

¡Identifíquese!

Ángel Herráez

En un mundo donde la marea de información amenaza cada día con inundarnos, cada vez es más importante disponer de formas de acceso rápido y de identificación inequívoca. En este contexto, abordamos en esta ocasión diversos elementos, todos relacionados con los identificadores digitales aplicables a publicaciones, personas o moléculas.

► DOI

El uso del *Digital Object Identifier* (DOI) ha aumentado rápidamente en los últimos años. Curioseando su origen, encuentro que su página en la Wikipedia en español aparece en marzo de 2007, mientras que en inglés se remonta a enero de 2004. La primera referencia de normalización de su sintaxis data de 2000. Tiene, pues, el DOI una larga historia, aunque tengo la impresión de que solo recientemente se ha convertido en vocablo habitual, se ha empezado a comprender el concepto y a ganar difusión en la comunidad de bioquímicos españoles. Quizá me engañe, pero preventivamente me propongo explicarlo en este artículo.

La razón de ser del DOI es disponer de un puntero, una referencia permanente a una pieza de información en internet –tal como un artículo de revista científica– empleando un formato corto y, sobre todo, que siga funcionando a pesar de los sucesivos rediseños de las sedes web de las revistas. Se podría pensar que la dirección URL de un artículo ya sirve a este propósito, pero a menudo esa ubicación sufre cambios, cada vez que la editorial reorganiza la estructura de su sitio web. A modo de metáfora, podríamos decir que el DOI siempre rastreará

esos cambios y nos llevará a la dirección URL vigente actualmente.

¿Cómo es posible esto? Por la existencia de un servicio centralizado de DOI (*doi resolver*) y el compromiso de las editoriales en comunicarle la dirección original y cualquier modificación que esta sufra ulteriormente. Cuando nosotros buscamos un cierto DOI, la petición pasa por el servidor <http://dx.doi.org>, que lo interpreta y redirige a la URL actual de la revista.

Si bien el auge del DOI se alcanzó con los artículos de revistas periódicas, se ha ido extendiendo su asignación a todo tipo de materiales: libros, datos, publicaciones oficiales (Unión Europea, OCDE), vídeo comercial, e incluso registros en bases de datos (caso de las estructuras de macromoléculas en PDB, como veremos más adelante).

Como utilidad práctica, se puede instalar un complemento en el navegador de internet (*Firefox* o *Chrome*, pero también los hay para *MacOS* y *Adobe Reader*),¹ que permiten escribir o pegar directamente el texto con formato doi:etc en la barra de direcciones, o bien lo reconocen en el texto de un documento y lo convierten en un enlace, con lo que se llega de un golpe al artículo sin pasar por el servidor o escribir la dirección http.

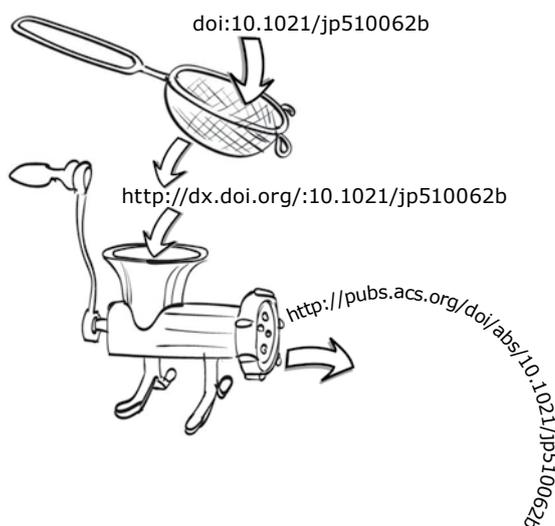
Cabe comentar que la primera parte del código (hasta la barra) identifica la editorial responsable del indexado, pero el formato del resto es una combinación de letras, dígitos y signos que varía ampliamente entre editoriales. Como en algunos casos la longitud del DOI termina siendo considerable, ya existe un servidor que proporciona un equivalente acortado (*shortDOI*[®]).

DOI

El formato recomendado es doi: seguido del identificador, por ejemplo
doi:10.1002/bmb.2006.494034042644

pero es también común verlo en su forma de hiperenlace o dirección a través del servidor,
<http://dx.doi.org/10.1002/bmb.2006.494034042644>





Tú también puedes tener tu DOI

Pero no solo las editoriales e instituciones pueden conseguir identificadores DOI. Han aparecido servicios que ofrecen la posibilidad de publicar en su servidor cualquier tipo de documento, gráfica, paquete de datos, etc. y conseguir un DOI para ello.

permisos relativos a derechos de autor⁴ para saber si podemos hacerlo. En el caso de material de nuestra autoría, necesitamos además considerar en qué términos lo vamos a ofrecer; Zenodo permite la publicación bajo diversas licencias y niveles de acceso, incluyendo tanto Creative Commons como otros. Asimismo, al

cobertura algo más amplia en cuanto a información. Define un «objeto digital» que incluye tanto el documento como alguna información sobre él, su identificador digital y otros metadatos. Para no entrar en tecnicismos, bastará decir que un documento puede quedar registrado en el servidor de Handle con un identificador del tipo hdl:etc, que redirigirá a la url donde esté ubicada su información.

Cabe mencionar que al menos uno de los complementos para el navegador de internet ya mencionados para DOI interpreta también textos con formato hdl:etc en la barra de direcciones (*CNRI Extension for Firefox*).¹

► **ISBN e ISSN**

Estos son identificadores mucho más conocidos, aunque en ocasiones mal interpretados. En ambos casos, se trata de códigos internacionales de identificación; el primero, para libros (publicaciones únicas); el segundo, para revistas (publicaciones periódicas). Y entrará al toro directamente: el ISBN no es un indicador de calidad, aunque nos hayan querido hacer creer eso.

El origen del ISBN (*International Standard Book Number*) –y su propósito– es comercial: permitir a un proveedor (librero) ubicar de forma unívoca un libro en el mercado internacional. No supone un registro del contenido del libro (por ejemplo, de cara a defender derechos de autor) ni una garantía de su calidad o reputación.

¿Qué aportas? El título, la fecha, los autores, el formato y poco más. Distintos formatos (tapa dura, blanda, CD-ROM, libro electrónico) o distintas ediciones requieren distintos ISBN. ¿Qué recibes? Un número y un código de barras equivalente.

DOI propio

Por ejemplo: doi:10.5281/zenodo.12620 corresponde a un programa de software publicado por su autor.



Como ejemplo, el servicio gratuito ofrecido por Zenodo² asigna un DOI a todos los materiales enviados a su servidor y disponibles de forma pública. También ofrece la posibilidad de agruparlos en colecciones personales. Se aloja en el centro de datos del CERN³ en Suiza y se financia con el proyecto europeo *OpenAIREplus* (EU FP7). Por ello, está enfocado preferentemente a la compartición en abierto de resultados de investigación, tanto positivos como negativos, en cualquier tipo de formato. Entre sus sugerencias se encuentran libros, capítulos, comunicaciones en congresos, artículos en revistas, patentes, versiones previas a la impresión, informes, tesis, notas técnicas, datos, figuras, dibujos, diagramas, fotos, *software*, grabaciones sonoras o en vídeo, materiales interactivos y lecciones. Dada su inspiración, Zenodo está además integrado de forma automática en sistemas de auditoría para la investigación financiada por la Comisión Europea (*OpenAIRE*) y, en un futuro, por otras agencias.

Obviamente, a la hora de publicar material debemos ser consecuentes con los

publicar documentos el sistema detectará de forma automatizada algunos elementos de información y los incorporará como metadatos.⁵

Finalmente, para usar Zenodo debemos registrarnos, lo que puede hacerse usando nuestra identidad de diversos servicios en línea, entre ellos ORCID.

► **Handle**

El sistema Handle⁶ supone un concepto y método análogos al de DOI, con una

HANDLE

Ejemplos:

<http://hdl.handle.net/10017/19767>
guarda una comunicación en un congreso

<http://hdl.handle.net/10017/203>
guarda una tesis

En ambos casos, se redirige a sendas páginas de la Biblioteca de la UAH⁷ con toda la información sobre los respectivos materiales, incluidos los documentos en sí.



¿Por qué digo que no indica calidad? Porque cualquiera puede conseguir un ISBN y aplicarlo a su libro, sin que este haya pasado por revisión ni por editores o editorial algunos. Al menos esto era así hasta hace unos años, cuando en España la solicitud podía hacerla un particular y era gratuita; posteriormente la ley ha cambiado y se cobra por la compra de lotes de números ISBN, lo que restringe el alta de los ISBN casi exclusivamente a la industria editorial. También hace años (en los tiempos del BioROM) me informaron en la oficina nacional responsable que solo podía solicitarse para un libro impreso, mientras que ahora hay discos y otros formatos que tienen ISBN. Pero estos cambios no alteran la filosofía, el concepto.

En cuanto al ISSN (*International Standard Serial Number*), se podría aplicar una interpretación similar. Bastará decir que es una forma útil para identificar sin ambigüedad una revista, en este mundo donde proliferan como roedores y es difícil que un nombre nuevo no se parezca a otros. Tened en cuenta que las publicaciones en formato electrónico llevan asociado un número (a menudo indicado como e-ISSN) diferente al de la revista impresa. Ha surgido también el ISSN-L que engloba en un solo identificador todas las versiones de formato de una misma publicación.⁸

► **ORCID**

Pasamos a otro ámbito: la identificación de personas. Concretamente, de autores de las publicaciones científicas. Como sabéis y posiblemente hayáis experimentado, la búsqueda de publicaciones de un cierto autor (o las tuyas propias) en las bases de datos puede sufrir tanto de falsos positivos como de falsos negativos. Las causas son diversas: mismos inicial y apellido para dos personas; uso del nombre completo o solo la inicial; en los autores hispanos y portugueses, uso inconsistente de uno o dos apellidos, o dos apellidos unidos por un guión; etc.

Se hace, por ello, conveniente disponer asimismo de identificadores únicos para los autores. Esto también lleva algún tiempo inventado y ahora su uso ya se está extendiendo y normalizando. Os hablaré del identificador ORCID (*Open Researcher and Contributor ID*); de nuevo, se trata de un sistema normalizado, centralizado, de uso libre y abierto, y apoyado por diversas instituciones y editoriales.

El sistema ORCID⁹ permite evitar ambigüedades, en un uso podríamos decir personal. Pero al mismo tiempo permite la gestión automatizada, proporcionando vías de enlace entre la identidad de una persona y su actividad científica. Finalmente, se integra en los sistemas de publicación (desde el envío del manuscrito hasta su aceptación y publicación), así como en procesos de petición de financiación.

¿Cómo funciona? Primero, eliges darte de alta en el sistema; se requiere tu nombre, dirección electrónica de contacto y una clave de acceso. A continuación, completas la información que juzgues conveniente para conformar tu «perfil»: nombres alternativos, país, ciudad, direcciones de correo, sedes web, formación académica, empleos o puestos de trabajo... Puedes elegir el nivel de privacidad de la información proporcionada. Inmediatamente el sistema te asigna un código identificador ORCID personal; la información que vayas añadiendo estará siempre recopilada en la página web asociada a ese identificador.

ORCID
 Por ejemplo, un identificador ORCID 0000-0002-9900-6845 está asociado a la página <http://orcid.org/0000-0002-9900-6845>



Finalmente, llegamos al meollo de la cuestión: construir un listado de tus publicaciones, proyectos, etc. La buena noticia es que no es necesario escribir toda la información: las publicaciones se pueden obtener de forma semiautomática y sus referencias quedarán guardadas en tu registro ORCID personal. El servidor puede conectar (si concedes el permiso) con diversos servicios externos, como *Europe PubMed Central*, *ScholarOne*, *Scopus*, *CrossRef*, *Elsevier*... y hará en cada uno una búsqueda con tu nombre; basta con que, en la lista de artículos que aparece, marques aquellos que confirmas son tuyos.

Por supuesto, puedes completar la colección añadiendo de forma manual los que no se hayan localizado, pero no serán muchos. Además de esto, ya es frecuente que cuando envías un manuscrito a publicar, la editorial de la revista te solicite tu código ORCID para que cuando el artículo llegue a aceptarse quede ya registrado con tu identidad digital; así, en tus publicaciones futuras ni siquiera tendrás que pasar por el proceso de búsqueda y confirmación descrito antes.

► **PDB id**

Terminamos con algunos toques moleculares; ya sabéis que me encanta este mundillo. Como es sin duda conocido, las estructuras moleculares (coordenadas de cada átomo en el espacio) que se generan en experimentos de resonancia magnética nuclear, cristalografía y difracción de rayos X, de electrones o de neutrones, criomicroscopía electrónica de alta resolución... están, para el caso de macromoléculas, almacenadas en Protein Data Bank.¹⁰

Esta base de datos tiene un elaborado proceso de recepción y validación de los datos, y asigna a cada entrada un código identificador que es permanente y permite identificar esos

datos (y su estructura 3D asociada) de forma unívoca. Un identificador PDB está formado por 4 caracteres alfanuméricos.

PDB
 Ejemplo de identificador PDB: 3rif que habitualmente se presenta escrito como 3rif.pdb o quizás pdbID:3rif o bien doi:10.2210/pdb3rif/pdb (así es, las estructuras de la base de datos también tienen un doi propio)



Trucos
 Si trabajas alguna vez con nuestro amigo Jmol, prueba a abrir la consola de guiones y escribir esto:
 load =3rif (carga desde PDB en Norteamérica) o
 load *3rif (carga desde PDB en Europa)
Voilà! Una pirueta efectista para ejecutarla en clase de forma casi improvisada mientras hablas del enzima o el receptor de turno



► InChI e InChIKey

Y ahora abordamos lo que me gusta llamar las «micromoléculas», es decir, todas aquellas que, por complejas que sean, no son «macro», no son poliméricas. Disponemos de otro identificador para su estructura química, sus datos y mucha más información. Podréis decir, claro, que ya tenemos el nombre normalizado de IUPAC, pero sabéis que no es sencillo de manejar, puede ser abstruso e incluso no ser único; además es casi imposible domar a los sistemas informáticos para que trabajen con él de forma eficaz en búsquedas, rutinas, etc.

Por eso se creó el *IUPAC International Chemical Identifier* (InChITM)^{11,12} que permite el tratamiento automatizado y será un identificador unívoco para cada compuesto químico, fármaco, metabolito, mensajero...

Como, para variar, muchos INCHI no son nada breves, se ha inventado el InChIKey, formado por una serie corta de caracteres alfanuméricos que resulta equivalente al (quizá) más descriptivo INCHI original.

Una estupenda solución para interconvertir nombre común, nombre IUPAC, INCHI, INCHIKEY, n° de registro CAS y otras muchas designaciones es el servidor CACTUS del NCI.¹³ Puedes incluso dibujar una fórmula y obtener en unos segundos su nombre o sus identificadores.

Terminamos aquí este recorrido por un muestrario de identificadores. Quizás este mes me haya desviado de los planteamientos docentes que inspiran esta sección pero, por otra parte, «educación universitaria» no dice a quién hay que educar, pues todos estamos aprendiendo continuamente. Además, nuestros alumnos también deben aprender este tipo de herramientas, y no es preciso ni quizá conveniente esperar a que estén graduados y las descubran por su cuenta. Ojalá os aproveche, pues, a vosotros y a vuestros alumnos. #

Ángel Herráez

BIOQUÍMICA Y BIOLOGÍA MOLECULAR,
DEPARTAMENTO DE BIOLOGÍA
DE SISTEMAS,
UNIVERSIDAD DE ALCALÁ

InChI

Ejemplos:

InChI=1S/C2H6O/c1-2-3/h3H,2H2,1H3 es el etanol

InChI=1S/C47H51NO14/c1-25-31(60-43(56)36(52)35(28-16-10-7-11-17-28)48-41(54)29-18-12-8-13-19-29)23-47(57)40(61-42(55)30-20 es el taxol

Y sus equivalentes INCHIKEY:

InChIKey=LFQSCWFLJHTTHZ-UHFFFAOYSA-N

InChIKey=RCINICONZNJXQF-MZXODVADSA-N



Trucos

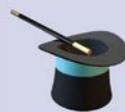
Si trabajas alguna vez con nuestro amigo Jmol, prueba a abrir la consola de guiones y escribir esto:

```
load «$InChI=1S/C3H4O3/c1-2(4)3(5)6/h1H3,(H,5,6)/p-1»
load $BSYNRYMUTXBXSQ-UHFFFAOYSA-N
```

Por cierto, también sirven estos:

```
load $thalidomide
load :tamoxifen
load :50-78-2
```

Que, respectivamente, se obtienen desde el *NCI Resolver* usando un código INCHI, un INCHIKEY o el nombre, y desde *PubChem* por nombre común y número CAS del *Chemical Abstracts Service*



► Bibliografía y notas

- International DOI Foundation (s.f.): *DOI® System Tools*. <https://www.doi.org/tools.html>.
- Zenodo (s/f): *Zenodo-Research-Shared*. <https://zenodo.org/about> (consultado 6 nov. 2015).
- Tus datos se guardarán en el mismo sitio que los *petabytes* que genera el Gran Colisionador de Hadrones (LHC). <http://home.cern/about/computing> (consultado 6 nov. 2015).
- Herráez A.: Mío, suyo... ¿nuestro? *Revista SEBBM* 2015; 183: 34-38.
- Herráez A.: Información entre bambalinas. *Revista SEBBM* 2015; 184: 30-34.
- Corporation for National Research Initiatives (s/f): *The Handle System*. <https://www.handle.net/> (consultado 6 nov. 2015).
- La Biblioteca Digital de la Universidad de Alcalá administra un prefijo hdl propio, pues utiliza para el almacenamiento y gestión de documentos el *software* DSpace, que incorpora el *CNRI Handle System*.
- The National Library of Finland (s.f.): *Linking ISSN (ISSN-L)*. <http://www.nationallibrary.fi/en/publishers/issn/issnl.html> (consultado 6 nov. 2015).
- ORCID, Inc. (s.f.): *ORCID: connecting research and researchers*. <http://orcid.org> (consultado 6 nov. 2015).
- a) Research Collaboratory for Structural Bioinformatics (s/f): *RCSB Protein Data Bank: an information portal to 113494 biological macromolecular structures*. <http://pdb.org/> (consultado 10 nov. 2015).
b) European Bioinformatics Institute, European Molecular Biology Laboratory (s/f): *Protein Data Bank in Europe: Bringing Structure to Biology*. <http://pdbe.org/> (consultado 10 nov. 2015).
- IUPAC (s.f.): *The IUPAC International Chemical Identifier (InChITM)*. <http://www.iupac.org/home/publications/e-resources/inchi.html> (consultado 6 nov. 2015).
- The InChI Trust (s.f.): *Find out about InChI*. <http://www.inchi-trust.org/> (consultado 6 nov. 2015). Incluye las presentaciones en vídeo «What on Earth is InChI?», «The Birth of the InChI», «The Googable InChIKey».
- NCI/CADD Group (s.f.): *Chemical Identifier Resolver*. Computer-Aided Drug Design Group, Chemical Biology Laboratory, National Cancer Institute, National Institutes of Health. <http://cactus.nci.nih.gov/chemical/structure> Documentación en http://cactus.nci.nih.gov/chemical/structure_documentation (consultados 6 nov. 2015).