A Uniform Resource Name (URN) Namespace for the Data Documentation Initiative (DDI)
draft-urn-ddi-00.txt

Abstract

This document describes the Namespace Identifier (NID) "ddi" for Uniform Resource Names (URNs) used to identify resources that conform to the standards published by the Data Documentation Initiative (DDI) Alliance (https://ddialliance.org/).

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79. Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on August 27, 2021.

Copyright Notice

Copyright (c) 2021 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.
Table of Contents

1. Introduction.................................................................3
2. Conventions used in this document.................................3
3. Specification.................................................................4
   3.1. Namespace ID........................................................4
   3.2. Registration Information.........................................4
   3.3. Declared Registrant of the Namespace..........................4
   3.4. Declaration of Syntactic Structure............................4
      3.4.1. Description.................................................4
      3.4.2. ABNF Grammar.............................................5
      3.4.3. Regular Expression........................................7
      3.4.4. Examples of DDI URNs......................................8
   3.5. Relevant Ancillary Documentation................................8
   3.6. Identifier Uniqueness Considerations..........................9
   3.7. Identifier Persistence Considerations........................9
   3.8. Process of Identifier Assignment..............................9
   3.9. Process for Identifier Resolution.............................9
   3.10. Rules for Lexical Equivalence................................10
   3.11. Conformance with URN Syntax................................11
   3.12. Validation Mechanism...........................................11
   3.13. Scope.............................................................11
4. Namespace Considerations................................................11
   4.1. URN Assignment Procedures........................................11
   4.2. URN Resolution/Delegation.......................................11
   4.3. Type of Resources to be Identified............................11
   4.4. Type of Services..................................................11
5. Community Considerations.................................................12
   5.1. Open Assignment and Use of Identifiers.......................12
   5.2. Open Operation of Resolution Servers........................12
   5.3. Creation of Software for Service Discovery...................12
6. IANA Considerations......................................................13
7. Security Considerations................................................13
8. References........................................................................13
   8.1. Normative References.............................................13
   8.2. Informative References...........................................15
9. Acknowledgments.............................................................17
Appendix A. Example DNS Records.........................................18
   A.1. Delegation of the URN Namespace "ddi".........................18
   A.2. Delegation of DDI Agencies......................................18
   A.3. DDI Services.......................................................18
Appendix B. Algorithm for DDI Service Discovery......................20
1. Introduction

This document registers a formal namespace identifier (NID) for Uniform Resource Names [RFC8141] associated with DDI resources in accordance with the process defined in [RFC8141].

The DDI Alliance is an international collaboration dedicated to establishing metadata standards and semantic products for describing social science data, data covering human activity, and other data based on observational methods. DDI specifications are free standards that can document and manage different stages in the research data lifecycle, such as conceptualization, collection, processing, distribution, discovery, and archiving. Documenting data with DDI facilitates understanding, interpretation, and use -- by people, software systems, and computer networks.


DDI URNs support reusability of DDI resources inside a single DDI instance and in a distributed network of DDI instances.

The DDI specification is developed and maintained by the DDI Alliance [DDIALL]. The DDI Alliance is a self-sustaining membership organization whose over 40 member institutions have a voice in the development of the DDI specifications.

2. Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

In this document, these words will appear with that interpretation only when in ALL CAPS. Lower case uses of these words are not to be interpreted as carrying RFC 2119 significance.

"DDI": Data Documentation Initiative. The single term is often used as a synonym for the DDI specification.
"DDI agency": An organization which maintains DDI resources.

"DDI Alliance": Alliance for the Data Documentation Initiative (DDI).

"RFC": The basic publication series for the IETF. RFCs are published by the RFC Editor and once published are never modified ([RFC2026] section 2.1).

3. Specification

This section provides the information required to register a formal namespace according to the registration procedure defined in [RFC8141]. The URNs conform to the syntax defined in [RFC8141].

3.1. Namespace ID

The Namespace Identifier (NID) "ddi" is requested.

3.2. Registration Information

Version 1
Date: February 23, 2021

3.3. Declared Registrant of the Namespace

Name: The Data Documentation Initiative Alliance (DDI Alliance)
Address: ICPSR, University of Michigan
        PO Box 1248
        Ann Arbor, MI 48106-1248
        USA
Website: ddialliance.org
Contact: Joachim Wackerow
E-mail: joachim.wackerow@posteo.de,
       secretariat@ddialliance.org

3.4. Declaration of Syntactic Structure

3.4.1. Description

The Namespace Specific String (NSS) of all URNs using the "ddi" NID is a globally unique identifier consisting of the DDI agency identifier (registration authority identifier), the identifier of the DDI resource (data identifier), and the version of the resource (version identifier) [DDIID]. This structure is according to the International Registration Data Identifier (IRDI) defined in ISO/IEC 11179 Information technology - Metadata registries (MDR) - Part 6: Registration, Annex A [IS11179].
A description of the DDI resource identification is available in the DDI Lifecycle (3.3) Technical Guide, section Identification [DDIID].

The DDI NSS has the following structure:

<agency-identifier>::<resource-identifier>::<version-identifier>

agency-identifier is the identifier of a DDI agency that maintains DDI resources. This identifier basically follows the rules of reversed domain names and is case-insensitive. This way, the DNS resolution of DDI agency identifiers is supported. The hierarchy of domains descends from the left to the right label in the name; each label to the right specifies a subdivision, or subdomain of the domain to the left. The left-most label of agency-identifier conveys the top-level domain. It SHALL be a country code corresponding to ISO 3166 alpha-2 codes [ISO3166] or another top-level domain maintained by IANA [TLD]. All two-letter top-level domains are reserved for current and future ISO 3166 codes. Assignment of identifiers for DDI agencies in the requested namespace is managed by the DDI Alliance (see section 3.8, on "Process of Identifier Assignment"). The next subdomain identifies the agency within that top-level domain. Further optional subdomains can follow. The top-level domain and possible subdomains are separated by the full stop character. The full stop character is not allowed within top-level domain names or subdomain names. The top-level domain and subdomains are composed from the limited set of characters for the preferred form of a DNS label ([RFC1035] section 2.3.1). The length of the label and the full name are restricted by DNS rules ([RFC2181] section 11). The agency identifier is case insensitive ([RFC4343] section 2).

resource-identifier is the identifier of a DDI resource of a DDI agency. The value MUST be unique in the scope of this DDI agency. The resource identifier is case sensitive.

version-identifier is the version of a DDI resource of a DDI agency. The value MUST be unique in the scope of this resource. The resource version is case sensitive.

3.4.2. ABNF Grammar

The following syntax specification for the complete URN uses the augmented Backus-Naur Form (ABNF) as described in [RFC5234].
ABNF Grammar

; Rules are case-sensitive, if not stated otherwise.
 ddi-urn = urn separator ddi separator ddi-irdi
; urn is case-insensitive, see [RFC8141].
 urn   = "urn"
; ddi is the URN namespace identifier.
; ddi is case-insensitive, see [RFC8141] section 2.1.
 ddi   = "ddi"
; ddi-irdi is the namespace specific string (NSS).
; ddi-irdi - international registration data identifier,
; see [IS11179] Annex A.2.
 ddi-irdi = agency-identifier separator
     resource-identifier separator
     version-identifier
; agency-identifier is case-insensitive. See [RFC4343] section 2.
; For allowed characters see [RFC1035] section 2.3.1.
; For length restrictions see [RFC2181] section 11.
 agency-identifier = top-level-domain
   *(sub-separator ddi-authority-id)
; length limit is 255 characters
 ; see section 11 of [RFC2181]
 top-level-domain = dns-label
 ddi-authority-id = dns-label
 ddi-sub-authority-id = dns-label
 dns-label = (ALPHA / DIGIT)
   [ *(ALPHA / DIGIT / "-")
       (ALPHA / DIGIT) ]
 ; length limit is 63 characters
 ; see section 11 of [RFC2181]
 resource-identifier = restricted-string
     *("/" restricted-string)
 version-identifier = restricted-string
     *("/" restricted-string)
 restricted-string = 1*(unreserved / sub-delims / ":")
; Definitions for unreserved and sub-delims from [RFC3986] 2.2.
 unreserved = ALPHA / DIGIT / ":" / "." / "-" / "~"
 sub-delims = "!" / "$" / ":" / ":" / ":" / ":" / ":" / ":" / ":" / ":" / ":" / ":"
 separator = "::"
 sub-separator = ":"
; ALPHA and DIGIT are actually defined in the ABNF specification.
; They are declared here for convenience purposes.
 ALPHA  = %x41-5A / ; uppercase letters
         %x61-7A ; lowercase letters
 DIGIT  = %x30-39 ; digits
3.4.3. Regular Expression

The used syntax is the XML Schema flavor which can be easily used in other flavors. These regular expressions implicitly anchor at the head and tail. The following regular expression syntax uses components (component names indicated by angle brackets, i.e. <component>) and is written in free-spacing mode for easier reading (the XML Schema flavor does not support that). Please note that use of multiple quantifiers in regular expressions can result in false outcomes due to so-called greediness. Therefore, there are separate regular expressions for the length restriction and other purposes for the components agency-identifier and dns-label.

```
<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ddi-urn := [Uu][Rr][Nn] : [Dd][Dd][Ii] : &lt;agency-identifier&gt; : &lt;resource-identifier&gt; : &lt;version-identifier&gt;</td>
<td>DDI urn syntax with components</td>
</tr>
<tr>
<td>agency-identifier := &lt;top-level-domain&gt; . &lt;ddi-authority-id&gt; (. &lt;ddi-sub-authority-id&gt;)*</td>
<td>Agency identifier syntax with components</td>
</tr>
<tr>
<td>agency-identifier := .{1,255}</td>
<td>This is a length restriction</td>
</tr>
<tr>
<td>top-level-domain := &lt;dns-label&gt;</td>
<td>Top-level domain syntax</td>
</tr>
<tr>
<td>ddi-authority-id := &lt;dns-label&gt;</td>
<td>DDI authority ID syntax</td>
</tr>
<tr>
<td>ddi-sub-authority-id := &lt;dns-label&gt;</td>
<td>DDI sub-authority ID syntax</td>
</tr>
<tr>
<td>dns-label := [A-Za-z0-9][A-Za-z0-9]*</td>
<td>DNS label syntax</td>
</tr>
<tr>
<td>resource-identifier := &lt;restricted-string&gt; (/ &lt;restricted-string&gt;)*</td>
<td>Resource ID syntax with components</td>
</tr>
<tr>
<td>version-identifier := &lt;restricted-string&gt; (/ &lt;restricted-string&gt;)*</td>
<td>Version ID syntax with components</td>
</tr>
<tr>
<td>restricted-string := [A-Za-z0-9_-~!$&amp;'()*+,;=@]+</td>
<td>Restricted string syntax</td>
</tr>
</tbody>
</table>
```
3.4.4. Examples of DDI URNs

The examples are taken from the DDI Lifecycle 3.3. documentation. Please note that the resource identifiers are simplified. In real applications, they are much longer for unique identification purposes. They don’t relate to DDI types like the examples might suggest.

URN of a Represented Variable

urn:ddi:us.ddia1:R-V1:1

The DDI represented variable identified by "R-V1" with the version "1" of the DDI agency "ddia1" located in the domain "us" [DDIEXRV].

URN of a Question Item

urn:ddi:us.ddia1:PISA-QS.QI-2:1

The DDI question item identified by "PISA-QS.QI-2" with the version "1" of the DDI agency "ddia1" in the domain "us" [DDIEXQU].

URN as Reference to a Controlled Vocabulary

urn:ddi:int.ddi.cv:AggregationMethod:1.0

The DDI controlled vocabulary identified by "AggregationMethod" with the version "1.0" in the scope of the DDI agency "ddi" and sub-agency "cv" in the domain "int" [DDICVAG].

3.5. Relevant Ancillary Documentation

An introductory article on DDI can be found at [DDIINTR].

Information on the DDI specifications (DDI-C, DDI-L, XKOS, Controlled Vocabularies, and SDTL) can be found in the standards section of the DDI Alliance website [DDIALL].

Information on domain names can be found in the relevant RFCs.

- For an overview, see [RFC1034].
- Regarding case insensitivity, see [RFC1035] section 2.3.3.
- Regarding syntax, see [RFC952] section "Grammatical Host Table Specification" B. and [RFC1123] section 2.1.
Regarding size limits, see [RFC1123] section 2.1 and [RFC1035] section 2.3.4.

3.6. Identifier Uniqueness Considerations

Assignment of identifiers for DDI agencies in the requested namespace will be managed by the DDI Alliance, which will ensure that the assigned DDI agency identifiers are consistent with the directives for unique identification of DDI agencies.

Assignment of URNs for resources of a DDI agency in the requested namespace will be managed by the respective DDI agency, which ensures that the assigned URNs are unique for the scope of the agency.

3.7. Identifier Persistence Considerations

Persistence of identifiers is dependent upon suitable delegation of resolution at the level of the DDI agencies, and persistence of DDI agency assignment. The persistence of the referenced resource is also the responsibility of the DDI agency.

3.8. Process of Identifier Assignment

Assignment of identifiers for DDI agencies in the requested namespaces is managed by the DDI Alliance. A registry for DDI agency identifiers ensures through an approval process that the syntax of agency identifiers complies with the associated rules [DDIREGI].

Assignment of URNs for resources of a DDI agency and sub-agencies of a DDI agency in the requested namespace will be managed by the respective DDI agency.

3.9. Process for Identifier Resolution

The DDI Alliance will promote a service discovery system for identifying available services connected to DDI agencies using the Domain Name System (DNS). A DNS request for a DDI agency within the domain ddi.urn.arpa is delegated by the DNS servers of the DDI Alliance to the DNS servers of the relevant DDI agency. The response is a list of available DDI services for the agency identifier under which the agency has assigned URNs. The approach is based on the Dynamic Delegation Discovery System (DDDS) [RFC3401] and especially the straightforward URI-enabled NAPTR (U-NAPTR) [RFC4848].

The DDI Alliance is responsible for operating or delegating resolution requests to the resolution servers of the relevant DDI agencies. DDI agencies are responsible for operating or delegating
resolution servers for the agency identifier under which they have assigned URNs.

Sample Sequence Diagram for receiving a list of DDI services from the example DDI agency "ddial".

Client | NS for | NS for | NS for | DDI services
| urn.arpa | ddialliance.org | example1.edu | for us.ddial |

1. The name server (NS) of IANA for the domain "urn.arpa." is reached with the request "ddial.us.ddi.urn.arpa." for the DDI agency "us.ddial".

2./3. The request is delegated to the name server for "ddialliance.org".

4./5. The request is delegated to the name server for "example1.edu" (domain of the DDI agency "us.ddial").

6. The server responds with a list of NAPTR records [RFC3403] pointing to available DDI services for the DDI agency "us.ddial".

7. The client selects an appropriate DDI service and sends a request for a DDI URN to this service.

8. The DDI service responds for example with a DDI object identified by the requested DDI URN.

See Appendix A for examples of name server records.

3.10. Rules for Lexical Equivalence

The DDI agency identifier basically follows the rules of domain names. Domain names are case-insensitive. Thus, the portion of the URN

urn:ddi:<agency-id>:
is case-insensitive for matches. The remainder of the identifier MUST be considered case-sensitive.

3.11. Conformance with URN Syntax

The Namespace Specific String conforms to the related section in [RFC8141]. It is composed from the limited set of characters for an URN NSS [RFC8141]. Percent encoding is not used.

3.12. Validation Mechanism

The DDI Alliance will promote development of software for validation purposes.

3.13. Scope

The scope is global.

4. Namespace Considerations

There is no available namespace that will allow one to uniquely identify and access DDI resources.

4.1. URN Assignment Procedures

See also above section on Process of Identifier Assignment.

4.2. URN Resolution/Delegation

See also above section on Process for Identifier Resolution.

It is RECOMMENDED to use sub-agencies for flexible administration. For example, delegation of URNs of a sub-agency to different servers would be easily possible.

4.3. Type of Resources to be Identified

The DDI specifications define resources at a granular level, many of which can be identified by a DDI URN.

4.4. Type of Services

Examples of potential services are listed below. The services and appropriate service tags need to be defined in future. The mentioned service tags are from [RFC2169].
5. Community Considerations

5.1. Open Assignment and Use of Identifiers

DDI agency identifiers can be registered at the DDI Alliance. The DDI Alliance will maintain a registry of the assigned values for the DDI agency identifier used in the NSS. Information may be obtained from the following address: secretariat@ddialliance.org.

DDI agencies assign URNs and potential sub-agencies within the scope of the assigned DDI agency identifiers.

See also above section on Identifier Uniqueness Considerations.

5.2. Open Operation of Resolution Servers

The DDI Alliance operates publicly accessible name servers for the delegation of DNS requests within the domain ddi.urn.arpa to DNS servers of DDI agencies.

5.3. Creation of Software for Service Discovery

The DDI Alliance will promote software for the resolution of DDI agency identifiers and service discovery. See also Appendix B for the resolution algorithm. A basic resolver library is available [DDIRESO].
6. IANA Considerations

This document defines an URN NID registration of "ddi". IANA is asked to register "ddi" in the URN Namespaces registry.

The registration for "ddi" in the "URN.ARPA" zone is planned. Requests for the domain ddi.urn.arpa will be delegated to the name servers of the DDI Alliance.

7. Security Considerations

This document introduces no additional security considerations beyond those associated with the use and resolution of URNs in general.

The security of the DNS-based resolution of DDI agency identifiers is only as good as the security of DNS queries in general. A full discussion of the security threats pertaining to DNS and possible solutions can be found in [RFC3833]. Further information on security considerations regarding U-NAPTR can be found in [RFC4848] section 6.

8. References

8.1. Normative References


8.2. Informative References


9. Acknowledgments

Many thanks to Arofan Gregory, Dan Smith, and Wendy Thomas from the DDI Alliance Technical Committee, and Peter Koch from DENIC (German Network Information Center) for discussion and input that led to this document.

The following software tools have been helpful in evaluating the ABNF grammar and the regular expressions: an ABNF parser [ABNFPFE], a tool that creates regular expressions from an ABNF grammar [ABNF2RS], and a tool that generates random strings that match an ABNF grammar [ABNFGEN].

This document was prepared using the Word template 2-Word-v2.0.template.dot [RFC5385].
Appendix A. Example DNS Records

The examples use NAPTR [RFC3403] and SRV [RFC2782][RFC3958] records. The values for the services and flags fields of the NAPTR records will be determined by the DDI application ([RFC3403] section 9.).

A.1. Delegation of theURN Namespace "ddi"

Example records defined at a.iana-servers.net and other authoritative name servers for the domain urn.arpa.

The empty flag indicates that the lookup is not terminal and the next probe to DNS is for more NAPTR records where the new domain is "registry.ddialliance.org".

; Delegation to name servers of ddialliance.org
; order pref flag service regexp replacement
ddi.urn.arpa.
    IN NAPTR 100 10 "" "" "" "" registry.ddialliance.org.

A.2. Delegation of DDI Agencies

Example records defined at registry.ddialliance.org for ddi.urn.arpa.

The empty flag indicates that the lookup is not terminal and the next probe to DNS is for more NAPTR records where the new domain is the DNS server of the relevant DDI agency.

; Delegation to name servers of subdomains in ddi.urn.arpa, i.e. DDI agencies.
; order pref flag service regexp replacement
ddia1.us.ddi.urn.arpa.
    IN NAPTR 100 10 "" "" "" dns.example1.edu.
ddia2.de.ddi.urn.arpa.
    IN NAPTR 100 10 "" "" "" dns.example2.org.
ddia3.gb.ddi.urn.arpa.
    IN NAPTR 100 10 "" "" "" dns.example3.ac.uk.

A.3. DDI Services

Example records defined at dns.example2.org for ddi.urn.arpa.

The "u" flag states that the rule is terminal and that the output is a URI which contains the information needed to contact that DDI service. The "s" flag states that the rule is terminal and that the output of the rewrite will be a domain-name for which an SRV record should be queried. See also [RFC4848] section 4.4.
The service N2R returns one instance of the resource identified by the given URN. That service is a repository of DDI resources available at http://repos.example2.org/N2R/; possibly a REST-based service. The service N2C returns a description of the resource identified by the given URN. That service is a registry of DDI resources available at registry-udp.example2.org port 10060.

U-NAPTR permits regular expressions of a form that does a complete replacement of the matched string with a URI, expressed as a constant string. With this limited form of regular expression ([RFC4848] section 2.2.), applications using NAPTR need not implement full regular expression parsers.

ddia2.de.ddi.urn.arpa.
;    order pref flag
 IN NAPTR 100 10 "u" "N2R+http" ( ; service
     
      "!*http://repos.example2.org/N2R/!" ;regex
     ; replacement
   )

 IN NAPTR 100 10 "s" "N2C+udp" ( ; service
     
      "" ;regex
     ; replacement
   )

; all subdomains in ddia2.de.ddi.urn.arpa.
*.ddia2.de.ddi.urn.arpa.

ddia2.de.ddi.urn.arpa.
;    order pref flag
 IN NAPTR 100 10 "u" "N2R+http" ( ; service
     
      "!*http://repos.example2.org/N2R/!" ;regex
     ; replacement
   )

 IN NAPTR 100 10 "s" "N2C+udp" ( ; service
     
      "" ;regex
     ; replacement
   )

;_service._protocol.name
;_TTL class SRV priority weight port target
_registry. udp.example2.org
_14400 IN SRV 0 0 10060 registry-udp.example2.org.
Appendix B. Algorithm for DDI Service Discovery

The description is based on the Dynamic Delegation Discovery System (DDDS) algorithm [RFC3402].

Application Unique String

The Application Unique String is a DDI URN.

First Well Known Rule

1. Extracting the characters between the second and third colon (the agency identifier).
2. Normalizing case of that string.
3. Reversing the order of the substrings separated by dots.
4. Appending the string ".ddi.urn.arpa" to the end to get a domain name.

Valid Databases

The DNS is specified as a DDDS Database for this application which uses the NAPTR DNS resource records to contain the rewrite rules for service discovery.

The DNS is queried for NAPTR records for the domain name which is the output of the First Well Known Rule.

Expected Output

The expected output is the information necessary to connect to authoritative server(s) (host, port, protocol, or URL) for an application service within a given DDI agency. The result is a list of terminal NAPTR records pointing to services available for the relevant DDI agency.

The application selects the appropriate service and contacts the service for the given URN.

The process can be optimized by an application cache for the NAPTR records of already requested DDI agencies.
Author’s Address

Joachim Wackerow
c/o DDI Alliance, ICPSR, University of Michigan
PO Box 1248
Ann Arbor, MI 48106-1248
USA

Email: joachim.wackerow@posteo.de