A Basic Comparative Framework for Evaluation of Digital Identifier Systems

Hamid Reza Khedmatgozar, Mehdi Alipour-Hafezi
Iranian Research Institute for Information Science and Technology (IRANDOC)
Iran
khedmatgozar@students.irandoc.ac.ir, alipour@irandoc.ac.ir

ABSTRACT: *In recent years, with the development of digital environment, one of the concepts that have been particularly important is digital identification of objects in this environment by digital identifier systems. The main objective of this study is to answer this important and basic question: How and on what framework can distinguish a digital object identifier system from other identifier systems? To achieve this goal, a comparative framework which consisted of 7 main characteristics of digital identifiers in three major domains (namely identifier features, digital coverage and comprehensiveness of scope) was proposed. Then, an illustrative comparison of 23 well-known worldwide identifiers was carried out. Results of this study revealed that six identifier systems of DOI, Handle, UCI, ARK, URN and PURL can be regarded as the best choices of digital identifiers in public areas identified in the digital space and four identifier systems of ORCID, MIRIAM, ISNI and NBN are the best options to be used in certain specialized areas. These results can help information technology specialists to identify and effectively use of digital identifier systems in actual and potential practical areas.*

Keywords: Digital Identifier Systems, Object, Digital Environment, Illustrative Comparison

Received: 17 May 2015, Revised 18 June 2015, Accepted 23 June 2015

1. Introduction

Digital object is a data structure whose principal components are digital materials or data plus a unique identifier for that material (Kahn and Wilensky, 2006). In other words, the so-called digital object can be used for any form of intellectual property appeared to be applied in the digital environment (IDF, 2015). In general, two main elements in the identification of digital objects are metadata and identifier (Arms, 2001).

Up to now, several definitions have been proposed for identifiers. For example, Kunze (2003) introduces an identifier as the association between one string and one thing. Using the term "*one*" for an identifier and for identified entity is what attracted the attention in all the definitions. This indicates a need for one-to-one relationship between an identified entity and identifier. This one-to-one relationship can be called uniqueness (Coyle, 2006). On the other hand, an identifier should not be changed and is not changeable (Clarke, 1994; Campbell, 2007). Paskin (1999) states that an identifier should have an unlimited lifetime, even

if the existence of identified entity is wiped out. He called this characteristic "*persistency*". Therefore, it can be said that uniqueness and persistency are the most important intrinsic characteristics of an identifier.

Now, an object with digital identity is referenced by the URL (Kunze, 2003). Over time, some risks of using the URL as an identifier have been detected. Impersistency (Prasad and Guha, 2005) and violations of uniqueness (Coyle, 2006), among those in practical areas like e-commerce (Sidman and Davidson 2001), citation analysis and cross referencing and linking (Simons, 2012) and digital rights management systems (Carreiro, 2010) are examples of these risks.

With this description, researchers have tried to provide solutions in order to create a unique identifier with the ability to have long-term persistency to overcome the above-mentioned risks. The main solution proposed for this problem is the use of indirect names instead of URLs; what worked for the DNS (Domain Name System) in stabilizing internet hostnames should work for digital object references (Kunze, 2003). Put simply, the proposed strategy is based on creating a system called digital identifier system for managing digital identifiers and a process called resolution. Resolution is identifier submission process based on a name to a network service (digital identifier system) and in return, receiving one or more pieces of current information related to the identified object, such as the location (URL) of the object. Resolution creates a level of managed redirection between the output and the identifier (IDF, 2015). Simplified process of this solution is shown in Figure 1.

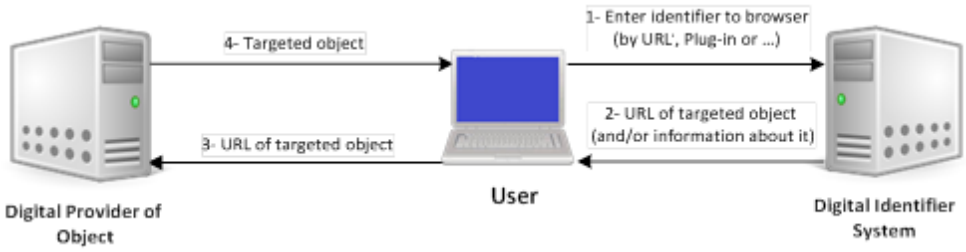

Figure 1. The simple process of resolution in digital identifiers as a solution for URL problems

But the main advantages and functions of this solution for digital object would be maintaining a permanent location information, standardizing and enriching metadata, facilitating content search, Securing copyright, increasing the volume of distribution, increasing efficiency, reducing costs and saving time, enabling exploitation of new business areas, and increasing the revenues related to digital content (Park et al., 2007). Paskin (2002) in his study pointed out some advantages of using these identifiers as increasing referrals and reducing the time to manage them to be saved, retrieving and distributing objects with identity in the digital space, persistency of access and enrichment of metadata. Attanasio (2003) mentioned solving the problems of digital right management with the use of these identifiers. In another study, Wynholds (2011) discusses the advantages of using these identifiers in fast and accurate binding of research data and their use in scientific publications.

Two main questions that should be asked in the face of the digital identifier are as follows: which of the available identifier systems are digital identifier systems? And on what framework and criteria can they be diagnosed? Answers to these questions can help Specialists, to design, implement, and also identify the strengths, and apply them effectively in the digital environment. Despite the fact that all over the world, there have been created many identifier systems to identify objects with identity in the digital environment, however, according to preliminary reviews, no study was found that determined criteria for separating them from physical environment identifiers Given these facts, the main objective of this study is to provide a framework for identifying authoritative digital identifier systems based on a comparative evaluation.

2. Research Methodology

To answer the question posed in the previous section, the comparative evaluation based on the principles proposed by Vartiainen (2002) was considered. He states that 4 principles should be identified in order to perform comparative evaluations:

• **Selecting the object for evaluation:** Specify the object and selection procedure of it. In this study, the object is the well-known worldwide identifier system in order to comparative evaluation of it with definition of the defined identifier systems.

• **The Level of Comparison:** Specify the scope, dominant principles, and similarity or difference levels of the objects. The scope

of this evaluation is limited to the proposed identifier systems in the world. All of the studied objects are identifier systems which are similar.

• **Conceptual comprehension:** Clearly define the concepts. The definitions of identifiers and digital identifiers were presented\ in the previous section.

• **Analysis of the findings of an evaluation:** Specify the method of analysis of the findings of evaluation. Given the purpose of the evaluation, the selected approach in this study is illustrative comparisons. In illustrative comparisons, evaluation units are compared indirectly and based on the proposed model or framework. The main use of illustrative comparisons is standardization and generalization of the use of the framework employed in the evaluation.

According to this principle, this study was considered in two main steps:

1- Designing the basic comparison framework: The framework was designed in two stages. The first stage was based on the literature review to include early identification criteria, and second stage included focus group interview with 4 experts to validate the chosen criteria. The experts were selected based on the purposeful sampling and their selection criteria were having expertise in the fields of information science and information technology as well as having familiarity with philosophy and digital identifier mechanisms. These criteria can be seen in Table 1. Some questions were designed for each criterion in order to give objectivity and measurability to the framework. Three statuses (namely approved, disapproved, and conditioned) were considered for each criterion based on the responses provided to the question(s) that measure it. According to Consensus of experts, the criteria for choosing a digital identifier system is that it at least has all selection criteria conditionally. Content and structural validity of the criteria and questions was also reviewed and approved.

2. Implementation of Comparative evaluation: After identifying 23 well known worldwide identifier systems, evaluation of each of these systems in each of the criteria was done by providing answers for the designed questions for each criterion. Scoring was applied in each criterion according to the answered approved states. Studying primary identifier systems, based on a deep and detailed investigation of articles, documents, standards, and websites of each using library method, was considered as the basis of judgment. To ensure the accuracy and precision of status determination of each system at all criteria, firstly, the evaluation was performed by two experts simultaneously and separately, and then preliminary results were finalized in some joint meetings.

3. Findings

The results of the performed comparative evaluation can be seen in Table 2. In this table the status of each of the systems is specified in each of the digital identifier criteria and the systems are sorted based on total points and compatibility with the defined criteria. According to the defined criteria described in the previous section and according to these results, systems that have conditional and approved statuses at all criteria can be introduced as best choices complied with the definition of digital identifier. These systems include: DOI, Handle, UCI, ARK, URN and PURL. However, some problems of URN system in single management of resolution mechanism, and PURL in proof of digital uniqueness and persistency in recognized resolvers can separate these two systems from four other systems in terms of performance.

At the second level, four systems of ORCID, MIRIAM, ISNI and NBN can be proposed as suitable options for digital identifier. The main problem of these four systems is at the seventh criterion, i.e. comprehensiveness of identification, which attempt to identify a specific part of objects with identity in the digital environment only in a particular field. ORCID system identifies scientific researchers and writers, MIRIAM system identifies biological models, ISNI system identifies contributors to creative works and those active in their distribution, and NBN system identifies library resources.

Good performance of ORCID and MIRIAM systems in creating mechanisms for direct access to the identified object using an identifier is also one of the most important features of them. In this criterion, i.e. digital actionability, there are considerations in ISNI and NBN systems such as lack of performing resolution mechanism under a single authority or performing this mechanism in a different and indirect way. Interoperability created between ORCID, ISNI and DOI can also be considered as a qualitative advantage of such identifier systems.

Another point which can be noted in the obtained results is that according to four main criteria of uniqueness, persistency,

digital uniqueness and digital persistency, Open URL system cannot be considered as an identifier system, and their provided address cannot be considered an identifier. This is in line with the findings of the studies by Warren (2005), Hakala (2010) and Paskin (2010). Warren (2005) states that Open URL is a syntactic combination for transferring identifiers and metadata along with the URL. He notes that due to the existing restrictions, it does not seem that Open URL is a suitable option to identify

Domain	Criterion	Definition	Questions	Resource
Identifier features	uniqueness	One to one correspondence between an identifier and the object itself in a physical environment. This means that an identifier must refer to a copy of targeted object that is located in the physical environment, and the targeted object also must have only one identifier.	Is identifier unique in its definition scope?	Clarke (1994), Paskin (1999), Coyle (2006), Campbell (2007),Lee and Stivila (2014)
			Is it possible to ensure that the identifier is not used for more than one item?	
	persistency	An identifier should remain unchanged throughout its life, and accessible and usable in the foreseeable future.	Does the identifier remain unchanged?	
			Is there any predicted life span for the system?	
Digital coverage	digital identification	All or part of the system is designed and implemented in order to identify objects that have been identified in the digital environment.	Does the system in its own identification area attempt to identify objects with identity in the digital environment?	Green and Bide (1997), Paskin (1999), Paskin (2000), Sidman and Davidson (2001), Kunze (2003), Prasad and Guha (2005), Paskin (2010), Pisanski and Aalbrg (2010), Duerr et al. (2011), Juty et al. (2012)
	digital uniqueness	One to one correspondence between an identifier and the object itself in a digital environment. This means that an identifier must refer to a copy of targeted object that is located in a specific address in the digital environment, and the targeted object also must have only one identifier.	In case of digital identification, is there any one to one relationship between an identifier and an object in a digital environment?	
	digital persistency	The system under investigation must provide this guarantee that long term relationship between identification number of the targeted object and the targeted object itself, or its related metadata is established.	Is this guarantee given by the system that there exists a permanent relationship between identification number of the targeted object and the targeted object itself, or its related metadata?	
	digital actionability	The system under study must have proposed a mechanism in order to provide accessibility to the desired object in the digital environment through the identifier itself or with the serial number or the defined way in it (through the resolver system).	In case of digital identification, is there any clear and focused mechanism proposed by the system for direct accessibility to the location of the desired object (not the metadata generated by the system) in the digital environment through it?	
Comprehensiveness of scope	globality	The system under study must be designed at the international level in order to identify different types of objects in different areas.	Does the system under study have the capability of investigating a variety of different objects in different areas internationally?	Paskin (1999), Coyle (2006), Campbell (2007), Lee and Stivila (2014)

Table 1. Basic Comparative Framework of Digital Identifier System evaluation

uniqueness of a document. Hakala (2010) also notes that although metadata along with Open URL can include an identifier, it does not necessarily guarantee its uniqueness and persistency. He considered probability of entering various kinds of identifier into the metadata along with Open URL as well as dependence of its persistency on persistency of the identifier as one of his main reasons. Paskin (2010) has also introduced Open URL in his study as a mechanism for transferring metadata and identifiers of a published item in order for context-sensitive linking and stated to be that Open URL must use digital identifier such as DOI in its associated metadata if it wants unique.

R	Symbol	System Name	1	2	3	4	5	6	7
1	ARK	Archival Resource Key	✓	✓	✓	✓	✓	✓	✓
2	DOI	Digital Object Identifier	✓	✓	✓	✓	✓	✓	✓
3	Handle	Handle	✓	✓	✓	✓	✓	✓	✓
4	UCI	Universal Content Identifier	✓	✓	✓	✓	✓	✓	✓
5	URN	Uniform Resource Name	✓	✓	✓	✓	✓	*	✓
6	PURL	Persistent Uniform Resource Locator	-	-	✓	*	*	✓	✓
7	ORCID	Open Researcher and Contributor ID	✓	✓	✓	✓	✓	✓	×
8	MIRIAM	Minimum information required in the annotation of models	-	-	✓	✓	✓	✓	×
9	ISNI	International Standard Name Identifier	✓	✓	✓	✓	✓	*	×
10	NBN	National Bibliography Number	✓	✓	✓	*	*	*	×
11	Open URL	Open URL	-	-	✓	×	×	✓	✓
12	OID	Object Identifier	✓	✓	✓	×	×	×	✓
13	ISAN	International Standard Audiovisual Number	✓	✓	✓	×	×	×	×
14	ISBN	International Standard Book Numbers	✓	✓	✓	×	×	×	×
15	ISMN	International Standard Music Number	✓	✓	✓	×	×	×	×
16	ISRC	International Standard Recording Code	✓	✓	✓	×	×	×	×
17	ISSN	International Serial Standard Number	✓	✓	✓	×	×	×	×
18	PII	Publisher Item Identifier	✓	✓	✓	×	×	×	×
19	SICI	Serial Item and Contribution Identifier	✓	✓	✓	×	×	×	×
20	CLEI	Common Language Equipment Identifier	✓	✓	×	×	×	×	×
21	GTIN	Global Trade Item Number	✓	✓	×	×	×	×	×
22	ISTC	International Standard Text Code	✓	✓	×	×	×	×	×
23	ISWC	International Standard Musical Work Code	✓	✓	×	×	×	×	×

Table 2. The Results of evaluation of 23 identifier systems based on the basic comparative framework

Guide: (1) identifier uniqueness, (2) identifier persistency, (3) digital identification, (4) digital uniqueness, (5) digital persistency, (6) digital accessibility, (7) comprehensiveness of identification, (✓) approval, (*) conditional, (×) disapproval.**Note:** Some systems are not defined in the physical environment in which the criteria 1, 2 have been replaced by the criteria 4 and 5.

4- Conclusion and Suggestions

The main objective of this study was to identify reliable digital identifier systems based on the illustrative comparison of the large number of identifier systems. The results obtained from the comparative evaluation, demonstrated that 7 main characteristics in three major domains, namely identifier features, digital coverage and comprehensiveness of scope, can be considered as the key criteria for recognizing digital identifier systems from other identifier systems. Then, based on the framework designed by 7 criteria, comparative evaluation was performed on 23 well known worldwide identifier systems. The results of this evaluation showed that six identifier systems of DOI, Handle, UCI, URN, ARK, and PURL are the best options for Introducing the Digital Object identifier, and four identifier systems of ORCID, MIRIAM, ISNI and NBN are also good choices for digital identification in certain specialized areas.

In general it can be said that introducing this integrated framework for the identification and differentiation of the digital identifier in this study, in addition to help experts of digital environment to identify the digital identifier systems versus other identifier systems based on valid comparative criteria, will enable them to effectively use various advantages of this identifiers in the actual and potential application areas, such as integration of digital content supply chain and network (Sidman and Davidson 2001), digital right management (Mooney, 2001; Chandrakar, 2006), cross referencing and linking (Galyani-Moghadam and URS, 2006), digital libraries (Wang, 2007) and citation analysis (Simons, 2012; Scopus, 2013). Accelerating the use of digital identifiers instead of physical identifiers in the digital environment, can also discover unidentified Potential application areas of these identifiers such as interactive television, digital museums, virtual tourism and E-Learning. Increase the number of systems studied, quality evaluation of digital identifier systems and focus on each specific application areas of these identifiers can be interest to researchers to address the restrictions of this study.

References

[1] Arms, W.Y (2001). Digital libraries. MIT press.

[2] Attanasio, P (2003). The use of DOI in eContent value chain: The case of Casalini Digital Division and mEDRA. White Paper, Multilingual European DOI Registration Agency. [Online]. Available: http://dx.doi.org/10.1392/BC1.0 (Accessed 27 March 2015).

[3] Campbell, D (2007). Identifying the identifiers. *In*: Proc. of the 7th international conference on Dublin Core and Metadata Application (DCMI 07), Doblin core, United States, p. 63-73. [online] http://dcpapers.dublincore.org/pubs/article/download/868/864 (Accessed 27 March 2015).

[4] Carreiro, E (2010). Electronic books: how digital devices and supplementary new technologies are changing the face of the publishing industry. *Publishing research quarterly* 26 (4) 219-235.

[5] Chandrakar, R (2006). Digital object identifier system: an overview. *The Electronic Library* 24 (4) 445-452.

[6] Clarke, R (1994). Human identification in information systems: Management challenges and public policy issues. Information Technology & People 7 (4) 6-37.

[7] Coyle, K (2006). Identifiers: Unique, persistent, global. *The Journal of Academic Librarianship* 32 (4) 428-431.

[8] Duerr, R., Downs, R., Tilmes, C., Barkstrom, B., Lenhardt, W. C., Glassy, J., Bermudez, L.E., Slaughter, P. (2011). On the utility of identification schemes for digital earth science data: an assessment and recommendations. *Earth Science Informatics* 4 (3) 139-160.

[9] Galyani Moghaddam, G., Urs, R.S (2006). Electronic Scholarly Journals: A Review of Technical Issues in Digital Environment. *Journal of Educational Media & Library Sciences* 44 (2) 235-245.

[10] Green, B., Bide, M (1997). Unique Identifiers: a brief introduction. Book Industry Communication/EDItEUR. [Online]. Available: http://www.bic.org.uk/files/pdfs/uniquid.pd (Accessed 27 March 2015).

[11] Hakala, J (2010). Persistent identifiers- an overview. KIM Technology Watch Report. [Online]. Available: http://www.persid.org/downloads/PI-intro-2010-09-22.pdf (Accessed 27 March 2015).

[12] IDF. 2015. DOI Handbook. [Online]. Available: http://www.doi.org/hb.html (Accessed 27 March 2015)

[13] Juty, N., Le Novère, N., Laibe, C (2012). Identifiers.org and MIRIAM Registry: community resources to provide persistent identification. Nucleic acids research 40 (D1) D580-D586.

[14] Kahn, R., Wilensky,R (2006). A framework for distributed digital object services. *International Journal on Digital Libraries* 6 (2) 115-123.

[15] Kunze, J (2003). Towards electronic persistence using ARK identifiers. *In*: Proc. of the 3rd ECDL Workshop on Web Archives. August 21st, 2003, Trondheim, Norway. [Online]. Available: http://bibnum.bnf.fr/ecdl/2003/proceedings.php?f=kunze (Accessed 27 March 2015).

[16] Lee, D. J., Stvilia, B (2014). Developing a data identifier taxonomy. *Cataloging & Classification Quarterly* 52 (3) 303-336.

[17] Mooney, S (2001). Digital object identifiers for eBooks: What are we identifying? *Publishing research quarterly* 17 (1) 29-36.

[18] Park, S., Zo, H., Ciganek, A. P., Lim, G. G. (2011). Examining success factors in the adoption of digital object identifier systems. *Electronic Commerce Research and Applications* 10 (6) 626-636.

[19] Paskin, N. (1999). Toward unique identifiers. *In: Proceedings of the IEEE* 87 (7) 1208-1227.

[20] Paskin, N. (2000). Digital Object Identifier: implementing a standard digital identifier as the key to effective digital rights management. The International DOI Foundation, Kidlington, Oxfordshire, United Kingdom.

[21] Paskin, N. (2002). Digital object identifiers. I*nformation Services and Use* 22 (2) 97-112.

[22] Paskin, N. (2010). Digital Object Identifier (DOI®) System. In: Bates,M. J., Maack,M. N. Encyclopedia of Library and Information Sciences, 3rd ed. Taylor & Francis, London, 1586-1592.

[23] Pisanski, J., •umer, M., Aalberg, T. (2010). Identifiers: bridging language barriers. In World Library and Information Congress: 76th IFLA General Conference and Assembly. Meeting, 93.

[24] Prasad, A. R. D., Guha, N. (2005). Persistent identifiers for digital resources. DRTC-HP International Workshop on Building Digital Libraries using Dspace, 7th – 11th March, DRTC, Bangalore.

[25] Scopus. 2013. Scopus: Facts and Figures. [Online]. Available: http://www.elsevier.com/__data/assets/pdf_file/0007/148714/scopus_facts_and_figures.pdf (Accessed 27 March 2015).

[26] Sidman, D., Davidson. T (2001). A practical guide to automating the digital supply chain with the digital object identifier (DOI). *Publishing research quarterly* 17 (2) 9-23.

[27] Simons, N (2012). Implementing DOIs for Research Data. *D-Lib Magazine* 18(5/6). [Online]. Available: http://dx.doi.org/10.1045/may2012-simons (Accessed 27 March 2015).

[28] Vartiainen, P (2002). On the principles of comparative evaluation. *Evaluation* 8 (3) 359-371.

[29] Wang, J (2007). Digital object identifiers and their use in libraries. *Serials review* 33 (3) 161-164.

[30] Warren, S. A (2005). DOIs and Deeplinked E-Reserves: Innovative Links for the Future. *Technical Services Quarterly* 22 (4) 1-17.

[31] Wynholds, L (2011). Linking to scientific data: Identity problems of unruly and poorly bounded digital objects. *International Journal of Digital Curation* 6 (1) 214-225.