flipped (geo-hydrology and hydro-geology). The functioning of collaboration networks in terms of a workflow, and the utility of standardized tools and components emerges from a set of case studies presented in the second group. The addition of a provenance language to the gamut of Semantic Web vocabularies is notable here. The third group considers the formalization of syntax and semantics in a community, approached in a number of ways. Top-down (“the experts”) vs. bottom up (“the people”) is a common contrast. The optimal point of adoption of standards (which should merely formalize community practice) is unclear: while some initiatives attempt to gain efficiencies by pushing responsibility for conformance back to source, this is not always achievable, and intermediaries are likely to have a role. The fact that the consensus may be continuously evolving is tackled in the fourth group, which discusses various approaches to formalizing dynamic knowledge. Finally, a challenge is presented: perhaps knowledge is best expressed informally, thus accommodating conflicts without breaking the system.

This book does not present a uniform theory, but it highlights some key considerations through a fascinating range of case-studies, mostly drawn from natural sciences but leavened with a couple of excursions into the humanities. Sharing of data has an associated requirement for propagation of provenance and uncertainty information. Formalization needs to accommodate distribution requirements and also support change in the consensus. Above all, collaborative knowledge development needs to be sensitive to the issue of reconciliation of different points of view, paradigms, and values. As a scientist whose (second) career has been based on almost daily participation in collaboration networks, I have reflected on why it appeared to work, and then on whether it really had, and what the alternatives are (Cox, Simons, & Walker, 2011). This book presents some frameworks for understanding this better. Any scientist who is not in their own ivory tower and is inclined to think about method will find it a stimulating and worthwhile read.

Simon Cox
CSIRO Land and Water in Melbourne, Australia
24th June 2014
Simon Cox trained as geophysicist, with a PhD from Columbia University. His work on informatics started in the early 1990s with GIS, moved quickly to the World Wide Web, and he became involved in metadata standards on the Dublin Core Advisory Council. Work on XML for mineral exploration data led on to the GeoSciML project and participation with the Open Geospatial Consortium, where he co-edited the Geography Markup Language standard. He developed Observations and Measurements as an OGC and ISO standard, which now forms the basis for operational systems in diverse fields including air-traffic, water data transfer, and environmental monitoring applications. He spent a year as a senior fellow at the EC Joint Research Centre in Italy working on integration of GEOSS and INSPIRE. He currently has leadership positions in the OGC, ISO/TC 211, and the Research Data Alliance. In 2006, he was awarded OGC’s Gardels Medal, and he presented the 2013 Leptoukh Lecture for the American Geophysical Union. Simon is currently based in CSIRO Land and Water in Melbourne, working on a variety of projects across environmental informatics, linked data, and semantics.

REFERENCES