Revista Española de Documentación Científica 47 (3)
julio-septiembre 2024, e388
ISSN-L:0210-0614, eISSN: 1988-4621
https://doi.org/10.3989/redc.2024.3.1491

ESTUDIOS / *RESEARCH STUDIES*

Hacia un ecosistema de aplicación de buenas prácticas de publicación de datos en la Web: mapeado de las DWBP como medio para la implementación y el cumplimiento de los principios FAIR, FACT y CARE

Morgana Carneiro de Andrade

Universidade Federal do Espírito Santo, Brasil
Correo-e: morganaandrade@hotmail.com.es | ORCID iD: https://orcid.org/0000-0002-4055-5882

María-José Baños-Moreno

ODILO, España
Correo-e: mbm41963@um.es | ORCID iD: https://orcid.org/0000-0001-9137-1330

Juan-Antonio Pastor-Sánchez

Universidad de Murcia, España
Correo-e: pastor@um.es | ORCID iD: https://orcid.org/0000-0002-1677-1059

Recibido: 02-05-23; 2ª versión 04-10-23; 3ª versión 07-12-23; Aceptado: 12-12-23; Fecha de publicación: 16-10-24

Cómo citar este artículo/Citation: Carneiro de Andrade, M., Baños-Moreno, M. J., Pastor-S*ánchez*, J. A. (2024). Hacia un ecosistema de aplicación de buenas prácticas de publicación de datos en la Web: mapeado de las DWBP como medio para la implementación y el cumplimiento de los principios FAIR, FACT y CARE. *Revista Española de Documentación Científica*, 47 (2), e388. https://doi.org/ 10.3989/redc.2024.2.1491.

Resumen: La preocupación por la publicación y reutilización de los datos publicados en la web ha promovido la proliferación de iniciativas orientadas a mitigar los problemas generados por la ausencia de calidad en estos (buenas prácticas y principios). Sin embargo, algunas se han publicado de forma aislada o no han ido acompañadas de herramientas que hicieran posible su explicitación. Se pretende avanzar hacia un ecosistema de buenas prácticas de publicación de datos en la web y, más específicamente, 1) identificar las relaciones existentes entre las Data on the Web Best Practices (DWBP) y los principios FAIR, FACT y CARE; y 2) presentar las DWBP como un medio para la implementación y cumplimiento de los principios indicados. Se realizó una revisión bibliográfica para la identificación de trabajos relacionados con las buenas prácticas y principios estudiados. Después, se analizaron estas iniciativas, mapeando las DWBP con los principios. Se identificaron las correspondencias entre las buenas prácticas y los principios FACT, FAIR y CARE, de modo que DWBP sirva como medio para la implementación y cumplimiento de dichos principios. Además, se identificaron varios bloques temáticos relacionados con la calidad de los datos en la web, especificando cuáles son tratados por las iniciativas analizadas y cuáles no. Las DWBP satisfacen en buena medida los principios FACT, FAIR y CARE. Aun así, tanto unas como otros no tratan todos los aspectos necesarios para asegurar la mayor calidad de los datos. Se propone la creación de un ecosistema de buenas prácticas para la publicación de datos en la Web, en el que DWBP, FAIR, FACT y CARE constituyen elementos esenciales, pero no únicos pues, como se ha indicado, no tratan aspectos esenciales para la calidad de los datos.

Palabras-clave: datos en la web, principios FACT, principios FAIR, principios CARE, data on the web best practices.

Towards an ecosystem for the application of good data publishing practices on the Web: mapping of DWBPs as a means for the implementation and compliance of the FAIR, FACT and CARE principles

Abstract: Concern about the publication and reuse of data published on the web has promoted the proliferation of initiatives aimed at mitigating the problems generated by the lack of quality in these (good practices and principles). However, some have been published in isolation or have not been accompanied by tools that would make their explanation possible. The aim is to move towards an ecosystem of good practices for publishing data on the web and, more specifically, 1) identify the relationships between Data on the Web Best Practices (DWBP) and the FAIR, FACT and CARE principles; and 2) present the DWBP as a means for the implementation and compliance of the indicated principles. A bibliographic review

was carried out to identify works related to the good practices and principles studied. These initiatives were then analyzed, mapping the DWBPs to the principles. Correspondences between good practices and the FACT, FAIR and CARE principles were identified, so that DWBP serves as a means for the implementation and compliance of these principles. In addition, several thematic blocks related to the quality of data on the web were identified, specifying which ones are treated by the analyzed initiatives and which ones are not. The DWBPs largely satisfy the FACT, FAIR and CARE principles. Even so, both do not address all the necessary aspects to ensure the highest quality of the data. The creation of an ecosystem of good practices for the publication of data on the Web is proposed, in which DWBP, FAIR, FACT and CARE constitute essential elements, but not the only ones since, as indicated, they do not address essential aspects for quality. of the data.

Keywords: data on the Web, FACT principles, FAIR principles, CARE principles, data on the web best practices.

Copyright: © 2024 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia de uso y distribución Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0).

1. INTRODUCCIÓN

En 2006, Berners-Lee proponía un conjunto de procedimientos que debían guiar la publicación de datos enlazados en la Web (*Linked Data*), un entorno en el que los datos podían conectarse a partir de la semántica asociada a ellos, permitiendo a personas y máquinas explotarlos (Berners-Lee, 2006, Bizer y otros., 2008). Con los años, han surgido diferentes propuestas, ya relacionadas con el paradigma de datos abiertos enlazados (*Linked Open Data*), identificándose diversos problemas y limitaciones. Entre estas propuestas, destacan las buenas prácticas en la publicación de datos en la Web. Una buena práctica *no es un consejo, una recomendación o una opinión. Es una norma universal, útil, documentada, sin valor numérico, que ha obtenido un consenso, que no está ligada a un país ni a una ley concreta* (Sloïm, 2021).

Por otro lado, destaca el concepto de principio, para Berrino Malaccorto (2023) es una práctica central que a la que gobiernos, organizaciones, sociedad civil y otras partes interesadas deben adherirse. Según se desprende de Wilkinson y otros. (2016), un principio precede a la implementación de la opción elegida para su consecución, es una guía. En este trabajo, una buena práctica es entendida como una forma de explicitar uno o varios principios, es decir, es una de las soluciones que hace posible el cumplimiento de un principio.

En un estudio exploratorio, se identificaron cuatro de las primeras buenas prácticas relativas a la publicación de datos en la Web (Andrade y otros, 2023): así, en 2006, inicialmente con la atención puesta en los datos del sector público, Berners-Lee presenta el sistema 5 star - Linked Open Data (LOD), donde se consigue mayor número de estrellas conforme se adoptan los procedimientos indicados (Berners-Lee, 2006). En 2011, se dan a conocer las Open Data Best Practices, un conjunto de 72 buenas prácticas para la publicación de datos abiertos (Romain, 2013, Romain y otros., 2012). El proyecto ha evolucionado, incluyendo, entre otros, listas de verificación para evaluar portales web, culminando en 2020 con la publicación de 240 normas de calidad de la web y actividades de certificación de calidad; en 2014, Goodman y otros (2014) publicaron las 10 Simple Rules for the Care and Feeding of Scientific Data, para la gestión de datos para la comunidad científica. Los autores especificaban que es recomendable proporcionar información para que los datos y el código fuente puedan ser reutilizados. Aunque se hace hincapié en los datos de investigación, las normas, definidas por los autores como "consejos prácticos", son generalmente aplicables a cualquier contexto; también en 2014, el World Wide Web (W3C) publicó las Best Practices for Publishing Linked Data (Hyland, y otros, 2014), con el propósito de facilitar la creación y publicación de datos gubernamentales como *Linked Open Data*, lo que ya evidencia una evolución hacia la publicación de datos en la Web. El documento recopila las prácticas más relevantes para la gestión y publicación de datos con alta calidad por parte de los gobiernos de todo el mundo.

Si bien estas propuestas son reseñables por su carácter histórico, en Andrade, y otros (2023) se destacaban otras buenas prácticas, pero también principios más recientes: Findable, Accessible, Interoperable and Reusable (FAIR), Data on the Web Best Practices (DWBP), así como Fairness, Accuracy, Confidentiality and Transparency (FACT) y Collective Benefit, Authority to Control, Responsibility, and Ethics (CARE). En el presente trabajo se analizan estas cuatro recomendaciones con el objetivo general de avanzar hacia un ecosistema de buenas prácticas de publicación de datos en la web. Más específicamente, se pretende 1) identificar las relaciones existentes entre DWBP y los principios FAIR, FACT y CARE, así como la ausencia de cobertura de aspectos relevantes en relación con la calidad de los datos en la web; y 2) presentar las DWBP como un medio para la implementación y cumplimiento de los principios indicados.

Los siguientes apartados se estructuran de la siguiente forma: la sección 2, Metodología, describe los procedimientos de recopilación de la bibliografía, análisis y alineamiento de las cuatro iniciativas objeto de estudio; la sección 3, Resultados y discusión, presenta, de forma resumida, las prácticas DWBP, y principios FAIR, FACT y CARE, algunas de sus características e implicaciones y cómo se relacionan, también se discuten los resultados obtenidos; la sección 4, Consideraciones finales y trabajos futuros, concluye con una síntesis de los descubrimientos principales y sugerencias para trabajos futuros.

2. Metodología

Se realiza una revisión integrativa sobre un conjunto de buenas prácticas y principios para la publicación de datos en la web. La revisión integrativa es una metodología de análisis que posibilita la síntesis de conocimientos y la aplicación de resultados de los estudios más relevantes de una práctica concreta (Souza, Silva & Carvalho, 2010). Este tipo de metodología permite resumir el estado actual del conocimiento sobre un tema específico e identificar lagunas que requieren una investigación adicional (Souza y otros, 2010; Valmorbida, y otros, 2018; Whittemore y Knafl, 2005), que combina datos de diferentes tipos de fuentes, incorpora conceptos; revisa teorías y evidencias, analiza problemas metodológicos y permite el abordaje cualitativo o cuantitativo de un tema en concreto (Whittemore & Knafl, 2005). La revisión integrativa es utilizada para apoyar (soportar) estudios basados en evidencias

De entre los modelos de revisión integrativa existentes, se optó por Botelho y otros (2011), que se organiza en seis etapas:

1. Determinación del tema objeto de estudio, las cuestiones de investigación, las bases de datos y los términos (ecuaciones de búsqueda) a utilizar en el análisis bibliográfico (Tabla I).
2. Definición de la muestra según criterios de inclusión y exclusión de artículos y estudios.
3. Preselección inicial de artículos y estudios a partir de la muestra obtenida.
4. Selección final y categorización del conjunto de artículos y estudios de trabajo.
5. Análisis e interpretación de resultados (análisis de las buenas prácticas, en este caso).
6. Presentación de la síntesis del conocimiento.

El periodo de cobertura se sitúa entre 2010 y 2022, ya que 2010 es el año en que se inicia la

Tabla I. Estrategias de búsqueda empleadas en el análisis bibliográfico

Tema	Base de datos	Ecuaciones de búsqueda
FACT principles	SCOPUS	"Fairness AND Accuracy AND Confidentiality AND Transparency"
		"Responsible Data Science" OR "Green Data Science"
	Web of Science	"Fairness AND Accuracy AND Confidentiality AND Transparency"
		"Responsible Data Science" OR "Green Data Science"
	Google s Scholar	"Fairness AND Accuracy AND Confidentiality AND Transparency" OR "FACT principles"
		"Responsible Data Science" OR "Green Data Science"
FAIR principles	SCOPUS	"FAIR principles" AND data
	WOS	"FAIR principles" AND data
	Google Scholar	"FAIR principles" AND data
CARE principles	SCOPUS	"CARE principles" AND "Indigenous data"
	WOS	"CARE principles" AND "Indigenous data"
	Google. Scholar	"CARE principles" AND "Indigenous data"
Data on the Web Best Practices	SCOPUS	"data on the web best practices"
	WOS	"data on the web best practices"
	Google. Scholar	"data on the web best practices"

divulgación de recomendaciones para la publicación de datos en la Web. En la estrategia de búsqueda se incluyeron artículos en inglés, español y portugués, indexados por las bases de datos *Web of Science* y *Scopus*. En cuanto a los criterios de inclusión y exclusión de documentos en la muestra, se tuvieron en cuenta artículos de investigación y de revisión de literatura, informes y documentos publicados por órganos de reconocimiento internacional que tratan sobre conceptos, interpretaciones o abordajes sobre los principios y prácticas. Como muchas veces la literatura gris no se recoge en las bases de datos bibliográficas, también se utilizó el motor de búsqueda Google Scholar. Quedaron fuera del análisis los estudios de caso, artículos sobre herramientas y modelos, así como aquellos documentos no disponibles a texto completo.

Después se realizó una categorización de los trabajos, dando como resultado una matriz que consta de los siguientes campos: autoría, identificación de la buena práctica, objetivo, metodología, síntesis de las recomendaciones/conclusiones. Para la gestión de las referencias bibliográficas se utilizó el software Zotero, que permite organizar los resultados de investigación en carpetas por términos y por bases de datos (*Scopus/Wos/Google Scholar*), así como eliminar las duplicidades.

A continuación, se analizaron cada una de las iniciativas objeto de estudio de este trabajo. Se presenta una breve descripción en el apartado 3.2.

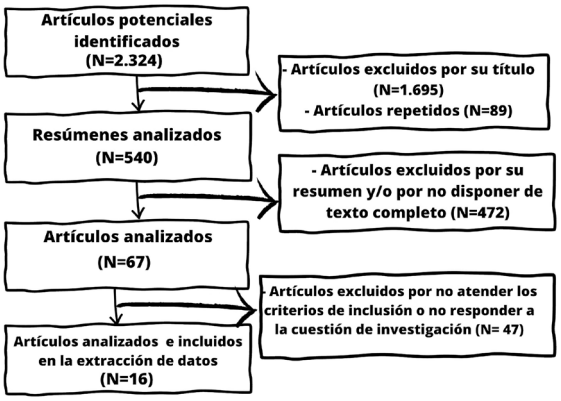
Posteriormente, se identificaron las relaciones entre DWBP y FAIR, FACT y CARE, esto es, cómo las buenas prácticas podían contribuir a la explicitación de los principios. Primero, fue elaborada una plantilla para cada una, compilando así los datos de los 16 artículos que componen la muestra. En el proceso, fueron considerados abordajes conceptuales, semánticos y prácticos referentes a las 4 propuestas. Es importante destacar que no en todos los trabajos es posible identificar de forma clara o explícita algunos conceptos, como ocurre, por ejemplo, con "nivel de formalización", "metadatos enriquecidos", "datos enriquecidos", "imparcialidad", "justicia" o "relaciones positivas", entre otros. En esta etapa, se tuvo en cuenta sobre todo el modelo de Buenas Prácticas (BP) presentado por la W3C (Lóscio y otros, 2017) que incluye el porqué de la buena práctica, resultado esperado, posible abordaje para su implementación y cómo probarlo, también evidencia su utilización y beneficios asociados. Como resultado, se obtuvo una tabla de afinidades de Buenas prácticas (BP) y principios y, finalmente, se establecieron las categorías temáticas que eran cubiertas o no por cada iniciativa.

3. RESULTADOS Y DISCUSIÓN

3.1 Revisión bibliográfica

Con la investigación bibliográfica se pre-seleccionaron 67 artículos alineados con la propuesta de estudio que, tras aplicar los criterios de elegibilidad y verificar la potencialidad de respuesta a la pregunta de investigación, se redujeron a 16 trabajos. La Figura 1 muestra un resumen de los procedimientos adoptados para la configuración de esta muestra.

Figura 1. Proceso de selección de los artículos



Como se observa en esta figura se produce una reducción significativa de la cantidad de trabajos analizados en relación con el total de documentos identificados inicialmente. Esto tiene que ver con el propio tema objeto de estudio, pues únicamente se trataron aquellas publicaciones que analizaban las prácticas DWBP, así como los principios FAIR, FACT y CARE. La casi totalidad de los artículos eliminados corresponde a trabajos que abordaban estas iniciativas como parámetros, o hacía referencia a su uso en la evaluación de portales, para el desarrollo de herramientas, métricas o modelos. Otro factor que influyó en el recorte fue la ambigüedad de los términos *FACT principles* y *FAIR principles*.

Los 16 artículos seleccionados fueron categorizados en cinco facetas: autoría, práctica(s)/principio(s) abordado(s), metodología empleada para la realización de los estudios y una síntesis de la conclusión (Tabla II).

Descripción de las iniciativas analizadas

A continuación, se presentan y describen las buenas prácticas y los tres principios estudiados.

Rev. Esp. Doc. Cient., 47(3), julio-septiembre 2024, e388. ISSN-L: 0210-0614. https://doi.org/10.3989/redc.2024.3.1491

Tabla II. Categorización de los artículos analizados.

Autores	Práctica	Objetivo	Metodología	Síntesis de la conclusión
Andrade y otros (2020)	FACT	Identificar orientaciones e iniciativas que viabilicen la implementación de las directrices FACT.	Exploratoria Bibliográfica Documental	Describe las iniciativas de la *Responsible Data Science* y de la *Green Data Science* y presenta definiciones de los términos imparcialidad, precisión, confiabilidad y transparencia, en consonancia con las directrices FACT, que buscan salvaguardar los derechos individuales y la ética en el acceso y reutilización de los datos.
Carroll y otros. (2020)	CARE	Presentar los principios CARE.	Descriptiva	Explican la lógica de los principios CARE y describen su relación los los FAIR. También presentan ejemplos de su aplicación.
Carroll y otros (2021)	CARE FAIR	Operacionalizar los principios CARE con los principios FAIR.	Descriptiva	Describe los principios CARE en el contexto de los datos científicos y los FAIR en el de los datos de indígenas; cómo se cruzan; implicaciones y próximos pasos para cambios que permitan la puesta en marcha de FAIR con CARE.
De Jong y otros (2019)	FAIR FACT	Explicar la estrategia y la implementación de la Infraestructura Europea de Investigación CLARIN alineada a los principios FAIR y FACT.	Descriptiva	Presenta la estrategia ejecutada por la CLARIN para adoptar los principios FAIR y alinearse con la iniciativa Responsible Data Science con la adopción de los principios FACT.
Delgado y otros (2021)	FAIR TRUST	Integrar los principios TRUST a los principios FAIR para garantizar la confiabilidad de los datos y de softwares.	Investigación aplicada	Presenta el mapeado de los principios TRUST para el flujo de trabajo de la "FAIRificación". Los autores destacan que la aplicación de los dos principios es fundamental para el desarrollo de softwares abiertos y confiables para el acceso a los datos de salud.
Henning y otros (2018)	FAIR	Presentar los principios FAIR y contextualizarlos en el área de las Ciencia da Información.	Revisión de la literatura	Presenta la fundamentación teórica y conceptual de los principios FAIR.
Jacobsen y otros (2020)	FAIR	Proporcionar subsidios para la implementación de los principios FAIR.	Descriptiva	Proporciona interpretaciones y consideraciones de los 15 principios FAIR.
Landi y otros (2020)	FAIR	Proporcionar orientaciones para la implementación de los principios FAIR.	Descriptiva	Propone la inclusión de los derechos individuales en la implementación de los principios FAIR y orienta hacia el procesamiento de datos personales basándose en principios éticos y legales.
Lóscio, Burle y otros (2017)	DWBP	Presenta un conjunto de buenas prácticas para la publicación de datos en la Web.	Descriptiva	Presenta 35 buenas prácticas para la publicación de datos, los beneficios alcanzados con su adopción y sugerencias de elementos de metadatos para el cumplimiento de esas prácticas.
Silva, y otros (2019)	FAIR	Identificar posibilidades de publicación de datos de investigación siguiendo los principios FAIR a partir de la adopción de las tecnologías de la Web Semántica, por medio del Linked Data.	Descriptiva exploratoria	Presenta un análisis de los principios FAIR con las Mejores Prácticas de Datos Enlazados (W3C).

Autores	Práctica	Objetivo	Metodología	Síntesis de la conclusión
Torino y otros (2021)	FAIR CARE	Discutir los principios FAIR y CARE para la gestión de los datos de investigación con seres humanos de comunidades específicas.	Descriptiva	Describe los principios FAIR y CARE y presenta prácticas referentes a los principios CARE, denominadas #SejaJUSTOeCUIDADOSO. Recomienda la adecuación de planes de datos a esas prácticas en todo el proceso de investigación y en el ciclo de vida de los dados.
van der Aalst (2016)	FACT	Responder a la pregunta de cómo las técnicas de *process mining* pueden contribuir a que los datos sean imparciales, confiables, precisos y transparentes.	Descriptiva	Introduce el concepto Green Data Science, un conjunto de técnicas de minería de procesos para garantizar la imparcialidad, la confidencialidad, la precisión y la transparencia de los datos.
van der Aalst y otros (2017)	FACT	Discutir las tendencias de investigación basadas en datos en el área de Ingeniería de Sistemas Empresariales (BISE).	Descriptiva	El artículo propone el acrónimo FACT a partir de los términos Fairness, Accuracy, Confidentiality y Transparency para representar los cuatro principales desafíos de la *Data Science Responsável*. Incentiva, además, las investigaciones sobre cómo incorporar elementos FACT en los modelos y requisitos, buscando alternativas innovadoras en el uso de los datos de forma segura y controlada y evitando los efectos negativos que los datos mal utilizados pueden tener en las vidas de las personas.
van der Aalst (2017)	FACT	Presentar cuatro desafios específicos da ciencia de datos: imparcialidad, precisión, confidencialidad y transparencia.	Descriptiva	Propone soluciones tecnológicas para proteger a individuos, organizaciones y a la sociedad contra los "efectos negativos de los datos".
Wilkinson y otros (2016)	FAIR	Presentar los principios FAIR.	Descriptiva	Además de describir la lógica de los principios FAIR y proporcionar ejemplos de implementaciones, posiciona los principios FAIR como forma de mejorar la capacidad de las máquinas para encontrar y usar los dados para su reutilización por individuos.
Wilkinson y otros (2016)	FAIR	Describir los principios FAIR.	Descriptiva	Detalla los principios FAIR y presenta implementaciones de estos.

3.1.1. Data on the Web Best Practices (DWBP)

Comenzaron a ser definidas por el grupo de trabajo del W3C en 2013, con el objetivo de estimular y posibilitar la expansión de la Web para el intercambio de datos (Lóscio y otros, 2017). Es la misma entidad que promueve la *Best Practices for Publishing Linked Data*.

Se trata de 35 buenas prácticas que abarcan aspectos relacionados con la publicación y el consumo de datos, sus formatos y acceso, identificadores y metadatos (Tabla 3). En Lóscio y otros, (2017) se describe su composición, así como motivación, resultados esperados de su implementación y cómo realizarla, entre otros. También se especifican indicaciones sobre las propiedades que deben adoptar los metadatos para cada buena práctica, así como ejemplos de casos de uso. La Tabla III muestra las buenas prácticas y las dimensiones en que se categorizan.

En gran medida, esta propuesta se centra en la aportación de información a través de metadatos. Diversos autores destacan la relevancia de utilizarlos para mejorar el rendimiento de las tecnologías computacionales sofisticadas (Landi y otros, 2020). Se pueden emplear independientemente del dominio y de la aplicación, y pueden complementarse con otras prácticas que aborden contextos diferentes.

Tabla III. Data on the Web Best Practices – DWBP

Dimensión	Buena práctica
Metadatos	BP 1 – Proporcionar metadatos (legibles por máquinas y humanos).
	BP 2 – Proporcionar metadatos descriptivos.
	BP 3 – Proporcionar metadatos estructurales.
Licencia de datos	BP 4 – Proporcionar información sobre la licencia de datos.
Origen de los datos	BP 5 – Proporcionar información sobre el origen de los datos.
Calidad de datos	BP 6 – Proporcionar información de la calidad de los datos.
Versión de los datos	BP 7 – Indicar la versión.
	BP 8 – Proporcionar un histórico de la versión.
Identificadores de datos	BP 9 – Usar URLs persistentes como identificadores de conjuntos de datos.
	BP 10 – Usar URLs persistentes como identificadores dentro de conjuntos de datos.
	BP 11 – Atribuir URLs para las versiones de los conjuntos de datos y series.
Formatos de datos	BP 12 – Usar formatos de datos estandarizados legibles por máquinas.
	BP 13 – Usar representaciones de datos que sean independentes de su localización.
	BP 14 – Proporcionar los datos en varios formatos.
Vocabularios de datos	BP 15 – Reutilizar vocabularios, dando preferencia a los estándares.
	BP 16 – Escoger el nivel de formalización adecuado.
Acceso a los datos	BP 17 – Proporcionar la descarga global (bulk download).
	BP 18 – Proporcionar subconjuntos para conjuntos de datos extensos.
	BP 19 – Usar la negociación de contenido para disponibilizar los datos en formatos múltiples.
	BP 20 – Proporcionar acceso en tiempo real.
	BP 21 – Proporcionar datos actualizados.
	BP 22 – Explicar por qué los datos no están disponibles.
APIs de acceso a datos	BP 23 – Hacer que los datos estén disponibles mediante API.
	BP 24 – Usar estándares WEB como base para construcción de APIs.
	BP 25 – Proporcionar una documentación completa para las APIs.
	BP 26 – Evitar alteraciones que afecten al funcionamiento de su API.
Preservación de datos	BP 27 – Preservar los identificadores.
	BP 28 – Evaluar la cobertura del conjunto de datos.
Feedback	BP 29 – Recolectar el feedback de los consumidores de los datos.
	BP 30 – Compartir el feedback disponible.
Enriquecimiento de datos	BP 31 – Enriquecer los datos mediante la generación de nuevos datos.
	BP 32 – Proporcionar visualizaciones complementarias.
Republicación de datos	BP 33 – Proporcionar feedback para el publicador original.
	BP 34 – Obedecer los términos de licencia.
	BP 35 – Citar la publicación original del conjunto de datos.

Fuente: Lóscio y otros, (2017).

3.1.2. FAIR Principles: Findable, Accessible, Interoperable and Reusable

En 2016, un grupo de profesionales e investigadores definieron estos principios para satisfacer las necesidades de la comunidad científica, aunque están siendo adoptados en diferentes dominios y comunidades. Los FAIR abordan cuestiones prácticas relacionadas con el intercambio y la distribución de datos que deben ser atendidas por los recursos digitales para su cumplimiento.

Se trata de 10 subprincipios (algunos subdivididos en otros sub-subprincipios) agrupados en cuatro categorías:

- **Localizable** (f*indable*): el recurso digital debe ser localizable, accionable, legible y procesable por humanos y máquinas. Para ello, debe proporcionar metadatos accionables por máquinas, identificadores persistentes y un almacenamiento en infraestructura indexada.
- **Accesible** (a*ccesible*): es posible mediante el uso de protocolos de comunicación, autenticación, niveles de acceso y mecanismos como los identificadores persistentes. Debe haber declaraciones claras y explícitas sobre la autorización de acceso y los datos protegidos.
- **Interoperable** (i*nteroperable*): implica que las máquinas puedan procesar e interactuar automáticamente con la información y los servicios. Para ello, se deben adoptar normas que garanticen la semántica, mediante el uso de vocabularios globalmente reconocidos, y protocolos que permitan el tratamiento de datos por parte de mecanismos de inteligencia artificial, entre otros.
- **Reutilizable** (r*eusable*): puede optimizarse mediante la descripción de los datos, información clara de su procedencia y de las condiciones de reutilización (Wilkinson y otros, 2016, Jacobsen y otros, 2020).

La Tabla IV presenta un resumen de estos principios y subprincipios. Varios de los principios FAIR hacen referencia a los "(meta)datos" (con "meta" entre paréntesis), porque el principio correspondiente se aplica tanto a los metadatos como a los datos.

Su cumplimiento se logra a través de la implementación técnica y/o semántica, lo que permite a

Tabla IV. Principios FAIR

Principios	Subprincipios
Localización	F1. A los (meta)datos se les asigna un identificador globalmente exclusivo y persistente.
	F2. Los datos deben ser descritos con metadatos enriquecidos (definidos por R1 más abajo).
	F3. Los metadatos incluyen clara y explícitamente el identificador de los datos que describen.
	F4. Los (meta)datos son registrados o indexados en un recurso de búsqueda.
Accesibilidad	A1. Los (meta)datos son recuperables por su identificador utilizando un protocolo de comunicación estandarizado.
	A1.1. El protocolo es abierto, gratuito y universalmente implementable.
	A1.2. El protocolo permite un procedimiento de autenticación y autorización cuando es necesario.
	A2. Los metadatos son accesibles, incluso cuando los datos no están disponibles.
Interoperabilidad	I1. Los (meta)datos usan un lenguaje formal, accesible, compartible y ampliamente aplicable para la representación del conocimiento.
	I2. Los (meta)dados utilizan vocabularios y/o ontologías que siguen los principios de FAIR.
	I3. Los (meta)dados incluyen referencias cualificadas para otros (meta)datos.
Reusabilidad	R1. Los (meta)datos se describen detalladamente con atributos precisos y relevantes.
	R1.1. Los (meta)datos se publican con una licencia de uso de datos clara y accesible.
	R1.2. Los (meta)datos están asociados con su origen de forma detallada.
	R1.3. Los (meta)datos cumplen los estándares comunitarios relevantes para el dominio.

Fuente: Traducido de Wilkinson y otros. (2016)

la inteligencia artificial manejar los recursos digitales (Jacobsen y otros, 2020, Wilkinson y otros, 2016) y al ser humano, comprender la información que se maneja.

En cuanto a la puesta en práctica de los principios, destacan la iniciativa de la Infraestructura Europea de Investigación CLARIN en el alineamiento de los principios FAIR y la *Responsible Data Science,* con la adopción de los principios FACT (De Jong y otros., 2019).

3.1.3. FACT Principles: Fairness, Accuracy, Confidentiality, Transparency

Derivan de una iniciativa del proyecto *Responsible Data Science*, fuertemente asociados "a los nuevos riesgos y abusos, intencionados o no, que permiten las potentes técnicas de análisis" de datos (van der Aalst, 2017). En este sentido, tienen como objetivo "[...] hacer frente a los desafíos éticos y legales, así como desarrollar técnicas de ciencia de datos, infraestructuras y enfoques responsables para procesos justos y precisos, y para los datos confidenciales y transparentes desde su concepción" (AIMS, 2017). Pretenden restringir las decisiones injustas, las revelaciones no deseadas, imprecisiones y la opacidad. Se trata de garantizar la calidad y la seguridad en el uso de los datos (Stoyanovich y Howe, 2018, van der Aalst, 2016), evitando la divulgación de información clasificada y que, al mismo tiempo, los resultados de los análisis informáticos puedan entenderse y aclararse para que sean indiscutibles y, por tanto, fiables. Para afrontar estos retos, van der Aalst (2017) aboga por una ciencia de datos responsable basada en cuatro principios:

- **Imparcialidad** (f*airness*): las decisiones y percepciones automatizadas no deben utilizarse para discriminar a ningún individuo o incluso una determinada acción desde un punto de vista legal o ético.
- **Precisión** (a*ccuracy*): está relacionado con la veracidad y la fiabilidad de los datos. En el caso de que los datos sean inexactos o estén incompletos, se debe permitir que se eliminen (derecho al olvido) y restringir su tratamiento, permitiendo su almacenamiento, pero rechazando el tratamiento de los datos personales.
- **Confidencialidad** (c*onfidenciality*): tiene que ver con el derecho a la intimidad que, a su vez, permite un tratamiento justo y transparente de cómo se accede, cómo se utilizan y cómo se manejan los datos personales, que pueden revelar información sensible (la identidad económica, social, cultural, mental, física, las opiniones políticas, los datos biométricos, entre otros) (van der Aalst, 2017, 2019; van der Aalst y otros., 2017). También está relacionado con la fiabilidad de cualquier información que pueda identificar a un individuo, tanto si es el resultado de combinaciones entre diferentes informaciones, como si se obtiene de forma autónoma (Mougiakou y Virvou, 2018).
- **Transparencia** (t*ransparency*): garantiza que todo el tratamiento de datos relacionados con la privacidad, incluidos los aspectos legales, técnicos, organizativos y de procedimiento, pueda interpretarse y reproducirse en cualquier momento. La información debe estar disponible antes, durante y después del tratamiento.

No se explicita qué procedimientos seguir para su consecución, por lo que dejan abiertas diferentes vías como, por ejemplo, a través de Reglamento General de Protección de Datos Europeo (*General Data Protection Regulation* - GDPR) o similar (Schlehahn y Wenning, 2019, van der Aalst, 2016).

Tabla V. Principios FACT

Principio	**Elementos**
F - Imparcialidad	Ciencia de los datos sin prejuicios - ¿cómo evitar conclusiones injustas, aunque sean ciertas?
A - Precisión	Ciencia de los datos sin conjeturas - ¿cómo responder a preguntas con un nivel de precisión garantizado?
C - Confidencialidad	Ciencia de los datos que garantice la confidencialidad - ¿cómo responder a preguntas sin revelar secretos?
T - Transparencia	Ciencia de los datos que proporciona transparencia - ¿cómo esclarecer las respuestas para que sean indiscutibles?

Fuente: van der Aalst (2017).

La tabla anterior muestra estos principios y las preguntas a las que responder para su cumplimiento, en el marco de que "cualquier uso de datos sesgado, que viole la privacidad o la confidencialidad, o que carezca de transparencia, puede distorsionar las conclusiones o romper las relaciones de confianza" (De Smedt y otros, 2018).

3.1.4. CARE Principles: Collective Benefit, Authority to Control, Responsibility, Ethics

Las bases de los principios CARE pueden encontrarse en los debates entre instituciones de países como Australia, Canadá y Estados Unidos sobre la Soberanía de Datos de los Indígenas y la Gobernanza de Datos de los Indígenas (Lovett y otros, 2019). Fueron propuestos en 2020 por el Grupo de Interés Internacional de Soberanía de los Datos de Indígenas de la *Research Data Alliance* (RDA) para la gobernanza de los datos y sobre los indígenas.

Esta iniciativa fue desarrollada para complementar a los FAIR y guiar en la inclusión de los pueblos indígenas en el procesamiento de los datos para mejorar el descubrimiento, acceso, uso, reutilización y atribución de datos en entornos actuales para esta comunidad. Estos 4 principios, descritos a continuación, son extensibles a otras comunidades que necesiten adoptar acciones de gobierno similares (Carroll y otros, 2020, 2021):

- **Beneficio colectivo** (*collective benefit*): establece que "los ecosistemas de datos deben ser diseñados y operados de manera que permitan a los pueblos indígenas beneficiarse de los datos".
- **Autoridad para controlar** (a*uthority to control*): la gobernanza de los datos sobre las tierras de los indígenas, los territorios, los recursos, los conocimientos y los indicadores geográficos debe ser llevada a cabo por los pueblos indígenas y por los organismos gubernamentales.
- **Rendición de cuentas** (r*esponsability*): establece que quienes trabajan con los datos de los indígenas deben compartir información sobre cómo se utilizan los datos para apoyar la autodeterminación y el beneficio colectivo.
- **Ética** (e*thics*): establece que los derechos y el bienestar de los pueblos indígenas deben ser una preocupación primordial durante todo el ciclo de vida de los datos y en todo el ecosistema de datos.

La Tabla VI presenta los cuatro principios CARE y sus subprincipios asociados.

Según Torino y otros, (2021), el cuidado y la justicia asociados a los datos de los pueblos indígenas pueden lograrse siguiendo los procedimientos presentados en la Tabla VII.

Los Principios CARE se centran no sólo en capacitar a los indígenas para que participen activamente en la reutilización de datos, sino también en la autonomía para tomar decisiones sobre sus datos. Se percibe la preocupación por garantizar la protección de los datos e intereses de los indígenas sin restringir la apertura, el intercambio y el aprendizaje automático por parte de los investigadores, gestores y usuarios, pero de forma justa y cuidadosa (Carroll y otros, 2021, 2020; Lovett y otros, 2019).

Tabla VI. Principios CARE

Principio	Subprincipio
Beneficio Colectivo	C1 – Para el desarrollo inclusivo y la innovación.
	C2 – Para mejorar la gobernanza y el compromiso de los ciudadanos.
	C3 – Para resultados equitativos.
Autoridad para controlar	A1 – Reconocimiento de los derechos e intereses.
	A2 – Datos para la gobernanza.
	A3 – Gobernanza de los datos.
Responsabilidad	R1 – Para relaciones positivas.
	R2 – Para expandir la competencia en la capacidad.
	R3 – Para lenguas y visiones de mundo indígenas.
Ética	E1 – Para minimizar daños y maximizar beneficios.
	E2 – Por justicia.
	E3 – Para uso futuro.

Fuente: Carroll y otros. (2020)

Tabla VII. Procedimientos propuestos en relación con los Principios CARE

Datos justos	Datos cuidadosos
Atención a la justicia y al cuidado con los datos	Foco en el desarrollo inclusivo
Uso de metadatos para la descripción del catálogo de datos y conjuntos de datos	Establecimiento de relaciones equitativas, de confianza, recíprocas y de respeto
Adopción de vocabularios estandarizados	Cumplimiento de los preceptos éticos y legales en la recogida, tratamiento, almacenamiento y disponibilidad de los datos
Adopción de identificadores persistentes	Identificación de los datos
Elección del entorno digital para el almacenamiento	Protección a los derechos, intereses, valores y la cultura
Definición y cumplimento de la licencia de uso	Participación en la gobernanza y control sobre sus datos
Definición del período de embargo	Mejoría en la representación de los datos
Procedencia de los datos	Capacitación para el uso de los datos
Versionado de datos	
Preservación de los datos	
Reconocimiento de créditos	

Fuente: Torino y otros, (2021).

3.2. Relaciones y alineamiento de DWBP, FAIR, FACT y CARE

Actualmente, las iniciativas analizadas se presentan como silos, con componentes aislados. La propuesta de este trabajo es establecer relaciones de equivalencias complementariedades o superposiciones entre sus elementos. En este sentido, FACT y CARE abordan la ética de los datos, junto con otros aspectos como la imparcialidad, la privacidad, la seguridad y la responsabilidad (van der Aalst y otros, 2017), cuya semántica puede variar dependiendo del dominio o de la comunidad. FAIR y DWBP, por su parte, no abordan la ética de los datos de forma explícita, aunque sí que aportan aspectos que pueden contemplarlos. La Tabla VIII presenta una visión global y resumida de los propósitos de cada propuesta.

Bajo esta noción de interrelación entre las iniciativas, se puede empezar a construir un ecosistema de buenas prácticas para la calidad de los datos en

Tabla VIII. Resumen de las iniciativas analizadas

Iniciativa	Síntesis de la iniciativa
DWBP	Proporciona directrices para la creación de datos basados en la Web que sean fácilmente detectables e interconectados, además de directrices técnicas y operacionales para publicar, compartir y reutilizar datos en la Web de manera sustentable.
FAIR	Destacan la importancia de que los datos sean localizables, accesibles, interoperables y reutilizables para hacer posible la e-ciencia y la innovación. Una vez alcanzados, sería posible que las máquinas hicieran un uso real de los datos. En términos sencillos, FAIR exige que "la máquina sepa lo que queremos decir".
FACT	Enfatizan la imparcialidad, la precisión, la confiabilidad y la transparencia de los datos para promover el acceso y utilización de forma ética, responsable y confiable. Mientras que FAIR aborda cuestiones prácticas relacionadas con la compartición y distribución de los datos, FACT se centra más en los desafíos científicos fundamentales.
CARE	Tienen como propósito beneficiar a los pueblos indígenas a partir de la gobernanza de los datos, la autoridad para controlarlos, responsabilidad y ética de los datos. Según Carroll y otros (2021), la adopción de estos principios "promueve la participación equitativa en la reutilización de datos, lo que producirá resultados más igualitarios". Estos principios, concebidos inicialmente para la comunidad indígena, se hacen extensivos a otras comunidades.

la Web, donde se exponen los puntos de intersección entre FAIR, FACT, CARE y DWBP, representados en la Figura 2 (destacados a través de distintos colores). El avance en esta perspectiva puede contribuir al desarrollo de instrumentos que hagan posible la implementación de las iniciativas estudiadas como un conjunto único.

A continuación, en la Tabla IX se explicitan las relaciones de los principios FAIR, FACT y CARE, tomando como elemento principal de alineamiento las DWBP, definidas inicialmente en la figura anterior. La tabla está estructurada de la siguiente manera: la primera fila presenta las buenas prácticas de las DWBP; las filas representan los principios de FAIR, FACT y CARE, organizadas en los tres bloques respectivamente. Las celdas destacadas en color representan la intersección entre una buena práctica y un principio, según corresponda, evidenciando cómo una o varias buenas prácticas pueden cubrir el alcance de un principio.

Para ello, se tomó como referencia la estructura presentada por la W3C sobre las DWBP (Lóscio y otros, 2017), que fue la más relevante para esclarecer algunos conceptos e implementaciones sobre buenas prácticas de calidad de los datos.

Los artículos de Henning y otros (2018), Jacobsen y otros. (2020), Landi y otros (2020), Silva y otro. (2019), Wilkinson y otros (2016a, b) y Lóscio y otros (2017) contribuyeron con interpretaciones y consideraciones sobre los principios FAIR, facilitando el establecimiento de correspondencias entre FAIR y DWBP.

A diferencia de los anteriores, FACT y CARE carecen de estudios para la definición de ontologías y adopción de patrones de metadatos. De ahí la necesidad de trabajar también con artículos sobre GDPR (Schlehahn y Wenning, 2019) y/o sobre gobernanza y soberanía de datos indígenas (Lovett y otros, 2019) para interpretarlos y alinearlos con DWBP.

Figura 2. Alineamiento de los principios FAIR, FACT, CARE, y DWBP

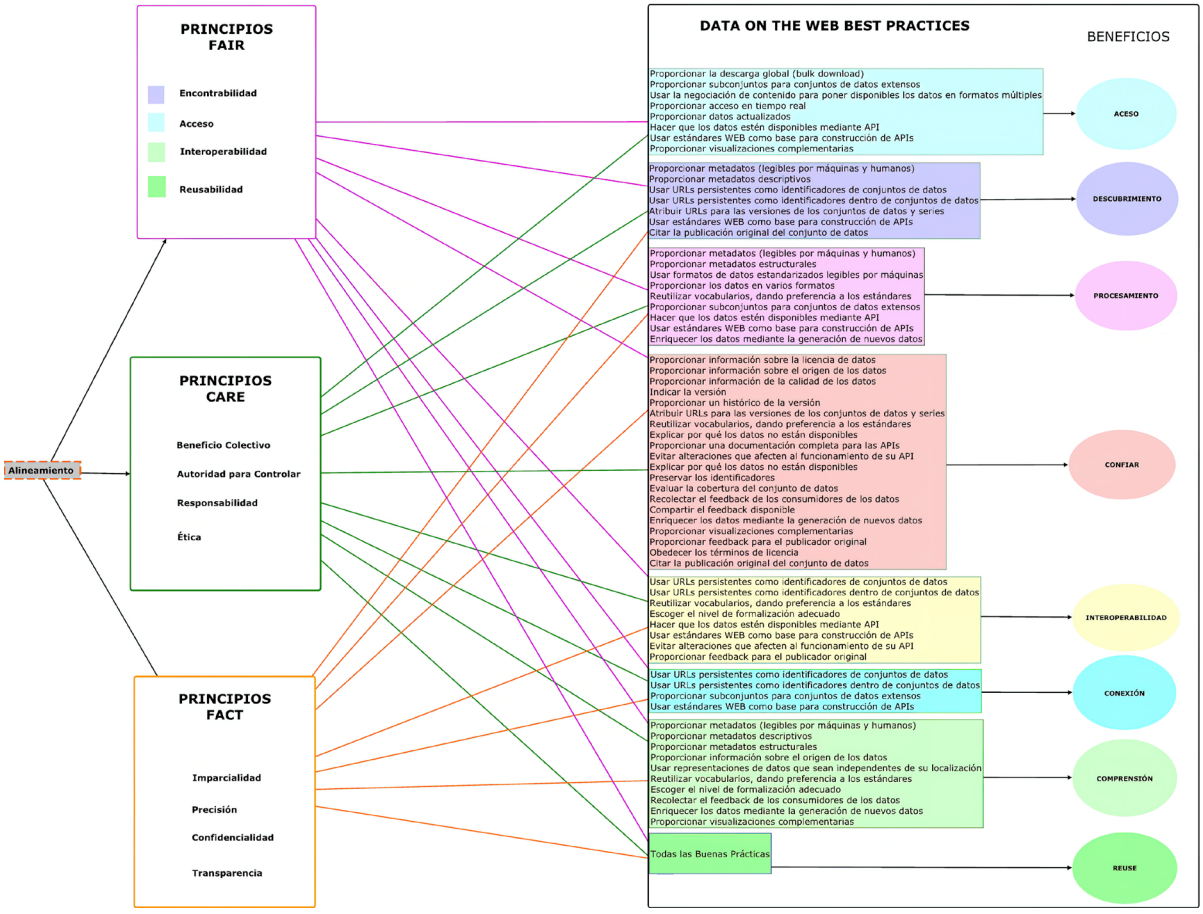


Tabla IX. Alineamiento de los principios y prácticas para la publicación de datos en la Web

	DATA ON THE WEB BEST PRACTICES (DWBP)	BP1 Proporcionar metadatos (legibles por máquinas y humanos)	BP2 Proporcionar metadatos descriptivos	BP3 Proporcionar metadatos estructurales	BP4 Proporcionar información sobre la licencia de datos	BP5 Proporcionar información sobre el origen de los datos	BP6 Proporcionar información de la calidad de los datos	BP7 Indicar la versión	BP8 Proporcionar un histórico de la versión	BP9 Usar URLs persistentes como identificadores de conjuntos de datos	BP10 Usar URLs persistentes como identificadores dentro de conjuntos de datos	BP11 Atribuir URLs para las versiones de los conjuntos de datos y series	BP12 Usar formatos de datos estandarizados legibles por máquinas	BP13 Usar representaciones de datos que sean independientes de su localización	BP14 Proporcionar los datos en varios formatos	BP15 Reutilizar vocabularios, dando preferencia a los estándares	BP16 Escoger el nivel de formalización adecuado	BP17 Proporcionar la descarga global (bulk download)	BP18 Proporcionar subconjuntos para conjuntos de datos extensos	BP19 Usar la negociación de contenido para poner disponibles los datos en formatos múltiples	BP20 Proporcionar acceso en tiempo real	BP21 Proporcionar datos actualizados	BP22 Explicar por qué los datos no están disponibles	BP23 Hacer que los datos estén disponibles mediante API	BP24 Usar estándares WEB como base para construcción de APIs	BP25 Proporcionar una documentación completa para las APIs	BP26 Evitar alteraciones que afecten al funcionamiento de su API	BP27 Preservar los identificadores	BP28 Evaluar la cobertura del conjunto de datos	BP29 Recolectar el feedback de los consumidores de los datos	BP30 Compartir el feedback disponible	BP31 Enriquecer los datos mediante la generación de nuevos datos	BP32 Proporcionar visualizaciones complementarias	BP33 Proporcionar feedback para el publicador original	BP34 Obedecer los términos de licencia	BP35 Citar la publicación original del conjunto de datos
FAIR	F1. A los (meta)datos se les asigna un identificador globalmente exclusivo y persistente									X	X	X																							
	F2. Los datos deben ser descritos con metadatos enriquecidos (definidos por R1 más abajo)	X	X	X																															
	F3. Los metadatos incluyen clara y explícitamente el identificador de los datos que describen	X	X	X						X		X																		X					
	F4. Los (meta)datos son registrados o indexados en un recurso de búsqueda												X		X											X									
	A1. Los (meta)datos son recuperables por su identificador utilizando un protocolo de comunicación estandarizado																			X					X	X									
	A1.1. El protocolo es abierto, gratuito y universalmente implementable																			X						X									
	A1.2. El protocolo permite un procedimiento de autenticación y autorización cuando es necesario																								X	X									
	A2. Los metadatos son accesibles, incluso cuando los datos no están disponibles																							X						X	X				
	I1. Los (meta)datos usan un lenguaje formal, accesible, compartible y ampliamente aplicable para la representación del conocimiento															X	X														X				
	I2. Los (meta)dados utilizan vocabularios y/o ontologías que siguen los principios de FAIR															X	X																		
	I3. Los (meta)dados incluyen referencias cualificadas para otros (meta)datos		X													X	X																		X
	R1. Los (meta)datos se describen detalladamente con atributos precisos y relevantes	X	X	X			X									X															X				

	DATA ON THE WEB BEST PRACTICES (DWBP)	BP1 Proporcionar metadatos (legibles por máquinas y humanos)	BP2 Proporcionar metadatos descriptivos	BP3 Proporcionar metadatos estructurales	BP4 Proporcionar información sobre la licencia de datos	BP5 Proporcionar información sobre el origen de los datos	BP6 Proporcionar información de la calidad de los datos	BP7 Indicar la versión	BP8 Proporcionar un histórico de la versión	BP9 Usar URLs persistentes como identificadores de conjuntos de datos	BP10 Usar URLs persistentes como identificadores dentro de conjuntos de datos	BP11 Atribuir URLs para las versiones de los conjuntos de datos y series	BP12 Usar formatos de datos estandarizados legibles por máquinas	BP13 Usar representaciones de datos que sean independientes de su localización	BP14 Proporcionar los datos en varios formatos	BP15 Reutilizar vocabularios, dando preferencia a los estándares	BP16 Escoger el nivel de formalización adecuado	BP17 Proporcionar la descarga global (bulk download)	BP18 Proporcionar subconjuntos para conjuntos de datos extensos	BP19 Usar la negociación de contenido para poner disponibles los datos en formatos múltiples	BP20 Proporcionar acceso en tiempo real	BP21 Proporcionar datos actualizados	BP22 Explicar por qué los datos no están disponibles	BP23 Hacer que los datos estén disponibles mediante API	BP24 Usar estándares WEB como base para construcción de APIs	BP25 Proporcionar una documentación completa para las APIs	BP26 Evitar alteraciones que afecten al funcionamiento de su API	BP27 Preservar los identificadores	BP28 Evaluar la cobertura del conjunto de datos	BP29 Recolectar el feedback de los consumidores de los datos	BP30 Compartir el feedback disponible	BP31 Enriquecer los datos mediante la generación de nuevos datos	BP32 Proporcionar visualizaciones complementarias	BP33 Proporcionar feedback para el publicador original	BP34 Obedecer los términos de licencia	BP35 Citar la publicación original del conjunto de datos
	R1.1. Los (meta)datos se publican con una licencia de uso de datos clara y accesible				X																													X	
	R1.2. Los (meta)datos están asociados con su origen de forma detallada		X			X		X	X			X																	X						X
	R1.3. Los (meta)datos cumplen los estándares comunitarios relevantes para el dominio		X				X									X	X												X		X			
FACT	Imparcialidad - Ciencia de los datos sin prejuicios		X	X						X	X	X		X		X																			
FACT	Precisión - Ciencia de los datos sin conjeturas		X	X	X		X						X	X		X	X				X	X							X						
FACT	Confidencialidad - Ciencia de los datos que la garantice		X	X	X		X																											X	
FACT	Transparencia - Ciencia de los datos que la proporciona	X	X	X	X	X	X	X	X															X		X		X		X	X		X		X
CARE	C1. Para el desarrollo inclusivo y la innovación	X	X	X			X			X	X	X	X	X	X	X	X	X	X	X		X		X	X	X	X	X	X	X	X				
CARE	C2. Para mejorar la gobernanza y el compromiso de los ciudadanos		X	X	X	X			X													X	X						X	X			X	X	X
CARE	C3. Para resultados equitativos		X			X								X		X													X	X	X		X		X
CARE	A1. Reconocimiento de los derechos e intereses			X	X						X																							X	X
CARE	A2. Datos para la gobernanza		X	X				X		X	X	X	X	X	X			X	X	X		X	X	X	X	X	X			X	X				
CARE	A3. Gobernanza de los datos				X																		X											X	
CARE	R1. Para relaciones positivas				X	X	X			X	X	X				X							X			X	X	X	X	X	X	X	X	X	X
CARE	R2. Para expandir la competencia en la capacidad				X																										X			X	X
CARE	R3. Para lenguas y visiones de mundo indígenas	X	X	X										X		X	X																		
CARE	E1. Para minimizar daños y maximizar beneficios		X	X	X											X	X													X	X		X	X	X
CARE	E2. Por justicia		X	X	X																													X	
CARE	E3. Para uso futuro	X	X		X	X				X	X	X																	X	X					X

Uno de los pocos estudios que analizan este tipo de correspondencias es Silva y otros. (2019). Los autores equipararon los subprincipios FAIR con los de DWBP. Sin embargo, contradiciendo los resultados del documento presente, no hallaron correspondencias para las siguientes prácticas: R1.3 (Los (meta)datos cumplen con los estándares de la comunidad para el dominio) y A1.2 (El protocolo permite un procedimiento de autenticación y autorización cuando es necesario). En la tabla anterior se muestran puntos de intersección entre BP2 (Proporcionar metadatos descriptivos), BP15

(Reutilizar vocabularios, dando preferencia a los estándares), BP16 (Escoger el nivel de formalización adecuado) y BP35 (Citar la publicación original del conjunto de datos) con R1.3; y BP23 (Hacer que los datos estén disponibles mediante API) y BP24 (Usar estándares WEB como base para construcción de APIs) con el principio A1.2.

Se identificó que la puesta a disposición de datos mediante API y el uso de los estándares Web facilita la autenticación y la autorización del uso de los datos. Es un aspecto que Romain y otros (2013) consideran importante para la seguridad de los datos. En relación con R1.3, el uso de metadatos descriptivos y vocabularios y niveles de formalización facilitan el intercambio de datos entre dominios, auxilian la recuperación de información y mejoran su consistencia.

Finalmente, para facilitar el entendimiento de cómo los principios y las DWBP pueden alinearse, se agruparon en las siguientes categorías temáticas:

- **Accesibilidad y localizabilidad**: los principios FAIR y las DWBP destacan la importancia de que los datos sean localizables, accesibles e interoperables. Las DWBP proporcionan orientaciones para lograr este propósito, especialmente en el contexto de la Web. Los principios CARE también ponen el énfasis en la accesibilidad de la comunidad indígena, especialmente como forma de empoderamiento. Los principios FACT cuando resaltan la transparencia, aspecto que contribuye directamente a los principios CARE.
- **Interoperabilidad y transparencia**: FAIR y las DWBP comparten el foco en la interoperabilidad que, a partir de las directrices proporcionadas por ambos, es posible alcanzar. También se subraya la importancia de que los datos sean reutilizables. CARE y FACT no abordan la interoperabilidad, pero es una cuestión esencial para cumplir el resto. La transparencia, aunque no se menciona explícitamente en FAIR y DWBP, está presente en la mayoría de sus directrices, a diferencia de lo que ocurre en los principios CARE e FACT, donde es asumida como concepto base.
- **Uso ético y responsable**: los CARE enfatizan las responsabilidades éticas de los datos, principalmente en lo que se refiere a la población indígena. Esto también es resaltado por los principios FACT, que destacan el uso responsable y ético de los datos. Sin embargo, a pesar de que los principios FAIR y las Buenas Prácticas posibilitan estos abordajes, estos aspectos no están claramente explicitados.
- **Gobernanza y control de los datos**: los principios CARE y FACT evidencian la preocupación por estas cuestiones, mientras que FAIR y las DWBP abordan aspectos relativos a la citación, al origen y a la descripción de los datos, también relacionados con dichos principios.
- **Reutilización de los datos**: las cuatro iniciativas tratan este tema. Los CARE destacan cómo se potencian los beneficios mediante el desarrollo inclusivo y la innovación, lo que está alineado con los principios FAIR. Facilitar la reutilización también es esencial para los FACT y DWBP para garantizar que los datos sean efectivamente reutilizados.
- **Derechos de autor y validez del archivo**: ninguno de estos asuntos fue tratado por ninguna de las cuatro iniciativas.

4. CONSIDERACIONES FINALES Y TRABAJOS FUTUROS

La apertura y la flexibilidad de la Web, así como la preocupación por las cuestiones técnicas y éticas, son retos tanto para los editores como para los consumidores de datos. Automatizar las condiciones de accesibilidad, conformidad con requisitos regulatorios, éticos y obligaciones de protección de datos es crucial para equilibrar la privacidad de las personas y los beneficios potenciales de la reutilización automatizada de datos para la investigación. Para los editores, se trata de gestiones relacionadas con la representación, descripción, organización y disponibilidad de los datos, mientras que los consumidores se enfrentan a incertidumbres como la privacidad, la fiabilidad y la responsabilidad en el uso de los datos. En este contexto, la formulación y disponibilidad de buenas prácticas para los editores permite que los datos alcancen niveles de calidad y confianza que permitan su uso y reutilización.

Existe un buen número de investigaciones centradas en los principios FAIR. Ahora bien, estos principios ponen el foco en los desafíos científicos fundamentales y están centrados en la calidad y seguridad de los datos. Las DWBP cubren más aspectos, particularmente en lo relativo a la retroalimentación derivada del uso de los datos.

Los principios FACT, más centrados en desafíos éticos y legales, complementan a los FAIR, especialmente en un contexto en el que el uso de métodos basados en datos, normalmente aplicados a grandes conjuntos de datos está aumentando (De Jong y otros., 2019). Casi todos los estudios asocian aplicaciones con algoritmos para el cumplimiento de los principios FACT. van der Aalst y otros (2017) destacan la necesidad de incorporarlos en

los modelos y requisitos, con la inclusión explícita de metadatos.

Por su parte, aunque los principios CARE han sido presentados recientemente, existen algunas propuestas para su implementación, incluso asociados a los FAIR (Carroll y otros., 2021; Torino y otros, 2021).

La identificación de las relaciones existentes entre los principios FAIR, CARE y FACT y las DWBP permite convertir estas buenas prácticas en un medio para la implementación y cumplimiento de los bloques de principios establecidos en el apartado anterior. Así se asegura que los datos son tratados de forma responsable, accesible y usable, facilitando el avance de la investigación científica, contribuyendo a la gobernanza de los datos y/o mejorando la accesibilidad y reutilización de los datos en la Web, dependiendo del grado de atendimiento de los principios.

Sin embargo, es necesario revisar y ampliar las DWBP, de forma que puedan satisfacer explícitamente algunos de los principios FACT y CARE, como si cubren, por ejemplo, las *Best Practices for Publishing Linked Data* (Hyland y otros, 2014). Se ha puesto en evidencia que existen lagunas relativas a las cuestiones éticas, la privacidad y la responsabilidad de los datos en las DWBP. Tampoco se han identificado trabajos en la muestra que presenten esquemas de representación con metadatos que aborden estos temas. Por otro lado, ni las buenas prácticas ni los principios recogen elementos específicos sobre derechos de autor ni validez del archivo de forma clara. La profundización en estos aspectos es necesaria para avanzar en una mejor publicación y reutilización de datos en beneficio de la sociedad. Es necesario completar y complementar las iniciativas analizadas con otras propuestas.

En este sentido se propone la creación de un ecosistema de buenas prácticas para la publicación de datos en la Web, en el que DWBP, FAIR, FACT y CARE constituyen elementos esenciales, pero no únicos pues, como se ha indicado, no tratan aspectos esenciales para la calidad de los datos. Esto permitirá la aplicación de una o de otra iniciativa (y de las buenas prácticas como vía de explicitación, entre otros) en función de las necesidades del usuario.

Finalmente, en lo que respecta al proceso de búsqueda de artículos sobre el tema analizado, destaca la dificultad para recuperar documentos que utilicen la expresión "buenas prácticas", así como la terminología relacionada en los campos de palabras clave o descriptores.

5. AGRADECIMIENTOS

La Dra. Morgana Andrade agradece a la Universidad Federal de Espírito Santo, donde es bibliotecaria, la oportunidad de realizar su postdoctorado y a la Universidad de Murcia por la acogida.

ACKNOWLEDGEMENTS

Dr. Morgana Andrade thanks the Federal University of Espírito Santo, where she is a librarian, for the opportunity to carry out her postdoctoral degree and the University of Murcia for the welcome.

6. DECLARACIÓN DE CONFLICTO DE INTERESES

Los autores de este artículo declaran no tener conflictos de intereses financieros, profesionales o personales que pudieran haber influido de manera inapropiada en este trabajo.

7. CONTRIBUCIÓN A LA AUTORÍA

Morgana Carneiro de Andrade: conceptualizacíon, investigación, metodología, Redacción – borrador original.

María José Baños Moreno: curación de datos, investigación, metodología, Redacción – revisión y edición.

Juan Antonio Pastor-Sánchez: Redacción – revisión y edición.

8. REFERENCIAS

AIMS. (2017). *Responsible data science. Ensuring, fairness, accuracy, confidentially, transparency*. Disponible en: https://aims.fao.org/fr/news/responsible-data-science-ensuring-fairness-accuracy-confidentially-transparency-fact.

Andrade, M. C., Baños-Moreno, M. J., y Pastor-Sánchez, J. A. (2023). Qualidade dos dados na web: revisão integrativa sobre diretrizes para publicação. *Perspectivas em Ciência da Informação*, 28, e40663.

Andrade, M. C., Gonçalez, P. R. V. A., Junior, D. W. B., Baptista, A. A., y Coneglian, C. S. (2020). Ciência responsável dos dados: Imparcialidade, precisão, confidencialidade, e transparência dos dados. *Informação y Informação*, 25(2), 26–48.

Berners-Lee, T. (2006). *Linked Data*. https://www.w3.org/DesignIssues/LinkedData.html

Berrino Malaccorto, R. (2023). *ODC's contribution to the Global Digital Compact – Report. Open Data and the Right to Privacy in a global digital strategy*. Disponible en: https://www.un.org/techenvoy/sites/www.un.org.techenvoy/files/GDC-submission_Open-Data-Charter.pdf.

Bizer, C., Heath, T., Idehen, K., y Berners-Lee, T. (2008). Linked data on the web (LDOW2008). *Proceedings of the 17th international conference on World Wide Web*, 1265–1266.

Botelho, L. L. R., de Almeida Cunha, C. C., y Macedo, M. (2011). O método da revisão integrativa nos estudos organizacionais. *Gestão e sociedade*, *5*(11), 121–136.

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, y otros. (2020). The CARE principles for indigenous data governance. *Data Science Journal*, 19(1), 1–12. DOI: https://doi.org/10.5334/DSJ-2020-043.

Carroll, S., Herczog, E., Hudson, M., Russell, K., y Stall, S. (2021). Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Scientific Data*, 8(1). DOI: https://doi.org/10.1038/s41597-021-00892-0.

De Jong, F., Maegaard, B., De Smedt, K., Fišer, D. y Van Uytvanck, D. (2019). Clarin: Towards fair and responsible data science using language resources. En Isahara H., Maegaard B., Piperidis S., Cieri C., Declerck T., Hasida K., Mazo H., Choukri K., Goggi S., Mariani, y otros (orgs.), *LREC - Int. Conf. Lang. Resour. Evaluation*, 3259–3264. European Language Resources Association (ELRA).

De Smedt, K., Jong, F., Maegaard, B., Fišer, D., y Van Uytvanck, D. (2018). Towards an open science infrastructure for the digital humanities: The case of CLARIN. En Tolonen M., Tuominen J., y Makela E. (orgs.), *CEUR Workshop Proc*., 2084, 139–151. CEUR-WS.

Delgado, J., Alvarez-Romero, C., Martínez-García, A., y Parra-Calderón, C. L. (2021). Approaches to the integration of TRUST and FAIR principles. En Di Nitto E., y Plebani P. (orgs.), *CEUR Workshop Proc.*, 2878, 33–37. CEUR-WS.

Goodman, A., Pepe, A., Blocker, A. W., Borgman, C. L., Cranmer, K., Crosas, M., Stefano, R. D., Gil, Y., Groth, P., Hedstrom, M., y otros (2014). Ten Simple Rules for the Care and Feeding of Scientific Data. *PLOS Computational Biology*, 10(4), e1003542. DOI: https://doi.org/10.1371/journal.pcbi.1003542.

Helliwell, J. R. (2019). FACT and FAIR with Big Data allows objectivity in science: The view of crystallography. *Structural Dynamics*, 6(5). DOI: https://doi.org/10.1063/1.5124439.

Henning, P., Ribeiro, C.J.W., Sales, L., Moreira, J., y Santos, L.O.B.S. (2018). Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. *Tend. Pesq. Bras. Cienc. Inf*., 11(1), 1-23. Disponible en: ttps://revistas.ancib.org/index.php/tpbci/article/download/458/443/.

Hyland, B., Atemezing, G., y Villazón-Terrazas, B. (2014). *Best practices for publishing linked data*. Disponible en: https://www.w3.org/TR/ld-bp/. World Wide Web Consortium.

Jacobsen, A., Azevedo, R. M., Juty, N., Batista, D., Coles, S., Cornet, R., Courtot, M., Crosas, M., Dumontier, M., Evelo, C. T., y otros. (2020). Fair principles: Interpretations and implementation considerations. *Data Intelligenc*e, 2(1–2), 10–29. DOI: https://doi.org/10.1162/dint_r_00024.

Landi, A., Thompson, M., Giannuzzi, V., Bonifazi, F., Labastida, I., da Silva Santos, L. O. B., y Roos, M. (2020). The "A" of FAIR – As Open as Possible, as Closed as Necessary. *Data Intelligence*, 2(1–2), 47–55. DOI: https://doi.org/10.1162/dint_a_00027.

Lóscio, B. F., Burle, C., y Calegari, N. (2017). *Data on the Web Best Practices.* Disponible en: https://www.w3.org/TR/dwbp/. W3C.

Lovett, R., Lee, V., Kukutai, T., Cormack, D., Rainie, S. C., y Walker, J. (2019). Good data practices for Indigenous data sovereignty and governance. *Good data*, 26–36.

Mougiakou, E., y Virvou, M. (2018). Based on GDPR privacy in UML: Case of e-learning program. *Int. Conf. Inf., Intell., Syst. Appl., IISA*, *2018-January,* 1–8. DOI: https://doi.org/10.1109/IISA.2017.8316456.

Romain, P. (2013). *A good pratices checklist for opendata*. Disponible en: https://www.w3.org/2013/04/odw/elie.pdf. OpQuast.

Romain, P., Larlet, D., Maire, S., y Sloïm, E. (2012). *2012 Open Data best practices Version 1.1*. Disponible en: https://res.cloudinary.com/opquast/image/upload/v1556788845/checklists/Opquast-Opendata_20120831_1.pdf. OpQuast..

Schlehahn, E., y Wenning, R. (2019). GDPR Transparency Requirements and Data Privacy Vocabularies. En E. Kosta, J. Pierson, D. Slamanig, S. Fischer-Hübner, y S. Krenn (orgs.), *Privacy and Identity Management. Fairness, Accountability, and Transparency in the Age of Big Data: 13th IFIP WG 9.2, 9.6/11.7, 11.6/SIG 9.2.2 International Summer School, Vienna, Austria, August 20-24, 2018, Revised Selected Papers.* DOI: https://doi.org/10.1007/978-3-030-16744-8_7.

Silva, L. C., Santarem Segundo, J. E., y Silva, M. F. (2019). Princípios FAIR e melhores práticas do Linked Data na publicação de dados de pesquisa. *Informação yAmp. Tecnologia,* 5(2), 81–103. DOI: https://doi.org/10.22478/ufpb.2358-3908.2018v5n2.44812.

Sloïm, É. (2021). *The end of the Opquast best practices?* Disponible en: https://www.opquast.com/en/the-end-of-the-opquast-best-practices/. Opquast.

Souza, M. T. de, Silva, M. D., y Carvalho, R. de. (2010). Revisão integrativa: O que é e como fazer. *Einstein (São Paulo)*, 8, 102–106.

Stoyanovich, J., y Howe, B. (2018 Nov 27). *Follow the data! Algorithmic transparency starts with data transparency*. Disponible en: https://ai.shorensteincenter.org/ideas/2018/11/26/follow-the-data-algorithmic-transparency-starts-withdata-transparency. [Fecha de acceso: 21 marzo 2019].

Torino, E., Vidotti, S. A. B. G., y Coneglian, C. S. (2021). #SejaJUSTOeCUIDADOSO: Princípios FAIR e CARE na gestão de dados de pesquisa. *Princípios FAIR aplicados à gestão de dados de pesquisa*, 201-2014. Rio de Janeiro: IBICT. DOI: https://doi.org/10.22477/9786589167242.cap15.

Valmorbida, S. M. I., Ensslin, S. R., y Ensslin, L. (2018). Avaliação de desempenho e contabilidade gerencial: revisão integrativa da literatura para superar as dificuldades de aplicação prática da avaliação de desempenho na gestão organizacional. *Revista Contabilidade, Gestão e Governança*, 21(3), 339-360.

Van der Aalst, W. M. P. (2016). Green data science: using big data in an" environmentally friendly" manner. En *18th International Conference on Enterprise Information Systems (ICEIS 2016),* 9-21. Roma: SciTePress. Disponible en: https://pdfs.semanticscholar.org/5889/68dd392ae93b1524aa7a491917d839bca050.pdf. [Fecha de acceso: 10 enero 2019].

Van der Aalst, W. M. P. (2019). *Responsible Data Science in a Dynamic World: The Four Essential Elements of Data Science*, 548. Nueva York: Springer.

Van der Aalst, W. M. P., Bichler, M. y Heinzl, A. (2017). Responsible Data Science. *Business and Information Systems Engineering*, 59(5), 311–313. DOI: https://doi.org/10.1007/s12599-017-0487-z.

Van der Aalst, W.M.P. (2017). Responsible Data Science: Using Event Data in a "People Friendly" Manner. En: Hammoudi, S., Maciaszek, L., Missikoff, M., Camp, O., y Cordeiro, J. (eds). *Enterprise Information Systems. ICEIS 2016. Lecture Notes in Business Information Processin*g, 291. DOI: https://doi.org/10.1007/978-3-319-62386-3_1.

Whittemore, R., y Knafl, K. (2005). The integrative review: Updated methodology. *Journal of advanced nursing*, 52(5), 546–553.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, y otros. (2016). Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*. DOI: https://doi.org/10.1038/sdata.2016.18.

Wilkinson, M., Dumontier, M., Aalbersberg, I., y otros. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018. DOI: https://doi.org/10.1038/sdata.2016.18.