The question “What is a digital library?” is certain to elicit a multitude of answers. In current information science circles, the definition of a working digital library (DL) includes a vast array of projects ranging from small digital photo collections in a proprietary software solution to the HathiTrust’s mass-digitized book collection. Falling within those extremes, one might add green open access institutional repositories (some using proprietary platforms, some open source), gold open access journal publishers such as Public Library of Science (PLoS), digital journal aggregations from commercial publishers, and any text, digital learning object or streaming media collection accessible to users via the Internet. The term “digital library” is, therefore, protean and the applications of the technology are numerous.

Theoretical Foundations for Digital Libraries attempts to formulate a much more programmatic approach to defining and designing digital libraries. The pedigree for the book is impeccable. As a leader in the field of digital library development since the 1970s, Edward Fox’s writings have provided a solid basis for the library and information science community. Co-authors Marcos André Gonçalves is a professor at Universidade Federal de Minas Gerais, Brazil, and Rao Shen works for Yahoo! Laboratories. Their combined expertise puts forth a strong case for their vision of a five-pronged approach to defining and designing the digital library.

The authors postulate that digital libraries can be better implemented and designed if they are more clearly defined by a uniform conceptual framework, in this case the five Ss, which stands for societies, scenarios, spaces, structure, and streams. These are further explained as “Social networks, Scenario-based design, geoSpatial databases, Structure-based approaches (e.g., databases, metadata, ontologies, XML), and data Stream management systems.” The authors then use this model to explain various aspects of designing and interpreting the digital library to several cohorts of readers, especially library and information science (LIS), computer science, and engineering students. The authors also contend that practitioners and researchers will find this work useful.

On one hand this novel approach does appear extremely useful for those new to the field. The first chapter provides an introduction to digital libraries, including an overview of the basic terminology used for DLs, and a summary of how a curriculum about digital libraries might be taught to LIS students. The first chapter then summarizes and provides examples of the 5S system. The second chapter presents a couple of case studies that allow students to apply the 5S system in order to deconstruct a digital library. The appendices provide detailed mathematical formulae, formalized digital library concepts, notations and definitions, proofs, and a glossary of terms.

Some issues with the text arise, however. First, the audience for the book occasionally appears to be unclear. While conceptual framework diagrams and a glossary will be welcomed by all readers with an interest in digital libraries, the mathematical formulae and proofs may not be as useful to lay users or to general practitioners. These proofs will likely satisfy only readers well-versed in computer science or engineering and with very strong backgrounds in mathematics.

As a conceptual framework, the 5S system appears to be extremely useful for helping to define the sometimes loose conceptual basis and the often balkanized definition of a digital
The 5S “meta-model” serves to provide a graspable abstraction of the world. In some ways, this is where the book both excels and stumbles. In attempting to define “streams,” for example, the authors provide multiple definitions, including natural (rivers), mathematical (i.e., bits, characters), dynamic (i.e., temporal and finite sequences), and static (text-based characters). All of this is useful, but at the same time a bit too vague to apply as a serious conceptual framework for a digital library.

In many parts of the text, as a result, it appears as if the authors are struggling to fit examples to the framework. It does not seem very helpful overall, especially when faced with some extremely complex formulae later in the text, to provide a photo that merely metaphorically defines the concept. Similar uses of images occur throughout the book, including an image of a building, a picture of outer space, and so on. This reviewer feels that the text’s space might better be served with diagrams more specific to digital libraries. Problems with image quality on the print version of the book were also noticeable, with several images appearing blurred or pixelated.

Overall, the 5S framework will be extremely valuable to the LIS and computer science student, but at times the framework may need to be more clearly explained and applied to actual real-world examples. The online companion site to the book (https://sites.google.com/a/morganclaypool.com/dlibrary) does seem to provide much supplemental material in the form of diagrams, lectures, and discussions that address this need. It is hoped that subsequent volumes of this series will hone in on these applications. On a final note, the audience for the book appears to be both the typical LIS and computer science student. This provides the reader with a somewhat unfocused result due to the simplified explanations of the 5S system juxtaposed with extremely high-level mathematical formulae, proofs, and formalized diagrams.

Reviewed by Andrew Weiss
Digital Services Librarian
California State University, Northridge
Northridge, CA