



## ¡El enmascarado no se rinde! Identificación y persistencia para la producción en Humanidades Digitales

[Mela Bosch](#), CAICYT-CONICET

**Resumen:** Se presenta la diferencia entre identidad e identificación en entidades materiales, se muestra una breve historia de los dispositivos de identificación de seres vivos y cosas. Se sigue con un análisis de los requisitos básicos de los identificadores. Se analiza el problema de la identificación de la producción intelectual y la aplicación de los identificadores numéricos internacionales basados en soportes a la producción en ambiente digital. El tema de la persistencia aparece con preeminencia por lo que se hace un recorrido de su problemática y características. Se concluye con la propuesta de buenas prácticas para la identificación de la producción en Humanidades Digitales.

**Palabras claves:** *autoría en investigación; identificación de la producción intelectual; identificadores persistentes; investigación en humanidades digitales, registros internacionales de producción intelectual*

### Indice

- [El enmascarado: caras y máscara, la identidad y la identificación](#)
- [De las personas a las cosas](#)
- [La identificación de las producciones no materiales](#)
- [La familia tradicional de los números de registro internacional](#)
- [La familia disfuncional de los identificadores uniformes](#)
- [La terapia de resolución](#)
- [¡No rendirse!](#)
- [Bibliografía](#)



## El enmascarado: caras y máscaras, la identidad y la identificación

Para iniciar esta reflexión proponemos un despeje de terminológico: El concepto de **identidad** refiere a los **aspectos esenciales de una entidad**, cuya forma y valor es expresada a través del conjunto de los rasgos particulares que diferencian a esa entidad de todas las demás, por lo que la primera condición necesaria pero no suficiente de una identidad es la **unicidad**. La **identidad tiene un valor ontológico**. En tanto la **identificación** es una **acción** que se aplica sobre una entidad para **representar** la identidad en forma unívoca respecto a esa entidad. La identificación opera por medio de **dispositivos de representación**, los **identificadores**, que varían en su complejidad y en sus soportes tecnológicos.

Así la identificación de las personas, desde la construcción del sujeto como individuo perteneciente a una sociedad, ha avanzado desde los primeros identificadores, los nombres propios, siguiendo por los documentos identificatorios, hasta la actual y casi fantástica identificación biométrica. Hoy es posible identificar unívocamente una persona y/o ser vivo: a través de huellas dactilares, características del ojo: iris y retina, otogramas, geometría de la mano, imágenes vasculares, composición química del olor corporal, ADN.

En el siglo XXI se ha logrado un altísimo nivel de acierto en la identificación aunque su aceptación social sea discutible y discutida por cuestiones éticas de privacidad, por su mayor o menor invasividad y/o por poder aplicarse en forma inadvertida por parte del sujeto identificado.

El nivel de acierto de una identificación se basa en el cumplimiento de un conjunto de **requisitos** básicos que deben reunir los **identificadores** tanto de seres vivos como de cosas:

**Univocidad:** en un mismo sistema no se identifica con los mismos atributos a una misma entidad.

**Universalidad:** todas las entidades del mismo género, especie, variedad y hasta niveles altamente específicos deben ser portadores del atributo o el conjunto de atributos identificatorios que identifica a esas entidades.

**Comparabilidad:** el atributo o el conjunto de atributos identificatorios han de ser mensurables cuantitativamente para poder ser comparables.

**Rendimiento:** el atributo o el conjunto de atributos identificatorios deben ofrecer un nivel de exactitud que asegure su funcionalidad.

**Aceptabilidad:** el atributo o el conjunto de atributos identificatorios deben ser reconocidos y aceptados por el sistema que realiza la identificación.

**Resistencia:** el atributo o el conjunto de atributos identificatorios no deben permitir el fraude o usurpación.

**Persistencia:** el atributo o el conjunto de atributos identificatorios deben poder **preservarse** es decir que permanezcan **inmutables** en el tiempo, espacio y condiciones físicas diversas los aspectos fundamentales que le permitan su reconocimiento y reutilización. (Villanueva, 2004)

## De las personas a las cosas

Como sabemos el dispositivo de identificación más antiguo en la humanidad es el nombre propio, que por supuesto, no cumple con muchos de los requisitos básicos como la univocidad, resistencia, rendimiento, ya que hay muchas personas con el mismo nombre y apellido, por ello estos identificadores se fueron complementando con conjuntos de atributos identificatorios, referencia al lugar de origen, luego en forma cada vez más simbólica, con documentos, en piedra, metal hasta impresos, virtuales y hoy con rasgos biométricos embebidos en ellos.

Luego de la primera guerra mundial los identificadores aplicados a las personas fueron impulsados por parte de los estados nacionales como parte de las políticas de control demográfico, social y productivo. (Deleuze, 1992)

De las personas pasamos a las cosas, las primeras identificaciones se hicieron en forma de inventarios de los reinos e imperios primero y luego de naciones. Para poder dar cuenta de ellos se perfeccionaron las unidades de pesos y medidas y los dispositivos de descripción que favorecieron la consolidación de las estructuras capitalistas en la edad moderna garantizando la circulación de bienes. (Kula, W. 1998)

Esto es así porque un aspecto importante relacionado con la identificación de una cosa es su posibilidad de **apropiación** por parte de otra entidad. La apropiación de cosas, y lamentablemente también de personas en muchos casos aún hoy, está ligada a su identificación.

En este contexto para los seres vivos que no son personas se aplican cada vez más formas de identificación que mezclan la identificación de personas y de cosas: desde las horribles marcas a fuego en personas y animales, en esclavos, en campos de concentración, en criaderos de animales, A los actuales códigos internacionales más o menos invasivos para identificación de bovinos, caprinos, equinos, con el objetivo de lograr la rastreabilidad de los productos derivados con fines comerciales y de seguridad alimentaria. (FAO, 2004)

Volviendo a las **cosas** se partió de considerarlos **productos** es decir objetos de un proceso aplicado sobre sus componentes que da como resultado una única unidad ensamblable y/o comerciable.





Con este enfoque y con el objetivo de asegurar la apropiación por medio de la identificación de las cosas en el siglo XX empieza el auge de los Identificadores universales: el Universal Product Code, UPC, código de 12 dígitos que se aplica a todo ítem de comercio se inició en 1932 y la versión actual se patentó en 1949.

En la segunda parte del siglo XX los identificadores universales de productos avanzaron sobre todos los ámbitos materiales gracias a su registro electrónico. Una versión más sofisticada del UPC fue propuesta por IBM en los años 70 con la tecnología de código de barras, complejizándose desde los años 90 con de matrices de códigos, el QR, que pueden embeber incluso identificadores digitales.

El estándar de identificación de productos está normalizado hoy a través del Electronic Product Code (EPC) que identifica cualquier objeto en el mundo y está desarrollado y mantenido por Global Standards 1, GS1, un consorcio de 120 corporaciones de empresas comerciales globales que mantiene el **Global Trade Item Number (GTIN)** que identifica los tipos de productos.

El EPC se ha ido sofisticando y especializando cada vez más para el comercio, es ahora el **Serialized Global Trade Item Number (SGTIN)** que combina el GTIN un único número serial de 96 bits que tiene embebido un **Radio Frequency Identifier RFID**, para transmitir la identidad de los objetos mediante ondas de radio permitiendo geolocalizar productos.

Entonces la digitalidad es en este momento un componente crucial en los dispositivos de identificación de entidades materiales: el EPC es un URI, sobre el que hablaremos luego, formado por un esquema de metadatos que identifica el objeto hasta su composición incluso molecular, el EPCglobal Tag Data Standard.

Estamos en un nivel de complejidad en la profundidad de la identificación tal que no se usan ya cadenas de números solamente, o una ficha de registro, sino un lenguaje normalizado con una gramática abstracta y una sintaxis concreta que permite una gran expresividad.

Además, desde la **identificación de los objetos** podemos seguir toda la **cadena de eventos** que se le asocian. Abraham Moles fue uno de los primeros en estudiar los eventos con los objetos con un enfoque sistémico son vistos como mediadores de relaciones sociales. Analiza un ciclo en ellos que se inicia con la **creación** seguida del **reconocimiento**, por medio de la identificación, se pasa a la **conservación**, cuando no es posibles conservarlos en forma inmutable se pasa a la **preservación** y finalmente llega la **reparación** o la **transformación**. (Moles, 1972)

En ese sentido con los objetos digitales y los digitalizados, es decir preservados, este ciclo se cumple a velocidades y dimensiones hasta ahora nunca vistas.





En resumen, en el **siglo XXI tenemos las herramientas técnicas para identificar en todo lo que nos rodea, a cada objeto en forma global y dentro de cada objeto material a sus componentes, a través de complejo conjunto de códigos de ítems articulados en un lenguaje de representación.** (Naciones Unidas, 2011)

Y en el caso de los objetos digitales esta **identificación** es el primer paso, el **reconocimiento** que nos permite luego seguir por los **ciclos de conservación, preservación, reparación, restauración y/o transformación.**

### La identificación de las producciones no materiales

Pero no es fácil identificar la producción de la invención humana, desde el siglo XIX se empezó a aplicar a las creaciones no materiales las mismas leyes de propiedad que a los objetos en el sistema capitalista, es lo que se llamó **propiedad intelectual**. La propiedad de los objetos no materiales fue Consagrada por el artículo 27 de la Declaración Universal de los Derechos Humanos, que contempla el derecho a beneficiarse de la protección de los intereses morales y materiales resultantes de la autoría de las producciones científicas, literarias o artísticas.

Esta es la enunciación general del derecho de **propiedad intelectual**, pero según avanzó la revolución industrial el concepto se adaptó a los tiempos: en el Convenio de París de 1883 se reconoció el concepto de **propiedad industrial**. Desde entonces las ideas e inventos se han patentado a través de oficinas nacionales, en realidad las **patentes no funcionan con una lógica de identificadores, son contratos entre quien realiza la patente y un estado nacional**. Este contrato, otorga al inventor el derecho exclusivo de impedir que otros fabriquen, utilicen o vendan el invento patentado durante un periodo de tiempo fijo, a cambio de que éste presente al público los detalles del invento.

Este contrato cubre algunos requisitos de identificador como univocidad, aceptabilidad, resistencia, rendimiento, pero además debe cumplir con requisitos propios específicos como **inventividad** que consiste en que la idea no sea deducible lógicamente del estado del arte en cuestión y los de **aplicabilidad o utilidad** de esa idea que significa que deberá representar una mejora económica para el estado de la técnica, es decir que el productor o el consumidor se beneficiarán de la aplicación industrial de la idea patentable, más allá de las discusiones relativas a los contratos de patentes, tenemos otro problema:

Según Convenio de París de 1883 y cuya última revisión es de 1979, **no se consideran invenciones los descubrimientos, las teorías científicas y los métodos matemáticos; el descubrimiento de materias existentes en la naturaleza, las obras literarias, artísticas, científicas o estéticas; los planes, las reglas y métodos de actividades intelectuales, de juego o programas de computador**



o el soporte lógico; los métodos terapéuticos o quirúrgicos para tratamiento humano o animal; y los métodos de diagnóstico.

Todo lo no patentable, lo que no es producción industrial, se subsume bajo el rótulo genérico de producción intelectual que internacionalmente está regida por el **Convenio de Berna para la Protección de las Obras Literarias y Artísticas de 1886**.

(<http://www.inpi.gov.ar/index.php?Id=107>;

[http://www.wipo.int/treaties/es/ip/paris/summary\\_paris.html](http://www.wipo.int/treaties/es/ip/paris/summary_paris.html))

## La familia tradicional de los Números de Registro Internacional

Vemos que entre producción industrial e intelectual, entre productos y creaciones, es evidente que hay grandes zonas grises, y más aún si consideramos que para lo que se llama producción intelectual la tendencia ha sido valerse de métodos de identificación que se aplican a los objetos materiales **identificando la materialidad sobre la que se manifiesta esa producción intelectual**.

Esto es así porque el primer movimiento fue identificar el **soporte** de la producción intelectual: libros, artículos de revistas, grabaciones.

Con el impulso de Unesco y de ISO a partir de los años 70 surge la **familia de Números de Registro Internacional**:

- ISO 2108: [International Standard Book Number](#) (ISBN)
- ISO 3297: [International Standard Serial Number](#) (ISSN)
- ISO 3901: [International Standard Recording Code](#) (ISRC)
- ISO 10957: [International Standard Music Number](#) (ISMN)
- ISO 15511: [International Standard Identifier for Libraries...](#) (ISIL)
- ISO 15706: [International Standard Audiovisual Number](#) (ISAN)
- ISO 15707: [International Standard Musical Work Code](#) (ISWC)
- ISO 21047: International Standard Text Code (ISTC) <http://www.istc-international.org/>
- ISO 27729: [International Standard Name Identifier](#) (ISNI)

Propongo observar el **ISSN**, está estandarizado por la norma ISO 3297 registra todo tipo de soporte de producción intelectual que tenga la característica de tener una **producción continua y actualizable sin que su fin esté predeterminado y esté puesta a disposición del público**.

**Comprende revistas, periódicos, boletines, actas, directorios, anuarios, memorias, informes de gestión, series monográficas, informes de gestión, todos ellos impresos o digitales, o solamente**



**digitales como sitios web no comerciales ni tampoco producción individual como blogs personales. Comprende también lo que se llaman desarrollos continuos e integrados como bases de datos académicas o institucionales. (ISSN, 2016)**

El registro como tal está constituido por elementos numéricos y cronológicos. Si una publicación se encuentra en varios soportes, cada uno de estos soportes debe tener un número ISSN, así como si cambia de idioma, o si hay variaciones que alteran los rasgos de identificación unívoca.

El número de registro internacional de **producción continua** permite una **identificación única, breve e inequívoca de un soporte de publicación que se proponga ediciones sucesivas, en forma más o menos larga en el tiempo con una periodicidad menor a 3 años.**

En síntesis podemos decir que los números de registro pueden funcionar como un número de documento, solo el número, sin foto, ni domicilio, lo cual no es poco ya que tiene algunas **ventajas**

- Permite ser citado con precisión.
- Permite su expresión en diferentes formatos para aplicaciones informáticas (Código de barras o QR)
- El ISSN es compatible e integrable con otros sistemas de identificación y enlace, tanto de la familia de Números internacionales como ISTC, ISNI, ISAN, como con otros identificadores persistentes como DOI, o con uso de PURL, Handle (sobre los que hablaremos más abajo)

Una última línea de desarrollo de estos números de registro es su articulación entre el número entero, ISBN, ISSN, etc. y el nivel de ítem, la parte más específica: Serial Item Contribution Identifier (SICI), que profundiza el ISSN, el Book Item Contribution Identifier (BICI) que profundiza el ISBN y el Publisher Item Identifier (PII) que se aplica a ambos, todos ellos codificables en el URN que veremos más adelante.

## La familia disfuncional de los identificadores uniformes

La localización en la producción intelectual también estuvo ligada a soporte, la signatura topográfica desarrollada por la bibliotecología que ha logrado a lo largo de su historia un alto nivel de precisión, indica un determinado libro que se encuentra en un determinado lugar. Así la identificación catalográfica se completa con la topográfica que da cuenta de la localización espacial en una biblioteca.





Pero cuando la biblioteca es universal y digital estamos ante problemas que Borges no pudo siquiera imaginar como la volatilidad y la movilidad.

A lo largo de los años se han acumulado iniciativas para estandarizar y normalizar la localización de una producción en el ambiente digital. Aquí nos proponemos repasar solo los rasgos más importantes a considerar para proponer prácticas para nuestra producción en Humanidades digitales.

Los **Números de Registro Internacional** cuando se aplican a **soportes analógicos** cumplen en un nivel **satisfactorio** con algunos de los requisitos básicos de un dispositivo de identificación tales como: **Univocidad, Universalidad, Comparabilidad, Rendimiento, Aceptabilidad y Persistencia y Resistencia.**

Pero en un **medio digital** el requisito de **persistencia** es el que ha resultado más cuestionado.

La persistencia requiere que los atributos identificatorios deban persistir en el tiempo, espacio y condiciones ambientales diversas, en esta condición ambiental podemos incluir la digitalidad: al llegar al **ambiente digital estos números de registro basados en soportes** presentan **inconvenientes, pero no significa no sean aprovechables.**

En lo que se refiere al tiempo los identificadores de la familia de números de registro internacional permiten indicar el **tiempo de inicio** de una publicación y en lo referente al espacio **NO se proponen registrar el espacio donde se encuentra un objeto.** Es decir que el rasgo de localización no está cubierto por estos identificadores

La herramienta básica para la identificación en el mundo digital son los **identificadores uniformes:** El más antiguo es el **Uniform Resource Locator (URL)** es un **Identificador de acceso** que no dice ninguna otra cosa sobre el objeto. Su sintaxis es una arborecencia por derecha que nos lleva hasta el lugar que buscamos: [Protocolo/ Nombre de dominio internacional/ Directorio/ Subdirectorío/archivo.extensión](#)

Nacieron con Internet en 1991 y en su primera enunciación se los consideraba universales. Un URL erróneo es el que nos devuelve el terrorífico: Error 404. El URL está vinculado al sustrato físico, el del acceso: **es una instrucción para un navegador web**, pero puede haber **movilidad** en los componentes de una página, opciones de menú, marcos, que no tengan su propia URL e incluso **conservarse la URL pero no el contenido** que buscábamos.

Con lo cual muchos de los requisitos de identificadores como univocidad o rendimiento o incluso resistencia no se cumplen, y el método que los debilita es el se usa también para ocultar la







identidad de las personas: porque se pueden realizar enmascaramientos para redireccionamientos para lo cual no solo es **importante saber dónde está algo sino qué es ese algo**.

Para ello tenemos otro miembro de la familia de los Uniform Resource: el **Uniform Resource Name (URN)** que fue desarrollado en 1997 por la Internet Engineering Task Force [IETF](#)

La norma , y recurre al viejo truco de usar un nombre y apellido. La preocupación del URN es decir **cómo se llama un objeto** prescindiendo de donde se encuentra. Así puede identificar cualquier tipo de objeto independientemente de su lugar, puede incluso ser un ISSN, un DOI, etc.

Por ejemplo `urn: ISSN:1668-0081`

El apellido es la categoría del objeto digital a la que se aplica el URN (en el ejemplo ISSN) , el **espacio de nombre, NID** y da la clave para la interpretación de la siguiente etiqueta NSS **Namespace Specific String**: que indica exactamente el nombre del recurso, que en este ejemplo es el ISSN que identifica el sitio web de CAICYT, pero NO indica como acceder a él. Quien señala como acceder, es el URL.

Con la **unión** de ambos, **URL** y **URN** tenemos el último de la familia Uniform Resource: todopoderoso **Uniform Resource Identifier (URI)** que es el **estándar para la identificación de recursos en la Web**.

El URI al reunir el acceso (URL) y el nombre (URN) del recurso para identificarlo dentro de un sistema, permite identificar recursos bajo una **gran variedad formatos y protocolos**, llamados genéricamente **esquemas de acceso** tales como HTTP, FTP, hasta Telnet o correos electrónicos localizables. Así el URI identifica el recurso ya sea en la web o en un sistema o en una red local y apunta directamente al recurso por medio de una subdirección, así se puede llegar a un nivel muy específico, a un fragmento incluso de un recurso al que apunta la dirección URL

El último niño de esta familia de identificadores es el **Internationalized Resource Identifier, IRI**, definido en 2005 que extiende el URI que estaba limitado al ASCII y puede contener el código universal de caracteres según [ISO 10646](#) que permite expresar los caracteres chinos, coreanos, cirílicos.

El URL, URN y finalmente URI e IRI son niños de la misma familia nacidos en momentos separados, ahora los encontramos totalmente articulados: la sintaxis genérica de la URI enunciada por la Internet Engineering Task Force define a la **URL como el subconjunto de localización de un URI** (<http://www.faqs.org/rfcs/rfc2396.html>)





Pero se trata de una familia disfuncional: estos identificadores **NO son funcionalmente intercambiables.**

Todos los URN y URL son URI. Pero no todos los URI son URN o URL. El URN tiene un esquema específico que es la parte inicial de un URI, pero no tiene instrucciones sobre cómo acceder al recurso identificado.

El URL, dice donde está, el URN que cosa es, el URI dice ambas. Pero muchas veces un URI puede estar formado solo por el segmento de URL o solo por el segmento de URN, estar incompleto o puede comprender otros objetos digitales no necesariamente descriptos por URL o URN, que identifican pero la idea es que igual sea un URI porque es la identificación que se posee. Además un URI puede no estar formado por un URL por ejemplo un contenido MIME que incluye formas de marcar un correo electrónico.

En la época de los primeros sistemas de inteligencia artificial John McCarthy imagina un sistema al que se le pregunta: ¿tiene hora? Y responde: Sí. Él dice un sistema es realmente inteligente cuando al preguntarle ¿tiene hora? Puede realizar lo que todos hacemos, es mirar el reloj y decir la hora. (McCarthy, 1969)

Las personas somos capaces de saltar la deducción pero aún no las máquinas. Así que esta familia de identificadores NO son accionables directamente, excepto los URL con sus riesgos del error 404 o los URI completamente formados con su URN.

Los URI son como decíamos de hecho el estándar en la web, y son fuertemente defendidos. El mismo Berners Lee en un mítico documento de 1998, convoca a su uso con unas reglas que ya son de oro:

*“What makes a cool URI?”*

*A cool URI is one which does not change.*

*What sorts of URI change?*

***URIs don't change: people change them”***

Berners-Lee, Tim. Cool URIs don't change, 3WC, 1998.

Entonces **los seres humanos tenemos que intervenir para resolver el problema**, la forma es **aplicando buenas prácticas para la construcción y mantenimiento de URI**. Pero los cambios a los URI no accesibles para los usuarios finales, estas **buenas prácticas son para la comunidad de webmasters** que son quienes las deben aplicar. Berners-Lee enunció claramente en aquellos años





las buenas prácticas profilácticas y de mantenimiento para la construcción de URIs, así como en lo referido administración de redes, comprenden: **incluir adecuadamente un URN, definir correctamente los nombres y metadata, no mover direcciones arbitrariamente sin realizar los mantenimientos**, etc.

Pero como sabemos, en una familia disfuncional el esfuerzo individual no alcanza, para hacer algo más funcional esta familia, es decir para disponer de lo que se llama **identificadores accionables** tenemos que recurrir a un apoyo terapéutico sobre sus miembros: los **sistemas de resolución**.

## La terapia de resolución

Para avanzar sobre este tema es importante diferenciar entre el **esquema de nombres** y los **sistemas de resolución** de esos esquemas de nombres.

Los **esquemas de nombres son políticas y procedimientos para denominar un espacio digital** en particular.

El **sistema de resolución registra y almacena los nombres en un servicio accesible a otras computadoras para que respondan a los usuarios finales**.

Estos agentes de externos de direccionamiento son necesarios porque los navegadores no pueden interpretar por sí solo estos nombres entonces necesitan un intermediario, así los URN a su vez tienen que tener encapsulados URL que los redirigiera, esto es lo que se llama resolución.

La resolución puede ser muy básica, simplemente indicando el nuevo lugar para enlazar, o bien por medio de la implementación de **un servidor de enlaces** que realiza los redireccionamientos en forma son transparentes al usuario. El software más conocido para esto es **PURL (Persistent URL)** desarrollado por la On Line Computer Library Center, OCLC.

Los **esquemas de nombres** se deben mantener **independientes** de los servicios de **resolución**, es la única forma de **garantizar la interoperabilidad**.

Cuando tenemos un servidor **PURL** la URL en **lugar de apuntar a la localización real** del recurso **apunta al servicio de resolución** que consulta una base de datos con la **localización actualizada del recurso**, así se enfrenta a la volatilidad de la URL. Pero claro, es fundamental que la organización responsable mantenga actualizada la base de datos. Por ello la **OCLC ofrece este servicio en forma arancelada** a terceros. Otras organizaciones pueden poner sus propios servidores de enlaces con el software de acceso abierto PURL, esto implica el desarrollo de una infraestructura específica.





Otro sistema de resolución es el **Handle** cuya arquitectura es diferente, es un **sistema distribuido de computadoras**, que almacena identificadores llamados handles.

En este caso el servicio de resolución genera **nombres indirectos** que quedan constantes aunque cambien los datos de origen.

Fue desarrollado por la Corporate for National Research Initiative, CNRI, de Estados Unidos. Al ser una arquitectura distribuida posee un modelo jerárquico, el Global Handle Registry, GHR, y en otro nivel los Servidores locales de Handle, LHS.

La GHR está operada por la CNRI y luego se establecen **autoridades nomencladoras**. Cada una establece un dominio para la atribución de nombres llamado **namespace**. En el ámbito de un namespace cada nombre es atribuido a un recurso único. Cada **autoridad nomencladora tiene su único registro en la GHR**. Completan esta arquitectura los servidores de handles que se ven como un navegador con reconocimiento de handle que ya llegan a nivel de la aplicación. A él puede acceder a las autoridades nomencladoras y por tanto a los nombres de manera transparente al usuario.

**Esta arquitectura distribuida es la que se usa actualmente asociada a software de repositorios como Dpace u otros repositorios que asignan un identificador a cada recurso depositado en él.**

Es también usada por un identificador muy difundido: **DOI, Digital Object Identifier**, es una aplicación de la arquitectura de handle pensada para gerenciar copyright permitiendo **controlar las transacciones entre lectores y editores**, y también se extiende al **comercio electrónico**. Se inició en 1997 y está administrado por la **International DOI Foundation** una institución constituida por la **Association of American Publishers**, que nuclea a **editores y empresas de software propietarios para representar los derechos autorales**. Funciona a través de agencias de registración que asignan los números y cobran los registros.

El **DOI puede identificar recursos en acceso abierto, pero debe pagarse su registro**, su ventaja es que es totalmente accesible con enlaces directos. En cuanto a su calidad de persistencia está garantizada por la arquitectura handle, y en cuanto a los otros aspectos de identificación requiere una estructura de metadatos. Para esto se establecen formas para asegurar la compatibilidad semántica entre las agencia registradoras a través de una ontología IDF. Otro beneficio de los DOI es la posibilidad de hacer referencias cruzadas a través de la base de datos de CrossRef <http://www.crossref.org/> que es la base de datos de la mayor agencia de registración de DOI.

Un último recurso terapéutico para esta familia disfuncional de **identificadores uniformes** de objetos digitales es la aplicación de una arquitectura que se **aplica solo a un miembro de la**





**familia, el URL**, enfrentando su debilidad ante situaciones de movimiento, borrado o reemplazo, **dándole la posibilidad de ampliar la información que devuelve, no solo el enlace localizador**. Se trata del **ARK, Archive Resource Key**, que **no utiliza redireccionamientos** monitoreados con PURL o nombres indirectos con Handles como DOI.

La idea John Kunze de la California Digital Library (CDL) que es quien implementó este identificador, es que no es necesario saber el nombre o la naturaleza del recurso (que es lo que agrega el URN en la conjunción URI), para asegurar el buen acceso a su contenido. Solo es necesario un buen servicio de resolución. (Kunze, 2010)

**ARK tiene el siguiente esquema:** [<http://NMAH/>]ark:/NAAN/Name[Qualifier]

Está formado por cinco componentes ordenados según el siguiente esquema:

La primera cláusula igual que todo URL:

http:// que es el protocolo

NMAH que es: Mapping Authority Hostport (normalmente un hostname),

La segunda parte:

La etiqueta ark: seguida de NAAN que significa Name Assigning Authority Number, seguido de Name que es el nombre que se le asignó al recurso. Finalmente en forma optativa se puede agregar un calificador, [Qualifier]

ARK está construido como un URL accionable encapsulando una única identidad global independiente del proveedor del servicio en el momento.

Cada **ARK** está formado por **tres cosas**: el **acceso al objeto**, la **metadata del objeto**, y la **declaración de quien provee el servicio**, de esa forma **un mismo objeto puede estar altamente replicado**, que es una de las características más importantes para asegurar su persistencia y que no la dan los handle ni el DOI.

ARK está constituido por la etiqueta ark: seguida de un URL. Para interrogar sobre él simplemente se pega el identificador en el navegador seguido el signo '?' que devuelve un breve registro de metadatos. Cuando se le agregan '??' devuelve los datos de quien provee el servicio. (Kunze, 2013)

El **ARK puede ser mantenido y resuelto sin necesidad de una arquitectura distribuida** como el Handle, en forma local usando el software de open source Noid (Nice Opaque Identifiers) o a





través de servicios como EZID o el resolvidor central N2T (Name-to-Thing), con lo que ofrece autonomía local e interoperabilidad global

Consideramos que este identificador es muy interesante porque **no está basado en una tecnología sino en una metodología.**

Con el aporte de Diego Ferreyra de Tecnología documental de CACYT hemos experimentado en su utilización. Se puede solicitar un ID para un recurso ante la Name Assigning Authority Number (NAAN) así el ARK depende de una agencia externa, pero también se lo puede implementar sin usar el agencia externa, por ejemplo este ARK que identifica un recurso dentro de COMCIENT, el Repositorio Institucional de CAICYT.

Ejemplo: <http://www.caicyt-conicet.gov.ar/comcient/ark%4a/16680081/>

Puede ir directamente en cualquier navegador, como enlace en citas bibliográficas y tiene una persistencia mayor que el simple URL ya que se puede mover el objeto ya que el navegador irá a buscar el ARK no la última parte del enlace.

**¡No rendirse!**

¿Por qué es importante identificar nuestra producción en Humanidades Digitales? Hemos comentado más arriba que la identificación es un dispositivo de control y como tal puede tener usos de apropiación y sujeción a estructuras de poder. (Deleuze, 1992)

En el caso de la producción de conocimiento tenemos diferentes líneas de acción. Una es la **identificación como política de los estados para visibilizar, estudiar y medir su producción**, es la línea de los Números Internacionales de Registro llevadas a cabo bajo el impulso de las **políticas socialdemócratas en Europa en los años 70.**

Otro accionar es la **identificación para facilitar y asegurar los intercambios comerciales**, es la línea la **Association of American Publishers**, que nuclea a editores y empresas de software propietarios para representar los derechos autorales, cuyo instrumento es el **DOI.**

También tenemos las acciones de las **comunidades de generación y uso de conocimiento**, instituciones gubernamentales, asociaciones, investigadores, editores no corporativizados, lectores. En este caso hay un movimiento social y colaborativo cuyo interés es la identificación con el objetivo, también de visibilizar y tangibilizar, pero también preservar y sobre todo **defender el acceso abierto al conocimiento.** Estamos en un momento importante en que entre las políticas de estado que pueden beneficiarnos o no y a la presión de empresas privadas que dan para





obtener a su vez información de con nuestras conductas individuales. Finalmente están las prácticas de la comunidad social, aquella que podamos consolidar e impulsar para plasmar con dispositivos como lo es la Ley Nacional de Repositorios y con esto logremos hacer la diferencia en procedimientos y políticas.

En ese sentido desde CAICYT estamos llevando adelante un vocabulario controlado sobre Biblioclastia. Hemos extendido el concepto a de biblioclastia, con una perspectiva de acceso abierto al conocimiento.

Y dentro de la Resistencia a la biblioclastia y la defensa del acceso al conocimiento hemos incluido la identificación de recursos.

Es una palabra con poco “gancho”, como indicó un periodista que me entrevistó hace poco, pero también pertenecen al registro culto palabras como femicidio, pedofilia, homofobia, cuyo peso social hoy se debe fundamentalmente a la indispensable presencia de quienes luchan contra ellas.

Para no rendirnos ante tanta complejidad proponemos:

- a. Establecer prioridades para identificar nuestra producción en humanidades digitales.
- b. Redundar, replicar, más vale que sobre y no que falte.

Entonces

1. Es indispensable recurrir a los tradicionalistas números de registros internacionales, siguen siendo sólidos, confiables y reconocidos. No permiten direccionar, pero sí identificar. Prioridad: asignar un ISSN a nuestras producciones: el cual luego podrá ser embebido en URN y en DOI. Puede registrarse una base de datos, un sitio web, etc.
2. Si en el interior de la producción considerada de actualización periódica tenemos otros objetos podemos embeber a otros números de registro según corresponda: International Standard Music Number (ISMN), International Standard Audiovisual Number (ISAN), International Standard Musical Work Code (ISWC)
3. Establecer nuestros derechos, si seguimos el camino de los derechos reservados no hay mayores obstáculos, sólo se debe pagar el DOI.
4. Si seguimos el camino de los derechos compartidos, que es el que más está en consonancia con gran parte de la producción hispana en Humanidades digitales, se puede pagar el DOI, con lo cual se debe cubrir este costo.
5. Realizar producción derivada, artículos de revistas, presentaciones en congresos, donde se pueden consignar estos números de registro o el DOI





6. Depositar nuestra producción derivada en la mayor cantidad de repositorios posibles, ya que los repositorios tienen sus propios handles o ARK, Por ejemplo:  
<http://www.rechercheisidore.org/> y también <http://eprints.rclis.org/>

## Bibliografía

Berners-Lee, Tom. (1998) Cool URIs don't change, W3.

<https://www.w3.org/Provider/Style/URI.html>

Deleuze, Gilles. (1992) Postscript on the societies of control. In: October, vol 59, pp 3-7.

[https://cidadeinseguranca.files.wordpress.com/2012/02/deleuze\\_control.pdf](https://cidadeinseguranca.files.wordpress.com/2012/02/deleuze_control.pdf)

FAO, Food and Agriculture Organization, (2004) Prácticas de identificación animal.

<ftp://ftp.fao.org/docrep/fao/010/y5454s/y5454s03.pdf>

Kula, Witold. (1998) Las medidas y los hombres, México: Siglo XXI.

Naciones Unidas. (2011) Sistema globalmente armonizado de clasificación y etiquetado de productos químicos, Cuarta edición revisada, ST/SG/AC.10/30/Rev.4.

[https://www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs\\_rev04/Spanish/ST-SG-AC10-30-Rev4sp.pdf](https://www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs_rev04/Spanish/ST-SG-AC10-30-Rev4sp.pdf)

ISSN, International Standard Number Organization. (2016) Manual, Paris.

<http://www.issn.org/understanding-the-issn/assignment-rules/issn-manual/>

Kunze, John A. (2010) Towards Electronic Persistence Using ARK Identifiers. California Digital Library. <https://confluence.ucop.edu/download/attachments/16744455/arkcdl.pdf>

Kunze, John A.; Rodgers, R. (2013). The ARK Identifier Scheme, Network Working Group, California Digital Library- University of California San Francisco <http://www.ietf.org/internet-drafts/draft-kunze-ark-18.txt>

McCarthy, J., and Hayes, P. J. (1969) Some philosophical problems from the standpoint of artificial intelligence. In Meltzer, B., and Michie, D., eds., Machine Intelligence 4. Edinburgh: Edinburgh University Press. 463-502. <https://web.archive.org/web/20130825025836/http://www-formal.stanford.edu/jmc/mcchay69.pdf>

Moles, A. (1972) Teoría de los objetos. Barcelona: Gili.

Pfleeger, Charles. (1989) Security in Computing, New Jersey: Englewood Cliffs.







Villanueva Cañadas, E.; Castilla Gonzalo, J. 2004. *Identificación en el sujeto vivo*. En: Medicina Legal y toxicología, Barcelona: Gisbert Calabuig.

## CV de la Autora

Prof. Mela Bosch. Directora por concurso del Centro de Información Científica y Tecnológica (CAICYT-CONICET). Donde además coordina el Laboratorio de Información de CAICYT. Lingüista, estudios en Ingeniería de Software en la Facultad de Informática de la Universidad Nacional de La Plata, UNLP, especializada en Manejo Documental en Empresas en la Biblioteca de la Universidad Jaume I de Barcelona. Ex profesora titular en el Departamento de Bibliotecología, UNLP. Docente de posgrado en varias universidades argentinas. En Milán desde 2002 hasta 2012 fue consultora en documentación en empresas e investigadora del proyecto Integrated Level Classification del capítulo italiano de International Society for Knowledge Organization (ISKO). Sus publicaciones se encuentran E-LIS y COMICIENT-CAICYT otros repositorios digitales.

