

**International Association of Sound
and Audiovisual Archives**



**Internationale Vereinigung der
Schall- und audiovisuellen Archive**



**Association Internationale d'Archives
Sonores et Audiovisuelles**



**Asociación Internacional de Archivos
Sonoros y Audiovisuales**



Published on *International Association of Sound and Audiovisual Archives* (<https://www.iasa-web.org>)

[Home](#) > 4: Unique and Persistent Identifiers

4: Unique and Persistent Identifiers

4.1 Introduction

4.1.1 A digital sound recording, whether stored on a mass storage system or on discrete carriers, must be able to be identified and retrieved. An item cannot be considered preserved if it cannot be located, nor linked to the catalogue and metadata record that gives it meaning. There is a need for every digital item to be unambiguously and uniquely named. In ensuring that the digital object is unambiguously and uniquely named the first step in the identification is to determine what is being named, and at what level.

4.1.2 All computer records by their very nature have some sort of system identifier that enables them to be stored without conflict. This identifier may be an acceptable public identifier, but more often than not such identifiers are system oriented and subject to change based on system requirements. There is a subsequent need for a persistent public identifier to maintain an item's accessibility, to ensure that it can be located and displayed by those who wish to use it so that citations and links made to it continue to provide access to it. There is also a requirement for that identifier to resolve to the item to which it refers regardless of where it has been stored or what its system identifier may have become.

4.1.3 The Resource Description Framework (RDF) standard is an important reference for the identification of digital objects (<http://www.w3.org/RDF/> ^[1]). RDF is based on the concept of identifying things using Web identifiers called URIs (Universal Resource Identifiers). The identification systems are based on two basic mechanisms. The first is the naming of an item by creating an identifier based on semantics or other rules of labelling such that the identifier will remain attached to the item. In the RDF standard, such identifiers are called URNs (Universal Resource Names). The second is the locator, which is organising a location system so that the item intended to be identified could be found from the locator. In the RDF standard, such identifiers are called URLs (Universal Resource Locator).

4.1.4 There have been many proposed schemes for naming a digital object, some specifically for audio or audiovisual objects, amongst them the EBU Technical Recommendation R99-1999 'Unique' Source Identifier (USID) for use in the <OriginatorReference> field of the Broadcast Wave Format (BWF). Such schemes are intended to provide a unique number within a particular community. Such schemes have not been successful in obtaining universal acceptance.

4.2 Persistent Identifiers

4.2.1 Even before the issue of digitisation made it critical, libraries, archives and audio collections generally have tended to develop systems with varying degrees of sophistication, which allow them to access their materials. These numbering systems, which tend to be unique within their own domain, can be incorporated into more universal naming schemes with the addition of a unique name for the domain or institution. This kind of structure allows maximum flexibility to an organisation in the local identification of its resources, whilst allowing the identifiers to be incorporated into a global system with the addition of an appropriate naming authority component. These persistent identifiers are for the user of the content to be able to identify a

work (as opposed to a file) which remains constant through time as a reference for that work regardless of how the file naming conventions have changed.

4.2.2 A Persistent Identifier (PID) is an identifier constructed and implemented such that the identified resource will remain the same independently of the location of its representation and independent of the fact that several copies are available at various locations. It means that the PIDs are URNs.

4.3 File Naming Conventions and Unique Identifiers

4.3.1 Care should be taken when discussing this subject to maintain the distinction between the persistent identifier used to refer to a work, and the file naming conventions. In many practical system there may well be links between the two. This section makes recommendations about file naming conventions. Data files managed in any given repository may include several types of data, not just audio. A Unique Identifier (UID) uniquely identifies a resource. This means that the identifier may change for the particular embodiment of the resource and each copy of the resource has its own ID. It consequently means that the UID are URL's. For the purposes of this discussion, file names will also be referred to as unique identifiers.

4.3.2 For linkages within and external to any system the unique identifier is the primary key to managing audio data and all of its associated files, e.g. the master copies, playback copies, compressed versions of playback copies, metadata files, edit lists, accompanying texts, images, versions of any one of those master files or derivatives. Therefore, unless the archive is using a system-assigned 'dumb' identifiers, it is vitally important that the unique identifier's structure is logically determined, clearly understood by those who have to apply it, and able to be read by people and machines. It is also important to reveal the connections between 'families' of data files: one commentator likens this connectivity to "the persistent 'thread' that enables resources to be re-tagged or re-stitched on the Web". Talking in terms of 'resources' rather than collections is an important underlying concept in these guidelines.

4.3.3 One of the most powerful ways of constructing an identification system that reveals those connections is to base it on the concept of Root ID (RID). The RID is the identifier of entity. All the files and folders involved in the representation of the entity will be derived from the RID by addition of prefixes and suffixes such as the creation of unique identifiers.

4.3.4 Regardless of whether identifiers have embedded intelligence or not, it is normal for computer-generated and computer-readable identifiers to have fixed length codes as the primary key. This offers the following advantages:

- 4.3.4.1 They enable rules to be established for creating new unique identifiers.
- 4.3.4.2 They guarantee unambiguous recognition in the system (and for users who know the rules).
- 4.3.4.3 They permit validation of the code or components of the code.
- 4.3.4.4 They support searching, sorting and reporting.

4.3.5 There has been a prolonged debate about the relative merits of dumb and intelligent or expressive unique identifiers. Most systems allocate a dumb identifier the moment data are saved. They are quickly applied, require no human intervention and their uniqueness is guaranteed. However, their randomness and arbitrariness means that other ways have to be found to show how the different files generated in the life-cycle of a digital resource connect. A better way to do this is by use of intelligent, expressive identifiers.

4.4 Identifier Characteristics

4.4.1 The following characteristics should be considered when developing a naming scheme:

- 4.4.1.1 Uniqueness, the naming scheme must be unique within the context of the organisation's digital resources and, if necessary, globally unique.
- 4.4.1.2 There should be a commitment to persistence; an organisation must have a commitment to maintain the association of the current location of the resource with the persistent identifier.
- 4.4.1.3 An identifier system will be more effective if it is able to accommodate the special requirements of different types of material or collections.
- 4.4.1.4 Although not absolutely critical, and not essential for machine generated persistent identifiers, a system will generally be more successful if it is easy to understand and apply, and if it lends itself to short and easy to use citations.
- 4.4.1.5 The identifier should be capable of distinguishing parts of an item, as well as versions and roles that a digital item might have. Relying on the file extension to distinguish a distribution copy from an archival copy is not advisable as the format may change over time, though the role remains the same (Dack 1999).
- 4.4.1.6 The identifier should permit batch renaming for ingestion into different content management systems.

© International Association of Sound and Audiovisual Archives (IASA)
Translation not permitted without consent of IASA Executive Board

[PRIVACY AND COOKIES](#) | [TERMS OF USE](#) | [CONTACT US](#) | [SITE MAP](#)



Source URL: <https://www.iasa-web.org/tc04/unique-and-persistent-identifiers#comment-0>

Links

[1] <http://www.w3.org/RDF/>