



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica
Superior d'Enginyeria
Informàtica

Escola Tècnica Superior d'Enginyeria Informàtica
Universitat Politècnica de València

Métricas basadas en datos

Trabajo Fin de Máster

Máster Universitario en Gestión de la Información

Autor: Isaac Hernández Arizo

Tutor: Antonia Ferrer Sapena

2015-2016

Resumen

El presente trabajo titulado “Métricas basadas en datos” pretende dar una visión de la información que disponemos para evaluar la actividad científica. Para ello se estudian los diferentes tipos de datos abiertos que existen y los indicadores que se han generado a lo largo del tiempo para medir el impacto y la reutilización de la investigación científica. La parte práctica ha consistido en realizar un análisis de varios repositorios de investigación científica y ver qué información nos ofrecen para medir la reutilización de sus datos.

Palabras clave: datos abiertos, datos abiertos de investigación, datos de gobierno abierto, altmetrics, figshare, zenodo.

Abstract

This paper entitled "data-driven metric" is intended to give an overview of the information available to assess scientific activity. To do that, we have studied different tips of open data and indicators exist that have been generated over time to measure the impact and reuse of scientific research. The practical part consisted of an analysis of various repositories of scientific research and see what information we provide to measure the reuse of data.

Keywords: open data, open research data, open government data, altmetrics, figshare, zenodo.

Tabla de contenidos

1.	Introducción	9
1.1.	Justificación.....	9
1.2.	Objetivos.....	12
1.3.	Asignaturas relacionadas.....	12
1.4.	Estructura de la memoria.....	13
2.	Estado de la cuestión.....	14
2.1.	Datos abiertos.....	14
2.1.1.	Beneficios de los datos abiertos.....	14
2.1.2.	Origen de los datos abiertos.....	15
2.1.3.	Valores públicos de los datos abiertos.....	16
2.1.4.	Contradicciones entre los datos abiertos y los valores públicos.....	16
2.2.	Datos abiertos de investigación.....	18
2.2.1.	Principales actores en el desarrollo del intercambio y métricas de datos.....	19
2.3.	Datos de gobierno abierto.....	21
2.3.1.	Beneficios de los datos abiertos.....	23
2.3.2.	Origen de los gobiernos abiertos.....	25
2.3.3.	Buscando un gobierno eficiente.....	26
2.3.4.	Información del Sector Público.....	27
2.3.5.	Diferencias entre ISP y los datos de gobierno abierto.....	28
2.4.	Datos abiertos de las empresas.....	29
2.5.	Legislación sobre reutilización.....	29
2.5.1.	Proceso de reutilización.....	30
2.5.2.	Legislación comunitaria.....	30
2.5.3.	Legislación española.....	31
2.6.	Casos prácticos.....	34

2.6.1. Ámbito nacional.....	34
2.6.2. Ámbito autonómico.....	35
2.6.3. Asociaciones independientes.....	36
2.7. Formatos de archivo.....	37
3. Métricas.....	39
3.1. Los índices bibliométricos en la investigación científica.....	39
3.2. Indicadores bibliométricos.....	40
3.2.1. Ventajas de los indicadores bibliométricos.....	41
3.2.2. Limitaciones de los indicadores bibliométricos.....	41
3.3. Tipos de métricas.....	42
3.3.1. Métricas tradicionales.....	42
3.3.2. Métricas actuales.....	45
3.4. Portales para medir el impacto de los datos de investigación.....	46
3.4.1. ImpactStory.....	46
3.4.2. Métricas de PLoS a nivel de artículo.....	47
3.4.3. PlumX.....	48
3.4.4. Altmetric.....	49
3.4.5. ResearchGate.....	51
3.4.6. Google Scholar.....	51
3.4.7. Microsoft Academic Search.....	52
3.4.8. Otros servicios.....	52
3.5. Requisitos de la citación de datos.....	52
3.6. Creación de la ficha para el análisis.....	55
3.6.1. Citación de datos.....	54
3.6.2. Asignaciones.....	55
3.6.3. Visualizaciones de página.....	55
3.6.4. Descargas.....	56

3.6.5. Enlaces a social media.....	56
3.6.6. Revisión por pares posterior a la publicación.....	57
4. Metodología.....	58
4.1. Discusión.....	58
4.2. Resultados.....	59
4.2.1. Repositorios por área científica.....	59
4.2.2. Repositorios que ofrecen referencia bibliográfica del dataset.....	61
4.2.3. Identificador.....	62
4.2.4. Visualizaciones de página.....	63
4.2.5. Descargas.....	64
4.2.6. Difusión en redes sociales.....	65
4.2.7. Número de veces que se difunde el dataset.....	66
4.2.8. Difusión en gestores bibliográficos.....	67
5. Conclusiones.....	69
6. Bibliografía.....	71
7. Anexos.....	75

1. Introducción

1.1. Justificación

El término Open Access o acceso abierto hace referencia al acceso a fuentes de información científica como monografías, revistas, tesis, manuales de práctica clínica y cualquier otro tipo de documento con formato electrónico. El acceso se efectúa a través de Internet, y se pueden realizar búsquedas, leer, recuperar documentos, obtener copias, imprimirlas y distribuir las mediante un sistema de acceso libre, permanente y gratuito (Sanz Valero y otros, 2007).

El acceso abierto supone un cambio de paradigma, al afirmar que la utilización de los conocimientos de la educación y la ciencia son libres y gratuitos para todo el mundo. Por tanto, no es el usuario/a del conocimiento quien cubre los gastos de publicación y distribución, sino los productores, sus instituciones, los proveedores o los intermediarios (Kuhlen, 2007).

Para que este cambio de modelo sea efectivo, deben ayudar las universidades, al igual que las agencias de financiación y los organismos responsables de las políticas públicas de I+D, que son los actores fundamentales en este ámbito (Abadal y otros, 2013).

Por otro lado, se debe nombrar la Open Access Initiative (OAI) o iniciativa de acceso abierto, que es una iniciativa comprometida con la calidad de los contenidos. Permite la revisión por pares, la impresión, la preservación de contenidos y otras características asociadas a la literatura académica tradicional. En este caso el lector, que en la mayoría de las ocasiones es el productor del conocimiento científico, no está obligado a pagar por acceder a la información. De esta manera se reducen las barreras para la investigación impuestas por los grandes grupos editoriales (Sanz Valero y otros, 2007).

Otra diferencia que encontramos son las condiciones que se imponen a los autores en materia de «propiedad intelectual». Normalmente, en el sistema tradicional, los autores están obligados a ceder los derechos de autor del manuscrito a las editoriales, pero la OAI busca el respeto de la propiedad intelectual y el mantenimiento del copyright por el autor (Sanz Valero y otros, 2007).

La OAI nace a través de iniciativas que persiguen almacenar información de calidad procedente de distintas colecciones y fuentes de conocimiento del Open Society Institute (OSI), fundación creada en 1993 por el investigador y filántropo George Soros, con tal de promover sociedades, que den soporte a proyectos relacionados con la educación, multimedia, salud pública, derechos de la mujer, reformas legales, sociales y económicas (Sanz Valero y otros, 2007).

El verdadero lanzamiento de la OAI se produce a partir de la reunión que mantuvieron en Budapest el 1 y 2 de diciembre de 2001, los representantes científicos de diferentes organizaciones, instituciones, asociaciones internacionales y representantes de diferentes países del mundo. Esta reunión, junto con otras anteriores, marcaba como principal objetivo aunar esfuerzos y unificar iniciativas separadas dentro de un plan estratégico que generará grandes ficheros de fuentes de conocimiento y de los logros recientes alcanzados, colocando las fuentes de información de OSI como la base de la OAI (Sanz Valero y otros, 2007).

De esta manera, surge el 14 de febrero de 2002, la Budapest Open Access Initiative (BOAI) como una directriz de principios, estrategias y compromisos que fue suscrita por 270 organizaciones y 3.400 particulares. En junio de 2004, el OIS dedica una partida presupuestaria importante para dar soporte al proyecto Open Access, desarrollándose herramientas como el Directory of Open Access Journals (DOAJ), software, guías, dotaciones presupuestarias destinadas a la organización de conferencias, seminarios, talleres, y otra partida para publicaciones. El éxito de la OAI, se produce gracias a la ayuda de OIS y BOAI, pero sobre todo, por la enorme acogida que obtuvo por parte de la comunidad científica internacional (Sanz Valero y otros, 2007).

Otro acuerdo importante es la declaración de Berlín, que fue aprobada en 2003. Esta declaración dio lugar al reconocimiento científico y político de los principios de acceso abierto en todo el mundo. La declaración de Berlín está conectada con la Budapest Open Access Initiative de 2001, con la declaración de ECHO (European Cultural Heritage Online) y con la declaración de Bethesda (Bethesda statement on open access publishing), aprobada en 2003 (Kuhlen, 2007).

Siguiendo el concepto internacional de acceso abierto, la declaración de Berlín establece que las contribuciones basadas en el “principio de acceso abierto” deben cumplir dos condiciones (Kuhlen, 2007):

1. Los autores de las publicaciones permiten a todos los usuarios el derecho de acceso gratuito, irrevocable y mundial, y la opción de copiar, usar, distribuir, transmitir y exhibir la publicación para cualquier propósito responsable, con el reconocimiento de autoría y el derecho de efectuar copias impresas para su uso personal.
2. Se depositará en un formato electrónico estándar una versión completa de la publicación y todos sus materiales complementarios, con una copia de la autorización mencionada antes, y se publicará al menos en un depósito online que utilice estándares técnicos aceptables (como los de Open Archive). El depósito deberá estar gestionado y mantenido por una institución o sociedad científica, una institución pública u otra organización comprometida con el “principio de acceso abierto”, la distribución sin restricciones, la interoperabilidad y la capacidad archivística a largo plazo.

También habría que explicar los conceptos de ruta dorada y ruta verde. El primero consiste en publicar nuestro artículo científico en una revista de acceso abierto y el segundo es que nuestro trabajo esté depositado en un repositorio de acceso abierto (Kuhlen, 2007).

En la línea de las iniciativas de lo *Open* nacen las iniciativas de datos abiertos. Cada vez más gobiernos alrededor del mundo están definiendo e implementando estrategias de “datos abiertos” con el fin de aumentar la transparencia, la participación y / o la eficiencia del gobierno. La idea que se extrae de estas estrategias es que la publicación de los datos del gobierno en un formato reutilizable puede fortalecer la participación ciudadana y el rendimiento de nuevas empresas innovadoras. Sin embargo, como estas estrategias sobre datos abiertos son relativamente nuevas, todavía no se conoce con certeza el efecto que provocan (Huijboom y Broek, 2011).

La información pública debe ser accesible, abierta e interoperable. Así que es necesario un esquema de la información y de los datos que permita la colaboración entre las administraciones públicas y de éstas con los ciudadanos. Es decir, gobierno o administración electrónica, transparencia, datos abiertos y participación forman un todo inseparable e interdependiente (Ramos Simón y otros, 2009).

En los últimos años se han generalizado los portales de datos públicos abiertos en los países de la Europa comunitaria. Este tiene su conexión directa con tres hechos que reseñamos brevemente (Ramos Simón y otros, 2009):

- La declaración que realizó Barack Obama en 2009 a favor del Gobierno abierto, donde defendía los principios de transparencia, colaboración y participación.
- La publicación de la W3C (W3C, 2009) en la que se describen los desafíos que plantea Internet para el desarrollo del Gobierno electrónico. En el documento se tratan los dos principales problemas que afectan a la información pública: la existencia de formatos comerciales que dificultan el acceso a los ciudadanos y los límites en la usabilidad de los datos estándares abiertos y legibles por ordenador.
- La publicación en 2011 de la Comunicación al parlamento europeo sobre los datos abiertos y la propuesta de modificación de la Directiva 2003/98/CE relativa a la reutilización de la información del sector público (Comisión Europea, 2011a; 2011b) que concluyen una larga política de apertura de datos y de impulso de un mercado de información europeo. También destacan la directiva 2003/4/CE sobre el acceso público a la información medioambiental y la directiva INSPIRE 2007/2/CE para el establecimiento de una infraestructura europea de datos espaciales.

Es evidente que los datos abiertos traen consigo unos beneficios. Suponen una ayuda para la creación de servicios y productos innovadores. Los datos abiertos, pueden servir para aportar conocimiento en la toma de decisiones, lo que genera riqueza en una empresa en particular y en consecuencia en la economía del país (Pintos y Marín, 2013).

Otro beneficio provocado por los datos abiertos es la reducción de costes, en especial para las empresas infomediarias. Si se consiguen las materias primas sin ningún coste o a un precio muy bajo nos estamos asegurando una mejor posición en el mercado lo que dará a nuestra empresa más posibilidades de subsistir. El sector infomediario está compuesto por las empresas que generan valor realizando servicios, aplicaciones o productos basándose básicamente en la información del sector público. Por último, no disponer de información del sector público puede producir una ventaja competitiva en las empresas extranjeras, perjudicando a las locales. Esto sería especialmente grave para las pymes (Pintos y Marín, 2013).

Según la Comisión Europea el tamaño del mercado de la reutilización de la información tiene un valor que oscila entre 10.000 y 48.000 millones de euros. A nivel oficial, se utiliza normalmente la cantidad de 27.000 millones que es el valor medio de la estimación. Este valor suponía el 0,25% del PIB agregado de la Unión Europea en 2010 (Pintos y Marín, 2013).

1.2. Objetivos

El principal objetivo de este trabajo final de máster es conocer los distintos tipos de métricas que se están utilizando actualmente para medir la reutilización de los datos. Para ello hay que conocer los distintos tipos de datos de los que estamos hablando. Son los siguientes:

- Datos abiertos de investigación
- Datos abiertos de la administración
- Datos abiertos de las empresas

Por otro lado, vamos a analizar diversos repositorios de datos de investigación para ver qué información nos dan sobre la reutilización de los datos que contienen. Por ejemplo, las métricas o indicadores como el número de visualizaciones, descargas o enlaces en redes sociales nos pueden informar del interés de esos datos.

En este momento es difícil conocer qué personas u organismos están utilizando los datos abiertos. Si no lo descubrimos, no podemos conocer el uso que hacen de ellos, ni el impacto que provocan en el resto de la sociedad. Tampoco los beneficios económicos o científicos que los datos abiertos pueden generar.

1.3. Asignaturas relacionadas

EXM (Explotación de datos masivos). En esta asignatura hemos estudiado brevemente la minería de datos que consiste en analizar y procesar datos para generar conocimiento.

FDI (Fuentes de datos e información). Nos sirvió para conocer tanto fuentes de información públicas como privadas. Especialmente importante es para este trabajo el tema 5 en el que se analizaban los productores de datos.

WSO (Gestión de datos: web semántica y open data). Una asignatura dedicada a estudiar que son los datasets, y conocer proyectos de datos abiertos como dato.gob.es, entre otros.

SOI (Sociedad de la información). En esta asignatura buscamos casos prácticos en que los datos ofrecidos por los gobiernos abiertos nos servían para mejorar la eficiencia de las ciudades, (convirtiéndolas en *smart cities* o ciudades inteligentes), y aumentando el nivel de vida de los ciudadanos.

PES (Planificación estratégica SI/TI). Una asignatura muy importante para entender conceptos como las barreras de entrada o de salida que puede haber en un determinado sector económico. También sirvió para ver lo necesario que es la información para tomar decisiones acertadas.

TII (Técnicas de investigación e innovación). Fue una asignatura interesante para saber cómo llevar a cabo una investigación. Vimos diferentes técnicas para recopilar información como el cuestionario, la entrevista personal o el grupo delphi.

BAN (Business analytics). Esta asignatura también se centra, al igual que PES, en el valor de la información para la toma de decisiones.

1.4. Estructura de la memoria

1. Introducción: se describen las razones por las que estudiar las métricas de reutilización de datos, los objetivos del trabajo, que asignaturas realizadas en el máster MUGI están relacionadas con el tema del proyecto y la estructura de la memoria.
2. Estado de la cuestión: se analizan conceptos como datos abiertos o datos abiertos de investigación. Todo ello para entender mejor que tipo de información estamos analizando en este trabajo final de máster.
3. Metodología: en este apartado veremos que es una métrica y analizaremos una serie de repositorios en búsqueda de los datos de uso que nos ofrecen.
4. Resultados: realizaremos el ejercicio explicado en el punto anterior y analizaremos los resultados.
5. Conclusiones: realizaremos unas reflexiones que hayamos extraído del trabajo realizado. También señalaremos las dificultades que hayamos encontrado.
6. Bibliografía
7. Anexos

2. Estado de la cuestión

2.1. Datos abiertos

Según el libro *Open data: Reutilización de la información pública* el concepto de “open data” o “datos abiertos” se basa en que empresas, administraciones o particulares ofrecen libremente los datos producidos en el desarrollo de sus actividades. Tal vez pensamos que esos datos no son de una naturaleza especial pero en cualquier caso debemos publicarlos de alguna manera porque puede que otro consiga elaborar un producto u ofrecer un servicio con el que producir un beneficio. No se pretende que el productor de los datos pierda los derechos sobre los mismos, pero sí se pueden definir unas licencias o permisos para describir el uso que se puede hacer de esa información. Algunas restricciones podrían ser la obligación de no manipular los datos o de citar la fuente de origen de los datos devolviendo de esta manera un beneficio al productor de los datos (Pintos y Marín, 2013).

Según el Open Data Handbook “Los datos abiertos son datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, como mucho, a la obligación de atribución y de compartirse de la misma manera en que aparecen”. Los aspectos más importantes del término “abierto” son (Open Data Handbook, 2016):

- Disponibilidad y acceso: los datos deben estar disponibles en su conjunto y con un coste razonable de reproducción. Preferentemente mediante la descarga en internet. Los datos también deben estar disponibles en una forma adecuada y editable.
- Reutilización y redistribución: los datos deben proporcionarse en condiciones que permitan la reutilización y redistribución, incluyendo la mezcla con otros datasets.
- Participación universal: cada persona podrá utilizar, reutilizar y redistribuir. No puede haber discriminación en los campos de actividad, personas o grupos. Tampoco están permitidas las restricciones de uso para ningún propósito.

Además de que los datos sean abiertos, necesitamos que sean interoperables. La interoperabilidad muestra la capacidad de diversos sistemas y organizaciones para trabajar juntos (inter-operar). En nuestro caso, queremos tener la posibilidad de trabajar con varios datasets y no con único conjunto de datos (Open Data Handbook, 2016).

2.1.1. Beneficios de los datos abiertos

En nuestra vida diaria debemos tomar decisiones y cuanta más información tengamos mejores serán nuestras elecciones. Con los datos abiertos se puede responder a preguntas como las siguientes:

¿Sabe usted exactamente cuánto dinero de sus impuestos se gasta en farolas o en la investigación del cáncer? ¿Cuál es la ruta en bicicleta más corta y más segura desde su casa al

trabajo? Por el camino, ¿cómo es el aire que vamos a respirar? ¿En qué parte de su país va a encontrar las mejores oportunidades de trabajo o el mayor número de árboles frutales por habitante? ¿Cuánto se puede influir en las decisiones de los temas que te interesan profundamente? (Open Data Handbook, 2016)

Las nuevas tecnologías hacen posible la construcción de los servicios que responden a estas preguntas de forma automática. Muchos de los datos que necesitas para responder a estas preguntas son generados por los organismos públicos. Sin embargo, a veces ocurre que los datos requeridos no están disponibles en una forma fácil de usar (Open Data Handbook, 2016).

2.1.2. Origen de los datos abiertos

No se sabe con exactitud cuándo se origina el fenómeno de open data pero si vemos que a lo largo de la historia han ocurrido hechos que siguen esta filosofía. Es el caso de Joseph Henry, Williams Sturgeon y el electroimán. En 1825 Sturgeon enrolló alambre de cobre alrededor de una barra de hierro con forma de herradura. Envío una señal eléctrica por la barra de hierro, lo que hacía que el hierro actuara como imán. Joseph Henry, añadió al invento de Sturgeon una batería eléctrica con la que consiguió construir un electroimán con el que levantar hasta 300 kg. Más tarde fue capaz de levantar una tonelada. Joseph Henry siguió la filosofía de datos abiertos, aunque no conocía el nombre. Descubrimientos posteriores como el relé y una primera versión del motor eléctrico han sido posibles gracias a que Joseph Henry no incluyó ninguna licencia sobre sus inventos (Pintos y Marín, 2013).

El telégrafo que diseñó Samuel Morse, con la ayuda de John Henry, el teléfono de Alexander Graham Bell, y cualquier otro aparato impulsado por un motor eléctrico, se basa en la tecnología compartida por John Henry. Parece muy evidente que la evolución científica y las telecomunicaciones no habrían evolucionado tanto si estos inventos hubieran sido ocultados por sus autores. Tim Berners-Lee también colaboró con la filosofía de open data, publicando la tecnología de la web de forma totalmente abierta (Pintos y Marín, 2013).

La revista BBC History Magazine publicó en agosto de 2010 un artículo sobre la actuación de la enfermera Florence Nightmale en la guerra de Crimea (1854-1856). Utilizó estadísticas del Ministerio de Defensa Británico sobre las bajas británicas y sus causas para realizar un gráfico que demostraba que la cifra de fallecidos en el campo de batalla era la séptima parte de los fallecidos en los hospitales. Con estos datos se mejoró la calidad de los hospitales, lo que sirvió para reducir la tasa de fallecidos en un 75%. Estos casos muestran una reutilización de los datos, pero en este último sin fines empresariales. El concepto de open data es reciente, aunque como podemos ver, esta ideología se seguía siglos atrás (Pintos y Marín, 2013).

En 1957 el Consejo Internacional para la Ciencia (ICSU) creó “un conjunto de centros que permitían clasificar, catalogar, almacenar y distribuir información entre los distintos centros y los investigadores”. La Administración Pública tardó unas décadas más, hasta 1998 para redactar el libro verde sobre la información del sector Público en la Sociedad de la Información (Pintos y Marín, 2013).

Y llegando a la actualidad nos encontramos con el discurso que hizo Barack Obama el 21 de enero de 2009, titulado “Memorandum for the heads of executive departments and agencies”. En él se dan tres directrices para promover mayor apertura y transparencia en el Gobierno de Estados Unidos. Son las siguientes:

- ❖ El gobierno debe ser transparente. La transparencia promueve la rendición de cuentas y proporciona información a los ciudadanos sobre lo que está haciendo su gobierno.
- ❖ El gobierno debe ser participativo. El compromiso público mejora la eficacia del Gobierno y mejora la calidad de sus decisiones.
- ❖ El gobierno debe ser colaborativo. La colaboración involucra activamente a los estadounidenses en el funcionamiento de su gobierno.

2.1.3. Valores públicos de los datos abiertos

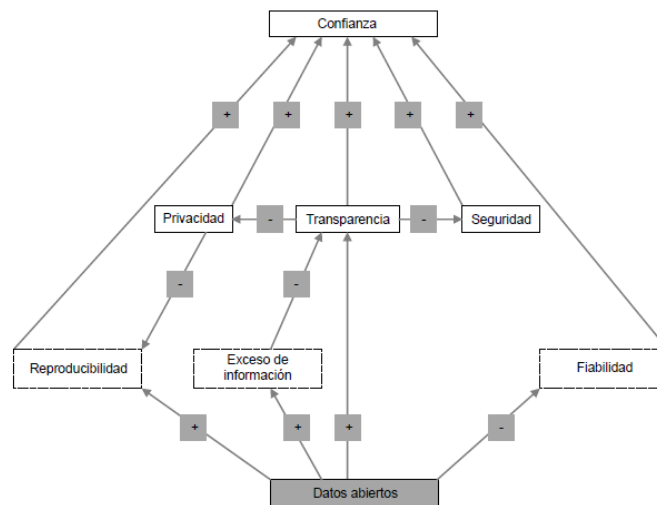
Los datos abiertos ofrecen a la sociedad cuatro valores (Meijer, Conradie y Choenni, 2014):

1. La transparencia es el valor central impulsado por los datos abiertos. La transparencia es vista como una condición necesaria para que un Estado democrático y de derecho funcione. Da legitimidad a las administraciones públicas y confianza a la población civil en su gobierno. La comunidad científica también ha pedido una mayor transparencia con sus propios datos de investigación.
2. El segundo valor es la confianza. Gracias a una mayor transparencia de los datos se llega a un aumento de la confianza.
3. El tercer valor es la privacidad. Los datos abiertos no deben contener información que revele ninguna identidad personal. En la Directiva de protección de datos de la Unión Europea se exige que si se procesan datos personales, debe hacerse de manera justa, legal y con fines determinados, explícitos y legítimos. Los fines para los cuales se procesan los datos deben ser especificados en el momento de la recogida de los datos.
4. El cuarto y último valor es la seguridad. La seguridad es un concepto derivado de la privacidad. Para evitar la divulgación accidental o malintencionada, modificación o destrucción de los registros y datasets, la seguridad de los datos es indispensable.

2.1.4. Contradicciones entre los datos abiertos y los valores públicos

Una vez identificados los valores más importantes de los datos abiertos, se analizan ahora las relaciones entre ellos. En la Figura 1 se muestran los valores de los datos abiertos y la forma en que supuestamente se relacionan entre sí y tres elementos intermedios como la reproducibilidad, el exceso de información y la fiabilidad. Las flechas representan la forma en que los valores públicos y los elementos intermedios se relacionan entre sí. Si los elementos se refuerzan mutuamente, se indicado con el signo +, si son inversamente proporcionales, se indica con el signo -. La intención de este gráfico es mostrar las aparentes contradicciones entre los valores públicos descritos anteriormente (Meijer, Conradie y Choenni, 2014).

Figura 1: Ilustración de las contradicciones entre datos abiertos y valores públicos



Fuente: Reconciling contradictions of open data regarding transparency, privacy, security and trust / Ronald Meijer, Peter Conradie y Sunil Choenni (2014, p5).

Al dar acceso a los datos de investigación y de gobierno abierto, los ciudadanos, políticos, periodistas, auditores y científicos obtienen oportunidades para controlar, verificar los datos, repetir los hallazgos de investigación o elaborar nuevos proyectos. Se supone que esto se traduce en el mantenimiento o aumento de la confianza. Sin embargo, hay un lado oscuro si el acceso a los datos se efectúa sin limitaciones ni restricciones. Este lado oscuro puede dar lugar a varias contradicciones en los valores de la política de datos abiertos. Son las siguientes (Meijer, Conradie y Choenni, 2014):

1. En primer lugar, los datos abiertos pueden entrar en conflicto con la privacidad. La apertura de los datos se ve obstaculizada seriamente cuando se utilizan datos sensibles. En los datos abiertos no pueden aparecer datos personales, aunque pueden convertirse en datos de carácter personal si los combinamos con otros datos disponibles públicamente. Incluso cuando se comparten datos con un alto nivel de agregación, el riesgo de poder deducir información sensible permanece. La apertura de datos sin tener en cuenta los riesgos para la privacidad, pueden dar lugar a violaciones de la intimidad con posibles consecuencias muy negativas para la confianza de los participantes en la investigación. Por otra parte, influye en la confianza que tiene la ciudadanía en la investigación o en el gobierno, y reduce la confianza de otras personas que registran sus datos. Por tanto, la falta de privacidad reduce la confianza.
2. En segundo lugar, los datos abiertos pueden entrar en conflicto con la seguridad. La revelación de la identidad en encuestas o datos administrativos podría ser utilizada por grupos privados o públicos para perjudicar a las personas, o a empresas comerciales. Por tanto, deben tenerse en cuenta las normas y procedimientos para el intercambio de datos en el dominio público.

3. En tercer lugar, los datos abiertos pueden entrar en conflicto con la transparencia debido al exceso de información. La sobrecarga de información se produce cuando la información recibida se convierte en un obstáculo en lugar de ser una herramienta útil. Por lo tanto, se considera que como las organizaciones gubernamentales poseen grandes volúmenes de datos sobre muchos temas, la apertura de estos datos puede provocar una sobrecarga de información.
4. Por último, los datos abiertos pueden tener efectos negativos en la confianza cuando se reutilizan los datos. Esto puede afectar a los datos proporcionados y también a la unión de esos datos con otras fuentes de datos abiertos. Se piensa que a medida que se abren los datos, se reduce el control gubernamental sobre la fiabilidad y la validez. Además, los datos de las bases de datos administrativas pueden ser malinterpretados o utilizados para señalar o acusar a determinados grupos. También se considera la mala interpretación de los datos como un desafío general de los datos abiertos.

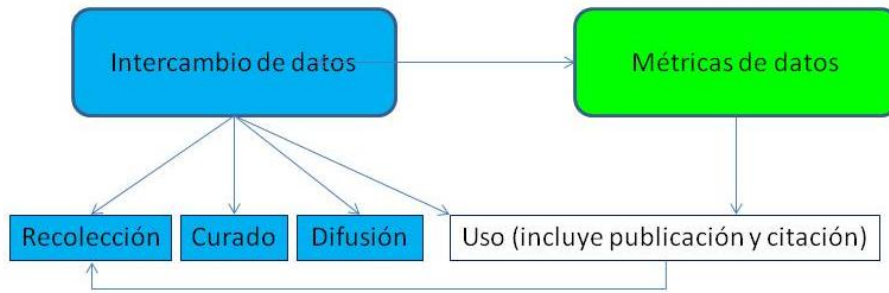
2.2. Datos abiertos de investigación

La investigación científica está relacionada con la producción, el análisis, el almacenamiento, la gestión y la reutilización de datos. El intercambio de datos ofrece beneficios importantes para el progreso científico y el avance del conocimiento (Costas y otros, 2013). Para Remedios Melero y Javier Hernández el acceso abierto a la ciencia significa acceder sin barreras a la literatura científica y poder compartir y reutilizar esos recursos generados en proyectos financiados con fondos públicos. El acceso restringido detiene o desacelera el ciclo de generación de conocimiento (Melero y Hernández San Miguel, 2014).

En el mundo de los datos de investigación existen cuatro conceptos que debemos conocer (Costas y otros, 2013):

1. "Data sharing" o el "intercambio de datos" es definido como la liberación de datos de investigación para ser usados por otras personas (Borgman, 2012). El intercambio de datos requiere la recolección sistemática, curado y difusión de los datos.
2. "Data citations" o "citación de datos". El concepto de citación de datos está vinculado con la idea de que los datasets deben publicarse al igual que otros tipos de productos académicos, siendo considerados también como productos de investigación de primer nivel (Lawrence, Jones y Matthews, 2011).
3. "Data publication" o "publicación de datos". La idea de la publicación de datasets refleja el modelo de publicación científica, a pesar de algunas críticas recibidas (Mayernik, 2012) ya que este modelo no se ajusta plenamente al carácter de intercambio y publicación de las bases de datos.
4. Las "data metrics" o "métricas de datos" están relacionadas con la publicación y la citación de los datos (podríamos incluir aquí las "altmetrics" en los datasets). Tanto la publicación de datos como la citación de datos pueden ser considerados como señales de uso de los datos. El uso de los datos puede generar nuevos datos, que pueden volver a la fase de recolección (ver Figura 2).

Figura 2: Vista esquemática del intercambio de datos –modelo de métricas de datos

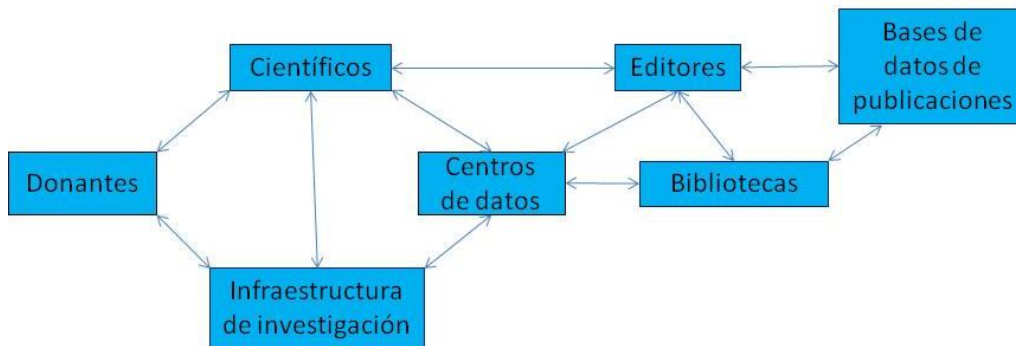


Fuente: *The value of research data / Costas y otros (2013, p7)*

2.2.1. Principales actores en el desarrollo del intercambio y métricas de datos

Contamos con muchos interesados con un papel importante en el intercambio de datos (Mayernik, 2012, p 13, 14) y, por tanto, en el desarrollo de métricas de datos. En la Figura 3 se presenta una visión esquemática de los principales grupos de interés y sus relaciones (Costas y otros, 2013).

Figura 3: Vista esquemática de los principales grupos de interés



Fuente: *The value of research data / Costas y otros (2013, p8)*

En la Tabla 1 se describen las principales partes interesadas, y su principal interés relacionado con el intercambio y las métricas de datos.

Tabla 1. Partes interesadas y su interés en el intercambio y las métricas de datos.

Principales interesados	
Interesado	Interés principal
Donantes	<ul style="list-style-type: none"> - Para financiar la recopilación de datos y sacar el máximo provecho de los fondos públicos dirigidos a la investigación - Para fomentar el uso de los datos - Para promover el reconocimiento y las recompensas por el intercambio y uso de los datos
Infraestructuras de investigación	<ul style="list-style-type: none"> - Desde la perspectiva de la financiación: Para financiar a largo plazo la conservación y el acceso - Desde la perspectiva de los proveedores: Para facilitar el uso de datos
Científicos	<ul style="list-style-type: none"> - El prestigio académico digital con el valor añadido del reconocimiento y la progresión de la carrera profesional - Para citar datos en publicaciones, y hacer que estén disponibles
Centros de datos	<ul style="list-style-type: none"> - Para estandarizar el almacenamiento y crear metadatos de forma global - Para realizar un seguimiento de la reutilización y promover las buenas prácticas científicas - Para efectuar el curado de los datos y recomendar las citas de los datasets
Editores	<ul style="list-style-type: none"> - Para tratar los datos que se proporcionan en las publicaciones presentadas. - Para reforzar y facilitar la citación de datos y las políticas y las normas de métricas de datos
Bibliotecas	<ul style="list-style-type: none"> - Para que los datos estén identificados y accesibles - Coordinar a los académicos con los centros de datos
Bases de datos de publicaciones	<ul style="list-style-type: none"> - Para enlazar publicaciones y citas de datos - Para permitir la publicación de datos y los recuentos de citas e indicadores

Fuente: The value of research data / Costas y otros (2013, p8)

En la siguiente figura aparecen algunos de los beneficios que se obtienen de compartir, reutilizar o transformar datos, basados en el manifiesto *Denton Declaration: An Open Access Data Manifesto* del año 2012 (Melero y Hernández, 2014):

Figura 4. Esquema basado en los principios de la Declaración de Denton



Fuente: Acceso abierto a los datos de investigación, una vía hacia la colaboración científica / Melero y Hernández San Miguel (2014, p6)

2.3. Datos de gobierno abierto

Según el *Open Data Goldbook for Data Managers and Data Holders*, los datos de gobierno abierto se refieren a la información recogida, producida o pagada por los organismos públicos y disponibles de forma totalmente libre para su reutilización para cualquier propósito. Los datos de gobierno abierto se publican bajo una licencia abierta y son libres de utilizarse en dominios privados y públicos (European Data Portal, 2016).

El gobierno abierto (open government, en inglés) es un concepto cada vez más extendido que se basa en tres pilares: transparencia, participación y colaboración. Se pretende alcanzar un modelo de democracia más participativo, algo más que el gobierno electrónico porque une las tecnologías de la información y la comunicación (TIC) con la máxima transparencia de la información gubernamental, además de la apertura a la participación y colaboración de los ciudadanos en todos los procesos de decisión y control. Su punto de partida es crear un mercado libre de datos generados por las instituciones públicas, que puede ser consultado y utilizado por cualquier ciudadano (Clabo y Ramos Vielba, 2014).

Para llegar a disponer de un gobierno abierto, hay dos requisitos imprescindibles. El primero es el acceso libre, abierto y gratuito a los datos y a la información relacionada (open data). En segundo lugar, facilitar vías a la ciudadanía para que participen (open action) (Coroan y Campos, 2011).

La clave de este tipo de gobierno es el acceso a los datos producidos por la Administración en el desarrollo de sus actividades. Dicho acceso se produce básicamente a través de internet, pero no se trata sólo de consultar sitios o páginas web, sino algo más. Por ejemplo, obtener los datos de la Administración, que anteriormente los ha estructurado y codificado de modo que sean reutilizables por los usuarios que accedan a ellos (Clabo y Ramos Vielba, 2014).

Por tanto, los datos de gobierno abierto se establecen como información pública que contiene una serie de características que facilitan su acceso y su reutilización. En concreto, han de cumplir los siguientes criterios (Sunlight Foundation, 2010):

a) Los datasets ofrecidos por el gobierno deben ser tan **completos** como sea posible. Es decir, que recojan toda la información disponible sobre un tema en particular. Toda la información en bruto de un dataset debe ser liberada al público, excepto la información que permita la identificación de las personas. Los metadatos que definen y explican los datos en bruto también se deben incluir, junto con las fórmulas y explicaciones de cómo se han obtenido los datos derivados. Si se hace, se permitirá a los usuarios entender el alcance de la información disponible y examinar cada elemento de los datos con el mayor nivel de detalle posible.

b) Los datasets publicados por el gobierno deben ser **datos de fuentes primarias**. Esto incluye la información original recogida por el gobierno, detalles sobre cómo se recaudan los datos y los documentos de fuentes primarias que expliquen la recolección de los datos. La difusión pública permitirá a los usuarios verificar que la información se recogió correctamente y se trató con precisión.

c) Los datasets publicados por el gobierno deben estar a disposición del público **en el momento oportuno**. Siempre que sea posible, la información recogida por el gobierno debe estar disponible tan pronto como se recolecta y se recoge. Se debe dar prioridad a los datos cuya utilidad es sensible al tiempo. Las actualizaciones de la información en tiempo real maximizan la utilidad que el público puede obtener de esta información.

d) Los datasets publicados por el gobierno deben ser tan **accesibles** como sea posible, definiendo accesibilidad como la facilidad con la que se puede obtener información, ya sea a través de medios físicos o electrónicos. Las barreras para el acceso físico incluyen requisitos para visitar una oficina en particular en persona o requisitos a cumplir con los procedimientos particulares. Las barreras para el acceso electrónico consisten en publicar datos únicamente a través de formas o sistemas que requieren tecnologías en el navegador (por ejemplo, Flash, Javascript, cookies o applets de Java). Por el contrario, proporcionar una interfaz para que los usuarios descarguen toda la información almacenada en una base de datos y los medios para hacer llamadas específicas de datos a través de una interfaz de programación de aplicaciones (API) facilita el acceso a los datos.

e) Las máquinas pueden **manejar** ciertos documentos mejor que otros. Por ejemplo, para las máquinas es muy difícil procesar las notas escritas a mano en un papel. El escaneo de texto a través del reconocimiento óptico de caracteres (OCR) da lugar a muchos errores. Por ejemplo, para las máquinas es muy difícil analizar un documento en formato PDF. Por lo tanto, la información debe ser almacenada en formatos de archivo que sean fáciles de analizar por las máquinas.

f) **No puede haber discriminación** hacia ninguna persona para acceder a los datos ni del uso que se le debe dar a los datos. Las barreras para la utilización de los datos pueden incluir requisitos de registro o de adhesión. Otra barrera posible es que sólo se permita el acceso a los datos a algunas aplicaciones. En su sentido más amplio, la no discriminación en el acceso a los

datos significa que cualquier persona puede acceder a los datos en cualquier momento y sin necesidad de identificarse o proporcionar ningún motivo para hacerlo.

g) Cuando una única empresa produce el programa que puede leer el archivo donde se almacenan los datos, el acceso a esa información depende del uso del programa de la empresa. A veces ese programa no está disponible al público, o está disponible, pero pagando una tarifa. Por ejemplo, Microsoft Excel es un programa de hojas de cálculo bastante utilizado pero de pago. En cambio, existen formatos alternativos en que se puede acceder a los datos almacenados sin necesidad de una licencia de software. La eliminación de este coste hace que los datos estén disponibles para un grupo más amplio de usuarios.

h) La imposición de “Términos de Servicio”, los requisitos de atribución o las restricciones a la difusión actúan como barreras para el uso público de los datos. La máxima apertura implica claramente que la información pública es una obra del gobierno y disponible sin restricciones de uso como parte del dominio público.

i) La capacidad de encontrar cierta información durante mucho tiempo se conoce como **permanencia**. La información publicada por el gobierno en línea debe estar disponible en línea permanentemente. Muchas veces, la información se actualiza, modifica o elimina sin ninguna pista del cambio que se ha hecho. Para un mejor uso del público, la información disponible a través de internet deberá permanecer en internet.

j) Uno de los mayores obstáculos para el acceso a la información es el **coste** impuesto al público - incluso cuando el coste es mínimo. Los gobiernos utilizan una serie de conceptos para calcular el precio por el acceso a sus propios documentos: costes de creación de la información; coste para recuperar información; coste por página o por coste de la investigación; coste de procesamiento; coste de la duplicación, etc. Evidentemente, la imposición de tarifas de acceso reduce el número de personas interesadas en acceder a la información.

2.3.1. Beneficios de los datos abiertos

Los datos de gobierno abierto son un potencial sin explotar. Al igual que cualquier iniciativa del dominio público, también implica gastos y el esfuerzo de los recursos internos. Entender mejor los beneficios de los datos abiertos puede ayudar a acelerar el compromiso alrededor de la iniciativa Open Data (European Data Portal, 2016).

Hacer que la información generada y recogida por entidades del sector público esté disponible y sea reutilizable es importante por muchas razones (European Data Portal, 2016):

- Ofrece a los ciudadanos una base de conocimientos fiables con respecto a las actividades de los organismos del sector público y del gobierno.
- Les permite participar en actividades de los organismos del sector público y por lo tanto participar activamente en las decisiones públicas.

- Representa el material inicial para que los agentes públicos o privados desarrollen nuevos servicios de valor añadido y los suministren a los ciudadanos.
- Es una de las tareas cruciales de la Agenda Digital para Europa “entregar beneficios económicos y sociales sostenibles de un mercado único digital basado en aplicaciones rápidas y ultrarrápidas de Internet y aplicaciones interoperables” (Kolodziejcki, 2013).

En la práctica, los beneficios de Open Government Data pueden variar según el tipo de actor involucrado. Estos actores se pueden dividir en 3 grupos principales: organizaciones gubernamentales, ciudadanos y re-usuarios. Encontramos diversos beneficios para cada uno de estos grupos de interés (European Data Portal, 2016).

En primer lugar, los propios gobiernos son uno de los principales re-usuarios de los datos que recogen ellos mismos. La práctica ha demostrado que los datos que publican los gobiernos, ellos mismos vuelven a usarlos, lo que se traduce en un ahorro de costes (European Data Portal, 2016).

Publicar datos abiertos permite el intercambio de información dentro de los gobiernos en formatos interoperables legibles por máquina, lo que se traduce en la reducción de costes en el intercambio de información y la integración de datos, la reducción de errores por tener una copia en lugar de varias, etc. Esto provoca una mejora en la gestión de datos, en términos de calidad y eficiencia, y en una reducción general de los costes administrativos (European Data Portal, 2016).

El análisis económico realizado por el portal de datos europeo estima que el ahorro de costes acumulados en los países de la Unión Europea será igual a 1.700 millones de euros en el año 2020. Hay más beneficios a tener en cuenta (European Data Portal, 2016):

- **La apertura de los datos puede optimizar su proceso interno.** Cuando los datos estén abiertos, ninguno de los miembros de la organización tendrá que pasar por un proceso interno para recibir unos datos en concreto. Muchas organizaciones han encontrado el beneficio de tener sus datos abiertos, simplemente porque se necesita menos tiempo para encontrar sus propios datos. Como hemos visto antes, es muy probable que la misma organización sea la principal re-utilizadora de sus datos.
- **No sólo la misma organización, sino también los ciudadanos se beneficiarán de una mejor – y tal vez más rápida - estructura de información interna.** Los procesos llevarán menos tiempo, los servicios podrán ser digitalizados, y los ciudadanos se beneficiarán de una mayor eficiencia y transparencia.
- **Si la infraestructura de datos de la organización se queda obsoleta, la iniciativa de datos abiertos podría ser una buena oportunidad para lograr un cambio interno.** Muchas organizaciones han aprovechado la oportunidad para rediseñar su infraestructura interna de datos y han incorporado la publicación de datos como una actividad principal en las instrucciones de trabajo.

- **El feedback de los usuarios, puede mejorar la calidad de los datasets.** El poder de la multitud, conocido como crowdsourcing, es una forma muy eficiente de poner en común los recursos para alcanzar un resultado, a veces sorprendente.

En segundo lugar, al publicar la ISP (Información del Sector Público), las acciones del gobierno se hacen más visibles. Este tipo de transparencia ayuda a gobiernos y a ciudadanos. Permite a los ciudadanos verificar las acciones del gobierno. A su vez, el conocimiento de los ciudadanos sobre el gobierno aumentará y se sentirán más fortalecidos gracias a un mayor acceso a la información. Este crecimiento puede estimular la democracia y la participación en el gobierno local (European Data Portal, 2016).

2.3.2. Origen de los gobiernos abiertos

Para conseguir que los estados alcancen una democracia real y que la ciudadanía confíe en sus dirigentes, la gestión de las administraciones públicas debe ser transparente para sus ciudadanos. Este concepto no es nuevo. De hecho, las primeras medidas para la transparencia y rendición de cuentas fueron introducidas en la Edad Media, cuando el Rey Juan de Inglaterra firmó en 1215 la Carta Magna, donde aparecía por primera vez el principio de consulta y consenso a un comité de barones para la inclusión de nuevos impuestos (Álvarez Espinar, 2014).

A parte de rendir cuentas ante la ciudadanía, aquellos gobiernos que buscan la participación pública en la toma de decisiones siguen políticas de gobierno abierto, u *open government* en inglés. Este concepto se utiliza habitualmente en las agendas parlamentarias de los gobiernos de todo el mundo, y aparece citado en el año 1957 en la publicación *The George Washington Law Review*, concretamente en un artículo titulado *Open Government Principle: Applying the Right to Know Under the Constitution* (Álvarez Espinar, 2014).

En los últimos años son muchos los programas políticos que incluyen el concepto de gobierno abierto como base para fomentar la confianza de la sociedad. Esta apertura del gobierno se basa en la transparencia de las políticas públicas y la promoción de una democracia activa y participativa (Álvarez Espinar, 2014).

Un avance muy importante en la apertura gubernamental ha sido la constitución en 2011 del Open Government Partnership (OGP), donde los máximos representantes políticos de 8 países firmaron un acuerdo para construir gobiernos más abiertos, más democráticos y transparentes. En la actualidad, más de 60 países se han suscrito los principios del OGP (Álvarez Espinar, 2014).

A nivel concreto, los países que suscribieron el OGP tratan de alcanzar por los menos dos de los grandes retos propuestos, entre los que se encuentran: la mejora de los servicios públicos, el incremento de la integridad pública, una mayor efectividad en la gestión de los recursos públicos, la creación de comunidades más seguras y el incremento de la responsabilidad de las empresas (Álvarez Espinar, 2014).

La Comisión Europea, siguiendo los valores del OGP, pretende fomentar el concepto de gobierno abierto en las agendas digitales que definen el futuro tecnológico de los estados miembros. Además, con la publicación del documento *A vision for public services* (Una visión de los servicios públicos), la Unidad de Servicios Públicos de la Comisión Europea, ofrece directrices para la implantación de los servicios prestados por las administraciones públicas, utilizando las TIC como mecanismo para ofrecer y compartir datos y documentos abiertos para su acceso y uso (datos abiertos), fijar servicios a disposición de la ciudadanía y empresas (servicios abiertos) y construir vías de diálogo con la ciudadanía para la mejora de la democracia (decisiones abiertas) (Álvarez Espinar, 2014).

Figura 5 Descripción visual del concepto de gobierno abierto



Fuente: Apertura y reutilización de datos públicos / Álvarez Espinar (2014, p10)

El Observatorio Nacional de las Telecomunicaciones y la Sociedad de la Información elaboró un informe denominado “Estudio de la demanda y uso de gobierno abierto en España”, en el que se resalta que en la sociedad española ya empieza a aparecer una demanda importante de transparencia y de participación ciudadana en los asuntos públicos. No define acciones concretas, pero indica que se irán definiendo a medida que se vayan creando nuevos servicios e iniciativas de gobierno abierto (Álvarez Espinar, 2014).

2.3.3. Buscando un gobierno eficiente

La Comisión Europea, mediante el Plan de acción sobre Administración electrónica 2011-2015 subraya la necesidad de abrir los gobiernos, incluyendo a la ciudadanía y a las empresas en las decisiones de los gobiernos, pero también destaca la oportunidad de aprovechar esta revolución de la información que estamos viviendo para mejorar en la efectividad y eficiencia de la sociedad, teniendo en cuenta que la información posee gran valor y la Administración dispone de información con un inmenso potencial social y económico (Álvarez Espinar, 2014).

La información de las iniciativas de gobierno abierto se libera principalmente porque se considera el acceso a la información como un derecho fundamental de la ciudadanía y aquella información generada y custodiada por las instituciones públicas pertenece en última instancia a quienes pagan sus impuestos. De acuerdo a esta afirmación, y al valor potencial que tienen los recursos de información públicos, muchos gobiernos han querido dar un paso más y buscar, además de la transparencia o involucrar a la sociedad civil en las políticas públicas, la mejora de la productividad de la propia Administración y ofrecer la información pública de forma abierta para que cualquier ciudadano pueda hacer uso de ella (Álvarez Espinar, 2014).

Cualquier administración pública gestiona ingentes cantidades de recursos y documentos digitales que podrían ser de gran valor para su reutilización por parte de terceros: información geográfica, datos estadísticos, información sobre el transporte, etc. Los usos son múltiples y variados, y aunque muchas veces la reutilización podría hacerse directamente sobre una imagen o un texto legal, habitualmente la información más fácil de reutilizar es aquella estructurada y que permite un procesamiento automático –datos–. La dificultad en el procesamiento de los datos condicionará de forma notable su reutilización, que es el objetivo primordial de la apertura de los mismos (Álvarez Espinar, 2014).

2.3.4. Información del Sector Público

Como información del sector público (o Public Sector Information, en inglés) se considera aquella información generada o custodiada por organismos públicos, propia de cualquier nivel territorial, cualquier sector económico y de cualquier tipo (documentos pictográficos, datos estadísticos, resultados de estudios o análisis, información sobre los servicios públicos, etc.), y que podría estar sujeta a licencias o derechos de autor. Asimismo, gran parte de esta información está sujeta a regulación estricta por tratarse de datos sensibles, como por ejemplo los clasificados bajo confidencialidad comercial y estadística, los datos personales, información pública o aquella información que afecte a la defensa o seguridad del Estado (Álvarez Espinar, 2014).

La información del sector público (ISP) es la amplia gama de información que los organismos del sector público recogen, producen, reproducen y difunden en muchos áreas de actividad mientras desarrollan sus tareas institucionales. La ISP puede incluir (entre otros) información social, económica, geográfica, catastral, meteorológica, turística, o de negocios (European Data Portal, 2016).

La ISP que se genera y se recoge de manera regular tiene un enorme potencial. El potencial económico de la ISP en la Unión Europea va más allá de miles de millones de euros al año, con un potencial para estimular la economía global y crear nuevos puestos de trabajo. El estudio económico realizado como parte del proyecto del Portal de Datos Europeo estima que el tamaño del mercado directo de datos sea de 75,7 millones de euros en 2020 y el número de puestos de trabajo de datos abiertos sea de casi 100.000 en 2020. El informe "Creating Value through Open Data" (Informe de creación de valor a través de datos abiertos) puede ser

encontrado aquí:

http://www.europeandataportal.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf (European Data Portal, 2016).

En 2013, en la cumbre del G8 se señaló la importancia de los datos de gobierno abierto mediante la creación de la Open Data Charter (Carta de datos abiertos). Este documento pone énfasis en el papel que los datos abiertos pueden desempeñar tanto en la gobernabilidad como en la estimulación del crecimiento. La carta define cinco principios que deben seguir las naciones que abren sus datos (European Data Portal, 2016):

1. Datos abiertos por defecto: todos los datos del gobierno serán publicados por defecto.
2. Calidad y cantidad: los datos deben ser liberados tan pronto como sea posible, en su formato original, y descritos en un lenguaje claro. Este principio denota la importancia de los metadatos y el feedback del usuario para mejorar la calidad.
3. Usables por todos: los datos serán publicados en formatos abiertos siempre que sea posible y sin coste ninguno para el usuario.
4. Liberar datos para la mejora del gobierno: los gobiernos compartirán su experiencia técnica y documentarán sus propias iniciativas de datos abiertos.
5. Liberar datos para la innovación: los gobiernos del G8 promoverán la alfabetización en datos abiertos y que la liberación sea en formatos legibles por máquinas.

La directiva relativa a la reutilización de la información del sector público (2003/98 / CE), también conocida como la Directiva ISP, proporciona un marco jurídico común para un mercado europeo de datos controlados por los gobiernos. En 2013 se elaboró una revisión de la Directiva ISP, la 2013/37 / UE (EUR-Lex, 2013). Las principales modificaciones son cambiar el modelo de coste para obtener información del sector público, la inclusión de ciertas instituciones culturales como organismos del sector público, el aumento de la transparencia en relación con el cálculo de las tasas, y el apoyo a formatos legibles por máquina y formatos abiertos (European Data Portal, 2016).

2.3.5. Diferencias entre ISP y los datos de gobierno abierto

Muchas veces se utilizan como sinónimos los términos ISP y datos de gobierno abierto, pero una definición estricta de la ISP de acuerdo con la Directiva ISP revelaría algunas discrepancias. Se debe tener en cuenta que tanto la Directiva ISP como el llamado movimiento Open Data proporcionan un conjunto de reglas y principios que pueden ponerse en práctica de una manera ligeramente diferente en diferentes países y diferentes marcos legales existentes. En resumen, la diferencia principal es que la ISP se refiere sólo a los datos en poder de los organismos del sector público, y que su reutilización puede ocurrir en determinadas circunstancias. Si la ISP está disponible bajo una licencia abierta, se llama datos de gobierno abierto. El término "datos abiertos" se refiere también a otros tipos de datos que no son del sector público y que están disponibles gratuitamente, por ejemplo, datos de social media (European Data Portal, 2016).

2.4. Datos abiertos de las empresas

Las empresas serán un miembro muy interesado en la producción de datos abiertos y su posterior publicación para ofrecer servicios de valor añadido. En el artículo “Open data, big data: ¿hacia dónde nos dirigimos?” de Antonia Ferrer y Enrique Sánchez publicado en el anuario ThinkEPI de 2013 se analiza el alcance que tendrán los datos abiertos para las empresas. Por desgracia la información encontrada ha sido muy escasa y no se ha podido profundizar en el tema.

2.5. Legislación sobre reutilización

Empezaremos viendo que es la reutilización de la información o de los datos. Cuando la información accesible pasa a ser reutilizable, es cuando se desvincula su finalidad del proceso administrativo y se convierte en una actividad privada. Toda información reutilizable ha de ser un documento accesible, pero no toda información accesible es información reutilizable. Para que la información sea reutilizable necesita cumplir tres requisitos (Mendo y otros, 2013):

a) Jurídicamente accesible: Libre acceso, sin restricciones, no sujeto a las limitaciones impuestas por el derecho de acceso (propiedad intelectual, secretos oficiales, secreto industrial y comercial, defensa nacional, secreto de sumario, protección de la intimidad... o restricciones administrativas temporales).

b) Técnicamente interoperable. Para alcanzar la interoperabilidad debemos utilizar estándares abiertos, para evitar la discriminación tecnológica, y que existan listados de activos de información.

c) Vinculación a una fuente de información pública. Un documento reutilizable del sector público procede de la actividad previa del organismo público. Por lo tanto, el documento reutilizable debe estar vinculado a una fuente de información pública, ha de tener un origen público.

Así pues, la información reutilizable se puede definir como información generada por organismos públicos, accesible para todos los ciudadanos, que puede ser sometida a un proceso de generación de valor mediante técnicas de interoperabilidad y que en todo su ciclo de gestión (creación, tratamiento, difusión y archivo) está identificada como fuente de origen público. Cuando una información originada por un organismo público inicia un proceso de generación de valor para una finalidad distinta de la que fue creado se crea un nuevo activo de información (Mendo y otros, 2013).

2.5.1. Proceso de reutilización

Si consideramos la reutilización como un proceso, vemos dos aspectos del mismo como son la transformación documental del activo para añadirle valor y las posibilidades de uso que se da al activo de información una vez transformado (Mendo y otros, 2013):

a) Transformación documental del activo

Durante el proceso de reutilización, para añadir valor, el activo de información puede ser sometido a varias transformaciones, como por ejemplo, el resumen, la agregación a otro activo de información o la unión de varios activos para crear uno nuevo. La reutilización siempre ha de conllevar un aumento del valor o de la utilidad. Este valor puede ser informativo, social, tecnológico, etc.

Cualquier aplicación o herramienta informática susceptible de almacenar información y recuperarla de forma individual o agregada, como bases de datos o repositorios, son dispositivos para reutilizar información. En este caso tenemos los datos, que aislados no tienen ningún valor, sólo alcanzan valor cuando se agregan y desagregan con otras categorías de datos.

b) Posibilidades de uso del activo

Una vez configurado el activo para otros fines, es importante aclarar que tipos de usos son constitutivos de reutilización. El responsable puede realizar las siguientes acciones con el activo reutilizado:

- Ser aprovechado por el reutilizador en otros procesos (por ejemplo, servicios de estudios).
- Cederlo de forma libre a cualquier usuario, permitiendo nuevas reutilizaciones.
- Utilizarlo con fines comerciales y no comerciales.

La normativa española de reutilización de datos públicos viene determinada tanto por la legislación nacional como por la comunitaria.

2.5.2. Legislación comunitaria

El punto de partida de la legislación comunitaria lo constituye la directiva europea 2003/98/CE relativa a la reutilización de la información en el sector público. El concepto general de reutilización se define en dicha directiva como “el uso de documentos que obran en poder de organismos del sector público por personas físicas o jurídicas con fines comerciales o no comerciales distintos del propósito inicial que tenían esos documentos en la misión de servicio público para la que se produjeron. El intercambio de documentos entre organismos del sector público en el marco de sus actividades de servicio público no se considerará reutilización”.

De acuerdo con datos.gob.es (2011), en la directiva 2003/98/CE se proponen cuatro medidas principales:

- a) Todos los documentos difundidos por organismos públicos podrán reutilizarse para fines comerciales y no comerciales, a no ser que estén sujetos a restricciones derivadas de la normativa de acceso.
- b) Los datos se deberán ofrecer en formatos legibles por máquina con tal de garantizar su reutilización.
- c) La ampliación del ámbito de aplicación de la directiva, llegando también a las bibliotecas, los museos y los archivos.
- d) La implantación de un modelo de costes reducidos, salvo excepciones justificadas, para acceder a los datos públicos.

En diciembre de 2011 se presentó la propuesta de directiva del Parlamento Europeo y del Consejo por la que se modifica la directiva 2003/98/CE, relativa a la reutilización de la información del sector público. También en 2011, la Comisión Europea envió una Comunicación al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones titulada: *Datos abiertos Un motor para la innovación, el crecimiento y la gobernanza transparente* (Mendo y otros, 2013).

En la directiva 2013/37/UE del Parlamento Europeo y del Consejo de 26 de junio de 2013, por la que se modifica la directiva 2003/98/CE relativa a la reutilización de la información del sector público, según Gontzal Gallo (2013), los aspectos más novedosos son:

- a) Se reduce el ámbito de aplicación. En concreto, la directiva no se aplicará a documentos no relacionados con la actividad administrativa o a aquellos que entren de lleno en el terreno de la confidencialidad comercial. La directiva tampoco se aplicará a los documentos relacionados con el derecho a la protección de datos ni a los creados por instituciones culturales, con la excepción de las bibliotecas, archivos y museos.
- b) Se añade la posibilidad de interponer recursos sobre decisiones de reutilización ante un órgano de revisión imparcial.
- c) Se exige a los estados que creen herramientas para buscar documentos reutilizables.
- d) Especialmente novedoso es la obligación de los estados de presentar a la Comisión, cada tres años, "un informe sobre la disponibilidad de información del sector público para reutilización, las condiciones que rigen su disponibilidad y las prácticas en materia de recurso".

2.5.3. Legislación española

En nuestro país disponemos de tres normas sobre reutilización de la información:

- LRISP (Ley 37/2007 sobre Reutilización de la Información del Sector Público)
- RDRISP (Real Decreto 1495/2011, por el que se desarrolla la LRISP)

- Real Decreto 4/2010 por el que se regula el Esquema Nacional de Interoperabilidad en el ámbito de la Administración Electrónica

LRISP (Ley 37/2007 sobre Reutilización de la Información del Sector Público)

Esta ley es la adaptación al ordenamiento jurídico español de la Directiva 2003/98/CE. En el artículo 3 de la LRSIP se define la *reutilización* como “el uso de documentos que obran en poder de las administraciones y organismos del sector público, por personas físicas o jurídicas, con fines comerciales o no comerciales, siempre que dicho uso no constituya una actividad administrativa pública. Queda excluido de este concepto el intercambio de documentos entre administraciones y organismos del sector público en el ejercicio de las funciones públicas que tengan atribuidas”.

En el segundo punto del artículo 3, se define *documento* como “toda información cualquiera que sea su soporte material o electrónico, así como su forma de expresión gráfica, sonora o en imagen utilizada. A estos efectos no se considerarán documentos los programas informáticos que estén protegidos por la legislación específica aplicable a los mismos”.

De acuerdo a esta ley, algunos documentos quedan excluidos de la reutilización por tener derechos de propiedad intelectual. Son los siguientes (Clabo y Ramos Vielba, 2015):

- De procedimientos administrativos en curso.
- Que afecten a la defensa nacional.
- Ajenos a la función de servicio público.
- Con derechos de propiedad intelectual o industrial.
- De radiodifusión sonora o televisiva.
- De instituciones culturales, educativas y de investigación.

También se recoge en esta ley la posibilidad del cobro de tasas y precios públicos, el procedimiento de tramitación de solicitudes de reutilización y el régimen sancionador. Pero no se tratan aspectos interesantes como el establecimiento de modalidades de puesta a disposición, las condiciones que deben cumplir las licencias y la prohibición de acuerdos exclusivos (Clabo y Ramos Vielba, 2015).

En el año 2015 hubo un cambio en esta norma. Es la ley 18/2015, de 9 de julio, por la que se modifica la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público.

RDRISP (Real Decreto 1495/2011, por el que se desarrolla la LRISP)

En este real decreto se desarrollan los preceptos de la LRISP teniendo en cuenta las diferentes aportaciones internacionales surgidas desde su aprobación. En esta norma se cita el Plan de Acción de la Unión Europea sobre Administración Electrónica para el período 2011-2015, la Declaración Ministerial de Granada de abril de 2010 y la nueva Agenda Digital Europea de mayo de 2010. A su vez, el real decreto también se enmarca en el conjunto de medidas que constituyen la Estrategia 2011-2015 del Plan Avanza 2 (Clabo y Ramos Vielba, 2015).

El RDRISP establece una serie de responsabilidades en las administraciones para poner a disposición los datos reutilizables, desde el punto de vista y administrativo. También fija las condiciones de reutilización, de modo que pone a disposición de los ciudadanos la información sin condiciones específicas, a excepción de los límites marcados por la propiedad intelectual e industrial de terceros (Clabo y Ramos Vielba, 2015).

Por último, tras definir el marco legal para el funcionamiento del Catálogo de Información Pública, disponible en datos.gob.es, dedica un apartado a los aspectos técnicos del proyecto que afectan a los requisitos de visibilidad, accesibilidad e interoperabilidad. Así, coloca como de obligado cumplimiento la Norma Técnica de Interoperabilidad de Reutilización de Recursos de Información, y anima a crear un espacio web de datos abiertos, a ser posible en sede.gob.es (Clabo y Ramos Vielba, 2015).

Real Decreto 4/2010 por el que se regula el Esquema Nacional de Interoperabilidad en el ámbito de la Administración Electrónica

En esta norma se citan los criterios y recomendaciones tecnológicas básicas para la publicación de información reunida por el sector público. A continuación se enumeran los puntos más importantes (Clabo y Ramos Vielba, 2015):

- Definir el esquema de identificadores de recursos uniformes (URI) para que se muestren los datos de forma única, fiable y persistente.
- Determinar los metadatos para describir los catálogos y sus conjuntos de datos.
- Definir la taxonomía común de sectores primarios.

Reutilización de datos públicos. Proyecto Aporta

En España la ley 37/2007 de 16 de noviembre de Reutilización de la información del sector público es la adaptación a nuestro ordenamiento la Directiva 2003/98/CE de 17 de noviembre. Ha sido una herramienta importante para el avance de la reutilización, al establecer un marco regulatorio general para esta actividad aplicable a todas las administraciones públicas españolas. La Ley pretende “armonizar la explotación comercial de la información del sector público y la publicación de todos los documentos de libre disposición que obran en su poder, al considerar que son un instrumento esencial para el desarrollo del derecho al conocimiento, que constituye un principio básico de la democracia” (Marcos Martín y Soriano Maldonado, 2011).

Junto a este marco legal, en España existen numerosos organismos públicos que ya han elaborado políticas favorables a la reutilización. Destacan la Dirección General del Catastro, el Instituto Geográfico Nacional o la Agencia Estatal de Meteorología (Aemet). También es importante destacar la labor de organizaciones sectoriales como la Asociación Multisectorial de la Información (Asedie), empresas infomediarias e iniciativas sociales como Pro Bono Público (Marcos Martín y Soriano Maldonado, 2011).

El proyecto Aporta ha ayudado muchísimo en el impulso que está recibiendo la reutilización en España en los últimos años, según ha señalado entre otros la OCDE. Reúne un conjunto de buenas prácticas que han tomado como referencia otros estados de la UE, que han analizado el ejemplo español para definir sus propias políticas en la materia (Marcos Martín y Soriano Maldonado, 2011).

2.6. Casos prácticos

2.6.1. Ámbito nacional

A continuación vemos una serie de portales nacionales sobre datos abiertos (Ramos Simón y otros, 2012):

Austria

El portal de Austria (data.gv.at) persigue objetivos como la mejora en la transparencia y en la visibilidad de los datos como su utilización por terceras partes (por ejemplo, las empresas infomediarias), y los datos originales se ofrecen en formato CSV, JSON, KML y GML.

Bélgica

El portal data.gov.be es la plataforma belga en la que se alojan los datos abiertos de los servicios públicos federales y flamencos y de las instituciones del gobierno federal y flamenco. Los usuarios de este portal pueden utilizar la información disponible en el portal de forma gratuita, libre y para fines personales. No obstante, en el registro de los datasets se proporciona información relativa al tipo de licencia para aquellos casos en los que existan restricciones o se requiera una autorización para la reproducción o utilización de los datos; los servicios o instituciones gubernamentales son los responsables de establecer las condiciones que regulan el acceso y utilización de los datos disponibles en el portal.

El portal belga está disponible en holandés, francés, alemán e inglés y dispone de más de 5000 datasets, a fecha de septiembre de 2016. Hay que decir que este número ha crecido notablemente dado que en octubre de 2012 este portal sólo contenía 105 datasets. Ofrece sus datos en formatos como Atom/RSS, CSV, HTML, MDB, TXT, XHTML, XLS o XML.

Dinamarca

El portal digitaliser.dk es un repositorio de información pública abierta y de normas de intercambio de datos y, además, un espacio creativo para personas, organismos y empresas involucradas en la digitalización del sector público. El portal funciona como una red social para

el desarrollo y el intercambio de conocimientos, y como foro público para la digitalización de la información pública en Dinamarca. Su objetivo es estimular el desarrollo y la adopción de herramientas abiertas para la gestión de contenidos digitales públicos.

Con esta herramienta, el gobierno danés ha creado un modelo de colaboración entre la comunidad tecnológica y el gobierno, para conseguir una comunicación más directa entre el sector público, los ciudadanos y las empresas, que ya no son receptores pasivos de información pública, sino que participan en el diálogo y el intercambio de conocimiento con el sector público.

La mayor parte de las fuentes de datos pertenecen a distintas organizaciones gubernamentales y una minoría pertenece a empresas privadas. El número de registros del que dispone actualmente es de 700 aproximadamente. En cuanto al formato de los datos, en general son datos estructurados en formatos como GeoJSON, HTML, JSON, KML, XML, XLS, HTML o PDF.

España

datos.gob.es es el portal de carácter nacional que organiza y gestiona el Catálogo de Información Pública de la Administración General del Estado de España (AGE). En él, se puede encontrar información general, materiales formativos y noticias sobre la reutilización de la información en España y en el resto del mundo; destacan el catálogo de datos y las aplicaciones (Ramos Simón y otros, 2012).

Este portal está dirigido a profesionales, empresas, gestores institucionales y ciudadanos, con la intención de mejorar el acceso a la información pública y promover la reutilización de la información en el sector público correspondiente a la Administración General del Estado y a otras altas instituciones estatales. Por otro lado, también hay que tratar el abundante número de portales autonómicos y locales (Ramos Simón y otros, 2012).

2.6.2. Ámbito autonómico

En España también han surgido varios proyectos a nivel autonómico para promover el gobierno abierto. Uno de los más importantes ha sido Irekia, el proyecto del País Vasco (Coroan y Campos, 2011).

En mayo de 2010, basado en el memorándum de Obama, el gobierno vasco puso en marcha el proyecto Open Data, con el objetivo de hacer públicos los datos del gobierno autonómico en formato reutilizable. Irekia se define como un “canal de comunicación directa entre la ciudadanía y la Administración a través de Internet” que se basa en los tres ejes principales del gobierno abierto: transparencia, participación y colaboración (Coroan y Campos, 2011).

También encontramos planes a nivel municipal. Algunos de los proyectos locales más interesantes de gobierno abierto son los desarrollados por los ayuntamientos de Gijón, Zaragoza o Barcelona (Coroan y Campos, 2011).

El ayuntamiento de Gijón ha puesto en marcha el proyecto Datos Abiertos de Gijón, orientado al gobierno abierto, para desarrollar los valores de transparencia, participación ciudadana, servicio y eficiencia. Una de sus primeras medidas fue la creación de un catálogo de datos donde proporcionar información específica sobre los datos que pueden ser reutilizados. También se han diseñado aplicaciones piloto con tal de alcanzar una mayor participación ciudadana, por ejemplo, la consulta ciudadana sobre qué datos se deberían publicar (Coroan y Campos, 2011).

2.6.3. Asociaciones independientes

Por último se muestran unos grupos que participan activamente en la promoción y mejora de las iniciativas de datos abiertos o de la transparencia política (Álvarez Espinar, 2014):

Open Knowledge Foundation (Fundación del Conocimiento Abierto): es una entidad sin ánimo de lucro registrada en Reino Unido, que promueve el conocimiento abierto desde distintas perspectivas para mejorar la gobernanza, la cultura, la investigación y la economía.

OpenKratio: término que proviene de la unión de la palabra 'open' (abierto, en inglés) y kratio, proveniente de la palabra demokratio (democracia, en esperanto) es un grupo constituido en Sevilla para dar respuesta a las inquietudes de un grupo de ciudadanos y ciudadanas motivados por introducir en la sociedad, y especialmente en las administraciones públicas, los principios del gobierno abierto y los datos abiertos. Este colectivo de ámbito estatal, pretende concienciar sobre estos temas a través de iniciativas de difusión, participación, colaboración y desarrollo de proyectos en el espacio público y político.

Fundación Ciudadana Civio: es una organización sin ánimo de lucro que intenta mejorar la democracia en España, con el objetivo de una transparencia gubernamental real y un libre acceso a los datos públicos, animando a la participación.

Access Info: es una organización de derechos humanos creada en 2006 y que pretende promover y proteger el acceso a la información en Europa y el mundo, considerado como un derecho esencial para la participación en la toma de decisiones del gobierno.

Transparencia Internacional (TI) es una organización internacional, no gubernamental y sin fines lucrativos, dirigida a combatir la corrupción a nivel nacional e internacional. TI-España publica regularmente varios índices de referencia que evalúan el estado de las administraciones públicas en España.

Ana Corojan y Eva Campos también tratan otra iniciativa de origen civil como es ¿dónde van mis impuestos? Es la versión española de la web británica Where Does My Money Go?, que pretende ofrecer una explicación sencilla sobre cómo se gasta el Estado los impuestos recibidos (Corojan y Campos, 2011).

2.7. Formatos de archivo

Tim Berners-Lee, inventor de la web y director del W3C, elaboró una clasificación para medir la calidad tecnológica de los datos abiertos, mediante un sistema basado en estrellas –con valores de una a cinco– de acuerdo al formato utilizado para representar los datos. Este esquema considera los siguientes casos (Álvarez Espinar, 2014):

★ Una estrella

En el nivel más bajo encontramos los:

- Datos o documentos disponibles en la web en cualquier formato.
- Distribuidos bajo una licencia abierta no restrictiva.
- Con formato no estructurado.
- Los datos o el documento se pueden visualizar en la web pero no procesar automáticamente.

Algunos ejemplos serían una imagen en formato JPG o PNG o un documento escaneado en formato PDF.

★★ Dos estrellas

Aquí entrarían los:

- Datos o documentos estructurados.
- Procesables automáticamente.
- Formato propietario (no abierto).

Un ejemplo sería una hoja de cálculo en formato Excel (con extensión .xls o .xlsx).

★★★ Tres estrellas

En este tercer nivel estarían los datos o documentos con un formato estructurado y abierto (no propietario). Por ejemplo, una hoja de cálculo en formato CSV (valores separados por comas) pero no en formato de Microsoft Excel.

★★★★ Cuatro estrellas

En el cuarto nivel aparecerían los documentos que pueden ser referenciados mediante direcciones web persistentes o identificadores de recursos uniformes (URI). También encontraríamos documentos que siguen formatos estándares y abiertos del W3C, como RDF o SPARQL para describir semánticamente la información.

Un ejemplo sería representar en el modelo RDF los edificios de un organismo público, con sus datos de contacto y ubicación. Otro caso son los datos atómicos a los que se puede acceder mediante direcciones web (URI).

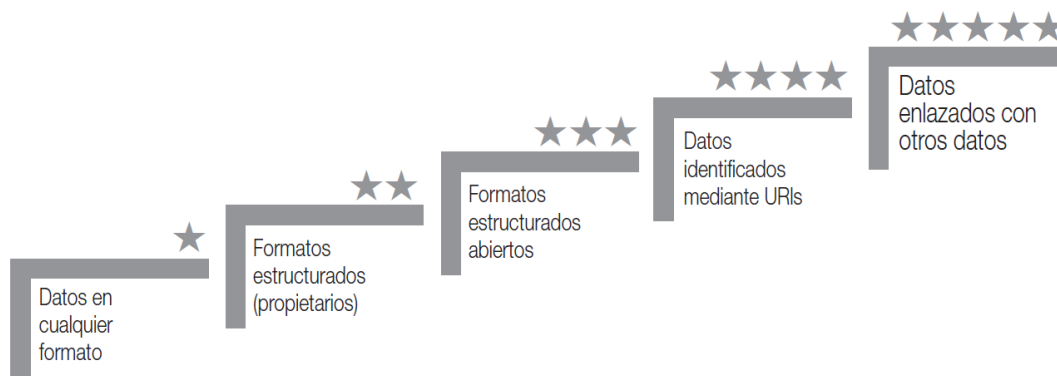
★★★★★ Cinco estrellas

En el nivel más alto tendríamos los datos enlazados con otros datasets para ofrecer contexto a la información. Esto permite que se establezcan relaciones semánticas entre la información enlazada.

Un ejemplo lo podríamos encontrar uniendo las descripciones de los edificios públicos del apartado anterior con enlaces a Geonames, una gran base de datos especializada en localizaciones geográficas.

Con estos enlaces se podría incluir una descripción al detalle de las localidades, regiones, o países y así tener acceso de forma directa a información socio-económica o toponímica de esos lugares.

Figura 6: Clasificación de iniciativas open data



Fuente: Apertura y reutilización de datos públicos / Álvarez Espinar (2014, p50)

La excelencia técnica –cinco estrellas– se consigue cuando los datos son enlazados con otros recursos en la Web mediante mecanismos semánticos, que ofrecen una interoperabilidad plena entre distintos sistemas, y permiten una posterior reutilización mucho más eficiente.

3. Métricas

3.1. Los índices bibliométricos en la investigación científica

Durante los últimos años del siglo XX se ha evidenciado una creciente necesidad de evaluar los resultados de la actividad investigadora, tanto de países o regiones como de centros o grupos de investigación. Obtener métodos objetivos para evaluar la calidad de la investigación es un deseo compartido y perseguido por los gestores de política científica de todo el mundo (Gómez Caridad y Bordons Gangas, 1996).

Una forma de valorar el producto de la investigación, consiste en analizar las publicaciones que genera y la publicación principal en la ciencia es el artículo científico (González de Dios, Moya y Mateos Hernández, 1997).

La herramienta utilizada para medir la actividad científica son los indicadores bibliométricos. Se utilizan, por un lado, para analizar el tamaño, crecimiento y distribución de la bibliografía científica (ya sean libros, revistas, patentes, etc.), para mejorar las actividades de información, documentación y comunicación científica, y por otro, para analizar los procesos de generación, propagación y uso de la literatura científica y la estructura y dinámica de los colectivos de investigadores que producen y utilizan dicha literatura (Sancho, 1990).

Los indicadores bibliométricos permiten manejar, clasificar y analizar grandes volúmenes de publicaciones científicas (Gómez Caridad y Bordons Gangas, 1996) y proporcionan información sobre los resultados del proceso investigador, su volumen, evolución, visibilidad y estructura. Por tanto, sirven para valorar la actividad científica, y la influencia (o impacto) tanto del trabajo como de las fuentes (Camps, 2008).

El uso de los indicadores bibliométricos para estudiar la actividad investigadora de un país o área se basa en que las publicaciones científicas son un resultado esencial de dicha actividad. Un nuevo conocimiento adquiere valor cuando se da a conocer y se difunde, porque sólo así puede contribuir al avance científico. Por ello, la publicación científica se convierte en un resultado importante y tangible de la investigación, y los indicadores bibliométricos adquieren validez como medida de la actividad científica (Camps, 2008).

La bibliometría y el empleo de sus indicadores, constituyen herramientas científicas necesarias, porque permiten la cuantificación de la ciencia en forma objetiva, pues se potencian con la explosión del conocimiento actual y su recopilación en bases bibliográficas de datos (Camps, 2008).

Aunque los estudios bibliométricos han recibido numerosas críticas en los últimos años, la mayor parte de ellas no cuestionan el valor de la bibliometría, sino el uso inadecuado que se hace de ella (Gómez Caridad y Bordons Gangas, 1996).

La bibliometría constituye un medio para situar la producción de un país con respecto al mundo, una institución en relación con su país y hasta los científicos en relación con sus propias comunidades (Camps, 2008).

Alan Pritchard publicó la obra “*Statistical Bibliography or Bibliometrics?*” en 1969 en la que utilizó por primera vez la palabra bibliometría (bibliometrics). Definió este nuevo concepto como la aplicación de métodos matemáticos y estadísticos a libros y otros medios de comunicación. Veinte años después, en 1989, Bertram Claude Brookes define la bibliometría como una disciplina propia de la actividad bibliotecaria, la cual se debe enriquecer con la estadística para refinar sus técnicas (Canales Becerra y Mesa Fleitas, 2003).

Una definición más reciente afirma que la bibliometría se centra en el análisis cuantitativo de las publicaciones científicas y académicas, incluidas las patentes. La bibliometría es parte del campo de la cienciometría: la medición de todos los aspectos de la ciencia y la tecnología, que podrán abarcar información sobre cualquier tipo de resultados de la investigación (datos, respuestas, software, interacciones científicas, financiación, comercialización de la investigación, y otros productos (Wilsdon y otros, 2015).

3.2. Indicadores bibliométricos

Un indicador es un parámetro que se utiliza en la evaluación de cualquier actividad. Lo habitual es utilizar varios y cada indicador se encarga de una parte del objeto a evaluar. La ciencia como es multidimensional requiere un conjunto de indicadores (Sancho, 1990).

En la obra “*The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*” se define un indicador como una cantidad medible que se reemplaza o sustituye por algo más difícil de medir y que se asocia con él sin medirlo directamente. Por ejemplo, el número de citas podría ser utilizado como un indicador del impacto científico de los artículos de revista a pesar de que puede haber un impacto científico que no genera citas. Del mismo modo, el recuento de programas de estudios en línea que mencionan un libro en particular puede ser utilizado como un indicador de su impacto educativo (Wilsdon y otros, 2015).

Isabel Gómez y María Bordons definen un indicador bibliométrico de la siguiente manera:

“Son datos estadísticos deducidos de las publicaciones científicas. Su uso se apoya en el importante papel que desempeñan las publicaciones en la difusión de los nuevos conocimientos, papel asumido a todos los niveles del proceso científico” (Gómez y Bordons, 1996).

Existen dos grandes grupos de indicadores bibliométricos: los indicadores cuantitativos de la actividad científica, donde estaría el número de publicaciones; y los indicadores de impacto, que tienen en cuenta el número de citas que reciben los trabajos. El número de publicaciones de un centro académico o región es un indicador útil para cuantificar su actividad científica.

Nos sirve para poder comparar la actividad de varios centros o territorios, y así disponer de una referencia en la que colocar nuestro nivel de estudio. También es interesante realizar seguimientos de la producción científica a lo largo del tiempo. Como indicadores de impacto encontramos el número de citas que reciben los trabajos y el factor de impacto de las revistas (Bordons y Zulueta, 1999). Nuestro trabajo está más relacionado con el segundo grupo que con el primero.

3.2.1. Ventajas de los indicadores bibliométricos

Gracias a los indicadores bibliométricos se puede determinar (Sancho, 1990):

- El crecimiento de un campo científico, conforme al número de trabajos publicados en ese campo.
- El envejecimiento de un campo científico, de acuerdo a la vida media de las publicaciones.
- El crecimiento de un campo científico, según al año de publicación de los documentos de ese campo.
- La producción académica de los autores o las instituciones, calculada por el número de publicaciones.
- La colaboración entre científicos o instituciones, obtenida por la cantidad de autores por trabajo o centros de investigación que trabajan.
- El impacto de las publicaciones en la comunidad científica internacional, calculado mediante el número de citas que reciben.

Históricamente el método más utilizado para evaluar la calidad de un trabajo o de un proyecto de investigación, o incluso la carrera de un investigador es la evaluación hecha por los propios investigadores o por pares (peer review) (Velasco y otros, 2012).

3.2.2. Limitaciones de los indicadores bibliométricos

Los indicadores bibliométricos cuentan con las siguientes limitaciones (González de Dios, Moya, y Mateos Hernández, 1997):

- El cómputo total de las publicaciones no proporciona idea de su calidad y también hay que tener en cuenta que las prácticas de publicación varían con el tiempo.
- El impacto de un trabajo muestra su eficacia y su valor, pero la falta de impacto no indica necesariamente la inutilidad del mismo.
- Los indicadores bibliométricos omiten la comunicación científica que no produce publicaciones. Esto promueve la división de información en la que se publican varios trabajos en lugar de uno, y la publicación de un mismo trabajo, con ligeras variaciones, en varias revistas distintas.

- Pueden haber investigadores denominados «de academia», los cuales están muy interesados en publicar para conseguir éxito en su carrera: el lema «publica o perece» es su regla de oro.
- La revisión por pares se puede ver manipulada por los intereses de los científicos que realizan las evaluaciones. Hay que reducir, tanto como sea posible, los sesgos existentes en las relaciones interpersonales (simpatías, antipatías, competencia, etc.) además de una adecuada selección de correctores. Se ha observado que la calidad de la revisión es notablemente superior si ésta se efectúa a ciegas. El experto debe tener en cuenta varios aspectos en la evaluación de un artículo: rapidez y puntualidad, actitud positiva e imparcialidad, rigor y proceder sistemático, y honestidad.
- Por último, los investigadores pueden tener preferencia por las disciplinas tradicionales en detrimento de las nuevas.

3.3. Tipos de métricas

3.3.1. Métricas tradicionales

El factor de impacto o índice de impacto

El factor de impacto (FI), fue ideado por Eugene Garfield en 1955 para identificar las revistas que debían estar en el Science Citation Index (SCI). Se publica anualmente en el Journal Citation Reports (JCR). El FI de una revista es el número de veces que son citados en un año los artículos de una determinada revista publicados en los dos años anteriores (Aleixandre Benavent, Valderrama Zurián y González Alcaide, 2007).

El factor de impacto es la primera medida objetiva y cuantificable de la valoración de una revista en el ámbito científico. Por esta razón, ha sido ampliamente aceptado por parte de la comunidad científica, aunque también tiene inconvenientes.

Ventajas del FI

Utilizar el factor de impacto tiene las siguientes ventajas: es universal, fácil de usar y de entender. En el año 2010 hay más de 10.500 revistas indexadas de más de 80 países y clasificadas en más de 200 categorías. También puede ser importante para los bibliotecarios o los especialistas en documentación a la hora de seleccionar las revistas para la colección (Velasco y otros, 2012).

Defectos del FI

Los problemas principales del FI son los siguientes (Velasco y otros, 2012):

- Un tiempo de cálculo de citas de dos años se debería utilizar sólo en áreas con un rápido envejecimiento de las publicaciones. Por eso, esas áreas son las que presentan valores más altos en el FI. Por ejemplo, la biología molecular o la genética. En el lado opuesto, las áreas de lento envejecimiento como la cirugía o la pediatría, una gran cantidad de citas tendrán una antigüedad superior a dos años y no se incluyen por lo tanto en el cálculo del factor de impacto. Para el análisis de revistas de esas disciplinas sería mejor utilizar un periodo más largo (entre 4 o 6 años).
- El factor de impacto de las revistas se utiliza para evaluar a los autores. En una revista nos podemos encontrar con artículos muy citados de un determinado autor y también con publicaciones de otro autor que no reciben citas. Los dos autores publican sus artículos en la misma revista pero sus publicaciones obtienen un impacto muy diferente.
- El factor de impacto de una revista se ve afectado por el tamaño de su comunidad científica. Es decir, cuanto más grande sea el tamaño de la comunidad científica más artículos “super citados” tendrá el área (Bordons y Zulueta, 1999).
- Algunos científicos afirman que algunos editores obligan a los autores a citar artículos de su propia revista. Se denominan “autocitas” y provoca que el factor de impacto de esas revistas aumente de forma poco ética. Si el porcentaje de autocitas llega al 20% la revista podría ser excluida de este índice.
- Las revistas con acceso electrónico obtienen mayor factor de impacto que las revistas que no lo tienen.

María Bordons y María Ángeles Zulueta añaden que el número de citas no mide la calidad o importancia del artículo. También nos avisan de que en una disciplina con una gran comunidad científica es más fácil obtener un alto factor de impacto (Bordons y Zulueta, 1999).

El factor de impacto tiene otros errores como que no separa las citas dirigidas a artículos de investigación de otros documentos menos importantes como las cartas o los editoriales. Además, en el SCI se prefiere incluir revistas que se publican en inglés. También es importante remarcar que el factor de impacto es mayor en áreas con un gran número de investigadores. Como conclusión hay que considerar que el factor de impacto mide la frecuencia con la que se citan artículos de una determinada revista en un año concreto y eso no implica que la revista sea de alto nivel (Aleixandre Benavent, Valderrama Zurián y González Alcaide, 2007).

Índice h

Este modelo consiste básicamente en ordenar las publicaciones de un autor de forma descendente según las citas recibidas por cada trabajo. Encontramos el índice h de un autor cuando la posición en la lista es mayor o igual al valor de la cita (Velasco y otros, 2012).

Ventajas del índice h

Este indicador, se dirige en principio a investigadores y ofrece las siguientes ventajas (Aleixandre Benavent, Valderrama Zurián y González Alcaide, 2007):

- a) Es una cifra que sólo puede aumentar o quedarse fija.
- b) Se dirige a los científicos de forma individual y permite comparar la producción de científicos de edades diferentes.
- c) Permite averiguar el rendimiento de un científico a medio plazo. Alcanzar un índice h de 20 es considerado un éxito para un investigador y un factor h igual a 35 o mayor correspondería a los mejores científicos.

Defectos del índice h

Los problemas o limitaciones de este indicador son las siguientes (Velasco y otros, 2012):

- La cantidad de citas que recibe un artículo no refleja su calidad, es decir, un trabajo práctico puede ser muy útil y no tener muchas citas. Además las “falsas investigaciones” pueden tener muchas citas.
- Los autores pueden publicar documentos científicos de disciplinas que no sean su especialidad y recibir muchas citas. Esto le ha ocurrido sin ir más lejos a Jorge Hirsch, cuyo artículo sobre el índice h ha recibido más de 2000 citas pero no está relacionado con su especialidad.
- Depende en gran manera de los años de trabajo del investigador. Cuanto más tiempo trabaje más trabajos serán citados.
- Al igual que el factor de impacto, el índice h depende de la cantidad de científicos que trabajen en un determinado área. Cuanto más grande sea la comunidad científica más artículos se citaran. Por tanto no se deben comparar los índices h de autores de diferentes disciplinas.

Un ejemplo práctico fue llevado a cabo por el grupo Scimago, de la Universidad de Granada. Esta asociación publicó en 2007 un trabajo en el que comparaba a investigadores nacionales de diferentes especialidades. En la siguiente tabla vemos que existen grandes diferencias en la productividad de los autores pero todos tienen el mismo índice h (6).

Tabla 2. Análisis del índice h de varios investigadores

Orden artículo (por nº citas)	Investigador 1	Investigador 2	Investigador 3	Investigador 4
1	289	34	150	12
2	142	33	72	10
3	113	29	26	9
4	62	26	20	8
5	47	25	18	7
6	34	23	11	6
7	5	6	1	0
8	4	6	1	0
9	3	4	0	0
10	1	4	0	0

Fuente: La utilización de los indicadores bibliométricos para evaluar la actividad investigadora / Berta Velasco y otros, (2012, p82).

3.3.2. Métricas actuales

En los últimos años ha aparecido un nuevo tipo de métricas denominadas “altmetrics” o indicadores alternativos. Son indicadores basados en la web 2.0 o web social. Se utiliza la denominación de alternativos porque miden la repercusión de la actividad científica de una forma no habitual. Se obtienen al contar los “me gusta”, de Facebook, menciones en blogs, número de tuits o veces que un artículo se guarda en un gestor de referencias. Estas medidas han cobrado especial relevancia y prueba de ello es la DORA (Declaración de San Francisco sobre Evaluación de la Investigación, en inglés, *San Francisco Declaration on Research Assessment*), donde se muestra la necesidad de mejorar la forma en que se evalúan los resultados de la investigación científica y de promover un cambio por una evaluación basada en el contenido científico del artículo y no en métricas relacionadas con la revista en que fue publicado ese contenido (Peralta González, Frías Guzmán y Gregorio Chaviano, 2015).

La actividad científica se ha beneficiado del nacimiento de plataformas que permiten compartir de manera libre todo tipo de información y enlazarse a través de la Web 2.0. Este es el caso de las redes sociales académicas, gestores de referencias online, repositorios de acceso abierto e índices de citas abiertos, como Google Scholar y Data Sharing. Los principales defectos de las altmetrics son (Peralta González, Frías Guzmán y Gregorio Chaviano, 2015):

- Algunos artículos científicos tienen muy poco impacto en las redes sociales, pero en otras fuentes (bases de datos e índices de citas) poseen resultados más visibles.

- Las fuentes de las que se extraen estos datos (facebook, twitter, mendeley) pueden ser herramientas inestables. En otras palabras, quizás dejen de existir en un periodo de tiempo corto, por lo que la información es efímera.

James Wilsdon y sus compañeros ofrecen la siguiente definición: Altmetrics son métricas no tradicionales que cubren además del número de citas, las descargas, las publicaciones en redes sociales y otras medidas de impacto de los resultados de la investigación. También se utilizan como sinónimos los términos “métricas alternativas” o “métricas a nivel de artículo”, y está relacionado con la webmetría, o la cibermetría, que miden las características y relaciones de los elementos en línea, como sitios web y archivos log. El ascenso de los nuevos medios sociales ha creado un flujo de trabajo adicional bajo la etiqueta altmetrics y son indicadores derivados de sitios web sociales, como Twitter, Academia.edu, Mendeley o ResearchGate (Wilsdon y otros, 2015).

3.4 Portales para medir el impacto de los datos de investigación

Alex Ball y Monica Duke describen los siguientes servicios para calcular el impacto de las publicaciones (Ball y Duke, 2015a):

Thomson Reuters Data Citation Index

En octubre de 2012, Thomson Reuters lanzó el Data Citation Index (DCI) como parte de la Web of Knowledge. Proporciona registros de manera que de un registro de un repositorio se pueden ver los registros de los estudios de los datos y los datasets en manos de ese repositorio.

En cada registro, el DCI muestra el número de veces que la entidad ha sido citada en la Web of Knowledge. Reconociendo la variedad de formas en que los datasets y los repositorios se pueden citar, el DCI cuenta no sólo las entradas en la lista de referencias, sino también las citas menos formales que se producen en otros lugares de los documentos académicos (por ejemplo, en el resumen o en los agradecimientos).

La selección para la inclusión en el DCI se realiza analizando el repositorio a nivel global y no sobre los datasets o estudios. Los criterios utilizados para la selección de repositorios incluyen la longevidad, sostenibilidad, actividad (en términos de nuevos datos que están siendo depositados), metadatos de los datos (preferiblemente en inglés, con enlaces a la literatura asociada, información de la financiación, etc.), y procedimientos de garantía de calidad.

3.4.1. ImpactStory

ImpactStory permite a los investigadores construir un perfil para presentar sus diferentes actividades académicas. Después de registrarse, los usuarios añaden a su perfil varios productos académicos, como artículos, presentaciones de diapositivas, videos, datos, o software. Se puede hacer al citar la URL de su perfil o identificadores que incluyen los ID de PubMed o los DOI. ImpactStory utiliza luego varios servicios externos para rastrear métricas relevantes para el impacto de sus recursos.

Algunas de las métricas utilizadas son específicas para el servicio que aloja el recurso. Por ejemplo, ImpactStory rastrea el número de veces que los videos en Vimeo y YouTube han sido vistos y los “me gusta” que han recibido, el número de veces que un repositorio de GitHub se ha bifurcado, y el número de veces que se han descargado los recursos del repositorio de datos Dryad, de Figshare, de las revistas de PLoS y de SlideShare. Otras métricas rastrean el interés en un recurso independientemente de donde esté alojado. El servicio puede consultar el número de citas en Scopus, el número de marcadores que aparecen en Mendeley y CiteUlike, y menciones en Facebook, Google+, Twitter, Wikipedia, y en publicaciones de blog. Las métricas se utilizan para elaborar informes sobre el interés mostrado en los productos del usuario, destacando los recursos más populares y proporcionando algunas estadísticas agregadas. Estos informes suelen ser enviados por correo electrónico al usuario con una frecuencia semanal.

ImpactStory funciona como una organización sin ánimo de lucro en Estados Unidos. Se desarrolló a partir de un proyecto denominado "total-impact", dentro del workshop “Beyond Impact” realizado en 2011. Desde 2012 ha recibido financiación de la Open Knowledge Foundation, la National Science Foundation, Jisc y la Fundación Sloan, y también está financiada por las inscripciones de los investigadores. En 2015 existía una suscripción anual que costaba 60 dólares y a partir del 15 de septiembre de 2016 habrá además, una suscripción mensual con un coste de 10 euros. Los datos recolectados por el servicio están disponibles de forma abierta, no restringida por terceros, y se pueden exportar ítems en concreto o perfiles enteros en cualquier momento. El código y la dirección del servicio también están abiertos. A través de un foro se pueden ofrecer ideas para un mejor funcionamiento del servicio, y los usuarios pueden votar por las ideas que les gustaría que los desarrolladores prioricen.

3.4.2. Métricas de PLoS a nivel de artículo

En 2009, la Public Library of Science (Biblioteca Pública de Ciencia) también conocida como PLoS lanzó su servicio de métricas a nivel de artículo (Article-Level Metrics). Esto compila un conjunto de indicadores de impacto de los propios sistemas de PLoS y otros servicios, y los pone a disposición tanto de forma visual y a través de una interfaz de programación de aplicaciones (API).

Las métricas recopiladas incluyen:

- Las estadísticas de uso (visualizaciones y descargas) de PLoS y PubMed Central;
- Las interacciones (comentarios, notas, ratings) en el sitio web de PLoS;
- Las citas identificadas por Scopus, Web of Science y otros;
- Las referencias que se hacen en las redes sociales como Twitter y Facebook, en varias plataformas de blogs, o en la Wikipedia.

Las métricas se muestran en las landings pages de los artículos de PLoS, y también pueden ser compiladas en informes personalizados.

PLoS liberó el código fuente de la aplicación de estas métricas en 2011. Fue utilizado como la base del (ahora suspendido) servicio ScienceCard, que proporcionaba una visión centrada en el autor de los datos. Mientras que las implementaciones del software hasta la fecha se han concentrado en los documentos, el software en sí es independiente del tipo de recurso, así que podrían ser aplicadas a datasets.

3.4.3. PlumX

PlumX es el principal producto de Plum Analytics, una empresa formada en 2011 y adquirida por EBSCO a principios de 2014. Su objetivo es ofrecer una visión más comprensible del impacto de la investigación no sólo en las citas, y para dar una idea del impacto que tienen los recursos en el período anterior a las primeras citas recibidas. El producto PlumX va más dirigido a las organizaciones que a los individuos, por lo que Plum Analytics tiene entre sus clientes a las universidades, las empresas, los editores y los financiadores, y los informes de rápido crecimiento desde su adquisición.

PlumX agrega información de una amplia gama de fuentes externas sobre el impacto de los productos de la investigación, incluyendo datasets y código fuente, así como las publicaciones más tradicionales. Las métricas se agrupan en cinco categorías:

- **Uso:** el número de veces que el recurso ha sido visto o descargado, el número de veces que se ha hecho clic en un enlace de Twitter o Facebook, el número de usuarios que contribuyen (en GitHub), el número de bibliotecas que guardan una copia;
- **Captura:** el número de veces que un recurso ha sido marcado como de interés (por ejemplo, marcado como favorito en Delicious, añadido a una librería de Mendeley, seguido, bifurcado o visto en GitHub);
- **Menciones:** el número de entradas de blog escritas sobre él, el número de observaciones formuladas al respecto (en Facebook, Slideshare, YouTube, etc.), el número de comentarios recibidos (en Amazon o Goodreads);
- **Social media:** el número de veces que el recurso ha sido promocionado (por ejemplo, por medio de “me gusta” en Facebook, “+ 1s” en Google+, votos en Reddit, tuits);
- **Citas:** el número de citas que el recurso ha recibido según Scopus, CrossRef, y otras fuentes.

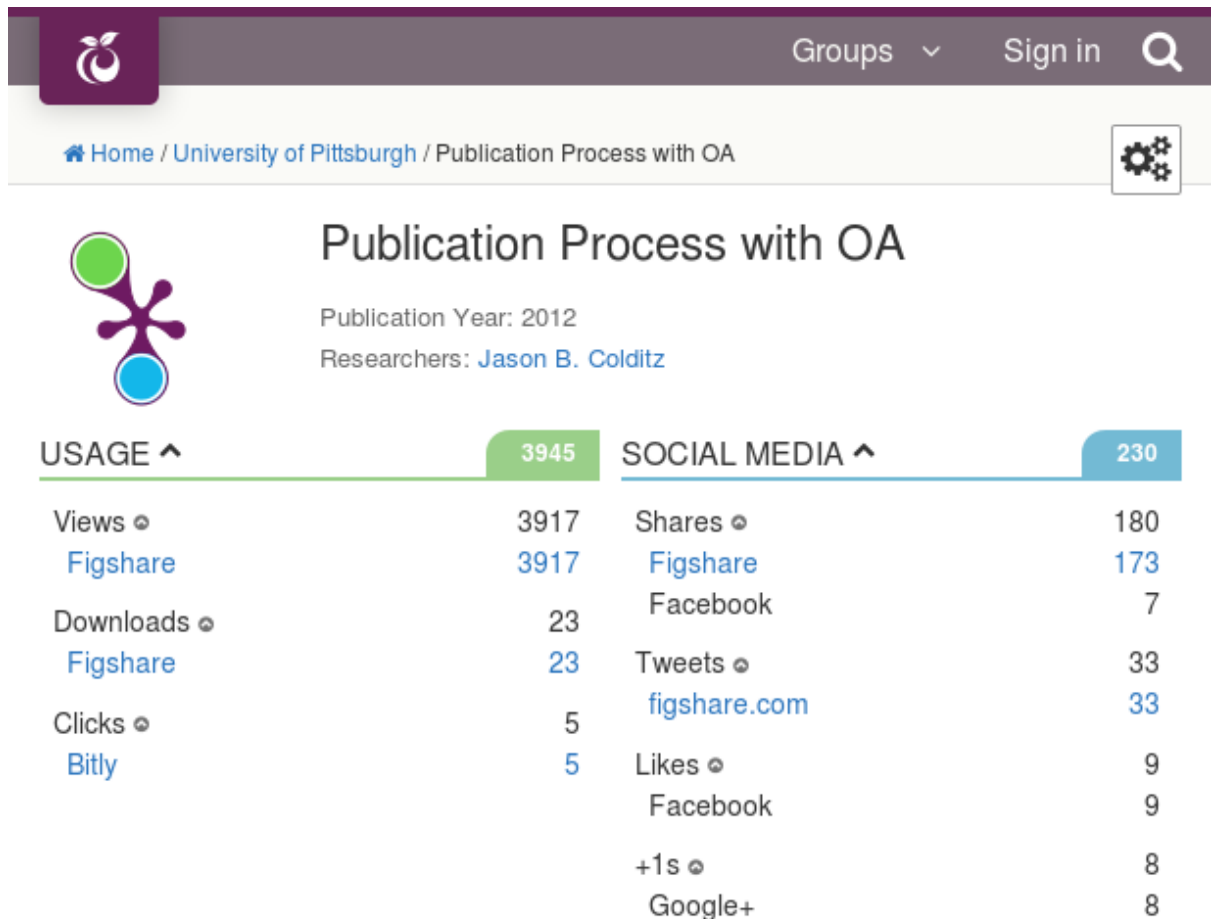
Los totales de estos indicadores se muestran en un cuadro de mando; gráficos de barras o diagramas circulares. También se puede incrustar en otros sitios web un resumen de Plum. La información que se muestra en el widget incrustado es personalizable y están disponibles los enlaces a la fuente original de datos.

Los investigadores pueden ayudar a sembrar la información disponible mediante la vinculación de sus perfiles PlumX a las cuentas que tienen en otros sistemas (por ejemplo, Slideshare, GitHub).

Ejemplo

Cuando Jason Colditz escribió una entrada de blog en una publicación de acceso abierto, él lo ilustró con un diagrama de flujo del proceso de publicación que había depositado en figshare. El perfil de PlumX de su institución rastreó el interés en la publicación como se muestra en la figura.

Figura 7: Difusión de la publicación de Jason Colditz según PlumX.



Fuente: How to track the impact of research data with metrics / Alex Ball y Monica Duke (2015a, p9).

3.4.4. Altmetric

Altmetric es un servicio centrado en el artículo que monitoriza diversas fuentes de menciones de artículos académicos. Estas fuentes son twitter, facebook, blogs, sitios de periódicos y documentos de políticas del gobierno. Sobre la base de esta información se calcula una puntuación dirigida a indicar la calidad y la cantidad de la atención recibida. La puntuación se calcula a partir del volumen de menciones, con cada mención ponderada de acuerdo a su fuente (los artículos de prensa tienen más peso que los tuits, por ejemplo), su autor y su público objetivo.

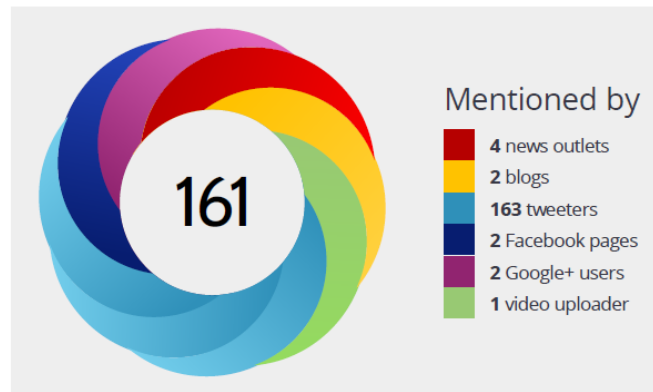
Los resultados de Altmetric se pueden consultar de cuatro maneras:

- Mediante un navegador web. Reúne toda la información recogida sobre todos los artículos. Los artículos y discusiones en torno a ellos se pueden explorar de forma individual con diferentes filtros. También están disponibles las alertas por correo electrónico, informes personalizados y la exportación de los datos. Pagando el precio de una institución se puede obtener la información de los artículos de esa institución.
- Mediante un bookmarklet. El bookmarklet es gratuito para los investigadores. Funciona dentro de un navegador y muestra altmetrics de artículos individuales mientras están siendo vistos. Actualmente está disponible para los artículos en PubMed y arXiv, y las páginas que contienen un DOI.
- Mediante la API. La API requiere una licencia comercial (aunque para las investigaciones no comerciales, apps y mash-ups existe una versión gratuita pero limitada). La API se basa en el protocolo HTTP y da acceso programado a los datos acerca de artículos y datasets recogidos por Altmetric.
- Mediante una ficha. La ficha es una visualización en forma de donut que muestra la puntuación altmetric en el centro y los bordes con códigos de color de acuerdo a la importancia de las menciones. Cada enlace a una landing page muestra el desglose de la puntuación de ese artículo. Se proporciona el código HTML para incrustar fichas en determinadas páginas web, por lo que los investigadores pueden agregarlas a sus páginas web personales o institucionales y pueden integrarlas en plataformas como los repositorios. Un uso importante lo realiza el PLoS Impact Explorer, que utiliza la puntuación altmetric para clasificar los artículos publicados por PLoS. La vista se puede configurar usando diferentes períodos de tiempo y subconjuntos de artículos. Para cada artículo se muestra la ficha de altmetric y también una selección de comentarios extraídos de social media.

Ejemplo

Las imágenes de un dinosaurio gigante 'Dreadoughtus schrani' fueron depositadas en figshare. Las diferentes menciones en blogs, tuits, noticias y otras fuentes fueron rastreadas con Altmetric, que colocó el artículo en el 5% superior de todos los artículos evaluados por ellos.

Figura 8: Menciones de un determinado artículo



Fuente: How to track the impact of research data with metrics / Alex Ball y Monica Duke (2015a, p10).

3.4.5. ResearchGate

ResearchGate es una red social para investigadores donde se anima a los usuarios a que añadan sus publicaciones a su perfil. Por su parte, el sitio web genera estadísticas sobre visitas, descargas y citas. Estas estadísticas se presentan al usuario mediante un cuadro de mando en su perfil del sitio web y se envían por correo electrónico de manera regular. Las estadísticas también contribuyen a las métricas de impacto del propio servicio, la puntuación de ResearchGate, junto con la actividad del investigador en los foros de ResearchGate y el número de usuarios que siguen su perfil.

Los investigadores pueden compartir los datos junto con las publicaciones. Además el sitio web cuenta con una herramienta de revisión denominada OpenReview. Eso permite al usuario publicar una revisión de cualquier documento que han leído, de acuerdo a una estructura que se centra en si los resultados del documento son reproducibles. Se pueden comparar muchas revisiones del mismo documento mediante una página de resumen. En el siguiente enlace vemos el análisis de un artículo de la revista Nature donde se valoran aspectos como la metodología empleada o las conclusiones alcanzadas: <http://tinyurl.com/ouc2orv>. Los usuarios también pueden solicitar una revisión de sus documentos.

3.4.6. Google Scholar

Google Académico es una versión del motor de búsqueda de Google dirigida a los trabajos académicos. Entre las facilidades que ofrece está la enumeración de las citas que cada obra ha recibido en las otras bases de datos. Esta información se genera de forma automática y por lo tanto es un poco aproximada, pero los investigadores pueden obtener un valor añadido del servicio mediante la configuración de un perfil y añadiendo sus trabajos académicos. Una vez hecho eso, se crea una página de perfil que muestra los trabajos junto a su respectivo número de citas. Además, la página ofrece otros tres datos: el recuento total de citas que las obras han

recibido, un índice h , y un índice i_{10} , siendo este último el número de obras que recibieron al menos 10 citas. Las tres medidas se proporcionan dos veces: la primera teniendo en cuenta todas las citas y la segunda contando sólo las citas de los últimos cinco años. Se proporciona un gráfico de barras de los recuentos anuales de citas.

Los usuarios que disponen de un perfil son capaces de mejorar la precisión del servicio al revisar las entradas que han elaborado (por ejemplo, garantizando que no hay duplicados o entradas que se fusionaron incorrectamente, corrigiendo la información bibliográfica) y añadiendo otros que Google puede haber pasado por alto.

3.4.7. Microsoft Academic Search

Microsoft Research ofrece un servicio similar llamado Microsoft Academic Search. También proporciona el número de citas de las obras de su base de datos, pero a diferencia de Google Scholar, crea páginas de perfil para todos los autores que descubre. Cada página de perfil, así como una lista de productos del autor, proporciona un recuento total de citas y un gráfico de líneas del número de citas por año. Los usuarios que se han registrado pueden corregir los errores que encuentran en los datos: no están restringidos a la edición de los datos relativos a ellos mismos o de sus publicaciones. Cualquier conflicto entre las correcciones se resuelve por Microsoft.

3.4.8. Otros servicios

PaperCritic es un sistema de revisión por pares abierto construido por Mendeley, el gestor de referencia y red social. Permite a los usuarios de Mendeley añadir revisiones de publicaciones en su propia colección de referencia o en la de los demás. También pueden evaluar las referencias de la publicación, la originalidad, la argumentación y la dificultad, e indicar si lo recomendarían a otros. Si bien es interesante como modelo para la revisión por pares posterior a la publicación, se desconoce si los datasets fueron revisados utilizando el servicio.

Crowdometer era un proyecto que usó el crowdsourcing para analizar y clasificar los tuits que contenían enlaces a documentos académicos. Los participantes utilizaron un subconjunto de la Citation Typing Ontology (CITO) para describir los tipos de enlaces realizados en el tuit. Un hallazgo importante fue que la mayoría de los tuits no contenían un sentimiento sobre el enlace que se compartía. El proyecto se llevó a cabo entre diciembre de 2011 y enero de 2012; mientras que su sitio web ya no está disponible, el código utilizado está disponible en GitHub.

3.5. Requisitos de la citación de datos

En el artículo “How to cite datasets and link to publications” se presenta un conjunto de elementos que deberían estar presentes en la cita bibliográfica de un dataset (Ball y Duke, 2015b).

Autor. El creador del dataset.

Fecha de publicación: puede ser la fecha en que el dataset se puso a disposición del público, la fecha en que terminaron todos los procedimientos para asegurar la calidad de los datos o la fecha en que terminó el periodo de embargo.

Título. Es el nombre del dataset citado, que puede incluir el nombre de una organización. Si un dataset es una subparte de otro dataset más grande también se pueden incluir los dos títulos.

Edición. El nivel o etapa de tratamiento del dataset, lo que indica si los datos son brutos o refinados.

Versión. El número de versión aumenta cuando los datos cambian.

URI. Una URI es utilizada para marcar un dataset.

Tipo de recurso. Ejemplos: "Base de datos", "dataset".

Editorial. La organización, ya sea alojando los datos o asegurando su calidad.

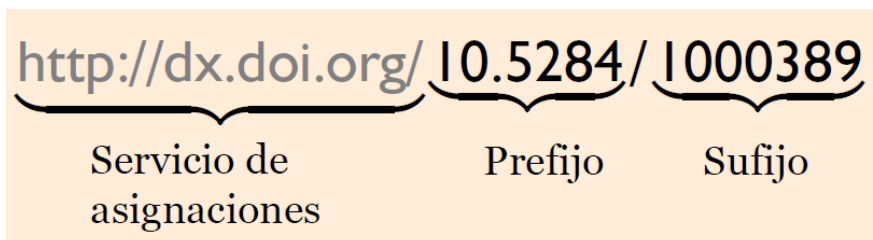
Huella digital única y numérica (Unique numeric fingerprint, UNF). Un sistema que se utiliza para asegurar que no ha habido cambios en el dataset desde el momento de realizar la cita.

Identificador. Un identificador para los datos, de acuerdo a un esquema persistente.

Ubicación. Una URL persistente desde la cual está disponible el dataset.

De los elementos anteriores, los más importantes son: el autor, el título, la fecha, la ubicación, y el editor. Estos dan el debido crédito, permiten al lector juzgar la pertinencia del dataset, permiten el acceso al mismo, y dan garantías de la calidad o la persistencia. En teoría, deben identificar de forma única el dataset; en la práctica, es necesario a menudo un identificador formal. La solución más eficiente es dar una ubicación que consiste en una asignación y un identificador. Véase la siguiente figura:

Figura 9: Anatomía de un DOI



Fuente: *How to cite datasets and link to publications* / Alex Ball y Monica Duke (2015b, p6).

Los identificadores son una parte muy importante en el desarrollo de las métricas. Si se refieren a un artículo disponemos del DOI de Crossref, para un dataset contamos con un DOI de DataCite y para un individuo podemos buscar su ORCID. Se recomienda prestar atención en la identificación de un subconjunto específico o componentes de un dataset. Tanto DataCite como CrossRef han establecido métodos para identificar partes en los documentos identificados por un único DOI, por ejemplo, partes de datos dentro de un único dataset publicado (NISO RP-25-201x-2A).

En el caso de los artículos académicos contamos con la norma COUNTER, la cual se usa para averiguar el número de veces que se ha accedido a la ficha de un artículo o se ha descargado. Desgraciadamente no existe una versión de esta norma para definir el impacto de los datasets. Muchos aspectos de la norma COUNTER podrían aplicarse también a los datos de investigación, como se hizo en el proyecto Making Data Count con alrededor de 150.000 datasets de investigación (NISO RP-25-201x-2A).

La norma COUNTER ignora los datos derivados de fuentes no humanas, ya sean piezas de software que recogen contenido de forma automática, o motores de búsqueda. Como los elementos no humanos son cada vez más frecuentes en la reutilización de datos científicos, esto ha de ser tenido para las estadísticas de uso de datos, debiendo ir un paso más allá del código COUNTER. Por ejemplo, un investigador podría desarrollar una aplicación que descargue automáticamente una copia nueva de un dataset. Es importante tener en cuenta esta forma de acceso y que no se descarte automáticamente. Además, como las API se vuelven cada vez más comunes, el acceso a los datasets de investigación puede llegar a ser más fragmentado y frecuente con el tiempo, y la métrica debe ser desarrollada para reconocer esas tendencias (NISO RP-25-201x-2A).

3.6. Creación de la ficha para el análisis

Vamos a describir las métricas utilizadas para la realización del análisis práctico del punto 4. Nos hemos basado en el artículo “How To Track the Impact of Research Data with Metrics” (Ball y Duke, 2015a):

3.6.1. Citación de datos

La citación de datos es la práctica de proporcionar una referencia formal y estructurada a un conjunto de datos en un trabajo académico, es decir, realizar una referencia bibliográfica completa de un dataset (NISO RP-25-201x-2A, 2016).

El modelo emergente más maduro para medir el impacto de los datos es paralelo a la publicación y citación de literatura. Ir más allá del mero intercambio de datos, donde los datos simplemente se ponen a disposición del público (por ejemplo, como archivos en un sitio web),

la publicación de datos implica que los datos han entrado en un marco para el control de su calidad, asegurando que son aptos para su reutilización, que se pueden buscar, encontrar, y se garantiza su accesibilidad a largo plazo. El dataset resultante ofrece la información bibliográfica suficiente para ser citado de forma fiable por otros artículos académicos. Estas citas se pueden contar de manera habitual para proporcionar evidencia del impacto del dataset.

Mientras la práctica de citar datasets está lejos de ser generalizada, y puede ser considerada como un objetivo alcanzable, algunas disciplinas están adoptando un enfoque de transición, donde las citas se colocan en un documento de datos. Es un documento que describe el dataset y su colección sin sacar conclusiones científicas de él. Dichos documentos podrán ser publicados en una sección especial de una revista, o en una revista dedicada a los datos, como la *Journal of Open Archaeology Data*. Las citas de documentos de datos pueden ser interpretadas como citas del dataset con el propósito de mostrar la evidencia del impacto.

En muchas disciplinas, sin embargo, el enfoque habitual es citar el primer documento para hacer uso de los datos, confiando en este documento para indicar si se han compartido los datos y cómo. Normalmente no es posible, al menos no sin el esfuerzo manual, para identificar si las citas a dichos trabajos deberán contar para el impacto de la argumentación y las conclusiones del documento, el dataset generado, o ambos. En tales disciplinas, por tanto, el número de citas es de poca ayuda como indicador del impacto de los datos, por lo que se deben utilizar los indicadores alternativos.

3.6.2. Asignaciones

Se utilizan esquemas para identificar de forma unívoca cada dataset. En otras palabras, hay servicios puente que unen los identificadores con una o más ubicaciones de Internet. Uno de esos esquemas es el Digital Object Identifier (DOI), para lo cual DataCite es la principal Agencia de Registro de datasets de investigación. Los propietarios de cuentas tienen acceso a la resolución de los datos para todo los DOI que gestionan. Estas estadísticas dan una idea de con qué frecuencia se han seguido las referencias al dataset.

3.6.3. Visualizaciones de página

Los servidores web registran cada interacción que tienen con un cliente, por lo que analizando los registros es posible contar de forma aproximada las veces que una página web ha sido abierta por un navegador. Algunos sitios además, pueden integrar en sus páginas código de JavaScript que notifica una solicitud de análisis cada vez que una página ha sido vista. De cualquier manera, esta estadística puede utilizarse para entender el nivel de interés en esa página.

Cuando los datasets están disponibles en línea, la mejor práctica es proporcionar una página web correspondiente mostrando un registro de catálogo para el dataset. Como mínimo se podría esperar que la página muestre el título del dataset, una mención de responsabilidad, una breve descripción y un enlace de descarga (o instrucciones sobre cómo acceder). El

número de veces que se ha visto la página del dataset da una indicación del nivel de interés en el dataset, y denota si el dataset es conocido o no.

3.6.4. Descargas

Los registros del servidor web también se pueden usar para contar el número de veces que un archivo de datos ha sido descargado. Esto indica un mayor nivel de interés en los datos que lo que se desprende de un recuento de páginas vistas, ya que implica un deseo de mirar los datos reales, pero la estadística por sí sola no revela el uso que podrían tener los datos descargados.

3.6.5. Enlaces a social media

Tal vez las métricas alternativas más cercanas a las referencias en artículos de revista son las que miden la actualidad del dataset en plataformas de social media. Si las personas empiezan a compartir o discutir un dataset con amigos, compañeros y el resto del mundo, hay una probabilidad de que les haya afectado de alguna manera, lo que significa que vale la pena investigar para ver la evidencia del impacto.

Twitter es una red social que permite a los usuarios enviar mensajes cortos conocidos como "tuits" a sus seguidores. Como se limitan a 140 caracteres, los tuits se prestan a la reacción inmediata y a breves sentimientos. Un tuit relacionado con una parte de una investigación puede contener un enlace, es decir, por ejemplo, un producto de una investigación, un sitio web del proyecto o una entrada de un blog que habla de él, acompañado de un comentario sobre el mismo. Detectar tuits relacionados con un dataset puede ser complicado, pero una posible estrategia de búsqueda es buscar por las menciones al identificador del dataset o enlaces a su página del catálogo. Una vez que se ha encontrado un tuit, para tener una imagen precisa del impacto que está teniendo el dataset es útil considerar si el tono del tuit es positivo, negativo o neutral, así como analizar citas de forma tradicional. También puede ser informativo rastrear cualquier conversación posterior- respuestas o comentarios ('retuits') -, ya que pueden indicar hasta qué punto otros están de acuerdo con los sentimientos del tuit inicial. Cabe señalar que existe una dimensión cultural por tuitear; por ejemplo, en algunas comunidades no se considera una actividad profesional para los académicos.

Los marcadores sociales y los servicios bibliográficos no se utilizan tan ampliamente como Twitter, pero son mucho menos "ruidosos" como fuente de la evidencia del interés en los productos académicos. Servicios como Mendeley, CiteUlike, Bibsonomy y Delicious permiten a los usuarios grabar los recursos en línea para su propia referencia o recomendar a otras personas. Si bien la funcionalidad ofrecida por cada servicio varía, normalmente se puede descubrir cuántos usuarios han marcado como favorito un enlace o un recurso en particular. Algunos, sobre todo Digg y Reddit, también proporcionan información sobre cuántos usuarios han votado a favor o en contra del recurso; esto puede dar una impresión de si el recurso está teniendo un impacto positivo o no.

Varias plataformas de blogs utilizan un sistema de enlaces traseros para rastrear conversaciones entre blogs. Cuando una entrada nueva escrita en el blog A trata una antigua entrada del blog B, el blog A envía una notificación al blog B. El blog B puede utilizar esta información para incorporar un extracto de la nueva entrada como un comentario sobre la antigua.

3.6.6. Revisión por pares posterior a la publicación

Una innovación propuesta en el campo de las comunicaciones académicas es el uso de la revisión por pares después de la publicación como método de control de calidad. Si bien todavía se tiene que establecer como una práctica generalizada, hay varios lugares donde dichas revisiones se pueden encontrar, como la Faculty of 1000 (tanto como parte integral de sus propias publicaciones y como un servicio de revisión de otra literatura) y PubPeer. Si bien el énfasis hasta la fecha ha sido en la revisión de los documentos de revista, se están dando pasos para aplicar el principio también a los datasets. A los documentos de datos presentados a la revista de datos Earth System Science Data, por ejemplo, se les da una breve revisión interna antes de ser publicados en su compañía Earth System Science Data Discussions, donde cualquiera es capaz de enviar una revisión. Sólo una vez que el documento ha superado satisfactoriamente esta fase pública de revisión se puede proceder a su publicación en la revista principal.

Aunque la revisión por pares posterior a la publicación se preocupa por la calidad en lugar del impacto, el texto o la naturaleza de esas revisiones pueden revelar que ha habido reutilización. Hay propuestas de que los archivos de datos pueden invitar a aquellos que han descargado y reutilizado sus datos a dejar respuesta en la landing page del dataset. Proporcionaría una fuente escalable de revisión por pares y puntos de vista en lo que hace reutilizable a los datos, sino que también proporcionaría una confirmación de que otros investigadores habían intentado utilizar los datos y, en caso de reutilización exitosa, que los datos han tenido un impacto.

4. Metodología

Para comprobar las métricas que se siguen en el estudio del uso de los datos abiertos se han analizado los resultados de los repositorios de datos abiertos de investigación. Para realizar el análisis, se ha hecho primero una revisión bibliográfica de las distintas métricas hasta el momento utilizadas por distintos teóricos para seleccionar los indicadores que se consideran más importantes. Posteriormente se ha hecho una selección de los repositorios a partir de la lista de repositorios indexados en el Data Citation Index de la Web of Science. En principio se iban a seleccionar únicamente los repositorios del área de conocimiento de ciencias sociales, pero el número de repositorios se reducía claramente, por lo que se decidió analizar todos los repositorios encontrados, dejando a un lado su disciplina académica.

4.1. Discusión

Hasta la fecha no existe una métrica plenamente establecida para el uso de los datos. La principal y más usada por parte de los distintos teóricos en la parte de los datos abiertos de investigación se basa en la cita de los datos que van asociados principalmente a una publicación.

Otros indicadores de tipo altmetrics son los que se están potenciando de acuerdo a la NISO RP-25-201x-2A.

Con todos estos elementos se ha compuesto la ficha de análisis basada en los siguientes indicadores:

- Área de conocimiento
- Cita bibliográfica
- Identificador
- Visualizaciones
- Descargas
- Difusión en redes sociales
- Difusión en gestores de referencias bibliográficas
- Veces difundido

Por lo que respecta a las métricas de los datos abiertos de la administración, quizás el modelo más desarrollado es el de Meloda (Abella, Ortiz-de Urbina Criado y De Pablos Heredero, 2014) en ella los indicadores principales son los de: obertura de los datos, facilidad o dificultad en el acceso a los datos, permisos para la reutilización y los modelos de datos. El acceso a este tipo de información en los portales de la administración, tal y cómo se ha señalado anteriormente no se ha realizado dadas las limitaciones del estudio.

4.2. Resultados

En la descarga de los principales repositorios del Data Citation Index, contábamos con una lista de 73 repositorios. Son los que aparecen en el anexo I. Siete de ellos fueron rechazados para el análisis; la mayoría por estar fuera de funcionamiento, presentan muchas dificultades para su manejo o están obsoletos. Los repositorios ignorados son los que aparecen señalados con un asterisco en el anexo. En el segundo anexo aparece la lista definitiva de repositorios analizados. Son 66.

Al principio contamos con una serie de 73 repositorios de datos. Se pueden encontrar en el anexo 1. Al confeccionar la lista definitiva, ocho de esos repositorios han sido ignorados, la mayoría por estar fuera de funcionamiento. En el anexo 2 se encuentran los 66 repositorios analizados que conforman la lista final. En esta segunda lista se han incluido repositorios de investigación que contengan datos, imágenes o textos.

El análisis de los repositorios en función del área de conocimiento seleccionada y de acuerdo a la ficha de análisis, se puede señalar lo que a continuación se desarrolla.

4.2.1. Repositorios por área científica

Si clasificamos los repositorios por área científica vemos que la categoría más estudiada es la biología que cuenta con catorce repositorios. Después nos encontramos con la categoría de "Ciencias sociales" en la que se encuentran nueve repositorios dedicados a varias ramas de las ciencias sociales. Como tercera categoría está la denominada "General" que está compuesta de ocho repositorios que recogen información de una gran cantidad de disciplinas, ya sean de ciencias sociales o científicas. También hay que destacar la gran diferencia que existe entre la biología (14) y otras categorías como medio ambiente (con 6 repositorios), geología (5) y química (4). Podemos ver que las ciencias sociales están menos representadas que las disciplinas científicas.

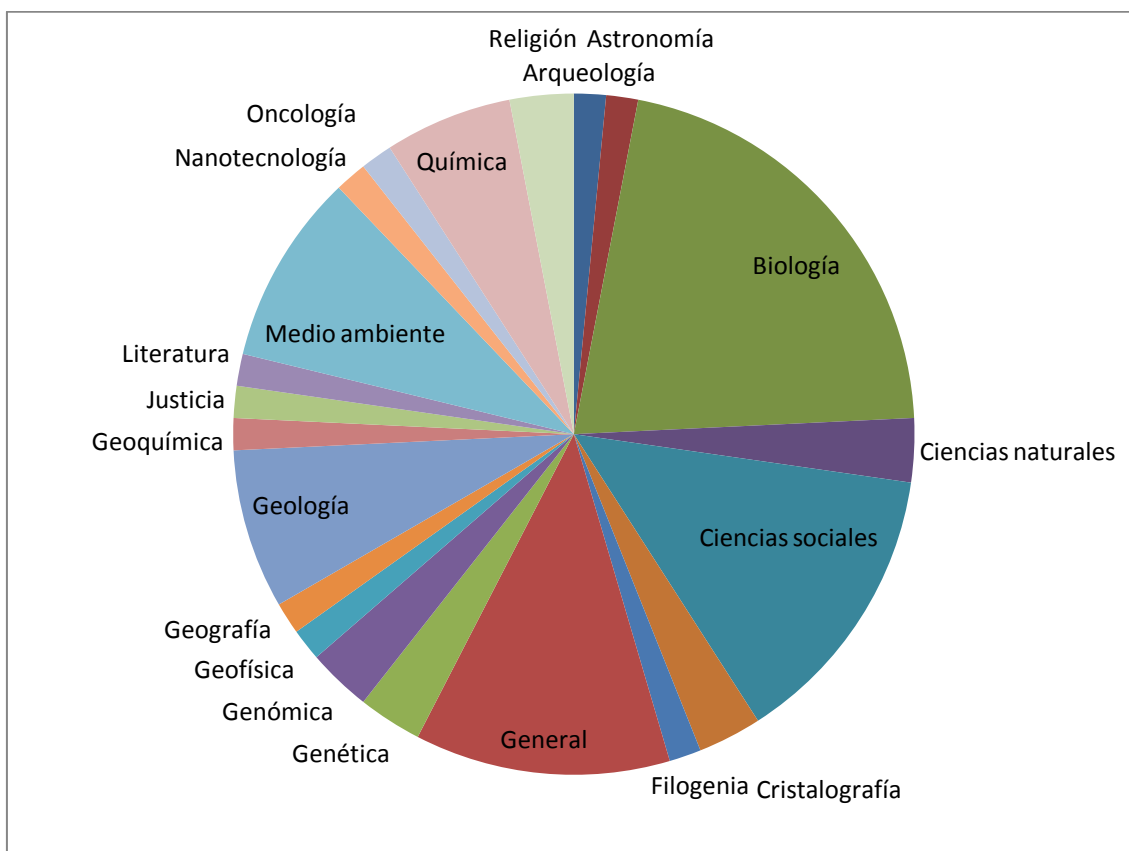
Tabla 3. División de los repositorios por área científica

Área de conocimiento	Nº repositorios
Arqueología	1
Astronomía	1
Biología	14
Ciencias naturales	2

Ciencias sociales	9
Cristalografía	2
Filogenia	1
General	8
Genética	2
Genómica	2
Geofísica	1
Geografía	1
Geología	5
Geoquímica	1
Justicia	1
Literatura	1
Medio ambiente	6
Nanotecnología	1
Oncología	1
Química	4
Religión	2

Fuente: Elaboración propia

Figura 10: Clasificación de los repositorios por área científica



Fuente: Elaboración propia

4.2.2. Repositorios que ofrecen referencia bibliográfica del dataset

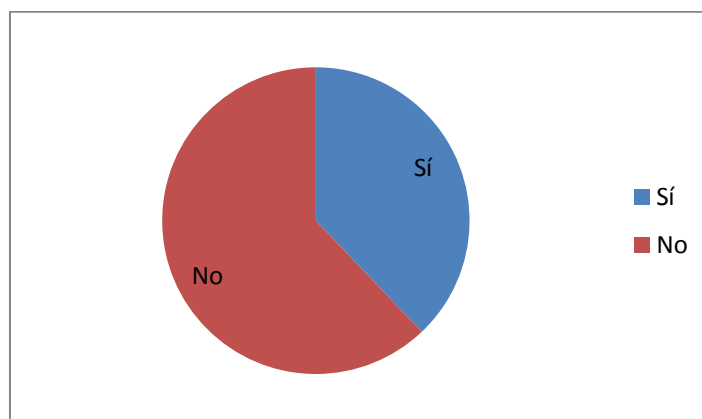
En este apartado vemos que los repositorios que no ofrecen una referencia bibliográfica de un dataset (41) son mayoría respecto de los que sí la ofrecen (25). Esto conlleva numerosos problemas a la hora de citar los datos de manera homogénea y que posteriormente exista un conteo real de las citas de cada uno de los datasets de manera individual.

Tabla 4. Repositorios que ofrecen referencia bibliográfica del dataset

Ofrecen cita del dataset	25
No ofrecen cita del dataset	41

Fuente: Elaboración propia

Figura 11. Repositorios que ofrecen referencia bibliográfica del dataset



Fuente: Elaboración propia

4.2.3. Identificador

En la figura número doce podemos observar que la fórmula más habitual para identificar un dataset es asignarle solamente una ID (23). Con el concepto de ID nos referimos a un código (numérico o alfanumérico) asignado por el propio repositorio para identificar de forma unívoca cada dataset. Llama la atención la diferencia que existe entre los repositorios que asignan un DOI al conjunto de datos (20+2+1) y los que ofrecen un handle (2+2). Entre ambos nos encontramos los repositorios que asignan una dirección web para localizar cada dataset.

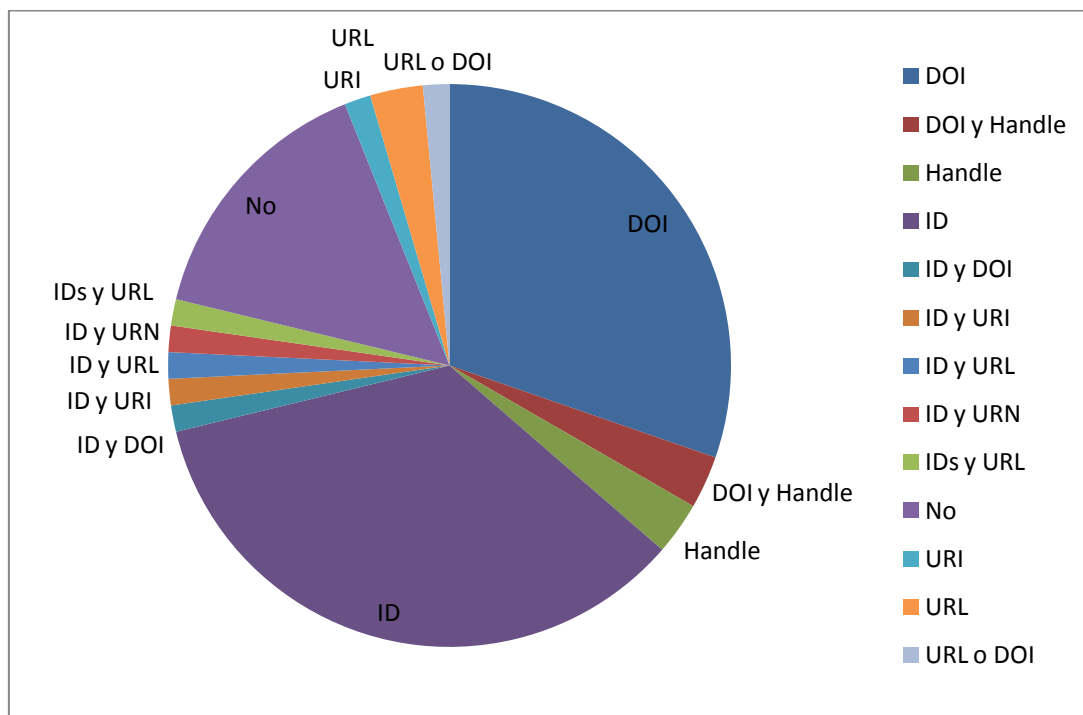
Tabla 5. Análisis del identificador que contiene el dataset.

Tipo de identificador	Número de repositorios
DOI	20
DOI y Handle	2
Handle	2
ID	23
ID y DOI	1
ID y URI	1
ID y URL	1
ID y URN	1

IDs y URL	1
No	10
URI	1
URL	2
URL o DOI	1

Fuente: Elaboración propia

Figura 12. Análisis del identificador que contiene el dataset.



Fuente: Elaboración propia

4.2.4. Visualizaciones de página

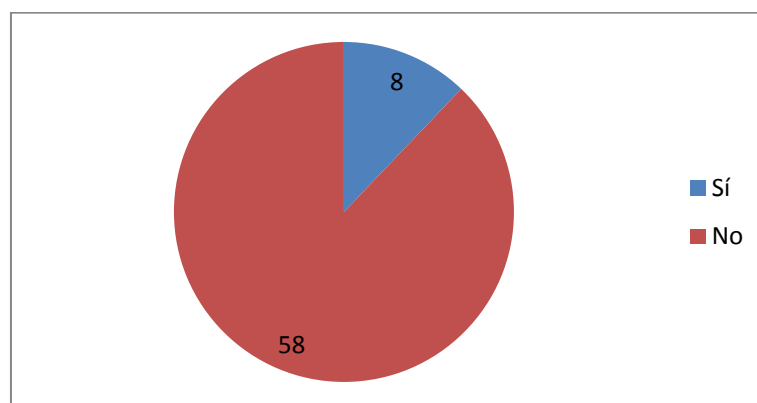
La gran mayoría de repositorios (58) no nos informan del número de veces que se ha visitado la página web con la descripción del dataset frente a ocho que sí lo hacen. Se conoce que la visita de un dataset no implica su uso, pero sí que es sintomático del interés que despierta entre la comunidad científica. Este es un dato que aunque no sea significativo para su uso, sí que es significativo para conocer los intereses de los investigadores en determinado tipo de información.

Tabla 6. Visualizaciones de página

Sí dispone de datos de visualización:	8
No dispone de datos de visualización	58

Fuente: Elaboración propia

Figura 13. Visualizaciones de página



Fuente: Elaboración propia

La gran mayoría de repositorios (58) no nos informan del número de veces que se ha visitado la página web con la descripción del dataset frente a ocho que sí lo hacen.

4.2.5. Descargas

Como en el apartado anterior la cantidad de repositorios que no ofrecen el número de descargas de los datos (60) son mayoría respecto de los que sí lo hacen (6). Este dato sí que es significativo en cuanto al posible uso que se hace de los datasets, el que no se ofrezca este tipo de información impide un análisis de los datos que interesan a los investigadores para su uso. El disponer de esta información podría motivar a otros investigadores a que aportaran sus datos al repositorio ya que tienen un interés real en la comunidad científica.

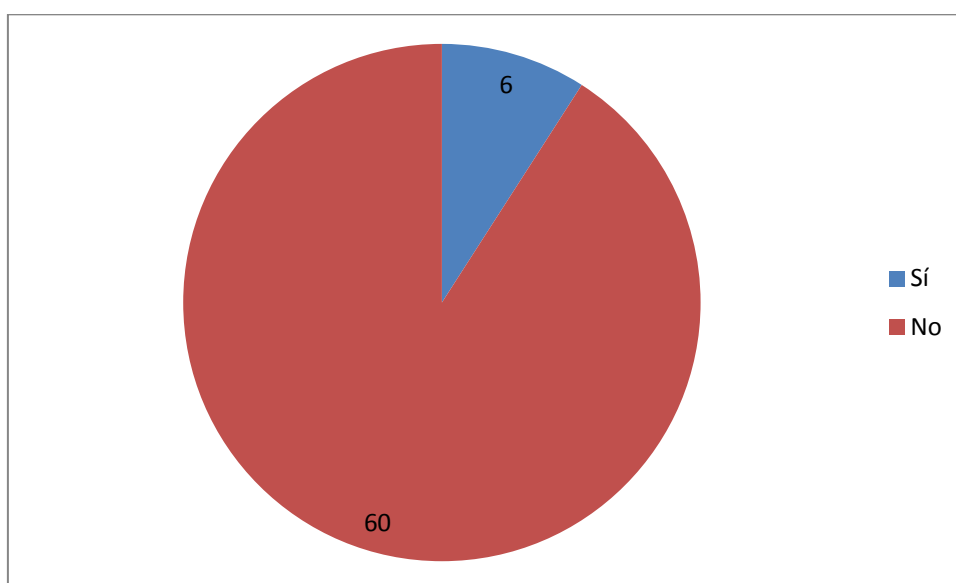
No ofrecer esta información lo consideramos una carencia importante por parte de la información facilitada por el repositorio.

Tabla 7. Descargas

Sí ofrece información de las descargas	6
No ofrece información de las descargas	60

Fuente: Elaboración propia

Figura 14. Descargas



Fuente: Elaboración propia

4.2.6. Difusión en redes sociales

Al igual que en los dos apartados anteriores, los repositorios que no ofrecen un mecanismo rápido para compartir los datos en redes sociales como facebook o twitter son mayoría. En este caso el número de repositorios que ofrecen esa posibilidad son 13 y 53 los que no lo hacen.

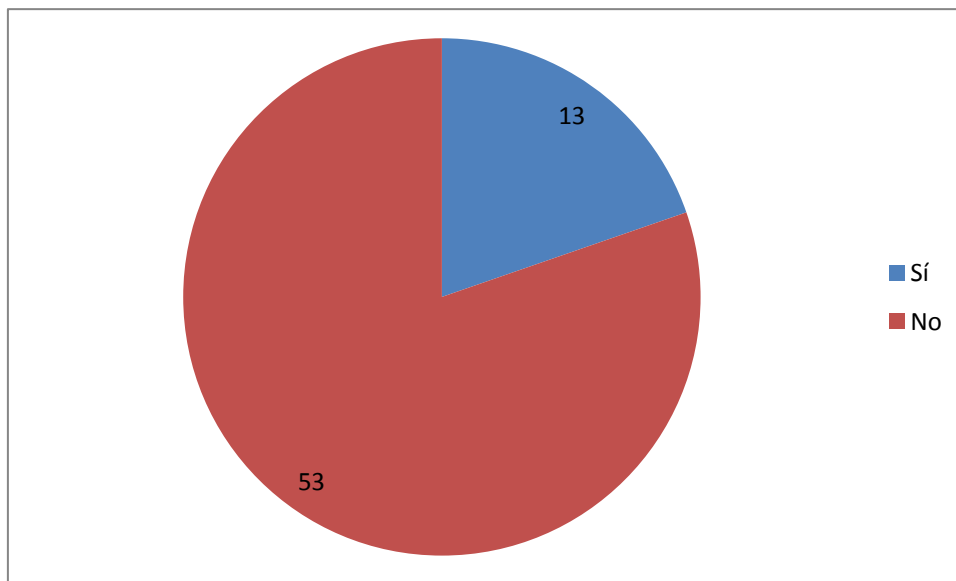
Esta circunstancia puede deberse a distintos motivos, uno de ellos es que en algunas áreas de conocimiento no se encuentra extendido el uso de este tipo de difusión en sus áreas de conocimiento, por lo que no consideran las redes sociales como posibles herramientas de difusión de sus estudios. Aunque este sea un motivo importante, esto no es un excusa para que sean tan pocos los repositorios que faciliten su difusión en las mismas dada la implantación que se está produciendo en el contexto científico.

Tabla 8. Difusión en redes sociales

Sí facilita la difusión en redes sociales	13
No facilita la difusión en redes sociales	53

Fuente: Elaboración propia

Figura 15. Difusión en redes sociales



Fuente: Elaboración propia

4.2.7. Número de veces que se difunde el dataset

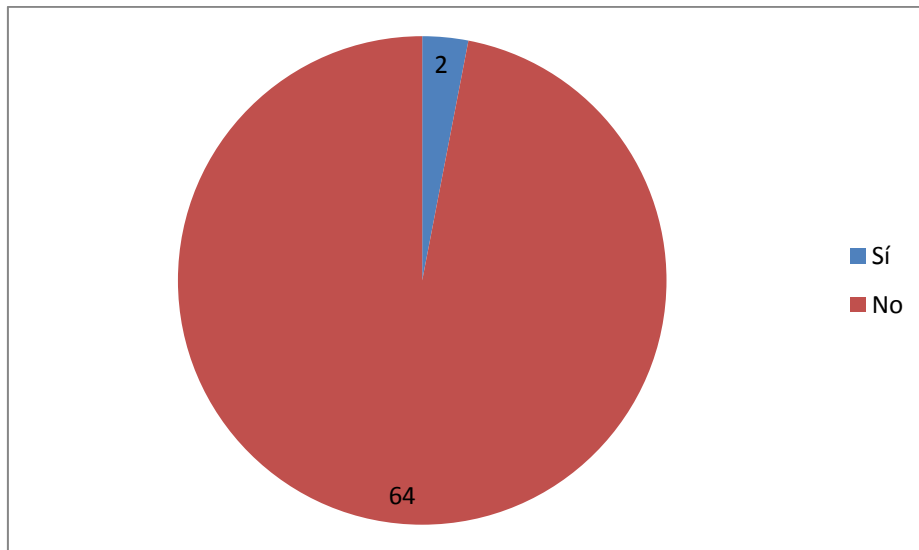
Este sería uno de los elementos que casi ningún repositorio lo cuenta. Únicamente dos repositorios muestran el número de veces que un dataset fue difundido o comentado en redes sociales.

Tabla 9. Veces difundido

Sí ofrece el número de veces que un dataset se difunde en las redes sociales	2
No ofrece el número de veces que un dataset se difunde en las redes sociales	64

Fuente: Elaboración propia

Figura 16. Veces que un dataset se comparte en redes sociales



Fuente: Elaboración propia

4.2.8. Difusión en gestores bibliográficos

En la figura 12 vemos que la cantidad de repositorios que no facilitan la agregación de la referencia bibliográfica a un gestor (55) es cinco veces más que el número de repositorios que lo hace (11).

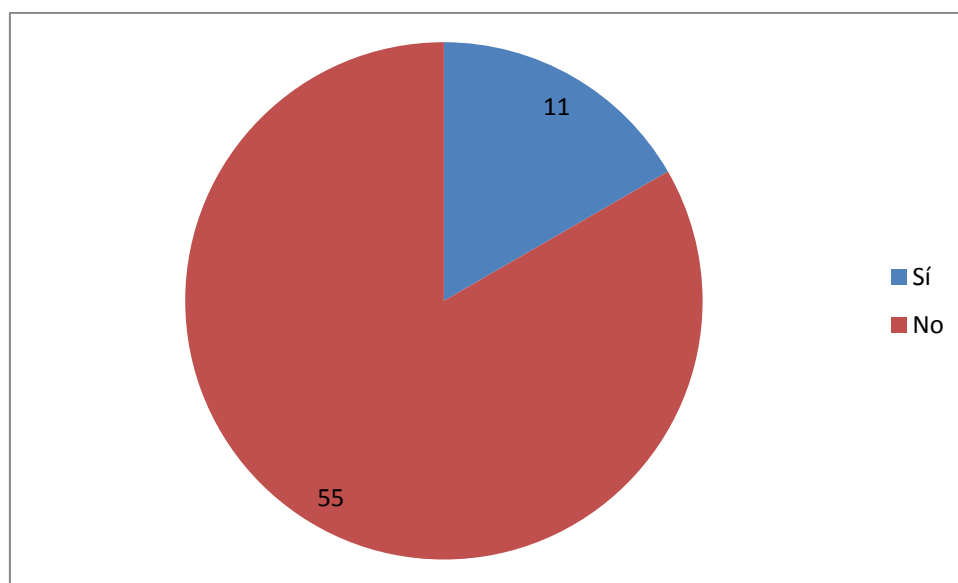
El hecho de que faciliten la incorporación de la cita a cualquier gestor de referencias bibliográficas es básico para que la cita sea homogénea y facilite al investigador el archivo del dataset para su posterior cita en su uso.

Tabla 10. Difusión en gestores bibliográficos

Sí facilita la difusión en gestores bibliográficos	11
No facilita la difusión en gestores bibliográficos	55

Fuente: Elaboración propia

Figura 17. Difusión en gestores bibliográficos



Fuente: Elaboración propia

En el anexo 3 se pueden ver las capturas de los repositorios analizados. Cada tabla contiene las imágenes de un repositorio diferente.

5. Conclusiones

Hemos realizado un trabajo relacionado con los datos abiertos. La primera parte ha consistido en realizar un estudio del arte sobre los diferentes tipos de datos abiertos que existen, bien sean del gobierno, de centros de investigación como las universidades y datos abiertos de las empresas. También hemos incidido en otros aspectos como la accesibilidad y formato de los datos además de analizar algunas plataformas que promueven la apertura de los datos. Luego hemos estudiado las diferentes métricas utilizadas a lo largo del tiempo para medir la calidad de la investigación científica. La primera métrica que se utilizó fue el factor de impacto de las revistas, mas tarde apareció el índice h para evaluar a los científicos o a las organizaciones. Estos dos índices tenían un problema común. Ambos utilizaban las citas recibidas por los artículos como único parámetro para evaluar la calidad de las publicaciones. También hemos analizado otros problemas de estos dos índices provocados por las diferencias existentes entre los campos científicos, lo que impedía valorar correctamente revistas o investigadores de diferentes disciplinas. Recientemente ha aparecido un nuevo tipo de métricas denominadas “altmetrics”, “métricas alternativas” o “métricas a nivel de artículo”.

En el aspecto práctico hemos realizado un análisis de una serie de repositorios del Data Citation Index. La idea inicial fue seleccionar aquellos repositorios relacionados con las ciencias sociales. El problema es que si restringíamos el estudio a las ciencias sociales el número de repositorios para analizar se reducía considerablemente. Por esta razón, nos quedamos con la lista de repositorios completa. El estudio ha consistido en analizar los diferentes indicadores que ofrecen esos repositorios para medir el uso que se hace de los datos que contienen. Algunos ejemplos son la referencia bibliográfica, el número de visualizaciones de la ficha del dataset o la cantidad de veces que se han descargado los datos.

La conclusión general de este análisis es la gran cantidad de trabajo que tienen por delante los repositorios. La referencia bibliográfica del dataset es ofrecida por 25 repositorios mientras que 41 no lo hacen. Es decir que ni siquiera un 40% de los repositorios analizados muestra esa información. Si nos fijamos en otros aspectos como las visualizaciones de página o las descargas el número de repositorios que ofrecen esos datos se reduce en gran manera. El dato de las visualizaciones sólo lo ofrecen 8 de los 66 repositorios analizados (un 12%) mientras que únicamente 6 repositorios (un 9%) nos informan de la cantidad de veces que se descarga un dataset.

En el apartado de la difusión, la cantidad de repositorios que nos ofrecen la posibilidad de compartir nuestros datasets en alguna red social (facebook, twitter, linkedin, google+) o en algún gestor bibliográfico (mendeley, endnote, citeulike) es bastante pobre. Sólo 13 repositorios (no llega al 20%) nos dan un botón rápido para compartir los datasets y en el caso

de los gestores bibliográficos únicamente 11 repositorios (16,6%) nos ofrecen esta opción. Si nos vamos a los repositorios que ofrecen el número de veces que se ha compartido un dataset en las redes sociales vemos que la cifra de los repositorios que lo hacen es casi insignificante. Sólo figshare y zotero ofrecen ese dato.

Otro de los puntos a analizar serían las “altmetrics”. Viendo el cambio tecnológico que se ha producido desde la mitad del siglo XX hasta la actualidad es lógico pensar que puede haber otra forma de medir la calidad científica de las publicaciones. Es cierto que gracias al avance tecnológico de estas décadas es más fácil y rápido calcular el factor de impacto de una revista, pero sigue siendo la misma métrica.

Por tanto, es necesario encontrar una nueva forma de evaluar la calidad de la investigación científica. Las “altmetrics” son una herramienta importante pero no pueden ser la única. En las métricas tradicionales podíamos encontrar problemas de interés o amiguismo entre los autores científicos y los revisores lo que provocaría una manipulación en dichas métricas. Las autocitas son otros inconvenientes que producirían el mismo efecto. Por otro lado, en las métricas alternativas encontramos problemas similares. Podemos compartir en las redes sociales algún dataset que no nos interese pero que haya sido publicado por un antiguo compañero, amigo o alguien conocido para que aumente su valoración. Es decir, tanto las métricas tradicionales como las actuales se pueden manipular por algún interés. Al menos en mi opinión las métricas alternativas complementan las métricas tradicionales pero no las sustituyen. Es decir, se deberían mantener los dos tipos.

Limitaciones del estudio

Actualmente no existen herramientas informáticas de libre acceso para el estudio de las métricas de citas y uso de datos de los distintos portales y repositorios. Por este motivo el análisis se ha basado en la observación y visita de los 73 repositorios de datos de ciencias sociales indexados en el Data Citation Index. Tampoco existe ninguna Universidad de nuestro país suscrita actualmente a esta base de datos.

Dada la limitación del estudio no se han podido comprobar si los indicadores comentados para los datos abiertos de la administración eran recogidos por los repositorios de datos de las distintas comunidades autónomas de nuestro país, así como del portal de datos abiertos del gobierno central.

6. Bibliografía

1. E. Abadal, C. Ollé Castellà, F. Abad García y R. Melero. (2013) Políticas de acceso abierto a la ciencia en las universidades españolas. *Revista española de documentación científica* (36)2; 1-15
2. A. Abella, M. Ortiz de Urbina Criado y C. de Pablos Heredero. (2014) Meloda, métrica para evaluar la reutilización de datos abiertos. *El profesional de la información* (23)6; 582-588
3. R. Aleixandre Benavent, J. C. Valderrama Zurián y G. González Alcaide. (2007) El factor de impacto de las revistas científicas: limitaciones e indicadores alternativos. *El profesional de la información* (16)1; 4-11
4. M. Álvarez Espinar (2014). Apertura y reutilización de datos públicos. *Govern Obert*; 2. 64 p.
5. A. Ball y M. Duke (2015a). How to track the impact of research data with metrics. Edimburgo: Digital Curation Centre: DCC How-to Guides. 15 p.
6. A. Ball y M. Duke. (2015b) How to cite datasets and link to publications. Edimburgo: Digital Curaation Centre: DCC How-to Guides. 15 p.
7. BOE (2007). Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público
8. BOE (2010). Real Decreto 4/2010, de 8 de enero, por el que se regula el Esquema Nacional de Interoperabilidad en el ámbito de la Administración Electrónica.
9. BOE (2011). Real Decreto 1495/2011, de 24 de octubre, por el que se desarrolla la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público, para el ámbito del sector público estatal.
10. BOE (2015). Ley 18/2015, de 9 de julio, por la que se modifica la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público.
11. M. Bordons y M. Á. Zulueta (1999). Evaluación de la actividad científica a través de indicadores bibliométricos. *Revista española de cardiología* (52); 790-800
12. C. L. Borgman (2012). The conundrum of sharing research data. *Journal of the american society for information science and technology* (63)6; 1059-1078
13. Á. Cabezas Clavijo y D. Torres Salinas (2010). Indicadores de uso y participación en las revistas científicas 2.0: el caso de PLoS One. *El profesional de la información* (19)4; 431-434
14. D. Camps (2008). Limitaciones de los indicadores bibliométricos en la evaluación de la actividad científica biomédica. *Colombia médica* (39)1; 74-79
15. H. Canales Becerra y M. E. Mesa Fleitas (2003). Bibliometría, Informetría, Cienciometría: su etimología y alcance conceptual. *Investigación Bibliotecológica*; 1-13
16. N. Clabo e I. Ramos Vielba (2015). Reutilización de datos abiertos en la administración pública en España y uso de licencias-tipo. *Revista española de documentación científica* 38(3); 1-14
17. Comision Europea (2011a). Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones: Datos abiertos (Un motor para la innovación, el crecimiento y la gobernanza transparente)

18. Comision Europea (2011b). Propuesta de Directiva del Parlamento Europeo y del Consejo por la que se modifica la Directiva 2003/98/CE relativa a la reutilización de la información del sector público
19. A. Corojan y E. Campos (2011). Gobierno abierto: Alcance e implicaciones. Fundación Ideas. 84 p.
20. R. Costas, I. Meijer, Z. Zahedi y P. Wouters (2013). The value of research data: Metrics for datasets from a cultural and technical point of view. Knowledge Exchange
21. datos.gob.es (2011). Nuevo paquete de medidas Open Data de la Comisión Europea. Disponible en: <http://datos.gob.es/content/nuevo-paquete-de-medidas-open-data-de-comision-europea>. Fecha de acceso: 23/05/2016
22. European data portal (2016). Open data goldbook for data holders and data managers: A practical guidebook for any organisation that wants to publish their data as open data. 81 p.
23. A. Ferrer Sapena, F. Peset y R. Aleixandre Benavent (2011). Acceso a los datos públicos y su reutilización: open data y open government. *El profesional de la información* (20)3; 260-269
24. A. Ferrer Sapena y E. A. Sánchez Pérez (2013). Open data, big data ¿hacia dónde nos dirigimos? *Anuario ThinkEPI*; (7); 150-156
25. A. Ferrer Sapena, E. A. Sánchez Pérez, R. Aleixandre-Benavent y F. Fernanda Peset (2016). Cómo analizar el impacto de los datos de investigación con métricas modelos y servicios. *El profesional de la información* (25)4; 632-641
26. G. Gallo (2013). Cambios en la Directiva de reutilización. Disponible en: <https://gontzalgallo.com/2013/06/27/cambios-en-la-directiva-de-reutilizacion/>. Fecha de acceso: 23/05/2016
27. I. Gómez Caridad y M. Bordons Gangas (1996). Limitaciones en el uso de los indicadores bibliométricos para la evolución científica. *Documento*; 21-26
28. J. González de Dios, M. Moya y M. A. Mateos Hernández (1997). Indicadores bibliométricos: Características y limitaciones en el análisis de la actividad científica. *Asociación española de pediatría* (47); 235-244
29. N. Huijboom y T. v. d. Broek (2011). Open data: an international comparison of strategies. *European journal of epractice* (12); 1-13
30. R. Kuhlen (2007). Open access: un cambio de paradigma para la puesta a disposición pública del conocimiento. El desarrollo en Alemania. *bid: textos universitaris de biblioteconomia i documentació* (18); 1-10
31. B. Lawrence (2011). Citation and peer review of data: Moving towards formal data publication. *The international journal of digital curation* (6)2; 4-37
32. C. Martín y S. L. Soriano Maldonado (2011). Reutilización de la información del sector público y open data en el contexto español y europeo. Proyecto Aporta. *El profesional de la información* (20)3; 291-297
33. M. S. Mayernik (2013). Bridging data lifecycles: Tracking data use via data citations workshop report. NCAR Library. 32 p.
34. R. Meijer, P. Conradie y S. Choenni (2014). Reconciling contradictions of open data regarding transparency, privacy, security and trust. *Journal of theoretical and applied electronic commerce research* (9)3; 32-44

35. R. Melero y J. Hernández San Miguel (2014). Acceso abierto a los datos de investigación, una vía hacia la colaboración científica. *Revista española de documentación científica* (37)4; 1-11
36. C. Mendo, L. F. Ramos, R. Arquero, F. d. Valle Gastaminza, R. Sánchez, C. Tejada, J. L. Peón, S. Cobo y A. Sala (2013). Del acceso a la reutilización, del dato al documento una visión conceptual de la información pública. *Revista española de documentación científica* (36)3; 1-12
37. NISO (2015). NISO RP-25-201x-2A Alternative Outputs in Scholarly Communications: Data Metrics
38. B. Obama (2009). Memorandum for the heads of executive departments and agencies, January 21, 2009
39. Open Knowledge International. The open data handbook. Disponible en: <http://opendatahandbook.org/>. Fecha de acceso: 12/05/2016
40. Parlamento Europeo (2003). Directiva 2003/98/CE del Parlamento europeo y del Consejo de 17 de noviembre de 2003 relativa a la reutilización de la información del sector público
41. Parlamento Europeo (2013). Directiva 2013/37/UE del Parlamento Europeo y del Consejo de 26 de junio de 2013 por la que se modifica la Directiva 2003/98/CE relativa a la reutilización de la información del sector público
42. M. J. Peralta González, M. Frías Guzmán y O. Gregorio Chaviano (2015). Criterios, clasificaciones y tendencias de los indicadores bibliométricos en la evaluación de la ciencia. *Revista cubana de información en ciencias de la salud* (26)3; 290-309
43. C. Pintos y J.L. Marín (2013). Open data: reutilización de la información pública. Innap Innova. 133 p.
44. L. F. Ramos Simón, R. Arquero Avilés, I. Botezan, S. Cobo Serrano, A. Sala Jiménez, R. Sánchez Jiménez y F. d. Valle Gastaminza (2012). De la reutilización de información del sector público a los portales de datos abiertos en Europa. *bid: textos universitaris de biblioteconomia i documentació* (29); 1-36
45. N. Robinson García, E. Jiménez Contreras y D. Torres Salinas (2015). Analyzing data citation practices using the Data Citation Index. *Journal of the Association for Information Science and Technology*; 1-12
46. E. G. Rodríguez (2013). La revisión editorial por pares: roles y procesos. *Revista cubana de información en ciencias de la salud* (24)2; 160-175
47. R. Sancho (1990). Indicadores bibliométricos utilizados en la evaluación de la ciencia y la tecnología. *Revista española de documentación científica* (13)4-5; 842-865
48. J. Sanz Valero, M. J. d. Agostino, L. D. Castiel y J. Veiga de Cabo (2007). La iniciativa Open Access, una visión de conjunto. *Medicina y seguridad del trabajo* (53) 207; 5-9
49. Sunlight Foundation (2010). Ten principles for opening up government information. Disponible en: <https://sunlightfoundation.com/policy/documents/ten-open-data-principles/>. Fecha de acceso: 21/05/2016
50. D. Torres Salinas y R. Ruiz Pérez (2015). Diez claves sobre métrica alternativa. *Unelibros* (30); 12-13
51. University of North Texas. Denton Declaration. Disponible en: <http://openaccess.unt.edu/denton-declaration>. Fecha de acceso: 14/07/2016

52. B. Velasco, J. M. Eiros, J. M. Pinilla y J. A. San Román (2012). La utilización de los indicadores bibliométricos para evaluar la actividad investigadora. *Aula abierta* (40)2; 75-83
53. W3C (2009). Improving access to government through better use of the web. Disponible en: <https://www.w3.org/TR/egov-improving/>. Fecha de acceso: 16/05/2016
54. J. Wilsdon, L. Allen, E. Belfiore, P. Campbell, S. Curry, S. Hill, R. Jones, R. Kain, S. Kerridge, M. Thelwall, J. Tinkler, I. Viney, P. Wouters, J. Hill y B. Johnson (2015). The metric tide: Report of the independent review of the role of metrics in research assessment and management. HEFCE. 163 p.
55. P. Wouters, M. Thelwall, K. Kousha, L. Waltman, S.d. Rijcke, A. Rushforth y T. Franssen (2015). The metric tide: Literature review (Supplementary report I to the independent review of the role of metrics in research assessment and management). HEFCE. 188 p.

7. Anexos

Anexo 1. Lista inicial de repositorios

1. 4TU.ResearchData
2. Archaeology Data Service
3. ArrayExpress
4. Australian Antarctic Data Centre
5. Australian Data Archive
6. Biological Magnetic Resonance Data Bank
7. British Antarctic Survey
8. British Atmospheric Data Centre
9. British Geological Survey
10. British Oceanographic Data Centre
11. caArray (*)
12. caNanoLab
13. Cancer Models Database (*)
14. Cancer GEnome Mine
15. CEH
16. Cell Centered Database
17. Codex Sinaiticus
18. Crystallography Open Database
19. Dataverse
20. DisProt
21. DrugBank
22. Dryad
23. eCrystals
24. Edinburgh DataShare
25. EMDatabank
26. EMAGE
27. Eurostat
28. Figshare
29. Finnish Social Science Data Archive
30. Forest Service Research Data Archive
31. GAVO Data Center
32. GEO
33. Greengenes
34. GWAS Central
35. HMDB
36. Infevers
37. Inter-university Consortium for Political and Social Research

38. MGDS
39. Michigan Corpus of Academic Spoken English
40. MiCroKiTS
41. miRBase (*)
42. Mouse Phenome Database
43. NASA Socioeconomic Data and Applications Center
44. NASA/IPAC Infrared Science Archive (*)
45. National Snow & Ice Data Center
46. NERC
47. NIST
48. nmrshiftdb2 (*)
49. NCDC
50. NGDC
51. Nucleic Acid Database (*)
52. Odum Institute Data Archive
53. Office for National Statistics
54. ORNL DAAC
55. PANGAEA
56. Protein Data Bank
57. PURR
58. Research Data Australia
59. RESID
60. Roper Center for Public Opinion Research
61. ShareGeo Open
62. South African Data Archive
63. Stanford Microarray Database (*)
64. The Association of Religion Data Archives
65. The Cell Image Library
66. The Proceedings of the Old Bailey
67. TIGER Line Shapefiles
68. TreeBASE
69. UK Data Archive
70. UK Reading Experience Database
71. UniPROBE Database
72. World Values Survey
73. Zenodo

Anexo 2. Lista de repositorios analizados

1. 4TU.ResearchData
2. Archaeology Data Service
3. ArrayExpress
4. Australian Antarctic Data Centre
5. Australian Data Archive
6. Biological Magnetic Resonance Data Bank

7. British Antarctic Survey
8. British Atmospheric Data Centre
9. British Geological Survey
10. British Oceanographic Data Centre
11. caNanoLab
12. Cancer GEnome Mine
13. CEH
14. Cell Centered Database
15. Codex Sinaiticus
16. Crystallography Open Database
17. Dataverse
18. DisProt
19. DrugBank
20. Dryad
21. eCrystals
22. Edinburgh DataShare
23. EMAGE
24. EMDatabank
25. Eurostat
26. Figshare
27. Finnish Social Science Data Archive
28. Forest Service Research Data Archive
29. GAVO Data Center
30. GEO
31. Greengenes
32. GWAS Central
33. HMDB
34. Infevers
35. Inter-university Consortium for Political and Social Research
36. MGDS
37. Michigan Corpus of Academic Spoken English
38. MiCroKiTS
39. Mouse Phenome Database
40. NASA Socioeconomic Data and Applications Center
41. National Snow & Ice Data Center
42. NERC
43. NIST
44. NCDC
45. NGDC
46. Odum Institute Data Archive
47. Office for National Statistics
48. ORNL DAAC
49. PANGAEA
50. Protein Data Bank
51. PURR



52. Research Data Australia
53. RESID
54. Roper Center for Public Opinion Research
55. ShareGeo Open
56. South African Data Archive
57. The Association of Religion Data Archives
58. The Cell Image Library
59. The Proceedings of the Old Bailey
60. TIGER Line Shapefiles
61. TreeBASE
62. UK Data Archive
63. UK Reading Experience Database
64. UniPROBE
65. World Values Survey
66. Zenodo

Anexo 3. Capturas del análisis de datos

4TU.ResearchData	
Referencia bibliográfica	<p>Citation style Datacite</p> <p>Mourik, V. (Vincent); Zuo, K. (Kun); Frolov, S.M. (Sergey); Plissard, S.R. (Sebastien); Bakkers, E.P.A.M. (Erik); Kouwenhoven, L.P. (Leo) (2012) Signatures of Majorana fermions in hybrid superconductor-semiconductor nanowire devices. Science Magazine. Dataset. http://dx.doi.org/10.4121/uuid:8bf81177-2f2b-49c2-aaf5-d36739873dd9</p> <p>select html code to copy</p> <p>Short link</p> <p>doi:10.4121/uuid:8bf81177-2f2b-49c2-aaf5-d36739873dd9</p> <p>select html code to copy</p> <p>Even shorter link convenient for social media, but may be slightly less durable</p> <p>http://doi.org/8vr</p> <p>select html code to copy</p>
Identificador	<p>Link to this Dataset [more citation options]</p> <p>doi: 10.4121/uuid:8bf81177-2f2b-49c2-aaf5-d36739873dd9 uri: http://dx.doi.org/10.4121/uuid:8bf81177-2f2b-49c2-aaf5-d36739873dd9</p> <p>copy/paste into your html code:</p> <pre>d</pre>

Archaeology Data Service	
Identificador	<p>Identifiers:</p> <p>[ADS] Depositor ID- NX07SE 3 [ADS] Import RCN- RCAHMS13-60886</p> <p>http://archaeologydataservice.ac.uk/archsearch/record.jsf?titleId=2333532</p>

ArrayExpress

Identificador	Accession E-GEOD-85381
---------------	---------------------------

Australian Antarctic Data Centre	
Referencia bibliográfica	Aadc, D. (2000, updated 2014) Automatic Weather Station Data from Casey Australian Antarctic Data Centre - CAASM Metadata (https://data.aad.gov.au/metadata/records/casey_aws) Rigual-Hernandez, A.S., Armand, L.K., Trull, T.W., Bray, S.G., Wilks, J., Lawler, K., Ballegeer, A., Connell, D.J. (2015, updated 2015) Diatom species of the Australian sector of the Southern Ocean Australian Antarctic Data Centre - doi:doi:10.4225/15/5563CAA53B5E0
Identificador	(https://data.aad.gov.au/metadata/records/casey_aws) doi:doi:10.4225/15/5563CAA53B5E0

Australian Data Archive	
Identificador	<p>LONGITUDINAL HOME CATALOGUE SERIES GUIDES VISUALISATION CREATING & USING DATA USER GUIDES NEWS ABOUT</p> <p>YOU ARE HERE: /LONGITUDINAL/01010</p> <p>SOCIAL SCIENCE Study Variables (5816) Related Materials (13) Log in to see your data access p</p> <p>HISTORICAL</p> <p>INDIGENOUS</p> <p>LONGITUDINAL</p> <p>QUALITATIVE</p> <p>INTERNATIONAL</p> <p>CRIME & JUSTICE</p> <p>Study Information Content</p> <p>Study Title Longitudinal Surveys of Australian Youth, 1995 cohort</p> <p>Series Name Longitudinal Surveys of Australian Youth</p> <p>Primary Investigator Australian Department of Education, Employment and Workplace Relations</p> <p>ADA ID au.edu.anu.ada.ddi.01010</p> <p>Abstract The Longitudinal Surveys of Australian Youth (LSAY) is a survey program that tracks groups of young Australians studying their transition from school to further education and work. It uses large, nationally representative samples</p>

British Antarctic Survey	
Identificador	GB/NERC/BAS/PDC/00620

British Atmospheric Data Centre	
Referencia bibliográfica	Citable as: Cooke, K. (2007): GASPOL: Atmospheric perfluoromethylcyclohexane (PMCH) measurements, Birmingham, UK. NCAS British Atmospheric Data Centre, <i>date of citation</i> . http://catalogue.ceda.ac.uk/uuid/89733439ebe97d8bd96abe65e8c298b9
Identificador	http://catalogue.ceda.ac.uk/uuid/89733439ebe97d8bd96abe65e8c298b9

British Oceanographic Data Centre	
Identificador	BODC Series Reference 1664499
Difusión en RRSS / GGBB	<p>Volume of the Sea-Bird SBE</p> <p>body</p> <p>43 sensor</p> <p>1</p> <p>Length of the</p> <p>Facebook</p> <p>Twitter</p> <p>Google+</p> <p>+</p>

caNanoLab

Identificador	doi:10.3791/3570 (2012).
---------------	--------------------------

Cancer GEnome Mine		
Identificador	Experiment Name	Niini et al.: Array comparative genomic hybridization reveals frequent alterations of G1/S checkpoint genes in undifferentiated pleomorphic sarcoma of bone.
	Accession	CG-EXP-154

CEH	
Identificador	Dataset identifiers https://catalogue.ceh.ac.uk/id/5d324260-22d9-11e4-8c21-0800200c9a66

Cell Centered Database												
Identificador	<table border="1"> <tr> <th colspan="2">Summary</th> </tr> <tr> <td>Project ID</td> <td>P1170</td> </tr> <tr> <td>Leader</td> <td>Maryann Martone (Other projects)</td> </tr> <tr> <td>Collaborators</td> <td>Diana Price, Andrea Thor, Masako Terada, Hiro Hakozaiki</td> </tr> <tr> <td>Project Description</td> <td>NeuroLucida tracing of filled Purkinje neurons</td> </tr> </table>		Summary		Project ID	P1170	Leader	Maryann Martone (Other projects)	Collaborators	Diana Price, Andrea Thor, Masako Terada, Hiro Hakozaiki	Project Description	NeuroLucida tracing of filled Purkinje neurons
Summary												
Project ID	P1170											
Leader	Maryann Martone (Other projects)											
Collaborators	Diana Price, Andrea Thor, Masako Terada, Hiro Hakozaiki											
Project Description	NeuroLucida tracing of filled Purkinje neurons											





Crystallography Open Database	
Identificador	<p align="center">Information card for 1001663</p> <p align="center">1001662 << 1001663 >> 1001664</p>


Dataverse	
Referencia bibliográfica	Dr. Sunil Kumar Mishra, 2016, "Oriental Approach to Environmental Protection and Perspective of Media", doi:10.7910/DVN/7DC4ON, Harvard Dataverse, V1
Identificador	doi:10.7910/DVN/7DC4ON , Harvard Dataverse, V1

DisProt							
Identificador	<table border="1"> <tr> <th colspan="2">General information</th> </tr> <tr> <td>DisProt:</td> <td>DP00047</td> </tr> <tr> <td>Name:</td> <td>Myelin basic protein</td> </tr> </table>	General information		DisProt:	DP00047	Name:	Myelin basic protein
General information							
DisProt:	DP00047						
Name:	Myelin basic protein						

DrugBank							
Identificador	<table border="1"> <tr> <th colspan="2">Identification</th> </tr> <tr> <td>Name</td> <td>Beclomet</td> </tr> <tr> <td>Accession Number</td> <td>DB00394</td> </tr> </table>	Identification		Name	Beclomet	Accession Number	DB00394
Identification							
Name	Beclomet						
Accession Number	DB00394						

Dryad	
Referencia bibliográfica	Afkhami ME, Stinchcombe JR (2016) Data from: Multiple mutualist effects on genome-wide expression in the tripartite association between <i>Medicago truncatula</i> , nitrogen-fixing bacteria, and mycorrhizal fungi. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.8f66s
Identificador	http://dx.doi.org/10.5061/dryad.8f66s

Difusión en RRSS / GGBB	<p>Afkhami ME, Stinchcombe JR (2016) Data from: Multiple mutualist effects on genome-wide expression in the tripartite association between <i>Medicago truncatula</i>, nitrogen-fixing bacteria, and mycorrhizal fungi. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.8f66s</p> <p style="text-align: right;">Cite Share</p> <p>Download the data package citation in the following formats:</p> <p>RIS (compatible with EndNote, Reference Manager, ProCite, RefWorks)</p> <p>BibTex (compatible with BibDesk, LaTeX)</p> <p>  reddit this!  Tweet  Me gusta  </p>
-------------------------	---

eCrystals	
Referencia bibliográfica	Citation: Smith, Paul D. and Coles, Simon J. and Hursthouse, Michael B. (2004) University of Southampton, Crystal Structure Report Archive. (doi:10.5258/ecrystals/906)
Identificador	10.5258/ecrystals/906
Difusión en RRSS / GGBB	 

Edinburgh DataShare	
Referencia bibliográfica	Bird, Tom. (2016). Serial serum AFP results for en HCC free patients in HCC surveillance cohort, 1999-2012 [dataset]. University of Edinburgh. http://dx.doi.org/10.7488/ds/1397 .
Identificador	<p>http://hdl.handle.net/10283/1991</p> <p>http://dx.doi.org/10.7488/ds/1397</p>

EMAGE	
Identificador	IDs MGI:1915867

EMDataBank	
Identificador	Ex vivo mammalian prions are formed of paired double helical prion protein fibrils. Terry C, Wenborn A, Gros N, Sells J, Joiner S, Hosszu LL, Tattum MH, Panico S, Clare DK, Collinge J, Saibil HR, Wadsworth JD <i>Open Biology</i> (2016) 6 [PubMed 27249641] [DOI]

Eurostat	
Identificador	<p>House price index (2010 = 100) - quarterly data</p> <p>Title: House price index (2010 = 100) - quarterly data</p> <p>Code: ei_hpp1_q</p> <p>Last update of data: 26/07/2016</p> <p>Last table structure change: 26/07/2016</p> <p>Number of values: 3623</p>

Figshare	
Referencia bibliográfica	Gutiérrez, Silvia (2016): materias_investigadores_mexicanos_alemania.png. figshare. https://dx.doi.org/10.6084/m9.figshare.3717366.v1
Identificador	Retrieved: 06 17, Aug 20, 2016 (GMT) https://dx.doi.org/10.6084/m9.figshare.3717366.v1

<p>Difusión en RRSS / GGBB</p>	
<p>Veces compartido</p>	

Finnish Social Science Data Archive	
<p>Referencia bibliográfica</p>	<p>Bibliographical Citation Konu, Anne (University of Tampere): School Well-being Profile 2015-2016: Personnel [dataset]. Version 1.0 (2016-07-20). Finnish Social Science Data Archive [distributor]. http://urn.fi/urn:nbn:fi:fsd:T-FSD3119</p>
<p>Identificador</p>	<p>Title School Well-being Profile 2015-2016: Personnel</p> <p>Dataset ID Number FSD3119</p> <p>Persistent identifier urn.nbn.fi/fsd:T-FSD3119</p>

Forest Service Research Data Archive	
<p>Referencia bibliográfica</p>	<p>Cole, Jason, A.; Johnson, Kristofer D.; Birdsey, Richard A.; Pan, Yude; Wayson, Craig A.; McCullough, Kevin; Hoover, Coeli M.; Hollinger, David Y.; Bradford, John B.; Ryan, Michael G.; Kolka, Randall K.; Weishampel, Peter; Clark, Kenneth L.; Showronski, Nicholas S.; Hom, John; Ollinger, Scott V.; McNulty, Steven G.; Gavazzi, Michael J. 2015. North American Carbon Program biometric database. 2nd Edition. USDA Forest Service, Northern Research Station. http://dx.doi.org/10.2737/RDS-2013-0008-2</p>
<p>Identificador</p>	<p>You are here: Data Archive Home > Data Catalog > RDS-2013-0008-2</p> <p>http://dx.doi.org/10.2737/RDS-2013-0008-2</p>

GEO	
<p>Identificador</p>	<p>DataSet Record GDS5664: Expression Profiles Data Analysis Tools Sample Subsets</p>

Greengenes	
<p>Identificador</p>	<p>GenBank ACCESSION: AY701553.1</p>

GWAS Central	
--------------	--

Identificador	GWAS Central Identifier	HGVST699
---------------	-------------------------	----------

HMDB		
Identificador	Record Information	
	Version	3.6
	Creation Date	2005-11-16 15:48:42 UTC
	Update Date	2016-02-13 02:06:54 UTC
	HMDB ID	HMDB00002

Infevers		
Identificador	Usual name Name as first published or submitted to Infevers. May be different from the HGVS edited protein and sequence names.	R169Q

Inter-university Consortium for Political and Social Research	
Referencia bibliográfica	Study Description Citation Inter-university Consortium for Political and Social Research. Census of Population and Housing, 1970 [United States]: Documentation for Census Bureau Restricted Data. ICPSR21981-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2009-08-10. http://doi.org/10.3886/ICPSR21981.v2
Identificador	Persistent URL: http://doi.org/10.3886/ICPSR21981.v2

MGDS	
Referencia bibliográfica	How To Cite This Data Francis, B., (2016). Raw ship-based Multibeam Sonar Data acquired during R/V Maurice Ewing expedition EW9304 (1993). Integrated Earth Data Applications (IEDA). doi: http://dx.doi.org/10.1594/IEDA/305352 .
Identificador	DOI 10.1594/IEDA/305352

Michigan Corpus of Academic Spoken English	
Identificador	

MiCroKiTS	
Identificador	MiCroKit ID MCK-DM-00096

Mouse Phenome Database	
Identificador	doi: 10.1186/1752-0509-5-43

NASA Socioeconomic Data and Applications Center	
Referencia bibliográfica	<p>Recommended Citation(s)*:</p> <p>Yale Center for Environmental Law and Policy - YCELP - Yale University, Center for International Earth Science Information Network - CIESIN - Columbia University, World Economic Forum - WEF, and Joint Research Centre - JRC - European Commission. 2012. 2012 Environmental Performance Index and Pilot Trend Environmental Performance Index. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H48913SG. Accessed DAY MONTH YEAR.</p>
Identificador	http://dx.doi.org/10.7927/H48913SG . Accessed DAY MONTH YEAR.
Difusión en RRSS / GGBB	 <p>ENW (EndNote & RefWorks)†</p> <p>RIS (Others)</p>

National Snow & Ice Data Center	
Referencia bibliográfica	<p>Armstrong, R., K. Knowles, M. Brodzik, and M. A. Hardman. 1994, updated 2016. <i>DMSPP SSM/I-SSMIS Pathfinder Daily EASE-Grid Brightness Temperatures, Version 2</i>. [Indicate subset used]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: http://dx.doi.org/10.5067/3EX2U1DV3434. [Date Accessed].</p>
Identificador	Distributed Active Archive Center. doi: http://dx.doi.org/10.5067/3EX2U1DV3434 . [Date Accessed].

NERC	
Referencia bibliográfica	<p>Citable as: University of Oxford; The Open University; Montabone, L.; Lewis, S.R.; Read, P.L. (2011): Mars Analysis Correction Data Assimilation (MACDA): MGS/TES v1.0. NCAS British Atmospheric Data Centre, 29 November 2011. doi:10.5285/78114093-E2BD-4601-8AE5-3551E62AEF2B. http://dx.doi.org/10.5285/78114093-E2BD-4601-8AE5-3551E62AEF2B</p>
Identificador	http://dx.doi.org/10.5285/78114093-E2BD-4601-8AE5-3551E62AEF2B

NIST	
Identificador	http://dx.doi.org/10.6028/NIST.IR.8105

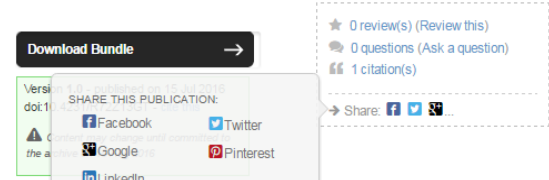
NCDC	
Identificador	<p>RESOURCE DESCRIPTION (data set id): noaa-lake-6064</p> <p>ONLINE RESOURCE (When Citing Data): http://www.ncdc.noaa.gov/paleo/study/6064</p>

Odum Institute Data Archive	
Referencia bibliográfica	<p>Obidziński, Artur; Kołaczowska, Ewa; Czurzycki, Wojciech; Marciszewska, Katarzyna; Mędrzycki, Piotr, 2014, "Rarity does not prevent modelling of invasive alien plants' abundance? Random Forest models for urban woodlands in Warsaw, Poland", http://dx.doi.org/10.15139/S3/11908 V1 [Version]</p>
Identificador	<p>Study Global ID doi:10.15139/S3/11908</p>

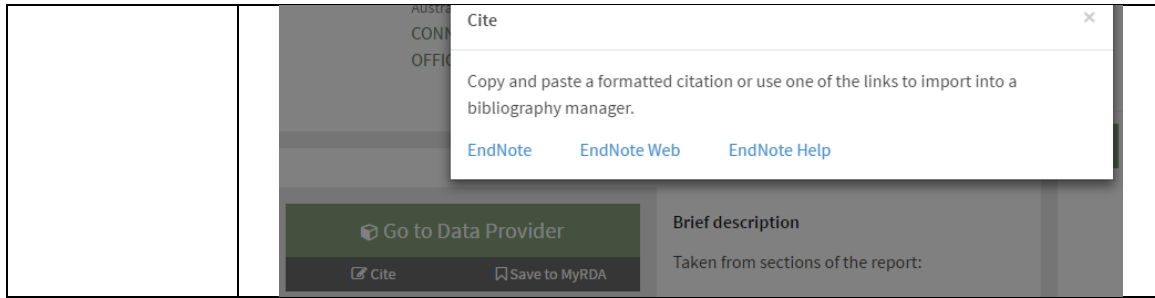
ORNL DAAC	
Referencia bibliográfica	<p>Smith, E. A., and W. L. Crosson. 1994. Bowen Ratio Surface Flux: Smith (FIFE). Data set. Available on-line [http://www.daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. Also published in D. E. Strebel, D. R. Landis, K. F. Huemmrich, and B. W. Meeson (eds.), Collected Data of the First ISLSCP Field Experiment, Vol. 1: Surface Observations and Non-Image Data Sets. CD-ROM. National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, U.S.A. (available from http://www.daac.ornl.gov). doi:10.3334/ORNLDAAC/22</p>
Identificador	doi:10.3334/ORNLDAAC/22

PANGAEA	
Referencia bibliográfica	<p><i>Citation:</i> Gervais, E (1987): Miocene dissolution facies and microfacies at Northwest Atlantic Deep Sea Drilling Project Site 93-603. doi:10.1594/PANGAEA.788868, <i>Supplement to: Gervais, Elisabeth (1987): Miocene dissolution facies and microfacies at Northwest Atlantic Deep Sea Drilling Project Site 603. In: van Hinte, JE; Wise, SW Jr; et al. (eds.), Initial Reports of the Deep Sea Drilling Project, Washington (U.S. Govt. Printing Office), 93, 511-524, doi:10.2973/dsdp.proc.93.109.1987</i></p>
Identificador	<p>doi:10.2973/dsdp.proc.93.109.1987</p> <p>Knap, Anthony H (2002): Manual of methods used in JGOFS. doi:10.1594/PANGAEA.67700</p> <p>Knap, Anthony H; Michaels, A; Close, A R; Ducklow, Hugh; Dickson, Andrew G (1996): Protocols for the Joint Global Ocean Flux Study (JGOFS) Core Measurements. <i>JGOFS, Reprint of the IOC Manuals and Guides No. 29, UNESCO 1994</i>, 19, 210 pp, hdl:10013/epic.27912.d001 ↗</p> <p>Knap, Anthony H; Michaels, A; Close, A R; Ducklow, Hugh; Dickson, Andrew G (1994): Protocols for the Joint Global Ocean Flux Study (JGOFS) Core Measurements. <i>IOC Manual and Guides, 29</i>, 170 pp, hdl:10013/epic.27912.d001 ↗</p>

Protein Data Bank	
Identificador	DOI: 10.2210/pdb5d2b/pdb

PURR	
Referencia bibliográfica	Merwade, V., Liuying Du, Nikhil Sangwan (2016). gSSURGO-based Floodplain Maps of Hawaii . Purdue University Research Repository. doi:10.4231/R72Z13GT
Identificador	doi:10.4231/R72Z13GT
Difusión en RRSS / GGBB	

Research Data Australia	
Referencia bibliográfica	<p>DataCite</p> <p>LEMON, RICHARD ; BROLSMA, HENK (2007): Survey report 1996/97 Voyage 3 Australian Antarctic Division Author - Richard Lemon / AUSLIG. Australian Antarctic Data Centre. URI : http://data.aad.gov.au/aadc/metadata/metadata_redirect.cfm?md=survey_1996-97_V3 http://data.aad.gov.au/aadc/metadata/metadata_redirect.cfm?md=survey_1996-97_V3</p>
Identificador	<p>Identifiers</p> <p>Local : survey_1996-97_V3 URI : http://data.aad.gov.au/aadc/metadata/metadata_redirect.cfm?md=survey_1996-97_V3 ↗ URI : https://data.aad.gov.au/aadc/metadata/metadata_redirect.cfm?md=AMD/AU/survey_1996-97_V3 ↗</p>
Difusión en RRSS / GGBB	<p>Survey report 1996/97 Voyage 3 Australian Antarctic Division Author - Richard Lemon / AUSLIG</p> <p>Australian Antarctic Data Centre</p> <p>BROLSMA, HENK (Principal investigator) LEMON, RICHARD (Processor)</p> <p>f t g View</p>



RESID	
Identificador	RESID:AA0027



Roper Center for Public Opinion Research	
Referencia bibliográfica	Siena Research Institute, Siena College. Siena Research Institute Poll: February 2015 Siena New York Poll, Feb, 2015 [dataset]. USSIENA2015-02NY, Version 2. Siena Research Institute, Siena College [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], accessed Aug-2-2016.

ShareGeo Open	
Identificador	<p>Rhododendron ponticum in Hatchford wood, Surrey</p> <p>Mostrar el registro completo del ítem</p> <p>Título: Rhododendron ponticum in Hatchford wood, Surrey</p> <p>Autor: Smith, Mike; Kingston University</p> <p>Resumen: Data Relating to the location of Rhododendron ponticum within Hatchford Woods, Within Ockham Common, Surrey.</p> <p>URI: http://hdl.handle.net/10672/69</p> <p>Fecha: 2010-09-14</p> <p>Format: ESRI Shapefile</p>



South African Data Archive	
Referencia bibliográfica	<p>dc.identifier.citation</p> <p>Statistics South Africa. Marriages and Divorces, 2013 [Computer file]. S0210. Pretoria: Statistics South Africa [producer] 2015. Pretoria: South African Data Archive, National Research Foundation [distributor], 2016.</p>
Identificador	<p>dc.identifier.uri</p> <p>http://hdl.handle.net/10956/204</p>





The Association of Religion Data Archives	
Difusión en RRSS / GBB	

The Cell Image Library	
Identificador	CIL:24788*

Difusión en RRSS / GGBB	<p>Licensing</p>  <p>Attribution Non-Commercial Share Alike: This image is licensed under a Creative Commons Attribution, Non-Commercial Share Alike License. View License Deed View Legal Code</p> <p> Share 4 <ul style="list-style-type: none"> CiteULike Facebook Twitter Imprimir Google+ Correo Gmail Más... (197) </p> <p>Comments:</p> <p>From: <input type="text"/></p> <p>Send  Add This Powered by Discussit</p>
-------------------------	--

The Proceedings of the Old Bailey	
Identificador	<p>JOSEPH OFFLEY, Deception > fraud, 26th June 1876.</p> <p>Reference Number: t18760626-271</p>

TreeBASE	
Identificador	URI:  http://purl.org/phylo/treebase/phyloWS/study/TB2:S19566
Difusión en RRSS / GGBB	

UK Data Archive	
Referencia bibliográfica	Office for National Statistics. (2016). <i>Monthly Wages and Salaries Survey, 2000-2016: Secure Access</i> . [data collection]. 8th Edition. UK Data Service. SN: 6702. http://dx.doi.org/10.5255/UKDA-SN-6702-8
Identificador	http://dx.doi.org/10.5255/UKDA-SN-6702-8
Difusión en RRSS / GGBB	<p>Catalogue</p> <p>UK Data Service data catalogue record for:</p> <p>OECD Social and Welfare Statistics, 1980-2015</p> <p>Related Studies Publications  Access</p> <hr/> <p>TITLE DETAILS</p> <p>SN: 4835</p> <p>Select the text above to add data citation in your outputs.</p> <p>Select citation format: <input type="text" value="APA"/> XML citation formats: CSL EndNote</p> <div style="text-align: right;"> <p>SHARE </p> <p>SHARE THIS PAGE</p> <p> Facebook</p> <p> Twitter</p> </div>

UK Reading Experience Database	
Referencia bibliográfica	Charles Allen, <i>Kipling Sahib</i> , (Great Britain, 2007), p. 221, http://www.open.ac.uk/Arts/reading/UK/record_details.php?id=32448 , accessed: 02 August 2016
Identificador	<p>Record ID: 32448</p> <p>Source: <i>Print</i></p> <p>Author: Charles Allen</p> <p>Editor: n/a</p> <p>Title: Kipling Sahib</p>

UniPROBE Database	
Identificador	UNIPROBE ACCESSION NUMBER UP00611

World Values Survey	
Difusión en RRSS / GGBB	

Zenodo	
Referencia bibliográfica	Weisberg, Joel et al.. (2016). Binary Pulsar PSR B1913+16 arrival time and associated files. Ref: Weisberg & Huang APJ 2016. Zenodo. 10.5281/zenodo.54764
Identificador	DOI DOI 10.5281/zenodo.54764
Difusión en RRSS / GGBB	Share
Veces compartido	03 June 2016 Dataset Open access Binary Pulsar PSR B1913+16 arrival time and associated files. Ref: Weisberg & Huang APJ 2016 Weisberg, Joel ; Huang, Yuping <div style="float: right; border: 1px solid #ccc; padding: 5px; text-align: center;"> See more details Tweeted by 2 </div>