Over the past three decades, shared scientific databases have evolved from the narrow confines of critically analyzed measurement tables (such as properties of materials) to collections of records covering a larger scientific scope (such as real-time data sets and imagery). Much of this evolution has happened concurrently with—and followed new capabilities emerging from—the evolution in information technologies.

Scientific records, especially in the past decade, have shifted from paper to digital form: memos and laboratory notebooks are giving way to email and electronic documents. These transformations occurred most in the areas of content and practice. Databases and research records contain not only a greatly expanded scope of content items, but also reflect new, more-complex uses and practices.

The Emergence of New Systems
To say that developments in information technologies have played significant roles as drivers of these evolutionary changes is a considerable understatement. Just consider the way scientific and engineering databases’ content and uses have changed. Thirty years ago, we used critically analyzed compendia of technical data that were periodically published as monographs. Data reports largely consisted of computer printouts—long tables that were arguably human-
readable but certainly not machine-readable.

Although cumbersome to use, these systems served us well as long as a desktop computer with information or tools into which we could ingrate archived data didn’t yet exist. Absent the need to search and re-search this data (often in different ways, as the analysis proceeded) or absent modeling tools that pretty much require input data to be machine-readable, these archival databases were adequate for science and engineering tasks. However, the empowerment that cheaper, more-accessible computers and computing tools afforded increased the demand for more timely and efficient access to archival scientific and engineering data. The Internet and its most famous child, the World Wide Web, have genuinely transformed expectations of what and how data and information should be accessible.

Think about how we stored and accessed laboratory measurement data in the research lab just a decade ago. Data from instruments, increasingly produced under computerized instrument control, were contained in files, often in proprietary formats. Generally, file names were the only digital identifiers of data sets, file dates the only digital metadata, and file system directories the only database structures for primary data. Metadata were mostly in paper notebooks and cross-referenced to their digital data counterparts via file names. If you were clever and disciplined enough to use descriptive file names when acquiring measurements, the names might actually be somewhat meaningful as data descriptors, permitting a measure of reflexive reference of data sets to their metadata.

Although tedious to use for retrieving and analyzing data, this system was workable to the extent that these primary data did not need to be shared with or interpreted by others, especially outside of a small research group and most especially with professionals beyond the field of expertise. Of course, there were exceptions (drug and chemical companies developed and used LIMS systems, for example), but these were largely restricted to cases in which scientific and engineering problems could be compartmentalized, and reports and papers could pass processed data on to others in the knowledge chain. Today, scientific and engineering problems that require groups of cross-disciplinary researchers to collaborate seamlessly in near real time are more the rule than the exception.

The Articles
Interestingly, it is computing progress that has brought databases and scientific work ever closer together. This conjunction is the motivation for this topical issue of Computing in Science & Engineering. In it, you’ll find four illustrative examples of database-related concepts that showcase emerging uses of evolutionary technologies that organize archival information.

“A concept overarching all the examples in this issue about modern databases, tools, and methods is the Open Architecture Information System (OAIS) reference model. OAIS provides a framework that addresses a full range of archival functions in which archival information system concepts can be understood and applied.

This unifying model offers a promise to standardize and rationalize the design and implementation of future archival collections. The editors of this magazine hope to treat this concept in an appropriate manner in a later issue, but if you’re interested in learning about it now, you’ll find an overview of OAIS at http://ssdoo.gsfc.nasa.gov/nost/isoas/overview.html. You can retrieve the latest draft of the model definition from www.classic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf.

Where to Go Next
A concept overarching all the examples in this issue about modern databases, tools, and methods is the Open Architecture Information System (OAIS) reference model. OAIS provides a framework that addresses a full range of archival functions in which archival information system concepts can be understood and applied.

This unifying model offers a promise to standardize and rationalize the design and implementation of future archival collections. The editors of this magazine hope to treat this concept in an appropriate manner in a later issue, but if you’re interested in learning about it now, you’ll find an overview of OAIS at http://ssdoo.gsfc.nasa.gov/nost/isoas/overview.html. You can retrieve the latest draft of the model definition from www.classic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf.

The editors of this magazine hope to treat this concept in an appropriate manner in a later issue, but if you’re interested in learning about it now, you’ll find an overview of OAIS at http://ssdoo.gsfc.nasa.gov/nost/isoas/overview.html. You can retrieve the latest draft of the model definition from www.classic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf.

The final article, “Re-Integrating the Research Record,” discusses tools to federate the results of scientific measurement, analyses, and problem solving from projects into novel knowledge bases that document a coherent record of research effort.

Norman Chonacky is a physicist and currently serves as a senior research scientist in Columbia University’s Earth and Environmental Engineering Department. His current research work is in methods of carbon sequestration, the design of real-time, low-cost, environmental sensors, and information systems for supporting collaborative, cross-disciplinary research practice. He has a BS in physics from John Carroll University and a PhD in physics from the University of Wisconsin, Madison. He belongs to the APS, the AAPT, and the AAAS. He serves on the editorial board for Computing in Science & Engineering and coedits its Technology News & Reviews department. Contact him at Columbia Univ., Mail Code 4711, New York, NY 10027; njc18@columbia.edu.