R

Resource Identifier

Greg Janée
Institute for Computational Earth System Science, University of California, Santa Barbara, California, USA

Synonyms

Document identifier; UUID; GUID; Uniform resource identifier; URI

Definition

In a networked information system, a *resource identifier* is a compact surrogate for a resource that can be used to identify, retrieve, and otherwise operate on the resource. An identifier typically takes the form of a short textual string. An identifier must be *resolved* to yield the associated resource.

Main Text

Resource identifiers can be broadly characterized as either *locations*, which identify resources by where they reside, or *names*, which identify resources by properties intrinsic to the resources [1]. This distinction is not absolute, and identifiers can exhibit characteristics of both classes. Nevertheless, the distinction is useful in defining the relationship between identifiers and resources. Consider:

Can two distinct, yet identical resources have the same identifier?

If a resource changes, must its identifier change?

If the answer to these questions is yes, then the identifiers should be considered names; if no, locations. To take two well-known examples, International Standard Book Numbers (ISBNs) are names, while HTTP URLs on the World Wide Web are locations.

Uniqueness

Uniqueness is the property that an identifier resolves to a single resource. The converse property – that every resource is identified by a single identifier, i.e., that identifier "aliasing" is avoided – is generally desirable, but is often not enforceable in systems that allow free generation of identifiers.

Broadly speaking, two approaches have been employed to guarantee uniqueness. The first is to incorporate into each identifier unique characteristics of the identified resource, for example a content-based signature, or characteristics of the context in which the resource and/or identifier system reside, for example a network address and timestamp. UUIDs incorporate both types of characteristics. The second approach is to acquire identifiers from an "authority" that maintains a centralized store of previously generated identifiers (identifier–resource associations are often stored as well). For scalability such systems are often arranged hierarchically so that a root authority, located at a well-known address, may delegate identifier generation and resolution requests to distributed sub-authorities. DNS and the Handle system are two well-known examples of this approach.

Persistence

Persistence is the property that an identifier continues to reference the associated resource over time. Strictly speaking, persistence is not a property of an identifier, or even a property at all; it's an outcome of the commitment of the operator of the identifier resolution system. A persistent identifier system is one that attempts to address known risks to persistence.

The risk of identifier breakage due to resource movement is universally mitigated by employing indirection: identifiers identify intermediate quantities which are maintained by resource owners to track current resource locations. In principle the indirection may be hidden from users, but for scalability reasons it is typically exposed. For example, the persistent uniform resource locator (PURL) system employs HTTP's redirection mechanism. The risk of breakage due to resource renaming has been mitigated in some systems by issuing so-called "semantics-free"

identifiers; for example, DOIs are strings of digits with no external referent. However, the benefit of this approach must be balanced by the inscrutability of such identifiers to humans. Other notable persistent identifier systems include OpenURLs, which identify objects by metadata constraints, i.e., by intrinic resource properties; "robust hyperlinks," which append content-based signatures to locations, specifically URLs; and archival resource keys (ARKs), which incorporate a protocol for obtaining resource persistence guarantees and policies.

Other Properties

Additional desirable properties of resource identifiers include global scope, global uniqueness, extensibility, machine readability, recognizability in text, and human transcribability [3]. Identifiers that are subject to transcription errors may benefit from having error-correcting codes incorporated into them.

Cross-references

- ► Citation
- ► Digital Signatures
- ► Distributed Architecture
- ► Object Identity

Recommended Reading

1. Jacobs I. and Walsh N. (eds.) (2004). Architecture of the World Wide Web, Volume One. http://www.w3.org/TR/webarch/
2. Hilse H.-W. and Kothe J. (2006). Implementing persistent identifiers: overview of concepts, guidelines and recommendations. London/Amsterdam: Consortium of European Libraries and European Commission on Preservation and Access. ISBN 90-6984-508-3. http://nbn-resolving.de/urn:nbn:de:gbv:7-isbn-90-6984-508-3-8
3. Sollins K. and Masinter L. Functional Requirements for Uniform Resource Names. IETF RFC 1737, 1994. http://www.ietf.org/rfc/rfc1737.txt