Book Review

Elyssa M. Gould

The Data Wrangler's Handbook: Simple Tools for Powerful Results. By Kyle Banerjee. Chicago: ALA Neal-Schuman, 2019. 164 p. $67.99 softcover (ISBN 978-0-8389-1909-5).

Banerjee has had a long-time goal to make technological concepts and skills accessible and comprehensive to a non-technical audience. This recent publication is no different. It expands on the author's purpose to have people adopt technological tools for everyday library projects. It does this by introducing the analogy of a technology cookbook. Practically speaking, this book highlights several recipes using their computers' native tools to solve ordinary problems when managing library data.

Librarians and information professionals have at their disposal powerful tools that come with every computer to help assemble and manage text-based data. Banerjee illustrates how these tools can tackle a wide variety of library data issues. Through the examples, he details how to break down complex issues thus managing smaller, more comprehensible tasks where these tools can be applied. Thanks to this process, the reader can gain confidence with these tools and think through complex data issues.

The book structures concepts and tasks to build on one another. The first two chapters "Getting Started with the Command Line" and "Command Line Concepts" are critical as they constitute the foundation from which Banerjee creates more sophisticated data manipulations in later chapters. These first chapters can be difficult for those with no previous technological knowledge. With that in mind, even those already familiar with these tools would find helpful information and a handy reference of solutions. Chapter Three, "Understanding Formats" by David Forero, describes file formats typically encountered in library work. Chapter Four, "Simplify Complicated Problems" addresses how to break down complex questions and problems into more easily understood ones. Chapters Five, Six, and Seven, "Delimited Text," "XML," and "JSON (JavaScript Object Notation)," continue to delve more deeply into common data formats in library work. Chapter Eight, "Scripting," introduces the concept of creating unique files to run commands from Chapter One. Chapter Nine, "Solving Common Problems," presents a recipe guide to solve common issues encountered with library data. Chapter Ten, "Conclusions," wraps up the handbook by listing additional tips and tricks.

This reviewer found it powerful to follow every example throughout the book. Looking for more information online on topics such as options for Unix commands or regular expressions gave this reviewer more confidence in the themes being discussed. Trying the examples as written, plus variations to see what would happen, encouraged a trial and error approach. As a result, this reviewer began to see the potential of how these tools could solve many library data issues encountered in technical services.

Working through the examples was not without its challenges. This reviewer would have appreciated more explanations for many of the examples in chapters One and Two, and more guidance on how to handle errors or where to find help. Banerjee underscores the importance of adopting a trial and error approach while working through the examples in the handbook. This approach could have been stressed even more in the first two chapters and introduction. That would be helpful, especially for the novice reader. Even with that in mind, those with no knowledge of command line tools might need extra time to work through examples. For those with little to no knowledge, it is still worth spending more time on chapters One and Two.

Overall, this book is a solid starting point. Thanks to its recipe style approach, it can be the focus of a study group, classroom, or self-study. It includes an index, glossary, useful commands, explanation of symbols, and commands to solve common issues. Each chapter provides numerous examples to work through. Although chapters One and Two are important, Chapter Four, "Simplifying Complex Problems," also stands out. Here, Banerjee writes: "It's important to keep in mind that what constitutes a specific data element depends on the specific task that you're working on" (36). It is the foundation from which to understand common data problems in library work and deconstruct these into smaller ones where your new tools can be applied. This chapter equips readers with a method to critically think through complex issues whether or not they are related to data. Chapter Four is the conceptual design of the kitchen where chapters One and Two are the tools found in that kitchen. This reviewer found it useful to spend as much time with Chapter Four as with the first two.

In this way, Banerjee provides a great departure to think about more than just library data problems. Many current handbooks on command line and its tools tend to be from the perspective of computer science. For those without any knowledge, these books can sometimes be

overwhelming. Banerjee here brings to librarianship a clear explanation of these tools as it relates to library or text-based data. The advantage is that this handbook creates a frame of reference to understand common library data issues. This shared framework allows the reader to compare how they currently approach these data issues and the solutions proposed by Banerjee. Moreover, the reference is shared in that all examples come from those that librarians work with on a daily basis, especially in technical services. Lastly, it acts as a starting point to better understand more complex literature on this subject from other fields such as computer science.

This handbook provides a good introduction to the command line and its tools. It is appropriate for a broad audience although it is tailored for those who work with library data. That is one of the best advantages of this book in that the examples are those that librarians might have already encountered or have used other tools to solve. Another bonus is that if the reader adopts a hands-on approach, they will have an even better understanding of these tools. Indeed, this is the book's strength. If the audience is willing to take the time to work through the examples, then they will reap the benefits—*Jennifer M. Eustis (jeustis@umass .edu), University of Massachusetts Amherst*