That modern scientists face a “data deluge” has almost become a cliché. Jim Gray and other top computer scientists describe how the techniques and technologies needed to perform data-intensive science now comprise a new “fourth paradigm” of research. This is a valuable insight, but to date, its emphasis has been on overwhelming data volumes, while it has underplayed the diversity of interdisciplinary data and the need to interrelate these data to understand complex systemic problems such as environmental change and its impact.

The International Polar Year (IPY) was a US$1.2 billion investment in polar research involving some 50,000 participants from 60 nations. It emphasized interdisciplinary collaboration and broad understanding of complex, intertwined, physical, biological, health, and social systems. It also generated many so-called research data collections, which are poorly addressed in current data systems. IPY and its data provide an excellent case study to understand the interdisciplinary dimension of the fourth paradigm.

In this paper, we discuss the experience of data scientists collecting, curating, and providing data from IPY to address interdisciplinary science goals. We find that while technology is a critical factor in addressing the interdisciplinary dimension of the fourth paradigm, the technologies developing for exa-scale data volumes are not the same as what is needed for extremely distributed and heterogeneous data. Furthermore, the greater challenges are more sociocultural than technical. We describe a vision of discoverable, open, linked, useful, and safe data and suggest the need for a rapid sociotechnical evolution in the overall science data ecosystem.