

Es necesario un identificador persistente además del DOI (Digital Object Identifier) para el sur global

CARLOS NORBERTO AUTHIER

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) (Argentina)

cauthier@conicet.gov.ar

DIEGO FERREYRA

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) (Argentina)

dferreyra@conicet.gov.ar

RESUMEN

Aborda la necesidad de otros identificadores persistentes para los artículos científicos en el sur global. Es posible afirmar en la actualidad que el DOI (Digital Object Identifier) se ha transformado en un estándar de alcance mundial para la identificación de los artículos científicos. Esta presentación muestra una experiencia desarrollada en Argentina con soporte del Consejo Nacional de Investigaciones Científicas y Técnicas, la cual está basada en un identificador persistente que no tiene costo para los editores científicos. El servicio está basado en el ARK (Archival Resource Key), desarrollado por John Kunze de la California Digital Library (<https://arks.org/>). Al 2022, un total de 67 revistas argentinas cuentan con este identificador con las mismas características y funciones operativas del DOI. La necesidad de desarrollar identificadores persistentes sin costo para los editores es aún más pertinente en el modelo de Acceso Abierto que no cobra APC (Article Processing Charge) a los autores.

PALABRAS CLAVE

Identificadores persistentes, PID, DOI, ARK.

Persistent identifiers, PID, DOI, ARK.

Acerca de ARK-CAICYT

Es abundante la bibliografía sobre los identificadores persistentes y aún más sobre el estándar actual, el Digital Object Identifier (DOI) y su importancia para la identificación de objetos digitales. Como principal fuente de documentación es posible recurrir al DOI handbook, que está disponible en la página de la organización (INTERNATIONAL DOI FOUNDATION, 2013). También hay bibliografía disponible del identificador Handle.Net (HNR) en su página oficial, en este caso un manual técnico (HANDLE, 2018). Además, entre los cuatro identificadores persistentes reconocidos por la International Federation of Library Associations and Institutions (IFLA), debemos mencionar el Persistent Uniform Resource Locator (PURL), que fue desarrollado e implementado por el Online Computer Library Center (OCLC).

Es posible consultar la información sobre los identificadores persistentes en el artículo de ZUMER (2008), donde se describen los principios que los rigen, los requerimientos, la funcionalidad, etc.

En el caso del identificador elegido para esta iniciativa, el Archival Resource Key (ARK), que es uno de los cuatro reconocidos por IFLA, es posible recurrir al desarrollador JOHN KUNZE (2003), quien explica las ventajas de este identificador. Este puede adjudicarse no solo a artículos científicos sino también a imágenes, textos, conjuntos de datos y herramientas de búsqueda. Ha sido implementado por la Biblioteca Digital de California (CDL) y al 2018 por 550 organizaciones y bibliotecas entre las que se cuenta la Bibliothèque Nationale de France.

¿Qué es el ARK?

¿Es posible, siendo el estándar para artículos científicos el identificador DOI implementar en Latinoamérica un identificador alternativo gratuito? De la misma manera que Latinoamérica ha sido pionera en la implementación del Acceso Abierto (en inglés, Open Access u OA), nos propusimos ante la demanda de los editores científicos en la Argentina que tenían dificultades para que sus instituciones abonaran el costo del DOI (U\$S 275 anuales más el costo por artículo), desde el Centro Argentino de Información Científica y Tecnológica (CAICYT), dependiente del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) de la República Argentina, una experiencia piloto para asignar un identificador único y persistente a los artículos científicos.

¿Por qué usar ARK en vez de DOI?

En primer lugar, las revistas científicas de acceso diamante suelen tener problemas para conseguir fondos, por lo tanto la ventaja del ARK es que no tiene el costo del DOI. Entre otras ventajas del ARK, es posible mencionar que se pueden crear identificadores sin metadatos, es posible crear un identificador incluso antes que exista un objeto digital, también es posible mantener el identificador privado mientras los datos y metadatos evolucionan. Luego puede asignarse un identificador adicional, como un DOI, un handle, etc. Además, el protocolo ARK permite crear identificadores más cortos, ya que las mayúsculas y minúsculas permiten cadenas más densas. Y, fundamentalmente, utilizar un identificador persistente con estructura abierta coherente con los valores de la mayoría de las organizaciones científicas latinoamericanas.

Relevamiento de la utilización de identificadores persistentes

Recientemente, se realizó un relevamiento de la utilización de identificadores persistentes basado en Directory of Open Access Journals (DOAJ) al 11 de mayo de 2022. En él es posible obtener los datos públicos de las revistas científicas en formato JSON (Java Script Object Notation), que son volcados semanalmente. El directorio incluye los datos de 130 países en 80 idiomas y al día indicado releva los datos de 17.602 revistas. DOAJ es un directorio en línea que brinda acceso a revistas revisadas por pares de alta calidad y acceso abierto. Todos los datos están disponibles gratuitamente. De los datos obtenidos es posible comprobar que un 63 % cuentan con el DOI como identificador persistente. Un número menor, no superior en ningún caso al 1 % cuenta con los identificadores ARK, Handle y PURL. Y 6.185 revistas, es decir un 35 % no cuentan con ningún identificador para artículos científicos.

La elección de ese directorio está basada en su cobertura mundial y el prestigio, y la posibilidad de descargar los datos completos de las revistas de manera gratuita. Por otra parte, los datos muestran la marginalidad del uso de otros identificadores. El más utilizado en segundo lugar es el handle (<https://www.handle.net/>). Tiene un costo menor que el DOI pero solo 49 revistas científicas lo utilizan, lo que representa un 0,27 % del total de revistas. En tercer lugar, se ubican las revistas que utilizan el ARK, identificador que, como se dijo, no tiene costo para las revistas, pero ellas deben instrumentar su propio resolvidor de identificadores. Solo 16 revistas científicas utilizan este identificador, en porcentaje un escaso 0,09 %. En cuarto lugar, cuatro revistas científicas utilizan el PURL, lo que representa el 0,03 % del total.

Identificadores persistentes

Los identificadores persistentes constituyen una infraestructura primaria para la representación, formalización, circulación y operacionalización del

conocimiento científico, ofreciendo mecanismos para la identificación no ambigua, persistente y funcional de los componentes intervinientes en las prácticas de investigación, desarrollo y comunicación científica. Permiten referenciar de manera formal y unívoca artefactos de todo tipo, ya sean constructos abstractos, entidades físicas, personas, instituciones o componentes de mediaciones comunicacionales.

Establecen las condiciones de posibilidad para que un componente del sistema científico pueda ser identificado, representado y utilizado, facilitando el reuso, citación y socialización de producciones, herramientas y resultados.

Catálogos, nomencladores, convenciones y estándares son ejemplos de las estrategias que a lo largo del tiempo desarrollaron las comunidades en su esfuerzo por designar de manera sistemática y no ambigua el conocimiento construido. El desafío que enfrentamos actualmente nos invita a diseñar y adoptar un esquema de identificadores persistentes en un medio digital y altamente integrado a través de redes informáticas globales. En tal sentido, una estrategia coherente con el actual contexto socio-técnico requiere un modelo que reúna las siguientes condiciones:

- **Referencia no ambigua en contextos digitales en red:** debe poder referenciar una entidad digital de manera no ambigua en el contexto global de Internet.
- **Resolución funcional:** debe poder asociarse con un mecanismo de resolución que garantice la disponibilidad y acceso hacia las entidades digitales referenciadas.
- **Persistencia:** la relación entre la referencia y la entidad digital referenciada debe mantenerse a través del tiempo.

Existen diferentes iniciativas y proyectos en desarrollo en la actualidad orientados a consolidar estándares y mecanismos capaces de cumplir con las condiciones antes mencionadas. Con el fin de establecer un instrumento de relevamiento, hemos establecido una serie de criterios de análisis y evaluación compatibles con una agenda de política científica situada en nuestras realidades y articular con modelos de ciencia abierta.

Los criterios de análisis a considerar son:

- **Sustentable:** capacidad para resultar sostenible desde el punto de vista organizativo y económico.
- **Técnicamente interoperable:** disponibilidad de un diseño tecnológico articular con las infraestructuras de red disponibles, los mecanismos de provisión de servicios automatizados y los estándares digitales de descripción de recursos
- **Tecnológicamente neutro:** condiciones para operar de manera no-dependiente con respecto a tecnologías de base específicas u otras limitaciones (datos, metadatos o servicios propietarios o restrictivos).
- **Políticamente fiable:** capacidad para evitar obstáculos legales (restricciones de acceso o bloqueo de servicios).
- **Portabilidad funcional:** condiciones para su utilización en diversos contextos y medios.
- **Alcance:** universo de entidades que es capaz de referenciar (objetos digitales, físicos, partes de objetos, etc.).
- **Documentación:** es un estándar documentado o un estándar de facto.
- **Soberanía:** modelo de gobierno que lo regula y administra.
- **Robustez:** arquitectura de persistencia utilizada.
- **Interoperabilidad diacrónica:** condiciones de articulación con infraestructuras previas o futuras.

El rol del CAICYT y Centro ISSN

En el CAICYT además funciona el Centro Nacional de ISSN, que otorga los números de identificación unívoca para las publicaciones seriadas argentinas. Por lo que en este desarrollo, elegimos vincular el estándar ARK con el estándar ISSN.

Cómo funciona

1. La revista científica argentina se pone en contacto con el CAICYT y este le asigna una cuenta de usuario con una contraseña con la cual podrá ingresar a través de la siguiente URL:
<http://id.caicyt.gov.ar/issn/>
2. Allí se le solicitarán los datos de cuenta de usuario para ingresar (correo electrónico, contraseña).
3. Una vez validado el usuario, accederá a la pantalla de ISSN asignados a dicha cuenta, donde se expondrá la lista de ISSN y sus nombres de revista.
4. Al hacer clic sobre el nombre de alguna de las revistas de la lista, se accederá a la pantalla de Lista de recursos asociados a dicho ISSN, y Formulario para nuevo recurso.
5. Esta pantalla dispone de dos secciones: **a.** Formulario para agregar un nuevo recurso/artículo asociándolo al ISSN actual. **b.** Lista de Recursos ya asignados a dicho ISSN.
 - 5a. En el campo "URL del recurso" se debe ingresar la URL del artículo de la revista sin ninguna alteración y luego hacer click en el botón "Generar ARK".
 - 5b. Luego de unos instantes la página se actualizará mostrando justo debajo de la URL ingresada el recientemente generado identificador ARK-CAICYT para dicha URL.

Requisitos de postulación

- La revista debe tener carácter científico o técnico.
- La publicación debe tener un ISSN de Argentina.
- Contar con un ISSN para la versión en línea (en caso de no tenerlo puede tramitarlo en el CAICYT).
- Estar en línea al momento de solicitar su inclusión en el proyecto en su URL definitiva.

- Tener dos años de antigüedad en su versión en línea.
- Contar con un sitio web oficial o institucional dedicado.
- Adherir a los principios de Acceso Abierto (http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/publications/policy_guidelines_oa_sp_reduced.pdf)
- No se incluirán revistas digitalizadas en un solo archivo PDF, deben mínimamente tener una tabla de contenidos y los artículos separados en PDF individuales.
- Los responsables editoriales y miembros de los comités deben proporcionar un correo electrónico.
- Sistema de arbitraje externo a la entidad editora.

Los datos requeridos para participar son:

1. ISSN de la revista
2. Nombre de contacto
3. Correo electrónico del contacto
4. CUIL del contacto

Publicaciones/usuarios del servicio de Identificadores persistentes ARK-CAICYT:

- *I+A Investigación + Acción*
- *Páginas de Filosofía*
- *Revista Argentina de Coloproctología*
- *Registros: Revista de Investigación Histórica*
- *Otra Economía*
- *Revista Latinoamericana de Estudios Rurales*
- *Diccionario de Científicos Argentinos "Dra. Cecilia Grierson"*
- *ARTILUGIO*
- *Revista TEFROS*
- *Cuadernos de Economía Crítica*
- *RASAL Lingüística*

- *Toma Uno*
- *AVATARES de la Comunicación y la Cultura*
- *Sudamérica: Revista de Ciencias Sociales*
- *Anuario Centro de Estudios Económicos de la Empresa y el Desarrollo*
- *Costos y Gestión*
- *Cuadernos Universitarios*
- *Cuadernos de Ingeniería*
- *OMNIA: Derecho y sociedad*
- *Teks del Sud: Cuadernos de Arquitectura y Diseño*
- *Crítica y Resistencias*
- *Mundo de Antes*
- *Revista Ucronías*
- *Argumentos: Revista de Crítica Social*
- *Revista de la Sociedad Odontológica de La Plata*
- *Ab Intus*
- *Contextos de Educación*
- *Cronía*
- *Temas y Problemas de Comunicación*
- *Cultura en Red*
- *Revista Argentina de Medicina (En línea)*
- *El Cardo (Paraná. En línea)*
- *Del Prudente Saber y el Máximo Posible de Sabor (En línea)*
- *Revista Latinoamericana de Antropología del Trabajo*
- *Sociedad y Religión*
- *(COORD) Coordinadas. Revista de Historia Local y Regional*
- *Revista del Instituto Internacional de Costos*
- *Educación y Vínculos: Revista de Estudios Interdisciplinarios en Educación*
- *Horticultura Argentina*
- *Chuy. Revista de Estudios Literarios Latinoamericanos*
- *Quintú Quimün. Revista de Lingüística*
- *(En)clave Comahue. Revista Patagónica de Estudios Sociales*
- *Anuario Pilquen. Sección Divulgación Científica*
- *Revista Pilquen. Sección Psicopedagogía*
- *Revista Pilquen. Sección Ciencias Sociales*
- *RAC - Revista Argentina de Comunicación*

- *Revista de la Escuela de Perfeccionamiento en Investigación Operativa*
- *Campo Universitario: Revista de Educación Superior*
- *Trabajo y Sociedad*
- *Revista Hormigón*
- *Revista Latinoamericana de Economía y Sociedad Digital*
- *Revista Enfermería Neonatal*
- *Revista Electrónica de Fuentes y Archivos (REFA)*
- *Revista de Ciencias Empresariales y Sociales*
- *Revista Científica de UCES (RCU)*
- *Desvalimiento Psicosocial*
- *Ratio Iuris. Revista de Derecho Privado*
- *Subjetividad y Procesos Cognitivos*
- *Conflicto Social*
- *Revista Educación en Biología*
- *Vivomatografías*
- *La Zaranda de Ideas*
- *Revista de Arqueología Histórica Argentina y Latinoamericana (RDAHAYL)*

Costo y arquitectura de un resoledor nacional

El costo del desarrollo es, aproximadamente, el que le demanda a un desarrollador senior 90 días de trabajo full time. Estos tiempos varían en el caso de un desarrollador semi senior (150 días) o junior (200 días). Las arquitecturas utilizadas fueron HTML en conjunción con CSS para el desarrollo frontend y JavaScript para el desarrollo del backend. Para minimizar los riesgos de seguridad se desarrollarán algoritmos de encriptación y servidores espejados para mantener las bases de datos.

Además, este desarrollo técnico fue acompañado de capacitaciones y asesoramiento a los usuarios. Actualmente se dispone de un tutorial disponible en línea (http://id.caicyt.gov.ar/docs/ARK_tutorial.pdf).

Para acompañar las capacitaciones se desarrollaron sucesivos materiales disponibles en línea, en el 2018: ARK como identificador persistente gratuito: una propuesta. En el 2022: Identificadores persistentes para la ciencia argentina: ARK-CAICYT.

Historia del ARK-CAICYT y futuros pasos

Durante el 2018 se realizó el diseño técnico y el desarrollo informático. En el 2019 se publicó el prototipo y se desarrolló el pilotaje automático y finalmente la implementación de la plataforma de gestión remota. En el 2022 se implementó de manera pública el backend de gestión de usuarios. Está previsto la implementación de API y el esquema de redundancia para el año 2023.

Conclusión

La Argentina ha demostrado que se puede implementar un identificador persistente sin costo para la ciencia con un proyecto que ya cuenta con más de sesenta revistas científicas, con todos los requisitos técnicos de los estándares internacionales. Es revelador el texto publicado por John Kunze en el blog de la página oficial del ARK en 2021, donde argumenta sobre los diez mitos persistentes sobre los identificadores persistentes (KUNZE, 2021), en el mito 9 argumenta: los PID deben estar centralizados. Falso. Cualquier PID con un núcleo único globalmente después del nombre de host de la URL es persistente. De hecho, si no puede ser atendido por otros hosts, no puede persistir ya que su destino está ligado a un host. Ningún nombre de host o protocolo dura para siempre. El ARK-CAICYT ha sido un gran impulso para la Ciencia Abierta en la Argentina que no podía costear el DOI. Puede replicarse en otros países del sur global con muy poca inversión de parte de los Estados. Estamos a disposición de ellos para compartir la iniciativa.

Bibliografía

- AUTHIER, CARLOS NORBERTO, FERREYRA, DIEGO ANDRÉS, Y BIGLIERI, HERNÁN. (2018). ARK como identificador persistente gratuito: una propuesta. COMCIENT: repositorio institucional del CAICYT especializado en información y comunicación científica. <http://www.caicyt-conicet.gov.ar/comcient/ark%3A/16680081/rscggh>
- BERMÈS, E. (2006). Des identifiants pérennes pour les ressources numériques: l'expérience de la BnF. <https://www.ifla.org/files/assets/pac/ipn/ipnn40.pdf>
- DIRECTORY OF OPEN ACCESS JOURNALS (DOAJ). <https://doaj.org/docs/public-data-dump/>
- DOI FOUNDATION. (2013). DOI (Digital Object Identifier): The Foundation. <http://www.doi.org/>
- FERREYRA, DIEGO ANDRÉS Y AUTHIER, CARLOS NORBERTO. (2022). Identificadores persistentes para la ciencia argentina: ARK-CAICYT. COMCIENT: repositorio institucional del CAICYT especializado en información y comunicación científica. <http://www.caicyt-conicet.gov.ar/comcient/ark%3A/16680081/rscggr>
- HANDLE (2018). Technical Manual Version 9 Preliminary edition. http://www.handle.net/tech_manual/HN_Tech_Manual_9.pdf
- KUNZE, J. (2021). Ten persistent myths about persistent identifiers. <https://arks.org/blog/ten-persistent-myths-about-persistent-identifiers/>
- KUNZE, J. (2003). Towards electronic persistence using ARK identifiers. In: Proceedings of the 3rd ECDL Workshop on Web Archives.
- ZUMER, M. (2008). Guidelines for National Bibliographies in the Electronic Age. <https://www.ifap.ru/pr/2008/n080616b.pdf>