Data-intensive research is producing ever-higher volumes of digital data, and pressures are mounting to make that data openly accessible. Research funding agencies, both in the United States and abroad, want to demonstrate a strong return on investment by making the results of federally funded research—both articles and data—openly accessible to the public. Therefore, managing data with long-term preservation and reuse in mind is important; however, many people tasked with managing research data, such as graduate students, have very little data management training and often employ inconsistent practices. The revolving door of incoming and outgoing graduate students also creates uncertainty about the data, its whereabouts, its status and sometimes even its accuracy.

To establish data trustworthiness, data management training programs are needed within higher education research institutions. Ideally, training for these skills would be integrated within the science, social science and engineering curricula so that students learn them within the context of their chosen fields. However, until this long-term goal is realized, institutions are introducing training programs, often from within the library, to educate researchers in effective data management practices.

What options are available to information professionals planning data management programs at their institutions? Several programs of varying approaches are already available, so it is not necessary to create a new training program from scratch. I briefly discuss here three representative programs that I used when developing training.

The first program is the New England Collaborative Data Management Curriculum (NECDMC) (http://library.umassmed.edu/NECDMC/index), which was developed by the Lamar Soutter Library at the University of Massachusetts Medical School in partnership with libraries from the Marine Biological Laboratory and Woods Hole Oceanographic Institution, Northeastern University, Tufts University and University of Massachusetts at Amherst. The NECDMC comprises seven modules. One module is an overview of the entire curriculum, and the other six modules cover different aspects of managing data such as data sharing, data preservation and metadata. PowerPoint slides and lecture content are available for download for each module. One strength of the NECDMC lies in its use of research case studies to demonstrate and apply the concepts in practice. These case studies cover fields such as medical, engineering and qualitative research. Another advantage of this program is its flexibility in delivery method and timeframe. For
those wanting a short overview session, the first module can be covered easily in 60-90 minutes. Alternatively, all of the modules can be the foundation of a one-day workshop or be expanded to fit a semester-long course.

The second program consists of the DataONE Data Management Education Modules (www.dataone.org/education-modules), which are a series of lessons covering tools and best practices for each stage in the DataONE Data Life Cycle. Like the NECDMC, this program is modular and can be used as the foundation for shorter or longer courses. Although the topics covered are similar, the DataONE modules cover more than the NECDMC modules. For example, the DataONE modules include topics such as workflows, data entry and quality control, which are very important for researchers, especially those within the earth and environmental sciences, the intended audience for DataONE resources. Also, they are designed to be used as a self-study course, so the content on the presentation slides is heavy. If used as the basis for lecture-style sessions, the slide content should be reduced to minimize the text.

The third program for teaching researchers data management best practices is the MANTRA (http://datalib.edina.ac.uk/mantra/) course developed by the University of Edinburgh. This program is modular, entirely web-based and uses the Xerte online learning environment, so the training modules are self-paced and interactive. The creators provide suggestions on where to begin the course based on role, such as research student, career researcher, senior academic or information professional. Each module provides an overview of the topic and provides videos, short quizzes and additional resources. Keep in mind this curriculum is geared towards researchers in the United Kingdom, so modification for the United States may be necessary in some parts.

At the University of Tennessee, I have offered graduate students from various scientific and engineering backgrounds a one-day data management workshop which was based entirely on the NECDMC. While this delivery method may be feasible for some, it was very preparation-heavy and ostensibly overwhelming for the students. I am applying these lessons to future workshops by designing a series of one-hour sessions on relevant topics and tailoring each to different disciplines.

I believe the long-term goal of injecting data management principles into the science, engineering and social science curricula is the most effective way for students, both undergraduates and graduates, to synthesize these skills. However, until the time such integrated data training exists, students will need to learn these skills elsewhere. Each institution will want a uniquely tailored program best suited to its environment; for example, you may decide a semester-long, for-credit course is most appropriate at your institution. Whatever your chosen method, I hope one or a combination of these three educational programs will help you to meet an immediate need on your campus.