



1 DDI Working Paper Series -- Best Practices, No. 10

2 **Subject**

3 Management of DDI 3.0 Unique Identifiers (2009-02-15)

4 **Document identifier:**

5 <http://dx.doi.org/10.3886/DDIBestPractices10>

8 **Authors:**

9 Nikos Askitas, Arofan Gregory, Maarten Hoogerwerf

10 **Editors:**

11 Arofan Gregory and Maarten Hoogerwerf

12 **Target Audience:**

13 Software developers, data archivists, database administrators

14 **Abstract:**

15 This best practice specifies the technical requirements and procedures for ensuring
16 well-managed identifiers for both internal use and external references. Management
17 of identifiers is divided into two situations: during creation of new DDI 3.0 objects
18 and during migration of older DDI versions into DDI 3.0 versions. These best
19 practices will be described from both procedural and technical perspectives.

20 **Status:**

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



23	TABLE OF CONTENTS	
24	INTRODUCTION.....	3
25	1.1 Problem statement	3
26	1.2 Terminology	3
27	2 BEST PRACTICE SOLUTION	4
28	2.1 Definitions	4
29	2.2 Best Practice behavior	4
30	2.3 Example.....	7
31	3 REFERENCES	8
32	3.1 Normative	8
33	APPENDIX A. ACKNOWLEDGMENTS	9
34	APPENDIX B. REVISION HISTORY	11
35	APPENDIX C. LEGAL NOTICES	12



36

37 **Introduction**

38 While DDI 3.0 provides enhanced functionality in managing identifiers, one must proceed
39 carefully to ensure that identifiers are constructed properly and managed consistently. This
40 best practice lays out some steps to follow when building new identifiers and migrating DDI
41 documents across versions.

42 **1.1 Problem statement**

43 All DDI 3.0 objects need to be uniquely identified to enable them to be referred to by other
44 DDI 3.0 instances. Since in DDI 3.0 the scope of all objects has increased from just a DDI
45 instance to the whole world, additional measures are needed to guarantee uniqueness
46 within this broader scope.

47 **1.2 Terminology**

48 The key words *must*, *must not*, *required*, *shall*, *shall not*, *should*, *should not*, *recommended*,
49 *may*, and *optional* in this document are to be interpreted as described in **[RFC2119]**.

50 Additional DDI standard terminology and definitions are found in

51 <http://www.ddialliance.org/bp/definitions>



52

53 **2 Best Practice Solution**

54 **2.1 Definitions**

55 Uniqueness: A Globally Unique Identifier is a special type of identifier used in applications to
56 provide a reference number that is unique in any context.

57 Migration: Migration in the DDI context refers to moving from a DTD to XML Schema in
58 terms of document structure; and from DDI 2 to DDI 3 as well as from DDI 3 back to DDI 2
59 in terms of porting content.

60 URN: A URN (Uniform Resource Name) is an Internet resource with a name that, unlike a
61 URL, has persistent significance - that is, the owner of the URN can expect that someone
62 else (or a program) will always be able to find the resource.

63 **2.2 Best Practice behavior**

64 The assignment of DDI 3.0 identifiers happens in two situations: (1) when creating new DDI
65 3.0 objects; and (2) when migrating older DDI objects.

66 **Creating New DDI 3 Objects**

67 DDI identifiers should be used primarily by dedicated applications. End users should not
68 have to deal with these.

69 DDI-aware applications should generate and interpret DDI URNs. DDI objects can be
70 identified in two ways: explicitly by the use of a URN and implicitly by interpreting the
71 identifying (inherited) element attributes: agency, version, and ID. The latter is automatically
72 possible; the former requires a URN to be specified for each element. Adding such URNs
73 has one advantage: clear identification of all elements. A drawback is that it creates a lot of
74 redundant information. We recommend not adding them to DDI, but instead enabling
75 software to be responsible for on-the-fly interpretation. See the Overview, Part 1, of the DDI
76 specification documents for more information on DDI URNs.

77 The uniqueness of DDI URNs is assured via a hierarchical system. DDI delegates the
78 responsibility for uniqueness to maintaining organizations. To this end, the maintaining
79 organization is to register a DDI agency name at the DDI Alliance. Such names are allowed
80 to contain semantic value such as the name or acronym of the maintaining organization.
81 Such names should be chosen with care: they may be questioned after organizational
82 changes.



Data Documentation Initiative

83 Some organizations maintain many collections of DDI objects. If such collections are not
84 tightly bound to the organization and are likely to be transferred to other organizations, then
85 a separate agency identifier should be considered. This may result in multiple agency
86 identifiers per organization. <http://tools.ddialliance.org/?lvl1=community&lvl2=agencyid>

87 This leads to the question of whether individual researchers can also apply for an agency
88 identifier.

89 When multiple organizations participate in a project, they should consider requesting a
90 special agency identifier because the project is likely to be maintained by different
91 organizations.

92 Once uniqueness outside the agency has been dealt with, the agency should make sure
93 that the uniqueness of the DDI 3.0 identifiers within the agency is guaranteed. Three
94 suggestions:

95 1. A data manager specifies a prefix for which he or she guarantees uniqueness

- 96 • Pro: User-friendly identifiers
- 97 • Con: Responsibility lies with user

98 2. A registry is maintained that ensures uniqueness

- 99 • Pro: User-friendly identifiers
- 100 • Con: Must maintain a registry

101 3. An algorithm is used to guarantee uniqueness (e.g., uuid)

- 102 • Pro: Automatic
- 103 • Con: Long user-unfriendly identifiers and increases in file size

104 One technique that can be used in combination with the above is separating, for example,
105 variables and variable schemes. If the variable scheme is unique, one only has to
106 guarantee the uniqueness of variables within this scheme (by simply enumerating them).
107 Another simplification would be to define mnemonics for element types, e.g., VS_ variable
108 schemes, CS_ for concept schemes, etc.

109 Once a DDI instance is created, its uniqueness needs to be checked by means of a general
110 purpose validation process. Other validations are to ensure that all required elements
111 (agency, ID, and version) are specified and that no circular references exist.



Data Documentation Initiative

112 The DDI should be validated in these ways:

- 113 • Identifiers within a scheme must be unique
- 114 • Schemes must be unique
- 115 • URNs must be unique
- 116 • URNs must be well-formed

117 See Part 1, Overview, of the DDI specification for more information.

118 **Migration from DDI 2.0 to DDI 3.0**

119 A special case is the migration of older DDI metadata to DDI 3.0.

120 A pragmatic approach is to do this in four separate phases:

- 121 1. Convert DDI 2.0 DTD to DDI 2.0 Schema if you have DTD-based documentation.
- 122 2. Transcribe all existing DDI into the DDI 3.0 format. Transcription means application
123 of the table for conversion available at: <http://www.ddialliance.org/sites/default/files/CorrespondenceMapping-spreadsheet.pdf>
124 This step is reversible.
- 125 3. Fill in the mandatory elements of DDI 3.0 such as variable schemes. If there are
126 mandatory elements in DDI 3.0 that are not in the conversion table, those elements
127 are guaranteed to be lost in going back from 3.0 to 2.0.
- 128 4. Use the advanced DDI 3.0 features.
 - 129 ○ E.g., identify any identical objects and assign correct DDI3 IDs to enable re-
130 use, to add concepts and concept-schemes, etc. In applying advanced
131 features of DDI 3.0, you disambiguate duplicate variable identifiers by using
132 variable schemes which should be unique. You have to encode these
133 schemes. This can be done by prefixing the variable with a scheme and a
134 separator. The separator is required to enable distinguishing between the
135 scheme and the variable when reversing this process again back to 3.0 (this
136 applies to concepts, universes, codes, etc.).

137 This would result in:

138 2.0 v1 --- DTD 2 Schema → 2.0 v2 --- transcribe → 3.0 v1 --- fill mandatory elements → 3.0
139 v2 --- advanced features → 3.0 v3

140 **Migration from 3.0 to 2.0**

141 A lot of existing tools can only work with DDI 2.0. How can you migrate 3.0 to 2.0?



Data Documentation Initiative

- 142 • Strip off everything that is in 3.0 that is not convertible into DDI 2.0 by means of the
143 table in Appendix 4 of Part I Overview. Obviously this cannot be undone when
144 migrating back. There will be information loss in this case.
- 145 • You have to encode the DDI 3.0 schemes into the existing variable. This can be
146 done by prefixing the original variable with this scheme and a separator. The
147 separator is required to enable distinguishing between the scheme and the variable
148 when reversing this process again back to 3.0. This prevents information loss (this
149 applies to concepts, universes, codes, etc.).
- 150 • There is a special case in DDI 3.0 when there are no variables but there are
151 elements that exist below the variable level in DDI 2.0 (e.g., question texts). A
152 solution to this problem is to generate dummy variables. Be aware that this does
153 introduce artificial/redundant variable information. A note can be made under the
154 variable to document the origin of this dummy variable. If you want to reverse this
155 migration, these dummy variables will appear in the DDI 3.0 instance.

156 Converting DDI 3.0 to DDI 2.0 can be done but loses the special DDI 3.0 functionality.
157 Therefore this migration should only be done for downwards compatibility with, e.g., tools.
158 Changes and maintenance should be done within the DDI 3.0 version if possible. Migration
159 back to DDI 3.0 is not recommended. In this case we advise to start with the originating DDI
160 3.0 version.

161

162 **2.3 Example**

163 See the DDI specification, Part 1, Overview, available at:

164 http://sourceforge.net/projects/ddi-alliance/files/Data%20Documentation%20Initiative/DDI%203.0%20%282008-04-28%29/DDI_3_0_2008-04-28_Documentation_XMLSchema.zip/download



165

166 **3 References**

167 See the DDI specification, Part 1, Overview, available at:

168 http://sourceforge.net/projects/ddi-alliance/files/Data%20Documentation%20Initiative/DDI%203.0%20%282008-04-28%29/DDI_3_0_2008-04-28_Documentation_XMLSchema.zip/download

169 **3.1 Normative**

170

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

173

174 OASIS, Best Practice, <http://www.oasis-open.org/committees/uddi-spec/doc/bp/uddi-spec-tc-bp-template.doc>, 2003

175



176

177 **Appendix A. Acknowledgments**

178 The following individuals were members of the DDI Expert Workshop held 10-14 November
179 2008 at Schloss Dagstuhl, Leibniz Center for Informatics, in Wadern, Germany.

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

181 Karl Dinkelmann, University of Michigan

182 Michelle Edwards, University of Guelph

183 Janet Eisenhauer, University of Wisconsin

184 Jane Fry, Carleton University

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

186 Arofan Gregory, Open Data Foundation

187 Rob Grim, Tilburg University

188 Pascal Heus, Open Data Foundation

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

190 Chuck Humphrey, University of Alberta

191 Jeremy Iverson, Algenta Technology

192 Jannik Vestergaard Jensen, Danish Data Archive (DDA)

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

194 Stefan Kramer, Yale University

195 Jenny Linnerud, Statistics Norway

196 Hans Jørgen Marker, Danish Data Archive (DDA)

197 Ken Miller, United Kingdom Data Archive (UKDA)

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

199 Ron Nakao, Stanford University



Data Documentation Initiative

- 200 Sigbjørn Revheim, Norwegian Social Science Data Services (NSD)
- 201 Wendy Thomas, University of Minnesota
- 202 Mary Vardigan, Inter-university Consortium for Political and Social Research (ICPSR)
- 203 Joachim Wackerow, GESIS - Leibniz Institute for the Social Sciences
- 204 Wolfgang Zenk-Möltgen, GESIS - Leibniz Institute for the Social Sciences



205

206 **Appendix B. Revision History**

207

Rev	Date	By Whom	What
0.9	2009-02-15	Stefan Kramer	Formatting changes (standard across all BPs).

208



209

210 **Appendix C. Legal Notices**

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

212

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

214

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

216 Attribution-Noncommercial-Share Alike 3.0 United States

217

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

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

220

221 You are free:

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

224

225 Under the following conditions:

- 226 • Attribution. You must attribute the work in the manner specified by the author or
227 licensor (but not in any way that suggests that they endorse you or your use of
228 the work).
- 229 • Noncommercial. You may not use this work for commercial purposes.
- 230 • Share Alike. If you alter, transform, or build upon this work, you may distribute
231 the resulting work only under the same or similar license to this one. For any
232 reuse or distribution, you must make clear to others the license terms of this
233 work. The best way to do this is with a link to this Web page.
- 234 • Any of the above conditions can be waived if you get permission from the
235 copyright holder.
- 236 • Apart from the remix rights granted under this license, nothing in this license
237 impairs or restricts the author's moral rights.

238

239 **Disclaimer**

240

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

245

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

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

249

250 **Legal Code:**

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