



1 ~~ÖÖÖY [\ ā * ÁÚæ ^! Û^! ā • ÄËÓ^ • óÚ! æ&æ^ • ÊP [ĚG~~

2 **Subject**

3 Work flows - Data Discovery and Dissemination: User Perspective (2009-02-15)

4 **Document identifier:**

5 ~~@j kãçã[æ! * Æ€H ì î ÖÖÓ^ • ú! æ&æ^ • €G~~

9 **Authors:**

10 Karl Dinkelmann, Michelle Edwards, Jane Fry, Chuck Humphrey, Ron Nakao,
11 Wendy Thomas

12 **Editors:**

13 Ron Nakao and Chuck Humphrey

14 **Target Audience:**

15 Metadata producers, funding agencies and councils, and institutions with
16 stewardship roles

17 **Abstract:**

18 Describes the best practices for metadata producers to provide end users with the
19 resources for data discovery and dissemination.

20 **Status:**

21 This document is updated periodically on no particular schedule. Send comments to
22 editor -- ddi-bp-editors@icpsr.umich.edu
23



24

25	1 INTRODUCTION	3
26	1.1 Problem statement	3
27	1.2 Terminology	3
28	2 BEST PRACTICE SOLUTION	4
29	2.1 Definitions	4
30	2.2 Best Practice behavior.....	5
31	2.3 Discussion	7
32	2.4 Example.....	8
33	3 REFERENCES	9
34	3.1 Normative	9
35	APPENDIX A. ACKNOWLEDGMENTS	10
36	APPENDIX B. REVISION HISTORY	12
37	APPENDIX C. LEGAL NOTICES	13



38 **1 Introduction**

39 DDI 3 facilitates the creation of metadata at a variety of starting points from the hypothesis
40 for a study through the capturing of legacy metadata. How and where one starts capturing
41 metadata depends upon the data being described, the application within which it is used,
42 and the organizational needs of the creators. The best practices on workflow provide
43 guidelines for setting up metadata creation processes within different environments,
44 identifying organizational and application features that impact the process structure,
45 addressing salient questions/issues in setting up the process, and determining the
46 implications of various starting points and process orders:

- 47 1. Metadata Creation Regarding Recoding, Aggregation, and Other Data Processing
48 Activities [see References section]
- 49 2. Archival Ingest and Metadata Enhancement [see References section]
- 50 3. Dissemination and Discovery: User Perspective (this document)

51 **1.1 Problem statement**

52 Each phase in the data life cycle represents a group of related processes. Within a stage,
53 specific processes or activities, when viewed collectively, represent a significant component
54 in conducting research. While some activities and products are intrinsic to each stage,
55 others flow across stages. For example, the design of an experiment or survey will be
56 integral to the Data Production stage, while data products emanating from this stage will
57 flow throughout the model. (Stewardship of Research Data in Canada: A Gap Analysis,
58 Draft [see References section])

59 The “End User” Model [see Discussion section] represents the user perspective on the data
60 life cycle. However, metadata emerges from discrete organizations over time with workflows
61 that often do not articulate cleanly from stage to stage across handoffs.

62 From the end user perspective, what are the best practices that metadata producers should
63 follow to provide potential data analysts with the resources they need for data discovery and
64 that archives need for dissemination?

65

66 **1.2 Terminology**

67 The key words *must*, *must not*, *required*, *shall*, *shall not*, *should*, *should not*, *recommended*,
68 *may*, and *optional* in this document are to be interpreted as described in **[RFC2119]**.
69 Additional DDI standard terminology and definitions are found in

70 [@d d , , Éãã|ã\) & É !* É\] É^~ãã }•](#)



71

72 **2 Best Practice Solution**

73 **2.1 Definitions**

74 Stewardship: In the context of this best practice, stewardship involves taking on custodial
75 responsibilities for a stage in the data life cycle.

76 Data life cycle: The stages of the data component of the research process, from study
77 conceptualization to data analysis and archiving, feeding back to earlier stages. This
78 process has often been depicted as linear, but the diagram embedded in this best practice
79 [see the End User Model below



Data Documentation Initiative

80 Figure 1: the End User Model] offers a different perspective on it, from the user's point of
81 view.

82 Citizenship: Being a contributing member of the full data life cycle and realizing that one is
83 part of a bigger scientific picture.

84 End user: Anyone using any product generated in the data life cycle. Examples include
85 research council/funding bodies, researchers, data producers, archivists, librarians, users,
86 registry managers, research analysts/authors.

87 Discovery: Strategies and processes used by the end user to locate and access products
88 (metadata, data, and other related information) of the data life cycle.

89 Dissemination: Data distribution with the aim of access by the end user to the products
90 (metadata, data, and other related information) of the data life cycle.

91 Data and knowledge repository: Places (may be virtual) where the products (metadata,
92 data, and other related information) of the data life cycle are located.

93 Knowledge transfer: The act of sharing the knowledge gained throughout the data life cycle.

94

95 **2.2 Best Practice behavior**

96 An underlying principle of metadata, data, and other information is that it can be shared and
97 accessible to a wider user community. The often-disjointed nature of the production of
98 metadata throughout the data life cycle can lead to fragmented metadata that can
99 undermine the achievement of this principle. While DDI 3.0 was designed to capture
100 metadata in a dynamic process for the purposes of informing the production process and
101 later stages in the model, including data discovery, analysis, and re-use, best practice by
102 metadata producers should ensure consistent and exhaustive coverage. Bear in mind that
103 these same features in DDI 3.0 can increase fragmentation. Metadata producers are
104 responsible for retaining the metadata integrity throughout the Data Life Cycle.

105 From the end user perspective, published metadata should be persistent, versioned, and
106 accessible. The quality and completeness of the content of the metadata should be
107 evaluated against the best practices of the wider DDI community, which should ideally be
108 addressed through certification.

109 Organizations and agencies involved in the production of metadata should seek certification
110 of their metadata creation processes. Certification should be based on best practices as
111 defined by the DDI community. Institutions with stewardship responsibilities, such as trusted
112 data and knowledge repositories, should establish certification criteria and processes.



Data Documentation Initiative

113 Funding agencies and councils should facilitate the adoption, oversight, and use of best
114 practices in metadata creation. Compliance and certification should be viewed as rewards in
115 and of themselves because they promote greater return on the initial investment in
116 metadata creation and reinforce the scientific method.

117 **Overarching Metadata Principles**

118 While there will always be new ways to use metadata, there are a set of important principles
119 that data and metadata producers should always keep in mind:

- 120 • Metadata may be used in ways not anticipated at the time of generation. Producers
121 need to look beyond their own goals and uses of their metadata, and be good
122 citizens to the general user community to support data discovery and dissemination.
- 123 • Related to the point above, metadata should never be discarded without careful
124 thought.
- 125 • Handoffs along the trajectory of the data life cycle (for example, the handoff from a
126 data producer to an archive) carry risks of metadata loss. It is best to be
127 conservative and to preserve and pass along all relevant metadata. See the best
128 practice on archival ingest [see References section].
- 129 • Documentation of data and metadata transformations over time is key to OAIS
130 compliance and to providing a way to track content changes and chain of custody.
131 See the best practice on metadata creation regarding recoding, aggregation, and
132 other data processing activities [see References section].
- 133 • Search and browse features are ubiquitous on the Web. In preparing metadata
134 content, one should always think ahead to how content will be used with these types
135 of finding aids.
- 136 • Controlled vocabularies used consistently can help users target the data and
137 metadata they need.
- 138 • DDI provides rich options for the description of coverage – topical, geographic, and
139 temporal. Coverage elements should always be populated as fully as possible.
- 140 • Concepts may be assigned in DDI starting at data conceptualization. Assigning
141 concepts at the question level is very helpful to data analysts seeking to understand
142 the rationale behind the question.
- 143 • All stages of data and metadata production may be relevant to the end user. For
144 example, an end user assessing data quality may want to know that the interviewers



Data Documentation Initiative

145 administering a questionnaire to a population whose first language is Spanish were
146 themselves fluent in Spanish.

147

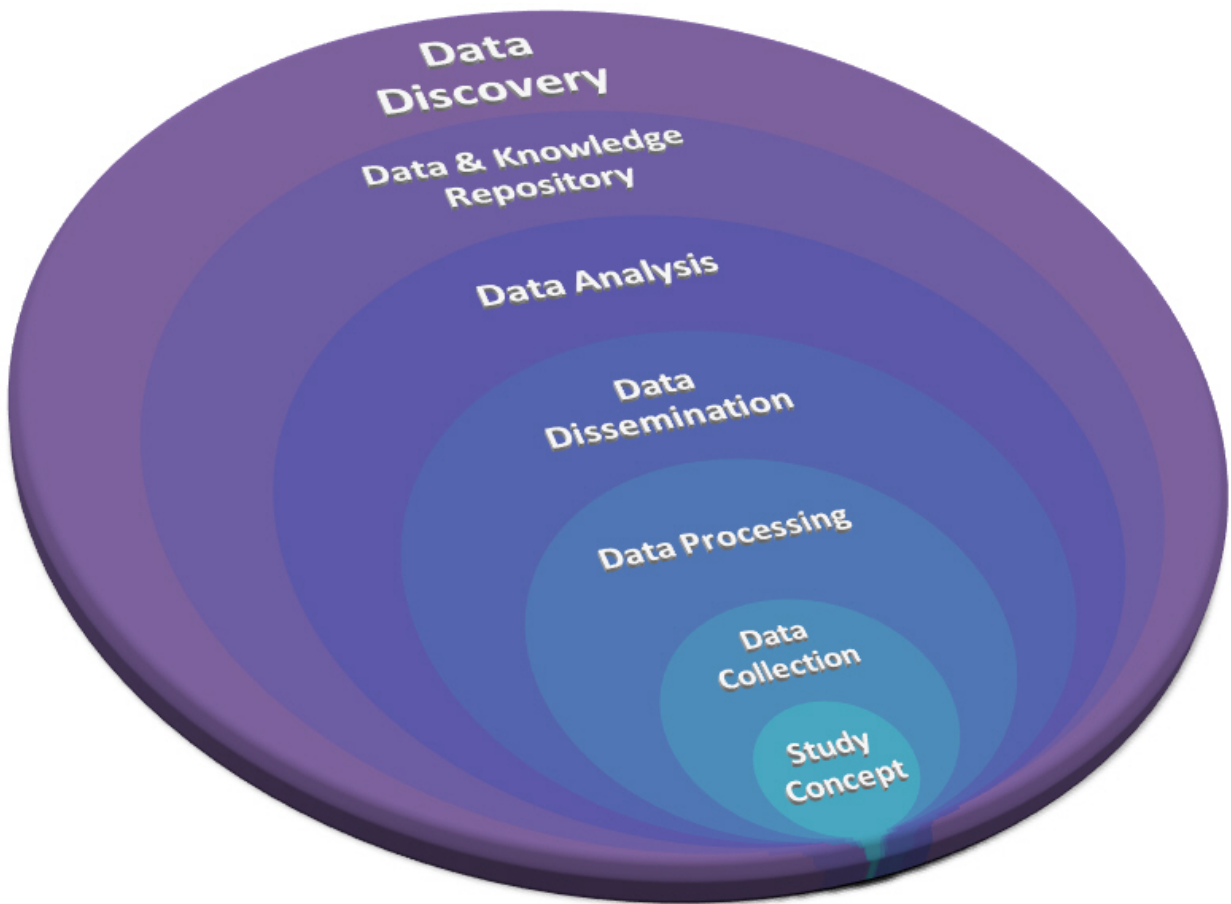
148 **2.3 Discussion**

149 The original Data Life Cycle Model used to develop DDI 3 represents a data production
150 perspective. The challenge is to determine whether the current model reflects the metadata,
151 data or other information needs of the end user for discovery or access. The End User
152 Model provides a view of the original Model from a user-centered perspective. For example,
153 in the current Model, production is often sequential, whereas end users will select relevant
154 products throughout the entire data life cycle based upon their particular needs (sequential
155 versus random access). Having comprehensive metadata -- in the fullness of description --
156 is essential for end users to locate what they require successfully.

157 A user-centered perspective is represented in the end user model below, which shows the
158 nested relationship of the different stages in the data life cycle.

159

160 Figure 1: the End User Model



161

162

163 The end user perspective should inform the metadata production throughout the data life
164 cycle. The end users may never see actual DDI metadata markup, yet they will be totally
165 reliant upon its existence to accomplish their goals. The workflow producing metadata within
166 each stage needs to take into account that the end user may unpredictably require
167 metadata, data, and other information from any stage. Thus, the “membranes” between life
168 cycle stages must be permeable.

169 Best Practices need to be developed for the metadata produced in each phase of the Life
170 Cycle to meet the discovery and dissemination requirements of the end users. However, the
171 Data Life Cycle phases do not have a one-to-one correspondence with the audiences that
172 produce metadata. Thus, gaps in the metadata may exist.

173 2.4 Example



Data Documentation Initiative

174 Two units of the University of Michigan’s Institute for Social Research – Survey Research
 175 Operations (SRO) and the Inter-university Consortium for Political and Social Research
 176 (ICPSR) – worked together to create interactive documentation for the Collaborative
 177 Psychiatric Epidemiology Surveys, which were harmonized. This documentation reflects not
 178 only what the user needs to know but also how the instrument looked as the interviewers
 179 administered it. There are links to interviewer aids and to other pertinent metadata,
 180 including the universe of respondents who answered a given question.

181 <http://www.icpsr.umich.edu/CPES/>

182 3 References

183 DDI Best Practice: Workflows for Metadata Creation Regarding Recoding, Aggregation and
 184 Other Data Processing Activities: [@j kâcâ\[æ! * F-ÈH ì î ðÖó^•ú:æ&ã^•€](#)

186 DDI Best Practice: Workflows - Archival Ingest and Metadata Enhancement:
 187 [http://www.ddialliance/bp/DDIBestPractices_Workflows-](http://www.ddialliance/bp/DDIBestPractices_Workflows-ArchivalIngestAndMetadataEnhancement.doc.PDF)
 188 [ArchivalIngestAndMetadataEnhancement.doc.PDF](http://www.ddialliance/bp/DDIBestPractices_Workflows-ArchivalIngestAndMetadataEnhancement.doc.PDF)

189 Stewardship of Research Data in Canada: A Gap Analysis, Draft: July 31, 2008, Research
 190 Data Strategy Working Group, <http://data-donnees.gc.ca/docs/GapAnalysis.pdf>

191 3.1 Normative

192
 193 [RFC2119] S. Bradner, Key words for use in RFCs to Indicate Requirement Levels,
 194 <http://www.ietf.org/rfc/rfc2119.txt>, IETF RFC 2119, March 1997.

195
 196 OASIS, Best Practice, [http://www.oasis-open.org/committees/uddi-](http://www.oasis-open.org/committees/uddi-spec/doc/bp/uddi-spec-tc-bp-template.doc)
 197 [spec/doc/bp/uddi-spec-tc-bp-template.doc](http://www.oasis-open.org/committees/uddi-spec/doc/bp/uddi-spec-tc-bp-template.doc), 2003

198

199

200 **Appendix A. Acknowledgments**

201 The following individuals were members of the DDI Expert Workshop held 10-14 November
202 2008 at Schloss Dagstuhl, Leibniz Center for Informatics, in Wadern, Germany.

203 Nikos Askitas, Institute for the Study of Labor (IZA)

204 Karl Dinkelmann, University of Michigan

205 Michelle Edwards, University of Guelph

206 Janet Eisenhauer, University of Wisconsin

207 Jane Fry, Carleton University

208 Peter Granda, Inter-university Consortium for Political and Social Research (ICPSR)

209 Arofan Gregory, Open Data Foundation

210 Rob Grim, Tilburg University

211 Pascal Heus, Open Data Foundation

212 Maarten Hoogerwerf, Data Archiving and Networked Services (DANS)

213 Chuck Humphrey, University of Alberta

214 Jeremy Iverson, Algenta Technology

215 Jannik Vestergaard Jensen, Danish Data Archive (DDA)

216 Kirstine Kolsrud, Norwegian Social Science Data Services (NSD)

217 Stefan Kramer, Yale University

218 Jenny Linnerud, Statistics Norway

219 Hans Jørgen Marker, Danish Data Archive (DDA)

220 Ken Miller, United Kingdom Data Archive (UKDA)

221 Meinhard Moschner, GESIS - Leibniz Institute for the Social Sciences

222 Ron Nakao, Stanford University



Data Documentation Initiative

- 223 Sigbjørn Revheim, Norwegian Social Science Data Services (NSD)
- 224 Wendy Thomas, University of Minnesota
- 225 Mary Vardigan, Inter-university Consortium for Political and Social Research (ICPSR)
- 226 Joachim Wackerow, GESIS - Leibniz Institute for the Social Sciences
- 227 Wolfgang Zenk-Möltgen, GESIS - Leibniz Institute for the Social Sciences



228

229 **Appendix B. Revision History**

230

Rev Date		By Whom	What
0.9 2009	-02-08	Stefan Kramer	Removed date from filename to accommodate linking. Began revision history tracking.
0.91	2009-02-15	Stefan Kramer	Added rev. date to Subject. Changed Abstract from question to statement. Removed reference to "Intellectual Property Rights section of the DDI Alliance." Changed sections above ToC to heading 3, rebuilt ToC for heading levels 1-2 only.

231



232

233 **Appendix C. Legal Notices**

234 Copyright © DDI Alliance 2009, *All Rights Reserved*

235

236 <http://www.ddialliance.org/>

237

238 Content of this document is licensed under a Creative Commons License:

239 Attribution-Noncommercial-Share Alike 3.0 United States

240

241 This is a human-readable summary of the Legal Code (the full license).

242 <http://creativecommons.org/licenses/by-nc-sa/3.0/us/>

243

244 You are free:

- 245 • to Share - to copy, distribute, display, and perform the work
- 246 • to Remix - to make derivative works

247

248 Under the following conditions:

- 249 • Attribution. You must attribute the work in the manner specified by the author or
250 licensor (but not in any way that suggests that they endorse you or your use of
251 the work).
- 252 • Noncommercial. You may not use this work for commercial purposes.
- 253 • Share Alike. If you alter, transform, or build upon this work, you may distribute
254 the resulting work only under the same or similar license to this one. For any
255 reuse or distribution, you must make clear to others the license terms of this
256 work. The best way to do this is with a link to this Web page.
- 257 • Any of the above conditions can be waived if you get permission from the
258 copyright holder.
- 259 • Apart from the remix rights granted under this license, nothing in this license
260 impairs or restricts the author's moral rights.

261

262 **Disclaimer**

263

264 The Commons Deed is not a license. It is simply a handy reference for understanding the Legal
265 Code (the full license) — it is a human-readable expression of some of its key terms. Think of it as
266 the user-friendly interface to the Legal Code beneath. This Deed itself has no legal value, and its
267 contents do not appear in the actual license.

268

269 Creative Commons is not a law firm and does not provide legal services. Distributing of, displaying
270 of, or linking to this Commons Deed does not create an attorney-client relationship.

271 Your fair use and other rights are in no way affected by the above.

272

273 **Legal Code:**

274 <http://creativecommons.org/licenses/by-nc-sa/3.0/us/legalcode>