

Manejo de Datos Abiertos en la Agricultura y Nutrición

Este curso de aprendizaje digital (e-learning) es el resultado de una colaboración entre socios de GODAN Action, incluyendo a **Investigaciones Ambientales Wageininen (WUR)**, **AgroKnow**, **AidData**, la **Organización de las Naciones Unidas para la Alimentación y la Agricultura** (FAO por sus siglas en Ingles), **El Foro Global sobre Investigaciones de Agricultura** (GFAR), y el **Instituto de los Estudios del Desarrollo** (IDS), **The Land Portal**, el **Instituto de Datos Abiertos** (IDI) y el **Centro Técnico de Agricultura y cooperación Rural** (CTA).

GODAN Action es un proyecto de tres años [por] el Departamento del Desarrollo Internacional del Reino Unido para capacitar a los que usan, producen, e intermediarios de datos para conectarse efectivamente con datos abiertos y maximizar la potencial por su impacto en los sectores de agricultura y nutrición. En particular, trabajamos para mejorar la capacitación, promover estándares comunes y mejores prácticas para medir el impacto. [www.godan.info]

Este trabajo está registrado con una licencia [CC BY-SA](#)



Unidad 2: Usando los datos abiertos

Lección 2.5: Referenciando a los datos



Foto por Niel Palmer (CIAT) licenciado bajo CC BY 2.0

Objetivos y metas de aprendizaje

Se puede utilizar a datos abiertos cuando

Después de estudiar esta lección debes saber cómo:

- Explicar la importancia de la citación
- Enumerar las características claves que provee la citación y explicar a todos
- Entender el valor de agregar identificadores persistentes en el flujo de trabajo del intercambio de datos
- Identificar las buenas prácticas existentes para el uso de identificadores persistentes

Contenido

Unidad 2: Usando datos abiertos

Lección 2.4: Datos abiertos en los ciclos de políticas

Objetivos y metas de aprendizaje

Lista de gráficas

Lista de tablas

1. Introducción
2. El caso para compartir datos científicos
3. Identificadores persistentes de datos en la web
 - 3.1. Proveedores DOI
4. Discusión

Referencias

Lista de Gráficas

Gráfica 1 El esquema de 5-estrellas para datos abiertos vinculados

Gráfica 2 DOI ejemplar para datos

Gráfica 3 Usando un DOI para conseguir un registro de bibliografía

1. Introducción

El compartir de investigaciones científicas es un proceso con una larga historia que se remonta a los años 1660 con la creación de la Sociedad Real en Inglaterra. La publicación de artículos de revista permitió que los científicos reclamen a sus hallazgos y compartir los resultados con otros. Siglos después, la publicación de artículos científicos no solamente se trata de reclamar a los hallazgos sino también para una medida de desempeño.

El modo clásico de distribuir a resultados científicos es por la publicación en revistas profesionales. Las revistas profesionales están revisadas por sus colegas y estrictamente controlados, que significa que el trabajo necesario para publicar no es insignificante.

La recompensa por tal trabajo es ser incluido en el “índice de citación”. El índice se usa para una evaluación de desempeño para los científicos. Esto quiere decir que el esfuerzo de publicar se recompensa ambos dentro de la comunidad y también como aspecto de la carrera y la promoción personal.

Por lo tanto, las citas sirven para los siguientes propósitos claves:

- Reclamar la pertenencia
- Crear un impacto
- La referencia consistente
- Facilitar del proceso del descubrimiento

Aunque los primeros dos son beneficios para el autor solamente, los últimos dos pueden ayudar con la reutilización. Así como al final de esta lección, las referencias pueden ayudar a conectar a la gente con la evidencia o material contextual. Las referencias por lo menos contienen los nombres de autores, lugar de la publicación y año de la publicación. Estos factores ayudan a la vez indicar el impacto, pero también la facilidad del descubrimiento, permitiendo que los usuarios encuentren la edición relevante de una revista y el número de página del artículo.

Juntos estos cuatro aspectos significan que los científicos tienen el incentivo de publicar investigaciones de alta calidad y son recompensados por hacerlo. La citación provee un mecanismo importante de confianza que demuestra a la audiencia que eres un autor responsable y has citado a trabajos de buena reputación en su campo.

2. El caso para compartir datos científicos

La publicación de artículos de investigación es una parte importante del método académico. Mas importante que el impacto personal, el método académico está diseñado para avanzar la enseñanza, investigación, y práctica de un campo de estudio académico.

El compartir de investigaciones científicas es esencial para avanzar a la enseñanza, investigación y práctica. Permite que científicos en un campo de estudio valide y construya sobre el trabajo de otros.

La validación y expansión de los hallazgos de investigaciones podría requerir el acceso a recursos escasos, como telescopios o agentes químicos, pero recientemente muchos han usado datos como un

input. Por lo tanto, ha habido llamados para el compartir de datos como parte del proceso de publicación.

En principio, los científicos están dispuestos a proveer datos, pero el trabajo extra necesario para procesar, documentar el contexto, y control de calidad muchas veces ni es apreciado o reconocido (Brase et al., 2015). Al contrario, la evidencia sugiere que los que hacen que los datos estén disponible recibieron más citas que otros estudios por lo cual los datos no fueron hechos disponibles, sugiriendo que los datos pueden incrementar el impacto (Piwowar, 2013).

La importancia de compartir los datos de la investigación ha significado que muchos han creado políticas y mandatos que obligan que publiquen sus datos de investigación.

Muchas de las revistas y editores ahora tienen requisitos de compartir datos que se debe cumplir. Puedes leer más sobre el tema aquí:

- La Política de Depósito de Datos¹ de *Ciencia*; busca abajo a la sección de ‘*Data and Materials Availability after Publication* (la disponibilidad de datos y materiales después de publicación)’
- Las Políticas del Compartir de Datos² de *Springer Nature*; como un ejemplo de una revista en particular, vea a la política de *Nature* sobre la Disponibilidad de datos, materiales y métodos.³
- El Servicio del Compartir de Datos⁴ de Wiley
- La Política de Replicación⁵ de Sage

Un problema con muchas de estas políticas es que los datos siempre están juntos a un artículo de la revista. Los datos en si no son objetos de investigación de “primera-clase” que se puede referenciar.

En el año 2004, la Biblioteca Nacional de la Ciencia y Tecnología de Alemania (TIB) asignó el primer Identificador de Objeto Digital (Digital Object Identifier o DOI) para datos científicos para poder hacer que conjuntos de datos científicos sean resultados de investigaciones citables.

El DOI es un identificador consistente que se usa para identificar a objetos únicos, estandarizado por la Organización Internacional de Estandarización (*International Organisation for Standardisation*, ISO). El DOI intenta proveer un identificador que sea relacionado a un objeto digital, como un URL. No como el ISBN o ISRC, que solamente son identificadores, un DOI es accionable e interoperable.

3. Identificadores persistentes de datos en la Web

La idea de “Identificadores accionables” fue concebido por Tim Berners-Lee como parte del esquema de datos abiertos vinculados.

Este esquema define el proceso de hacer que datos estén abiertos y vinculados en la web. Como parte del esquema, el cuarto paso tiene que ver con los identificadores. El principio general del cuarto paso es

¹ <http://www.sciencemag.org/authors/science-editorial-policies>

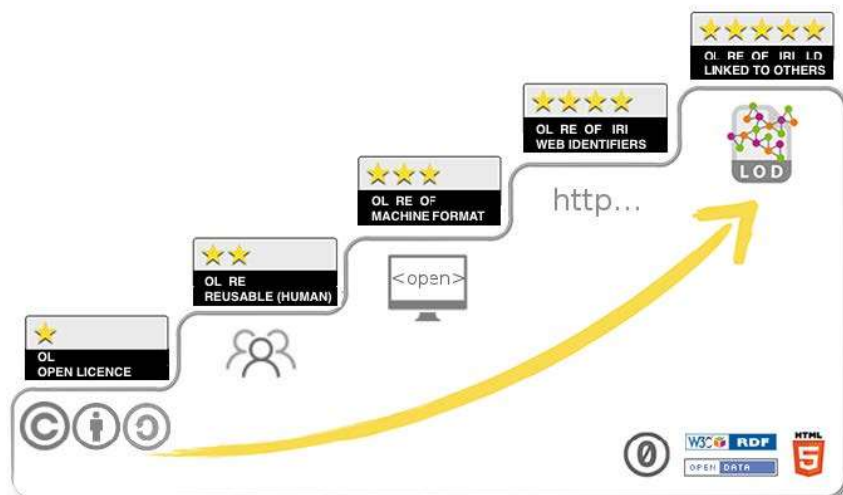
² <http://www.springernature.com/gp/authors/research-data-policy/?countryChanged=true>

³ <http://www.nature.com/authors/policies/availability.html>

⁴ <https://authorservices.wiley.com/author-resources/Journal-Authors/licensing-openaccess/open-access/data-sharing.html>

⁵ <https://us.sagepub.com/en-us/nam/journal/big-data-society#ARTICLETYPES>

para usar identificadores basados en la web (URI) para denotar a cosas. Así la gente puede hallar más información sobre sus cosas y guiarles a través del uso de un URI relacionable.



[Gráfica 1] left to right

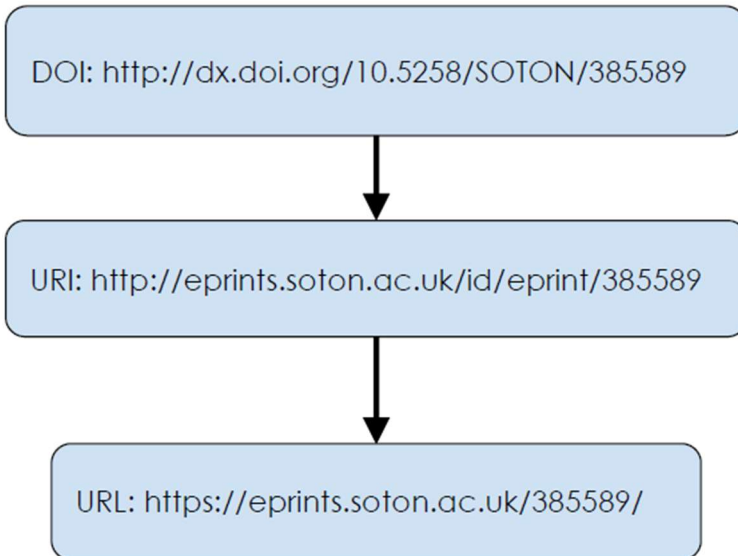
[LA – Licencia Abierta] [OL RE – Reutilizable (por humano)] [OL RE FM – Formato legible por máquina]
[OL RE FM IW – Identificadores por la web] [OL RE FM IW VO – vinculado con otros]

Gráfica 1: El esquema de 5-estrellas para datos abiertos vinculados

Disimilar a los libros, que usan un código ISBN como identificador, un identificador basado en la web no requiere un catálogo para vincular el identificador al objeto. Pero, los identificadores basados en la web pueden ser creados por cualquier persona, en cualquier dominio, y así no tiene el mismo nivel de autoridad.

La idea de un identificador de objeto digital era para proveer un identificador oficial y que podría ser vinculado con otros objetos. Básicamente, un DOI es un identificador persistente basado en la web que apunta al objeto digital en la web, sin importar su ubicación. Aunque cuando cualquier persona puede poner un objeto digital en la web, los DOI solamente están disponibles para objetos que han sido aprobados por ser de alta calidad. Esencialmente el DOI está intentando replicar el mismo impacto que tenga un índice de citas de una revista, pero en la web (Data Citation Synthesis Group 2014).

La gráfica 2 muestra como un DOI referencia a un conjunto de datos en la web. En este caso el conjunto de datos está disponible en un repositorio institucional. Este repositorio provee ambos un identificador único para el conjunto de datos (URI) en adición a una página legible por humanos sobre el conjunto de datos (URL). El DOI es simplemente otro identificador para el conjunto de datos que apunta al URI actual. El propósito del DOI es para proveer un identificador persistente y oficial para cualquier objeto digital sin importar la ubicación del objeto en la web.



[DOI: <http://dx.doi.org/10.5258/SOTON/385589>]

[URI: <http://eprints.soton.ac.uk/id/eprint/385589>]

[URL: <http://eprints.soton.ac.uk/385589/>]

Grafica 2: DOI ejemplar para datos

En adición a proveer un modo de vincular estos identificadores, el servicio web de DOI también puede proveer metadatos directamente al objeto digital por la negociación del contenido, como se muestra en la gráfica 3.

3.1. Proveedores DOI

Hay varias maneras en que un DOI puede estar vinculado a un base de datos:

- Figshare⁶ proveerá un DOI para cualquier obra depositada, que incluye a datos
- Zenodo⁷ también provee DOI para cualquier tipo de resultado de la investigación, incluso a conjuntos de datos
- DRYAD⁸ provee DOI para datos sometidos juntos a artículos (por un costo, que incluye al almacenamiento, la curación, el archivamiento y una revisión para buenas prácticas.

Hay un número de repositorios específicos al dominio que proveen un DOI para datos. La revista Scientific Data⁹ (Datos Científicos) mantiene una buena lista de repositorios que puedes mirar.

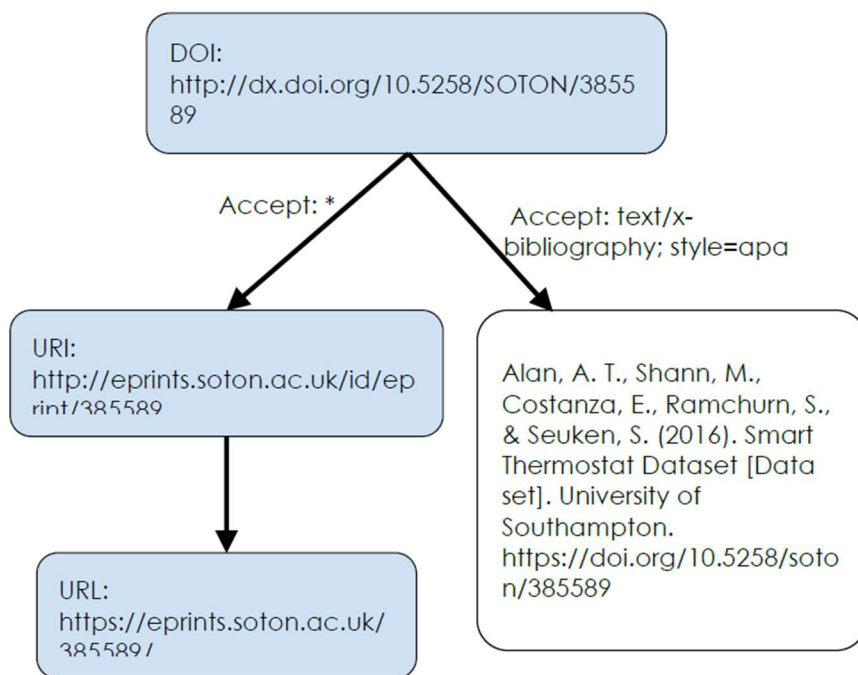
⁶ <https://figshare.com>

⁷ <https://zenodo.org>

⁸ <http://datadryad.org>

⁹ <https://www.nature.com/sdata/policies/repositories>

Escoger un proveedor del DOI dependerá en circunstancias particulares. Ambos *Figshare* y *Zenodo* son servicios “gratis-para-usar” por ejemplo, mientras que *Dryad* cobra por el servicio que ofrecen para cubrir costos actuales de almacenamiento y archivamiento.



[DOI: <http://dx.doi.org/10.5258/SOTON/385589>]

Accepta:*Acepta: texto/bibliografía-x; estilo de citación= apa

[URI:<http://eprints.soton.ac.uk/id/eprint/385589>]

[Alan, A. T., Shann, M., Costanza, E., Ramchurn, S., & Seuken, S. (2016). Smart Thermostat Dataset [Dataset]. University of Southampton. <https://doi.org/10.5258/soton/385589>]

[URL: <https://eprints.soton.ac.uk/385589/>]

Gráfica 3: Usando un DOI para conseguir un registro de bibliografía

Dryad acepta a datos relacionados con publicaciones; si no está asociado con un artículo académico entonces no lo aceptarán. *Figshare* y *Zenodo* aceptarían cualquier resultado de una investigación, sea vinculado con una publicación o no. En ese sentido *Figshare* y *Zenodo* son aplicables generalmente para cualquier resultado de investigación.

Es posible que los datos están más fáciles de hallar en *Dryad* o Repositorios Específicos-a-Dominio (*Domain-Specific Repositories*, DSRs) que en otros como *Figshare* y *Zenodo*. A lo mejor los datos tendrán un formato en otras maneras estándares y hallado más fácil en línea o por otras herramientas si están archivados en *Dryad* u otros DSR. Así es más probable de animar la reutilización.

Figshare es una entidad comercial con fin de lucro, *Zenodo* esta operado por CERN y estaba apoyado por el proyecto *OpenAIRE* en un punto, mientras que *Dryad* es una entidad sin fines de lucro apoyado por subvenciones de investigación y por cuotas de membresía para organizaciones.

4. Discusión

Libros y artículos de revista han beneficiado por mucho tiempo de una infraestructura que hace que sean fáciles de citar, un elemento clave en el proceso de investigación y discurso académico. *Datacite* (el grupo detrás del DOI) cree que objetos digitales deben ser citables en la misma manera.

Los DOI de *Datacite* están diseñados para:

- Apoyar a la atribución y crédito adecuado
- Apoyar a la colaboración y reutilización de datos
- Permitir la reproducibilidad de hallazgos
- Fomentar progreso de la investigación más rápida y eficiente, y
- Proveer el modo de compartir datos con investigadores futuros

Para investigadores la habilidad de publicar y citar objetos digitales y los objetos de primera clase ofrece incentivos a publicar datos abiertos de alta calidad y alto impacto. Plataformas como *Figshare* lo hacen cada vez más fácil publicar tales datos y crear identificadores persistentes de datos.

Referencias

Brase, J., Sens, I. and Lautenschlager, M. 2015. The tenth anniversary of assigning DOI names to scientific data and a five year history of DataCite. *D-Lib magazine* 21 (1/2).

<http://dx.doi.org/10.1045/january2015-brase>

Data Citation Synthesis Group (2014). Joint Declaration of Data Citation; Principles. Martone M. (ed.), FORCE11, San Diego, CA, USA. Available at: <https://www.force11.org/group/joint-declaration-data-citation-principlesfinal>

Piowar, H.A. and Vision, T.J. 2013. Data reuse and the open data citation advantage. *PeerJ* 1 (2013): e175.