

A uniform digital identifier for scientific articles from manuscript through citation

✉ Ernesto G Rodríguez, Raúl A Valdés, Marta E Ferrer, Dania A Peña

Editorial Elfos Scientiae,
Centro de Ingeniería Genética y Biotecnología, CIGB
Ave. 31 e/ 158 y 190, Cubanacán, Playa, CP 11 600, La Habana, Cuba
E-mail: ernesto.galban@cigb.edu.cu

FOCUS

ABSTRACT

A uniform digital identifier (UDI) is presented for scientific manuscripts from editorial receipt through bibliographic citation. The UDI replaces conventional pagination and allows citing the manuscript upon acceptance, remaining as article's metadata after the assignation of bibliographic information (year, volume and issue). It provides explicit information on the journal, type of article, manuscript editorial ID code, acceptance date, and, if peer-reviewed (i.e., original research) the period in months of that process. Additionally, the language of publication can be added as ISO 639-1. Significantly, it provides uniform tracking of manuscripts from submission, functioning as locator for all the article versions, and increases the explicit information on editorial processing provided when citing the article in bibliographic references. The last one will favor the bibliometric analysis of editorial peer-review directly from reference lists, adding value to them and also to bibliographic database metadata.

Keywords: persistent identifier, uniform digital identifier, journal article, electronic publication, scientific and technical publications

Biotecnología Aplicada 2013;30:142-144

RESUMEN

Identificador digital uniforme para los artículos científicos desde el manuscrito hasta la citación. Se propone un identificador digital uniforme (IDU) para manuscritos científicos desde la recepción editorial hasta la citación bibliográfica. El IDU reemplaza la paginación tradicional y permite citar el manuscrito tras su aceptación, y permanece como metadato del artículo después de asignarle a este la información bibliográfica (año, volumen y número). Este identificador proporciona información explícita sobre la revista, el tipo de artículo, el código de identificación de procesamiento editorial, la fecha de aceptación y, en caso de que el artículo sea evaluado por pares, el periodo de revisión en meses. Adicionalmente, se puede indicar el idioma del artículo mediante el código ISO 639-1. De forma significativa, el IDU permite el seguimiento uniforme de los artículos durante todo su ciclo de vida desde el envío, funciona como localizador de todas sus versiones e incrementa la información explícita sobre el procesamiento editorial, disponible como metadato al citarlo dentro de listas de referencias bibliográficas. Este último aspecto favorecería el análisis bibliométrico de los periodos de revisión editorial por pares a partir de las listas de referencias, y les añadiría valor a estas y a sus metadatos depositados en las bases de datos bibliográficas.

Palabras clave: identificador persistente, identificador digital uniforme, artículo de revista, publicación electrónica, publicaciones científicas y técnicas

Introduction

The generation of information persistent identifiers (PIs) has characterized the last 20 years of scientific literature and information sciences. Among the most widely known and established variants are the uniform resource identifier (URI) [1], the digital object identifier (DOI) [2, 3], the publisher item identifier (pii) [4] and more recently the application identifiers (AIs) [5]. The first two are intended for locating digitally the manuscript or the article, the pii for tracking the manuscript during the editorial process and AIs to interconnect physical objects with business related information. In the case of URI, DOI and pii, all of them are based on previous PIs as the international standard serial number (ISSN)[6] for periodicals or the international serial book number (ISBN) [7].

PIs have not been limited to locate electronic contents, but has also expanded to other types of identifiers for uniform tagging of scientific article metadata as the author ID for authority control [8] (e.g., ResearcherID [9], and more recently, the Scopus Sciverse's ORCID [10]) and articles version identifiers as CrossMark [11].

PIs aimed at identifying scientific manuscripts interrupted in bibliographic guidelines and speedup scientific articles consultation and citation as e-prints upon editorial acceptance. However, bibliographic guidelines have derived in a myriad of citation variants which are optional and the citation format generally differs for the same article based on journals' instructions for authors and bibliographic style guidelines. This required the availability of tracking systems for article's versions. At the same time, some of these identifiers require payment to be assigned (DOI), and others (pii) are restrained to a publishing consortium. For a more complicated picture, many journals have incorporated, rather than migrated to, digital publishing, with authors mostly favoring citation of the printed version of the article as the most traditional one. This has led strictly electronic journals to use unique codes for the articles instead of pagination (e.g., PLoS journals [12]) and exclusively to identify the article during the editorial process (the case of pii), all of them coexisting at a given time.

Simultaneously, bibliographic formats have evolved

1. Green B, Bide M. Unique Identifiers: a brief introduction. London: Book Industry Communication ; 1999. Available from: <http://www.bic.org.uk/files/pdfs/uniqueid.pdf>

2. International DOI Foundation. The DOI® system. Available from: www.doi.org

3. Paskin N. Digital Object Identifier (DOI®) System. In: Bates MJ, Maack MN, editors. Encyclopedia of Library and Information Sciences. 3rd ed. England: Taylor & Francis; 2010. p. 1586-92.

4. Paskin N. Publisher Item Identifier as a means of document identification. Kidlington: Elsevier Science Ltd; 2003. Available from: <http://web.archive.org/web/20031013073003/http://www.elsevier.nl/inca/homepage/about/pii/>

5. GS1 Application Identifiers. Brussels: GS1; c2012. Available from: http://www.gs1.org/barcodes/technical/application_identifiers

to use those identifiers for citation. For example, in the Vancouver citation style [13], a standard journal citation provides: i) author names, limited to six; ii) article title; iii) journal name; iv) publishing date (year), sometimes the electronic immediate publishing date which is optional; and v) archivist metadata (volume, issue and pagination). All these characterize the act of publishing and indicate the location of the article within a journal collection. Nevertheless, pagination has lost its original aim to some extent in strictly electronic journals, since the order of a given resource within a digital collection is determined by location rather by sequence, and is searched selectively rather than sequentially. Furthermore, unless consulting the journal table of contents, the most probable way to find the article would be metadata-based retrieval or through journal website searches.

Noteworthy, PIs for scientific articles aid on tracking them at the preprint stage, but lack explicit information on the source, even when used as locator (DOI). This brings a space to optimize the information expected or useful for researchers and editorial boards when consulting the reference lists of articles, to make them more meaningful.

In order to satisfy all the abovementioned requirements, a new system for bibliographic citation of scientific articles, denominated uniform digital identifier (UDI) is stated here.

Results and discussion

After analyzing all the alternatives, and upon study of bibliographic style guidelines, a new system was proposed which unifies all the points of views on editorial processing information, that provided to readers on the nature of citation and aid on locating the articles. The structure of the identifier was established taking into account the general properties of manuscripts processed and published in the *Biología Aplicada* journal, and further challenged with the bibliographic requirements of the Vancouver style guidelines.

The uniform digital identifier (UDI) structure is shown in the figure. It comprises from left to right: the initials identifying journal name, a letter indicating the journal's section, the sequential manuscript ID number, acceptance date (for peer-reviewed articles, or the receipt date for the non-peer-reviewed ones), two digits for the extent in month of the peer-review process, and finally, the two letter 639-1 ISO code for language identification [14].

The journal section letter and the sequential manuscript ID number supports tracking the manuscript from receipt to citation, unifying the life cycle of the article, a connection traditionally lost after publishing. The sequential number was set to three digits, regarding that submissions would never exceed 999 manuscripts for a given journal per month. The two-digit standard acceptance date (month and year) coincides with the publication date in the case of strictly electronic journals, and eliminates citation mistakes due to different acceptance and publication years. Traditionally, these data remain internal in the manuscript and are troublesome to retrieve, depending on consultation for bibliometric analyses. In this way, the meaning of reference lists is expanded for bibliometrics, and supports the related editorial processing times becoming

metadata. Moreover, it is also possible to calculate corrected Price indexes.

The inclusion of the peer-review period in months brings two parameters: i) the extension of peer-review, required for editorial processing calculations; ii) the character of the article (reviewed or not, in this last case, 00) and, therefore, if citable; and also supports calculating submission date.

The UDI was structured modular to indicate the original character of the article on its present form, directly from the list of references, in which will reinforce the dominance of bibliographic databases and allow automated statistics. Moreover, it aids on the assignation of authorship by cross-field match of author names to articles' metadata, helping on authority control. It can also be created for existing journal collections, cost-free.

Once established, the UDI was challenged with the Vancouver bibliographic style. It is consistent with the coexistence of both article's versions: accepted and published; in the last, it is compatible with the full bibliographic format. For this, the traditional print pagination was replaced by the UDI. In case of printed editions, the page number can be indicated in the table of contents or even added as a modular number.

Example citations are:

Accepted preprint

Rodríguez EG, Valdés RA, Ferrer ME, Peña DA. A uniform digital identifier for scientific articles from "manuscript" through "citation". *Biotechnol Apl*. BAF339-041300EN.

Bibliographic citation

Rodríguez EG, Valdés RA, Ferrer ME, Peña DA. A uniform digital identifier for scientific articles from manuscript "through" citation. *Biotechnol "Apl"*. 2013; 30(2):BAF339-041300EN.

This format provides all the three relevant publication dates from the single citation act: submission, online publishing and archivist recording (bibliographic data). The code provides a probability of 9038952 non-overlapping combinations, including time related parameters, as calculated for *Biología Aplicada* (8 article types/sections \times 999 sequential ID submission codes per month \times 87 years left from the 21st century \times 13 values available for peer-review period including 12 months and also the 00 value for non-reviewed articles or under open peer-review if unpublished). The peer-review period can exceed 12 months, regardless first submission or not. In the case of resubmission, it can be indicated by an additional R letter prior to the acceptance date, which can also provide information on iterative peer-review, something missing in bibliographic citation styles enforced.

The UDI may also avoid metadata dispersion, something that is not completely eliminated with the optional use of pii and DOI when they replace the entire bibliographic information, and at the same time, the article can be tracked easily by open or database-directed bibliographic searches. Very significantly, it provides information on the article's editorial processing at those two publication stages, which can be visually inspected and also automatically harvested. Moreover, it solves the limitations of

6. What is an ISSN? Paris: Centre International de l'ISSN. c2008 [cited 2013 May 3]. ISSN International Centre; [1 screen]. Available from: <http://www.issn.org/2-22635-What-is-an-ISSN.php>

7. International ISBN Agency [Internet]. London: International ISBN Agency. c2012 [cited 2013 May 3]; The agency [1 screen]. Available from: <http://www.isbn-international.org/>

8. Enserink M. Scientific publishing. Are you ready to become a number? *Science*. 2009;323(5922):1662-4.

9. ResearcherID. New York: Thomson Reuters. c2012 [cited 2013 May 3]. Available from: <http://www.researcherid.com>

10. Haak LL, Fenner M, Paglione L, Pentz E, Ratner H. ORCID: a system to uniquely identify researchers. *Learn Publ*. 2012; 25(4):259-64.

11. Meyer CA. Distinguishing published scholarly content with CrossMark. *Learn Publ*. 2010;24(2):87-93.

12. PLoS. Public Library of Science. [cited 2013 May 3]. Available from: www.plos.org

13. International Committee of Medical Journal Editors. Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Sample References. [updated 2010 Jul 27; cited 2013 Jan 21]. Available from: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=22109

14. International Organization for Standardization. ISO 639-1:2002. Codes for the representation of names of languages - Part 1: Alpha-2 code. Available from: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=22109

current bibliographic formats on lacking data about the citation character of the document source, which is not indicated uniformly in the title (for example, for letters and editorials within brackets). In this sense, the UDI is highly valuable by indicating not only the citation character of the references, but also supporting its classification in primary or secondary sources of information within the lists of references, relevant for establishing the intellectual base of the article considered. The lack of all the information included in the UDI as part of reference lists has limited a unified study of scientific literature and the integrative characterization of bibliometric trends since the intellectual background perspective. For article types, a more homogeneous convention could be established by limiting R for reviews and O for original research articles, respectively (Figure), the other article type one-letter definitions being more flexible based on the journal sections.

Additionally, the UDI does not discard any previous bibliographic element, integrating all of them, and, at the same time, allows version location for the article since it is assigned by the journal. It also unifies the multi-language character of publications as *Biología Aplicada*, for articles simultaneously published in two or more languages, but retaining the single article identity.

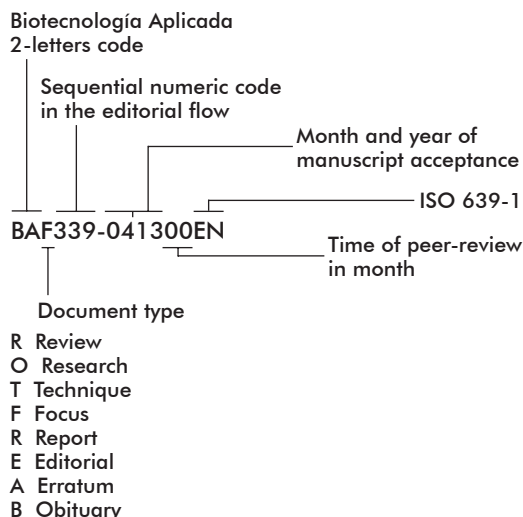


Figure. Structure of the uniform digital identifier (UDI) for scientific articles from the manuscript through citation.

In summary, the future practice on using UDI would mark a valuable contribution to information sciences, by unifying the whole submission-publishing-citation life cycle of scientific articles.