VOLUME TWO

CURATING RESEARCH DATA

A Handbook of Current Practice

With 30 case studies contributed by practitioners in the field

BY LISA R. JOHNSTON

Curating Research Data

Volume Two: A Handbook of Current Practice

by Lisa R. Johnston

with 30 case studies contributed
by practitioners in the field

Association of College and Research Libraries
A division of the American Library Association
Chicago, Illinois 2017

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences–Permanence of Paper for Printed Library Materials, ANSI Z39.48-1992. ∞

Cataloging-in-Publication data is on file with the Library of Congress

Copyright ©2017 by the Association of College and Research Libraries.
All rights reserved except those which may be granted by Sections 107 and 108 of the Copyright Revision Act of 1976.

Printed in the United States of America.

21 20 19 18 17 5 4 3 2 1

Cover image
Copyright: kentoh / 123RF Stock Photo (http://www.123rf.com/profile_kentoh)

Table of Contents

ix .....Acknowledgments

xi .....Foreword

xii..... Data Curation across the Data Life Cycle
xiv..... Recommended Further Reading
xvii .... Appendix: Summary of *Curating Research Data: A Handbook of Current Practice*
xxii..... Notes
xxiii ... Bibliography

1 .....Preliminary Step 0: Establish Your Data Curation Service

4..... Summary of Step 0: Establish Your Data Curation Service
4..... Notes
7..... Bibliography

11 ....Step 1.0: Receive the Data

11..... 1.1 Recruit Data for Your Curation Service
Develop a Communications Plan
Communication Vehicles to Reach Your Target Audiences
Marketing Strategies for Your Repository
15..... 1.2 Negotiate Deposit
Deposit Interview
Pre-Deposit Interview in Action

18 CASE STUDY—A Case Study for Organizing and Documenting Research Data
Qianjin (Marina) Zhang and Sara Scheib

21..... 1.3 Transfer Rights (Deposit Agreements)
Addressing Legal Issues in Deposit Agreements
Deposit Agreements in Action

24 CASE STUDY—Legal Agreements for Acquiring Restricted-Use Research Data
Kaye Marz

28..... 1.4 Facilitate Data Transfer
29 ..... 1.5 Obtain Available Metadata and Documentation

30 CASE STUDY—Strategies for Acquiring Metadata That Facilitates Sharing and Long-Term Reuse of Research Data
Ho Jung S. Yoo, Juliane Schneider, and Arwen Hutt

34 CASE STUDY—Challenges with Quality of Data Set Metadata in a Self-Submission Repository Model
Amy Koshoffer, Carolyn Hansen, and Linda Newman

42..... 1.6 Receive Notification of Data Arrival
43..... Summary of Step 1.0: Receive the Data
45..... Appendix 1.0 A: Directory Structure Used in "A Case Study for Organizing and Documenting Research Data"

46 ..... Appendix 1.0 B: Sampling of Information Requested in the Consultation Request Form
47..... Appendix 1.0 C: Sampling of Metadata Requested in the Collection Description Form
48 ..... Appendix 1.0 D: Descriptive Metadata for a Data Set Using Resource Description Framework (RDF)
50 ..... Notes
52..... Bibliography

55 ....Step 2.0: Appraisal and Selection Techniques that Mitigate Risks Inherent to Data

55..... 2.1 Appraisal
Appraisal Questions to Consider
Appraisal Criteria in Action

57 CASE STUDY—Scientific Records Appraisal Process: US Geological Survey Case Study
John Faundeen

60 ..... 2.2 Risk Factors for Data Repositories
Detect Potential Disclosure Risk
Steps to Mitigate Disclosure Risk
Understanding Risk Factors: Two Case Studies

67 CASE STUDY—Considerations for Restricted-Use Research Data
Kaye Marz

72 CASE STUDY—Human Subjects Data in an Open Access Repository
Christine Mayo and Erin Clary

74..... 2.3 Inventory
74..... 2.4 Selection
75..... 2.5 Assign
76..... Summary of Step 2.0: Appraise and Select
77 ..... Notes
80 ..... Bibliography

85 ....Step 3.0: Processing and Treatment Actions for Data

85..... 3.1 Secure the Files
86 ..... 3.2 Create a Log of Actions Taken
87..... 3.3 Inspect the File Names and Structure
File Relationships and Directory Structure
File-Naming Conventions
Understanding File Structures in Action

88 CASE STUDY—Just-in-Time Data Management Consultations
Jennifer Thoegersen

90 ..... 3.4 Inspect the Data
Software Needed by Data Curators

Attempt to Understand and Use the Data
Questions to Help Understand the Data
Documenting Computational Environments
Detecting Hidden Documentation in Action

94 CASE STUDY—Finding Metadata in R: Where to Look
Alicia Hofelich Mohr

100 CASE STUDY—Helpful Commands for Exporting Metadata from Statistical Software Packages SPSS, Stata, and R
Alicia Hofelich Mohr

104 .... 3.5 Work with Author to Enhance the Submission
Readme Files: Documentation for Any File Format
Working with Authors to Curate Their Data

108 CASE STUDY—Losing Research Data Due to Lack of Curation and Preservation
Andrew S. Gordon, Lisa Steiger, and Karen E. Adolph

115 .... 3.6 Consider the File Formats
Microsoft Excel: A Use Case in Format Preservation
Proprietary Files Formats Conversions
Preserving Data Files in Action

119 CASE STUDY—Preserving 3D Data Sets: Workflows, Formats, and Considerations
Archaeology Data Service

123 .... 3.7 File Arrangement and Description
Compression
Contextual Concerns with Data in Action

125 CASE STUDY—Legacy Research Data Management for Librarians: A Case Study
Tina Qin

127 .... Summary of Step 3.0: Processing and Treatment Actions for Data
128 .... Notes
130 .... Bibliography

133 ..Step 4.0: Ingest and Store Data in Your Repository
133 .... 4.1 Ingest the Files
Digital Repository Infrastructure Options
Functional Requirements for a Data Repository
Data Curation Without a Locally Hosted Repository

139 CASE STUDY—Data and a Diversity of Repositories
Karen S. Baker and Ruth E. Duerr

Repository Ingest Workflows for Data
Repository Ingest Workflows in Action

147 CASE STUDY—Standardization and Automation of Ingest Processes in a Fully Mediated Deposit Model
Juliane Schneider, Arwen Hutt, and Ho Jung S. Yoo

152 CASE STUDY—Dryad Curation Workflows
Erin Clary and Debra Fagan

155 .... 4.2 Store the Assets Securely
156 .... 4.3 Develop Trust in Your Digital Repository
Additional Factors That Influence Trust in Digital Repositories

158 CASE STUDY—Making the Case for Institutional Data Repositories
Cynthia R. H. Vitale, Jacob Carlson, Amy E. Hodge, Patricia Hswe, Erica Johns, Lisa R. Johnston, Wendy Kozlowski, Amy Nurnberger, Jonathan Petters, Elizabeth Rolando, Yasmeen Shorish, Juliane Schneider, and Lisa Zilinski

162 CASE STUDY—Making the Case for Disciplinary Data Repositories
Jared Lyle

164 .... Summary of Step 4.0: Ingest and Store Data in Your Repository
164 .... Notes
169 .... Bibliography

173 ..Step 5.0: Descriptive Metadata

173 .... 5.1 Create and Apply Appropriate Metadata
Dublin Core Metadata Schema in Action
Challenges with Capturing Metadata in Data Repository Submission Forms
Usability Testing of Data Repository Submission Forms
179 .... 5.2 Consider Disciplinary Metadata Standards for Data
Disciplinary-Specific Metadata in Action

181 CASE STUDY—Modern Metadata at Scale: CLOSER Case Study
Jon Johnson and Gemma Seabrook

184 CASE STUDY—Beyond Discovery: Cross-Platform Application of Ecological Metadata Language in Support of Quality Assurance and Control
Jon Wheeler, Mark Servilla, and Kristin Vanderbilt

187 .... Summary of Step 5.0: Descriptive Metadata
188 .... Appendix 5.0 A: Screenshots from the Data Repository for the University of Minnesota (DRUM) Submission Form
191 .... Appendix 5.0 B: Screenshots from the Sevilleta LTER Program's Instance of the Drupal Ecological Information Management System (DEIMS)
195 .... Notes
196 .... Bibliography

199 ..Step 6.0: Access

199 .... 6.1 Determine Appropriate Levels of Access
201 .... 6.2 Apply the Terms of Use and Any Relevant Licenses
Terms of Use for Data
Licenses and Copyright Notices for Data

204 CASE STUDY—Why Does Dryad Use Creative Commons Zero (CC0)?
Debra Fagan and Peggy Schaeffer

205 .... 6.3 Contextualize the Data
Contextualization with Visualization

207 CASE STUDY—The Geographic Storage, Transformation and Retrieval Engine (GSToRE): A Platform for Active Data Access and Publication as a Complement to Dedicated Long-Term Preservation Systems
Karl Benedict

Contextualization with Metadata
Linking to Related Articles

211 CASE STUDY—Pre-Metadata Counseling: Putting the DataCite relationType Attribute into Action
Elise Dunham, Elizabeth Wickes, Ayla Stein, Colleen Fallaw, and Heidi J. Imker

216 .... 6.4 Increase Exposure and Discovery
217 .... 6.5 Apply Any Necessary Access Controls
219 .... 6.6 Ensure Persistent Access and Encourage Appropriate Citation
Assigning Persistent Identifiers in Action

219 CASE STUDY—A Participant Agreement for Minting DOIs for Data Not in a Repository
Susan M. Braxton, Bethany Anderson, Margaret H. Burnette, Thomas G. Habing, William H. Mischo, Sarah L. Shreeves, Sarah C. Williams, and Heidi J. Imker

223 .... 6.7 Release Data for Access and Notify Author
Track Effort Throughout the Process
Gather Feedback
225 .... Summary of Step 6.0: Access
226 .... Notes
228 .... Bibliography

231 ..Step 7.0: Preservation of Data for the Long Term

231 .... 7.1 Preservation Planning for Long-Term Reuse
Practical Approaches to Digital Preservation

234 CASE STUDY—Balancing Digital Preservation Standards with Data Reusability
Christine Mayo

236 .... 7.2 Monitor Preservation Needs and Take Action
Tools for Active Monitoring and Management
Data Preservation Strategies
Data Preservation in Action

237 CASE STUDY—Archiving Three Waves of the National Survey of Families and Households
Chiu-chuang (Lu) Chou and Charlie Fiss

245 CASE STUDY—From KAPTUR to VADS4R: Exploring Research Data Management in the Visual Arts
Robin Burgess

249 .... Summary of Step 7.0: Preservation of Data for the Long Term
249 .... Notes
251 .... Bibliography

253 ..Step 8.0: Reuse

253 .... 8.1 Monitor Data Reuse
Impact Measure for Tracking Data Reuse
Data Citations

255 CASE STUDY—Facilitating Good Data Citation Practice
Christine Mayo

Altmetrics for Data

258 CASE STUDY—Reuse of Restricted-Use Research Data
Arun Mathur, Johanna Davidson Bleckman, and Jared Lyle

261 .... 8.2 Collect Feedback about Data Reuse and Quality Issues
Facilitate the Peer Review of Data
Data Review in Action

263 CASE STUDY—Enabling Scientific Reproducibility with Data Curation and Code Review
Limor Peer

267 .... 8.3 Provide Ongoing Support as Long as Necessary
268 .... 8.4 Cease Data Curation
269 .... Summary of Step 8.0: Reuse
269 .... Notes
272 .... Bibliography

277 ..Brief Concluding Remarks and a Call to Action

279 .... Notes
280 .... Bibliography

281 ..Bibliography

303 ..Biographies

303 .... Author Biography
303 .... Case Study Author Biographies