

Global Digital Format Registry (GDFR)

Identifiers

Version 1.0.1
Status: DRAFT
Issued 2006-09-18

1 Introduction

The Global Digital Format Registry (GDFR) provides sustainable services to store, discover, and deliver important representation information about digital formats. A format is the set of syntactic and semantic rules for serializing an abstract information model, an expression of exchangeable knowledge. The format of a digital object must be known in order to interpret the information content of that object properly. Without knowledge of its format, a digital object is merely a collection of undifferentiated bits. Thus, format typing is fundamental to the effective use, interchange, and preservation of all digitally-encoded content.

The wide diversity and rapid pace of adoption and abandonment of digital formats present an ongoing problem for long-term preservation efforts. The purpose of the GDFR is to address this concern by providing a sustainable resource for managing format-critical representation information necessary to the preservation function. The GDFR is a distributed network of independent, but cooperating registries using a common data model and communications protocol to synchronize their holdings of format representation information. Widespread redundancy is an important component of sustainable service.

2 Identifiers

The GDFR will maintain persistent unambiguous bindings between globally unique identifiers and managed resources necessary for the operation of the GDFR.

2.1 Functional requirements

The GRID scheme will meet the following requirements:

- The GDFR identifier scheme must be adaptable to all identifiers necessary for the operation of the GDFR.
- Globally unique GDFR identifiers must be capable of being minted in the context of the decentralized and distributed GDFR network.
- Each minted GDFR identifier must be permanently bound to a single resource.
- GDFR identifiers must not be used to convey semantic information about the resources being identified.
- GDFR identifiers must be easily transcribable in a variety of international contexts.
- GDFR identifiers are for purposes of *identification* only; they are not intended to be dereferenced for purposes of retrieving the identified resource.

2.2 Specifications

The GRID scheme will meet the following specifications:

- The length of GDFR identifiers is not limited.
- GDFR identifiers are formed in a subset of the ISO 646 (ASCII) character set.

- GDFR identifiers will be defined in the “info” URI scheme using the “gdfrr” namespace.

2.3 Syntax

The GDFR identifier syntax can be expressed in Augmented Backus-Naur Form (ABNF), as defined by RFC 4234, as follows:

```

identifier = type "/" type-specific
type      = gtoken
type-specific = *( gtoken / "/" )
gtoken    = 1*gchar
gchar     = alpha / digit / "-" / "." / "_"
alpha     = "A" - "Z" / "a" - "z"
digit     = "0" - "9"

```

The `type` component of the identifier indicates the type of the resource being identified. The following identifier types are defined:

<i>Type</i>	<i>Description</i>
<code>cls</code>	Identifier type for a GDFR format classification
<code>fmt</code>	Identifier type for a digital format registered in the GDFR
<code>nod</code>	Identifier type for a node registered in the GDFR network

Additional syntactic restrictions may apply to the `type-specific` component for specific identifier types.

3 “info” URI scheme registration

GIDS identifiers can be expressed as “info” scheme URIs through the registration of an appropriate namespace, “gdfrr”.

<code>info-namespace</code>	<code>gdfrr</code>
<code>namespace-title</code>	Namespace of Global Digital Format Registry (GDFR) identifiers
<code>namespace-authority</code>	
<code>organisation-name</code>	Harvard University Library
<code>organisation-address</code>	90 Mt. Auburn Street, Cambridge, MA 02138, USA
<code>organisation-uri</code>	http://hul.harvard.edu
<code>contact-person</code>	Stephen Abrams
<code>contact-email</code>	stephen_abrams@harvard.edu
<code>contact-tel</code>	+1 (617) 495-3724
<code>contact-uri</code>	
<code>identifier-syntax</code>	
<code>description</code>	<p>The syntax for GDFR identifier is given in ABNF (RFC 4234) form:</p> <pre> identifier = type "/" type-specific type = gtoken type-specific = *(gtoken / "/") gtoken = 1*gchar gchar = alpha / digit / "-" / "." / "_" alpha = "A" - "Z" / "a" - "z" digit = "0" - "9" </pre>
<code>uri</code>	
<code>identifier-normalization</code>	
<code>description</code>	Identifiers are case insensitive.

	uri	
identifier-examples		
	native	
	info-uri	
namespace-information		
	description	The Global Digital Format Registry (GDFR) provides sustainable services to store, discover, and deliver important representation information about digital formats. The GDFR will maintain persistent unambiguous bindings between globally unique identifiers and managed resources necessary for the operation of the GDFR. The type component of the identifier indicates the type of resource being identified. Three types are currently defined: “cls”, identifying GDFR classifications; “fmt”, identifying formats registered in the GDFR; and “nod”, identifying nodes registered in the GDFR network.
	uri	
service-mechanism		
	uri	

References

- About “INFO” URIs: Frequently Asked Questions*, May 24, 2006 <<http://info-uri.info/registry/docs/misc/faq.html>>.
- Brown, Adrian, *The PRONOM PUID Scheme: A scheme of persistent unique identifiers for representation information*, DPTP-02, Issue 1, 10 November 2005
<http://www.nationalarchives.gov.uk/aboutapps/pronom/pdf/pronom_unique_identifier_scheme.rtf>.
- Crocker, D. and P. Overell, *Augmented Backus-Naur Form*, RFC 4234, October 2005
<<http://www.ietf.org/rfc/rfc4234.txt>>.
- ISO/IEC 646:1991, *Information technology – ISO- bit coded character set for information interchange*.
- Van de Sompel, H., T. Hammond, E. Neylon, and S. Weibel, *The “info” URI Scheme for Information Assets with Identifiers in Public Namespaces*, RFC 4452, April 2006 <<http://www.ietf.org/rfc/rfc4452.txt>>.
- Weibel, Stuart, *“info” URI Registry Policy*, Draft, 2003-12-03 <<http://info-uri.info/registry/docs/misc/INFO+URI+Registry+Policy+Issues+2003-12-03.rtf>>.