

CONTROL DE LA INFORMACIÓN AUTORAL Y ORGANIZACIONAL PARA OPTIMIZACIÓN DEL ACCESO Y VISIBILIDAD CIENTÍFICA

*Thamyres Vieira dos Santos**

Universidade de São Paulo

Giovana Deliberati Maimone

Universidade de São Paulo

Resumen: Este artículo considera el alto volumen de información científica disponible en los medios digitales y la relevancia de los datos de autores e instituciones representados de forma estandarizada, con el objetivo de resaltar la importancia de la estandarización y el control de las autoridades personales y organizacionales para optimizar el acceso a la visibilidad de la investigación científica. Utiliza el método inductivo, realizando una búsqueda bibliográfica en bases de datos nacionales e internacionales con el fin de argumentar y fundamentar el contexto sobre el problema del control de autoridad. Los resultados muestran que el control de autoridad debe seguir principios y reglas encaminados a estandarizar los puntos de acceso en los catálogos bibliográficos, hecho que impacta directamente en la visibilidad de la investigación, de sus autores y de sus instituciones, otorgándoles el crédito de realización y facilitando el contacto entre pares y el establecimiento de posibles alianzas. Es evidente que un estándar establecido a lo largo del flujo de trabajo de las revistas científicas podría minimizar las variaciones en la representación de los nombres de los investigadores y sus instituciones, además de que la autenticidad de esta información debe ser visualizada y validada mediante el correcto tratamiento y estandarización de documentos.

Palabras clave: ciencia de la información; comunicación científica; control de la información; control de autoridades; recuperación científica; visibilidad científica.

Title: CONTROL OF AUTHOR AND ORGANIZATIONAL INFORMATION TO OPTIMIZE ACCESS AND SCIENTIFIC VISIBILITY.

Abstract: This article considers the high volume of scientific information available in digital media and the relevance of data from authors and institutions represented in a standardized way, with the aim of highlighting the importance of standardization and the control of personal and organizational authorities to optimize access to visibility of scientific research. It uses the inductive method, carrying out bibliographical research in national and international databases in order to argue and base the context on the problem of authority control. The results show that authority control must follow principles and rules aimed at standardizing access points in bibliographic catalogs, a fact that directly impacts the visibility of the research, its authors and its institutions, giving them credit for carrying out and providing contact between peers and the establishment of possible partnerships. It is evident that a standard established throughout the workflow of scientific journals could minimize variations in the representation of the names of researchers and their institutions, in addition to the fact that the authenticity of this information must be visualized and validated through the correct treatment and standardization of documents.

Keywords: information science; scientific communication; information control; authority control; scientific retrieval; scientific visibility.

Copyright: © 2023 Servicio de Publicaciones de la Universidad de Murcia (Spain). Este es un artículo de acceso abierto distribuido bajo los términos de la licencia Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0).

1 INTRODUCCIÓN

La recuperación de información científica pasa por varias discusiones y matices, ya sean derivados de la variación de la indexación de las bases de datos o del propio manejo de las herramientas de búsqueda. Ante tantas posibilidades derivadas del desarrollo de las tecnologías de la información y la comunicación (TICs), es fundamental que los contenidos científicos, su organización, representación, recuperación y uso se contemplen también en los estudios que pretenden optimizar su difusión, de cara al desarrollo científico. Es fundamental reconocer que la difusión de los contenidos científicos, especialmente a través de las revistas científicas electrónicas puestas a disposición por las bases de datos, se vuelve más dinámica, y esta característica contribuye considerablemente al acceso y al desarrollo de la ciencia.

* thamyres.vieira@hotmail.com

Cuando se considera el gran volumen de información científica en los medios digitales y se discute el acceso a la literatura por parte de la comunidad académica, se debe prestar atención a la forma en que los autores y las instituciones están representados en las fuentes de información y cómo esto impacta directamente en el acceso y la visibilidad de la investigación.

Normalmente, cuando hablamos de investigación científica solemos pensar en términos generales en la productividad de la ciencia y sus campos de estudio. Sin embargo, de manera particular, cuando las publicaciones son el centro de atención, dos componentes importantes del universo de la investigación son relevantes: los autores y las instituciones (Haak, Meadows y Brown, 2018). Estos dos actores deben ser cuidadosamente considerados en su representación, considerando su implicación directa en la identificación y recuperación en las fuentes de información.

Gasparyan *et al.* (2016) afirman que los problemas de autoría no son nuevos en el contexto editorial y que probablemente ya existían antes de la publicación de la primera revista científica, la *Philosophical Transactions of the Royal Society* en 1665. Durante décadas, el uso incorrecto de abreviaturas, nombres y apellidos afectó a investigadores de todo el mundo (Gasparyan *et al.*, 2016). Bourne (1977) llevó a cabo un estudio sobre la recuperación de información científica en bases de datos que mostró altos índices de errores ortográficos y cómo estos repercuten en los resultados de la búsqueda.

Fox (2016) comenta que saber el nombre de alguien no es suficiente para conectarlo con una persona real. Es bastante común que los investigadores tengan homónimos, personas con el mismo nombre que también se dedican a la investigación científica, y quienes no los tienen pueden tenerlos en el futuro. Considerando esta hipótesis, al buscar un nombre en una fuente de información bibliográfica, se recuperarán las publicaciones de todos los autores con ese nombre. La situación también es delicada si los investigadores tienen diferentes nombres completos, pero el mismo apellido y las mismas iniciales. También es común que en las normas bibliográficas para citas y referencias existan lineamientos de que la cita se haga solo por el apellido del autor y en las referencias este también aparece ligado solo a la primera letra del nombre. Sin embargo, este uso es particularmente serio cuando se consideran nombres japoneses y coreanos (Strotmann; Zao, 2012).

La formación del joven investigador, desde los cursos de licenciatura, debe orientarles para que, al iniciar su carrera académica, adopten una forma nominal única para ser utilizada en las publicaciones científicas actuales y futuras, con el fin de reducir los impactos negativos por los cambios y su consiguiente recuperación. La adopción y difusión de directrices claras en cuanto a los nombres institucionales más allá de los de autoría también debería formar parte de las preocupaciones de las universidades e instituciones de investigación, para presentar la forma correcta en que sus estudiantes, investigadores y profesores deben mencionar el nombre de la institución en sus afiliaciones.

Los esfuerzos en la creación y mantenimiento de manuales metodológicos trazan las normas que debe seguir la comunidad científica en cuanto al mejor uso de los nombres de los autores y las instituciones, sin embargo, si no hay control en su seguimiento poco se avanza. Fonseca (1973), aborda los problemas de comunicación de la información científica y, aunque es una obra de los años 70, incluso antes de la consolidación del entorno digital, señala la importancia de la normalización y el establecimiento de reglas para la descripción de las informaciones.

[...] no debe permitirse que los canales de comunicación sean tumultuosos y estén bloqueados por la información transmitida según los caprichos personales de los autores, las tradiciones obsoletas de los editores e incluso las limitaciones de los establecimientos de impresión. Al igual que los conductores deben conocer las leyes de tráfico y los vehículos deben contar con determinados equipos de seguridad, los autores y editores tienen la obligación de conocer las normas de comunicación científica y los vehículos de esta comunicación deben estar normalizados para que la información contenida en ellos pueda llegar más rápidamente al conocimiento de los interesados (Fonseca, 1973, p. 55-56).

La falta de estándares en las autoridades puede repercutir directamente en las evaluaciones de investigación que suelen realizar las instituciones sobre la producción científica de los investigadores o del profesorado. Este tipo de impacto afecta directamente al establecimiento de mejores prácticas, objetivos, políticas y financiación para fomentar la investigación. Los indicadores científicos, utilizados como insumo en las evaluaciones de productividad y calidad de la investigación, además de significar la implicación de sus autores e instituciones, también les confieren visibilidad, amplificada, actualmente, con el uso de redes y perfiles académico-científicos en algunas fuentes de información, contribuyendo a la correcta identificación de los investigadores y sus investigaciones, promoviendo también la certificación de la autoría.

Desde este contexto, el objetivo de esta investigación es evidenciar la importancia de la estandarización y el control de las autoridades personales y organizativas para optimizar el acceso a la visibilidad de la investigación científica,

hecho que pasa inevitablemente por las cuestiones de representación descriptiva de la información, las normas y los estándares.

El trabajo empieza con la contextualización temática y los objetivos en esta breve introducción, seguida de la presentación de los aspectos metodológicos propuestos para el tratamiento del tema. Posteriormente, se presenta un apartado específico sobre el marco teórico planteado, finalizando con las consideraciones finales y las referencias.

2 METODOLOGÍA

El método es la selección de un itinerario de investigación, con el fin de alcanzar un resultado o lograr un objetivo determinado. Son procedimientos realizados de forma organizada que conducen a un fin. El método inductivo fue elegido para esta investigación porque:

[...] es un procedimiento de razonamiento que, partiendo del análisis de datos particulares, avanza hacia nociones generales. En este caso, presenta una forma ordenada de razonar a partir de datos singulares hacia una verdad general (Fachin, 2001, p. 30).

Se analizan las normas y estándares nacionales e internacionales necesarios para representar y controlar la información científica del autor y las organizaciones con el fin de normalizar la inclusión de la autoría en las bases de datos y su disponibilidad en el entorno web, con el objetivo de lograr una correcta visibilidad científica. La investigación se puede caracterizar como bibliográfica ya que se analizan las principales aportaciones teóricas sobre el tema para argumentar la problemática. Según Severino (2007, p. 122), la investigación bibliográfica se lleva a cabo mediante:

[...] el registro disponible, derivados de investigaciones anteriores, en documentos impresos como libros, artículos, tesis, etc. Utilizamos datos de categorías teóricas ya trabajadas por otros investigadores y debidamente registradas. Los textos se convierten en fuentes de los temas a investigar. El investigador trabaja a partir de las aportaciones de los autores de los estudios analíticos contenidos en los textos.

En este sentido, la revisión bibliográfica se realiza con la búsqueda de materiales en las siguientes bases de datos: Scopus, Web of Science, PubMed y BRAPCI. Durante las lecturas también se consultaron otras fuentes que resultaron importantes para fundamentar la argumentación. Por lo tanto, se infiere que los análisis y los resultados están explícitos y hablan directamente con el fundamento teórico.

3 CONTROL DE AUTORIDAD

Considerando la existencia de personas o instituciones, que tienen nombres similares o idénticos, se hace evidente que en algún momento puede haber confusiones nominales, si este es el único criterio que se toma como válido para caracterizarlas.

El problema de la ambigüedad de los nombres de los autores consiste, en general, en dos situaciones:

- (a) la homonimia, que exige distinguir entre diferentes personas con el mismo nombre y;
- (b) las diversas variaciones del nombre de la misma persona (Mugnaini *et al.*, 2012, p. 265).

En este sentido, tenemos el concepto de control de autoridad como "[...] un proceso que se basa en la creación y mantenimiento de puntos de acceso autorizados y estandarizados que se utilizan para recuperar materiales bibliográficos [...]" (Rosado; Dias, 2019, p. 6), y que pueden estar disponibles en fuentes como catálogos, repositorios, bibliotecas digitales o bases de datos; siendo que el punto de acceso puede ser un nombre o término con el que se identifica un registro bibliográfico. A su vez, dicho registro puede definirse como un conjunto de datos que describen un recurso bibliográfico (International Federation of Library Associations and Institutions, 2016). Sandberg y Jin (2016) afirman que los catalogadores crearon registros de autoridad durante décadas, lo que dio lugar a enormes bases de datos con nombres de autor estandarizados como la *Library of Congress Authorities*.

La descripción bibliográfica basada en procesos estandarizados que buscan garantizar la asertividad de la recuperación de la información asegura resultados más efectivos, considerando que el acceso correcto es una de las principales funciones de un servicio de información.

La consistencia que pretende el control de autoridades ayuda a garantizar que los recursos bibliográficos se recuperan correctamente cuando se busca un término determinado por el usuario que tiene una necesidad de información específica. Assumpção, Santos y Zafalon (2017a) mencionan que, en el contexto de los catálogos digitales, los recursos bibliográficos pueden ser recuperados por los diferentes términos que aparecen en el registro, pero a lo largo de la historia de la catalogación, debido a limitaciones prácticas y tecnológicas, la recuperación solo era posible por el título del registro, sus responsables o por materia. Así, más directamente, el control de la autoridad puede definirse como:

[...] el proceso de unificar, mediante el uso de un formato estandarizado, los puntos de acceso de los catálogos automatizados y también mostrar las relaciones entre los diferentes puntos de acceso. En otras palabras, se trata de normalizar los nombres de personas, entidades, títulos uniformes o materias, que pueden constituir los puntos de acceso principales o secundarios de un catálogo automatizado. Su propósito es facilitar la identificación y recuperación de los documentos almacenados, evitando la confusión causada por homónimos, sinónimos o la variedad de nombres con los que se puede denominar a una persona, entidad, obra, tema o concepto (Herrero Pascual, 1999, p. 121, traducción nuestra).

Teniendo en cuenta el planteamiento de la Norma ANSI/NISO Z.39.19:2005 Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies, si hay muchas variaciones de nombres propios (personas, instituciones, organizaciones, lugares y títulos) sin ningún tipo de control, la recuperación de información bibliográfica no será eficiente (Rosado; Dias, 2020). En este sentido, y dilucidando que una de las funciones del catálogo es permitir que los materiales se encuentren por autoría, Cutter destaca:

La lista de autores del catalogador, mantenida por orden alfabético, evita la duplicación de trabajo. Es un registro de la forma de un nombre (completo) que se ha adoptado, con una nota de las autoridades consultadas y sus variaciones (Cutter, 1904, p. 133, traducción nuestra).

La primera edición del código de catalogación de la American Library Association (ALA) se publicó en 1908, con el título *Catalog Rules: Author and Title Entries*, y en 1949 se publicó la segunda edición, ninguna de las cuales contemplaba la creación de registros de autoridad (Assumpção, Santos y Zafalon, 2017b).

En cuanto a *Anglo-American Cataloguing Rules*, 2nd edition, revisión de 2002 (RCA2r), que, aunque haya sido creada en el contexto de la automatización de las bibliotecas, fue desarrollada a partir del catálogo de fichas (Assumpção, Santos y Zafalon, 2017b). Los autores también mencionan que esto llevó a la adopción de algunas convenciones de los registros en las fichas:

No existe un método o formato prescrito a nivel nacional para transcribir los datos de las autoridades en las fichas, pero se utilizan ampliamente varias convenciones. La primera convención, y la más importante, es situar el encabezamiento autorizado en la parte superior de la ficha, de la misma forma que aparece en las fichas de catalogación. Debajo se encuentran las remisivas. Las remisivas Ver [see from] son indicadas por el prefijo x y las remisivas Ver también [see also from] son indicadas por el prefijo xx. Debajo de ellas son dadas cualesquier notas. A continuación, se indican las posibles notas. Por último, se indica el nombre o las iniciales del catalogador, la fecha de elaboración del registro y la signatura o título de la publicación que ha dado lugar a la elaboración del registro de autoridad (Burguer, 1985, p. 23-24 apud Assumpção, Santos y Zafalon, 2017b, p. 94).

Los Principios Internacionales de Catalogación da *International Federation of Library Associations and Institutions* (IFLA), en los cuales el AACR2r se basa, indica que el nombre preferido utilizado como punto de acceso para la entidad debe elegirse según el nombre que la identifique "[...] ya sea porque se encuentra con mayor frecuencia en las manifestaciones o porque es un nombre ampliamente aceptado y apropiado para los usuarios del catálogo [...]" (International Federation of Library Associations and Institutions, 2016, p. 9). Otra norma establecida por la IFLA para nombrar a las personas establece que cuando el nombre tiene varias palabras, es decir, nombre y apellido, la elección del primero de ellos para ser un punto de acceso autorizado debe seguir las convenciones adoptadas por el idioma y el país más asociados a esa persona (International Federation of Library Associations and Institutions, 2020).

Siguiendo la IFLA en el capítulo 22 del AACR2r, entre varias reglas, debe garantizarse que la elección de la base del encabezamiento de una persona se haga por el nombre por el que se le conoce habitualmente, ya sea su nombre real, seudónimo, apodo, iniciales, título nobiliario u otro apodo; esta determinación debe considerar la principal fuente de información de sus obras, publicadas en su lengua o país de residencia/actividad. Para los casos de cambio de nombre, el AACR2r determina que se elija el nombre más reciente, excepto en los casos en que se crea que la forma antigua del nombre persistirá como la más conocida; y para los nombres idénticos, sugiere añadir como último elemento del encabezamiento información como las fechas de nacimiento o muerte (Federação Brasileira de Associações de Bibliotecários, 2004). Rosado y Dias (2020) comenta que las normas establecidas por este capítulo de las AACR2r son muy detalladas, con consideraciones en cuanto a las variaciones ortográficas y de idioma, ya que pretenden cubrir varias especificidades para que la elección de los nombres no sea arbitraria, por lo tanto, es una obra que debe ser considerada para la construcción de un control de autoridad.

Incluso con los catálogos digitales, que permiten la búsqueda y localización de un mayor número de información, la comunidad científica sigue encontrando variaciones en los nombres de los autores, ya sea por el desconocimiento de la importancia de esto o por la variedad de fuentes de información en las que el investigador puede estar registrado con los metadatos de diferentes maneras. Brandt (1988) comenta la relación entre los catálogos digitales y el control de la

autoridad en una época en la que el acceso digital aún no estaba tan extendido como ahora, pero revela la preocupación del autor por la gestión de los datos en un entorno en línea:

A medida que más y más bibliotecas migran a los catálogos en línea, el control de autoridad se ve estrechamente como una parte de la gestión de datos que garantiza la coherencia e integridad de las entradas, así como la interrelación de los registros que comparten un encabezamiento. Cuando se aplica a las bases de datos bibliográficas, el control de autoridad es algo más que una herramienta utilizada para establecer y relacionar encabezamientos: puede "enlazar" los registros de forma automática y literal. Por ello, el control de la autoridad puede considerarse una herramienta esencial para la administración de datos (Brandt, 1988, p. 297, traducción nuestra).

Teniendo en cuenta el creciente volumen de información en el contexto digital, se hizo evidente la necesidad de técnicas para optimizar el tiempo de procesamiento de todos estos contenidos, incluyendo el proceso de representación para su posterior recuperación. Clack (1988) afirma que cuando se introdujo el ordenador en la realidad de las bibliotecas, surgió una percepción errónea y exagerada de que todos los problemas bibliográficos se habían resuelto y que no era necesario garantizar la integridad de los registros. Esta visión equivocada contribuyó en gran medida a esta realidad llena de variaciones en los nombres de personas e instituciones sin control en las diversas fuentes de información. Pronto se comprobó que los catálogos en línea presentaban fallos en la recuperación de la información y esto apuntaba al control como posible solución.

El aumento del uso de archivos en el entorno digital ha estimulado los esfuerzos por crear normas, directrices y modelos para ampliar el intercambio de datos bibliográficos y de autoridad a nivel internacional, entre ellos el modelo *Functional Requirements for Authority Data* (FRAD) (Assumpção, Santos y Zafalon, 2017a). Antes de hablar sobre el FRAD, es fundamental mencionar que en 1998 la IFLA publicó el *Functional Requirements for Bibliographic Records* (FRBR). Sobre ellos, se puede decir que:

El primer objetivo en el que se centró el desarrollo de FRBR fue proporcionar un marco claramente definido y estructurado para relacionar los datos contenidos en los registros bibliográficos con las necesidades de los usuarios. Es decir, FRBR analizó los datos que el usuario necesitaba para realizar la búsqueda bibliográfica, así como la información que el usuario podía esperar encontrar en el registro (Mey y Silveira, 2009, p. 17).

Con el fin de ampliar este modelo para abarcar también los datos de las autoridades, la IFLA creó en 1999 el *Working Group on Functional Requirements and Numbering of Authority Records* (FRANAR) que publicó el FRAD en 2009 (IFLA Working Group on Flanar, 2013). Mey y Silveira afirman que:

Este modelo conceptual fue diseñado para crear un marco de referencia claramente definido, estructurado para relacionar los datos de los registros de autoridad con las necesidades de los usuarios de dichos datos, y para ayudar a evaluar el potencial de intercambio internacional de datos de autoridad más allá del universo de las bibliotecas (Mey y Silveira, 2009, p. 38).

Aunque la publicación de FRBR ha desencadenado el desarrollo de varios estudios que intentaron utilizar este modelo conceptual como modelo de datos para los catálogos digitales, las iniciativas fueron sólo experimentales y FRBR no se consolidó como referencia para el desarrollo de catálogos (Assumpção; Santos y Zafalon, 2017a).

El *Machine Readable Cataloguing* (MARC) tiene una relación directa con el control de las autoridades. Como señalan Mey y Silveira (2009), MARC apareció en 1965/1966, en el contexto de los avances informáticos de la década de 1960, desarrollándose a través de una cooperación entre la Library of Congress (LC) de Estados Unidos y la British Library. Se trata de un protocolo de identificación para el intercambio de información y permite estructurar e identificar los datos de tal forma que pueden ser reconocidos por la computadora. Permite y facilita la implementación de los estándares de control de autoridad establecidos por AACR2.

En 1969, teniendo en cuenta la experiencia con el formato MARC, LC lanzó el MARC II y en la década siguiente ya era posible encontrar diversas variaciones del formato en otros países; sin embargo, en 1976, teniendo en cuenta la preocupación por el control de autoridades en los catálogos digitales, LC lanzó la versión preliminar del documento *Authority work: the creation, use, maintenance, and evaluation of authority records and files*. En 1981 publicó la primera edición del documento *Autoridades: un formato MARC* y hasta hoy la LC mantiene actualizado este formato, que hoy se denomina Formato MARC 21 para datos de autoridad (Assumpção; Santos y Zafalon, 2017b).

Y en 2004, según Mey y Silveira (2009), se inició el proceso de revisión de las AACR2r que indicaba varias modificaciones para una tercera versión del código, lo que resultó en el desarrollo de su sucesor al año siguiente, el *Resource Description and Access* (RDA). A pesar de tener una fuerte relación con AACR2r, el RDA fue diseñado para el entorno digital y tiene un alcance más amplio (Oliver, 2011).

RDA es un estándar desarrollado y mantenido por el Joint Steering Committee for the Development of RDA, compuesto por representantes de instituciones fundamentalmente del mundo anglosajón (Library of Congress, American Library Association, Australian Committee on Cataloguing, The British Library, Canadian Committee on Cataloguing, CLIP: Chartered Institute of Library and Information Professionals, y la Biblioteca Nacional de Alemania).

En este contexto, aún tratándose de estándares diseñados para el entorno digital, es indispensable comentar que el *Online Computer Library Center* (OCLC) fue pionero en el control de autoridades y en el establecimiento de identificadores únicos para los autores, ya que el *Virtual International Authority File* (VIAF) es consecuencia de los esfuerzos de OCLC por brindar una base de autoridad internacional (Fox, 2016). VIAF se creó en 2003, como un proyecto de bibliotecas nacionales operado por el OCLC para unir los ficheros de autoridad virtual de aquellas bibliotecas. El proyecto fue iniciado por la Biblioteca del Congreso de los Estados Unidos y la Biblioteca Nacional de Alemania, a las que se unieron varias otras bibliotecas. Cada autor que aparece en el VIAF tiene su cadena única, asignada a través de una URI, que lo identifica y permite recuperar todas las formas de su nombre, nombres relacionados, países de publicación y publicaciones propias (Fox, 2016). *Uniform Resource Identifier* (URI) es un tipo de protocolo de transferencia de hipertexto con el que el usuario puede pegar la dirección en la barra de su navegador de internet para luego acceder al recurso que representa el identificador.

Hay varias iniciativas y normas que discuten la importancia del control de autoridad y cómo la falta de estandarización impacta directamente en la recuperación de información por parte de los usuarios. Lo mismo ocurre con la descripción y consiguiente recuperación de la información científica por parte de la comunidad. La literatura en el área de la Ciencia de la Información, como se ha visto, se ha preocupado por la consolidación de los nombres de las autoridades en los catálogos durante algún tiempo, pero con la consolidación de los catálogos digitales y la información disponible de una manera más dinámica y amplia, debido al uso de las fuentes informativas digitales, las soluciones propuestas no parecían seguir las necesidades cada vez más emergentes. Algunas soluciones, como el uso de identificadores persistentes para las autoridades o el uso de algoritmos para desambiguar diferentes variaciones de nomenclatura, aunque no están directamente relacionadas con la Ciencia de la Información, están aliadas con el proceso de control de autoridades en el entorno digital.

3.1 Formatos estandarizados

El aumento exponencial de la cantidad de información disponible en los entornos digitales, combinado con la ampliación del acceso de una parte de la población a los medios de comunicación, confirió la posibilidad de crear y poner a disposición contenidos con una intensidad y velocidad sin precedentes. Cuando se habla de contenidos digitales, no solo se trata de los producidos por los medios sociales o por Internet en general, sino también de los contenidos académico-científicos e industriales, por ejemplo, que ahora se difunden y consultan ampliamente en la red. La preocupación por la falta de normas y la normalización de la información en el entorno digital debería repercutir directamente en su correcta difusión. Dias señala que:

La normalización es una característica esencial de la actividad humana desde los albores de la civilización, habiendo sido esencial en el desarrollo del lenguaje hablado y escrito. Otros aspectos de la normalización de la actividad humana en épocas pasadas pueden observarse cuando el comercio entre los pueblos primitivos exigía el establecimiento de medidas estandarizadas de peso, tamaño y formas de pago (Dias, 2000, p. 137).

La normalización de la actividad humana es intrínseca a la organización de la sociedad y sus relaciones. Encontrar normas y principios facilita el intercambio de información y, en consecuencia, repercute en la forma de resolver y tratar cuestiones comunes a la vida cotidiana también en relación con aspectos empresariales, económicos o sociales. Las normas son capaces de garantizar el funcionamiento ordenado de determinadas actividades y situaciones.

La investigación científica necesita una normalización para garantizar su recuperación efectiva por parte de la comunidad académica. Meadows (1999, p. 120) comenta que la estructura de las publicaciones científicas ha mejorado mucho a lo largo de los años y que una parte considerable de los trabajos publicados no presentaban referencias y, cuando lo hacían, no se insertaban muchos detalles sobre ellas, lo que implicaba su no localización. Y para ello, se utilizan las normas técnicas que se pueden definir como aquellas que tienen como objetivo principal normalizar y estandarizar los productos y servicios apuntando a su calidad, considerándose como resultado de un proceso que indica características de los productos y servicios evaluando su calidad y fiabilidad (Crespo y Rodrigues, 2011).

La tendencia a promover la estandarización de la presentación, como ya se ha señalado, refleja las crecientes presiones ejercidas sobre la comunicación científica, concretamente como consecuencia de su rápida expansión, que ha dificultado aún más la localización de la información pertinente por parte de los lectores. La normalización ayuda en este sentido (Meadows, 1999, p. 120).

La información científica producida y publicada depende del apoyo técnico de las normas y estándares establecidos por organizaciones dedicadas a este fin. Según Rodrigues, Lima y García (1998), desde la década de 1960, las organizaciones internacionales que trabajan en el ámbito de la documentación científica, como la Organización de las Naciones Unidas (UNESCO), la Federación Internacional de Información y Documentación (FID), *International Federation of Library Associations and Institutions* (IFLA), *International Organization for Standardization* (ISO) y *International Science Council* (ICSU) prescriben normas.

En Brasil, la Asociación Brasileña de Normas Técnicas (ABNT), creada en 1940, es la organización dedicada a la normalización, basada en las normas ISO. Las normas de la ABNT centradas en la producción académico-científica forman parte del ámbito del Comité Brasileño de Información y Documentación (CB 014).

Entre las normas de la ABNT centradas en la normalización de los trabajos académico-científicos, cuatro de ellas proporcionan directrices específicas sobre la inserción del nombre de los autores del documento, que se muestran en el siguiente cuadro abajo:

Norma	Orientaciones sobre la inserción del nombre de los autores del documento
ABNT/NBR 6022 – Información y documentación – Artículo en publicación periódica científica impresa	se debe insertar el nombre del autor: nombre, abreviado o no, y apellido; cuando hay más de un autor, los nombres pueden aparecer en la misma línea, separados por una coma, o en líneas separadas; debe incluir un breve currículum de cada autor, con enlaces institucionales y direcciones de contacto (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2018a).
ABNT/NBR 6023 – Información y documentación Referencias – Elaboración	<ul style="list-style-type: none"> - los autores deben ser indicados por el último apellido (en letras mayúsculas), seguido del nombre y otros apellidos, abreviados o no, según lo que aparezca en el documento; - los autores deben tener sus nombres separados por punto y coma y espacios; es conveniente normalizar el nombre y los apellidos de los autores, en caso de que aparezcan de forma diferente en distintos documentos; - si el documento tiene hasta tres autores, deben indicarse todos ellos - si el documento tiene cuatro o más autores, es conveniente indicarlos a todos, pero se permite indicar el primer autor seguido de la expresión et al.; - los autores que tengan nombres hispanos, compuestos, con grado de parentesco o con apellidos con prefijos, deben seguir los ejemplos: GARCÍA MÁRQUEZ, Gabriel; ASSAF NETO, Alexandre; ESPÍRITO SANTO, Miguel Frederico de; D'AMBROSIO, Ubiratan; - cuando la responsabilidad del trabajo sea conjunta, la inscripción se hará con el nombre del responsable, seguido de la abreviatura del tipo de participación (organizador, editor, coordinador, etc.) en minúsculas; - las obras publicadas bajo seudónimo deben llevar el nombre del seudónimo en la referencia; - las obras publicadas bajo un psicógrafo deben llevar el nombre del espíritu en primer lugar; - en el caso de las obras adaptadas, el adaptador debe aparecer en primer lugar; los trabajos bajo la responsabilidad de empresas, organizaciones, asociaciones, etc., deben indicarse de la forma conocida o como si estuvieran resaltados en el documento, ya sea de forma completa o abreviada; - es conveniente estandarizar los nombres de las entradas de las personas jurídicas, en caso de que aparezcan de forma diferente en distintos documentos - cuando la responsabilidad del documento corresponda a una institución gubernamental, el nombre irá precedido del nombre del organismo superior; - cuando la institución vinculada a un organismo superior tiene un nombre específico, la inscripción se realiza directamente por su nombre; - cuando la institución sea homónima, deberá indicarse al final, entre paréntesis, la unidad geográfica que la identifica; - las obras resultantes de eventos se introducen por el nombre del evento cuando se desconoce la autoría del documento, la entrada se realiza por el título (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2018b).

ABNT/NBR 10520 – Información y documentación – Citas en documentos – Presentación	- las citas, las llamadas por los apellidos del autor, institución o título deben ir en mayúsculas y minúsculas, y cuando van entre paréntesis deben ir en mayúsculas; - considerando el sistema autor-fecha, la indicación de la fuente se hace por el apellido de cada uno de los autores del documento o por el nombre de la entidad responsable hasta el primer signo de puntuación, seguido de la fecha de publicación del documento y de la página de citación; en el caso de la citación directa, se separan por una coma y entre paréntesis (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2002).
ABNT/NBR 12225 – Información y documentación – Lombo – Presentación	- el nombre del autor debe estar impreso en la misma dirección que el lomo - si el documento tiene más de un autor, los nombres se imprimirán uno debajo de otro en el lomo, separados por signos de puntuación, espacios u otros signos gráficos, permitiendo la abreviación u omisión de los nombres. (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2004).

Cuadro I. Orientaciones de las normas sobre la inserción de los autores del documento. Fuente: Datos de investigación (2021).

La ABNT enfoca su uso en trabajos académico-científicos de manera puntual y con el objetivo de seguir un patrón. Su amplio alcance y uso en la comunidad académico-científica son relevantes para suscitar espacios de reflexión sobre la inserción de datos estandarizados de organizaciones y autores en las referencias bibliográficas desde la redacción del trabajo por parte de los autores. Generalmente, la norma es el primer material con el que los autores, sean nuevos estudiantes o investigadores, entran en contacto con la necesidad de estandarizar las referencias utilizadas en su trabajo, incluyendo el nombre de sus autores. Este primer contacto es importante para consolidar prácticas a lo largo de la carrera científica, por lo que las normas, de alguna manera, logran evidenciar la importancia de la estandarización. En este sentido la tendencia es que los investigadores sean cada vez más conscientes de la relevancia del control de autoridad.

Al comprobar las pautas previstas por las normas para la inserción de los nombres de los responsables, se puede observar la orientación común en cuanto a los apellidos de los autores, en mayúsculas, y los nombres, abreviados o no, según lo que figura en el documento. Si bien las normas son enfáticas en la forma en que se deben insertar los autores, su indicación para la inserción de los nombres de acuerdo a lo que está en el documento puede crear variaciones debido a la forma en que este documento original fue diagramado y publicado, aunque la norma también establece que un autor con diferentes formas de ingreso puede tener su nombre normalizado. La decisión de seguir el documento que tiene en sus manos para insertar los nombres de los autores tal y como aparecen en él, hace que los editores y diagramadores de los documentos científicos académicos tengan la responsabilidad sobre el control de las autoridades. Lo mismo ocurre con las obras a cargo de instituciones, ya que la norma orienta a que se haga por su nombre conocido o por la forma que aparece en el documento.

Otro punto a destacar es la orientación respecto a la colocación de todos los autores a la hora de elaborar las referencias bibliográficas del documento científico: aunque es conveniente indicar todos los autores de la referencia, se permite indicar solo el primero seguido de la expresión *et al.* (Associação Brasileira de Normas Técnicas, 2018). Dado que la norma permite incluir solo al primer autor, es fundamental reflexionar sobre las consecuencias de omitir autores en las referencias de los documentos y cómo esto repercute en su recuperación y en la ocultación de la autoría.

El formato Vancouver, ampliamente utilizado por las publicaciones académico-científicas en el área de las Ciencias de la Salud, fue adaptado por la *National Library of Medicine* (NLM) de Estados Unidos a partir del trabajo del *International Committee of Medical Journal Editors* (ICMJE), que tuvo su primera reunión en 1978 en la ciudad de Vancouver, Canadá (Hoppen, Magnus y Vanz, 2018). La NLM publica *Citing Medicine*, una guía que ayuda a los autores a referenciar sus publicaciones y a las revistas a establecer normas. La norma de Vancouver, a diferencia de la ABNT, no utiliza el sistema de citas autor-fecha, sino un modelo numérico. De acuerdo con lo que establece la *Citing Medicine*, los nombres de los autores se referencian siempre con los nombres abreviados y los apellidos siempre descritos en minúscula (Patrias y Wendling, 2007), diferente a las directrices de ABNT.

Bellini *et al.* (2006) en un estudio sobre las peculiaridades de cada área de conocimiento en la normalización de los documentos, argumentan que en las ciencias humanas se utiliza una norma que indica la posibilidad de insertar los nombres de los autores en su totalidad considerando que sus documentos tienen un mayor alcance, en cuanto al área temática, y por lo tanto aumenta la posibilidad de especialistas homónimos. En el área de Ciencias de la Salud, según la argumentación, las especialidades son más restringidas y los autores tienen la costumbre de publicar en grupo, lo que haría que el cruce entre el apellido del autor y la especialidad sea suficiente para encontrar información y producciones

de los autores. Aun con las peculiaridades de cada área, es necesario que se optimicen las estrategias que permitan una rápida y eficaz identificación de los autores de las producciones académico-científicas, sin necesidad de realizar mayores cruces para ello.

La norma de *American Psychological Association* (APA), como su nombre indica, se utiliza ampliamente en los trabajos académico-científicos en el campo de la Psicología. Fue creada en 1929 en Washington y surgió con el objetivo de mejorar la calidad de las referencias citadas (Hoppen, Magnus y Vanz, 2018). Al igual que la ABNT, la APA utiliza el sistema autor-fecha para realizar las citas y refiere los nombres de los autores por sus apellidos, en minúsculas, y sus nombres abreviados.

Las normas ABNT, Vancouver y APA son solo algunos ejemplos adoptados a nivel nacional e internacional que tratan de la descripción de los autores, según sus particularidades y que no siempre convergen entre sí. La comunidad científica, comprometida en contribuir al desarrollo de la ciencia, debe ser guiada a prestar atención a las normas con un propósito más profundo que el simple cumplimiento de los criterios exigidos por las instituciones de educación superior o las publicaciones científicas.

[...] la entropía o desorden que se observa en la comunicación entre científicos no resulta sólo del exceso de medios, sino de la falta de normalización y estandarización adecuada de estos medios. Los canales de comunicación científica son bloqueados por vehículos que, a pesar de ser excesivos, continúan produciéndose sin respetar las recomendaciones internacionales y las normas nacionales que regulan la materia (Fonseca, 1973, p. 55).

Las instituciones o grupos encargados de organizar y mantener estándares deben discutir y contar con profesionales que puedan hacer aportes efectivos sobre la importancia de establecer estándares y las mejores alternativas frente a problemas relacionados con la recuperación de información. Es indispensable que, en la búsqueda de soluciones a estos problemas, estos profesionales puedan llegar a discusiones sobre el tema que son recurrentes en la comunidad y en la literatura científica, como el tema del control de autoridad, ampliamente discutido desde hace años y que puede contribuir en gran medida a un efectivo establecimiento de formatos bibliográficos estandarizados.

En este sentido, la relevancia del tema del control de autoridad debe ser clara para estas instituciones o grupos, contribuyendo para que los formatos estandarizados estén en línea con los mismos estándares y preocupaciones derivadas del mismo. Es mínimamente preocupante cuando una norma busca garantizar solo su formato estándar, es decir, el formato de la referencia bibliográfica, y se vuelve indiferente a todos los elementos que pretende estandarizar en esa referencia, desde sus autores hasta el nombre de la revista científica.

Es útil que los investigadores comprendan la verdadera razón de cuidar los detalles "formales" o "normativos" de su trabajo, para que pueda ser recuperado como referencia, y que sus argumentos y citas sean también correctamente identificados y útiles para sus compañeros. También es prudente que los investigadores se orienten sobre la gestión de su tiempo de investigación, en cuanto al tiempo de escritura y el tiempo a dedicar a la normalización, aunque actualmente están disponibles varios gestores de referencias, como EndNote, Mendeley, Zotero, etc. que pueden hacer el proceso de normalización un poco menos arduo, aunque no eximen al autor del trabajo de una mirada cercana.

La tarea de orientación y concienciación de la comunidad científica en cuanto a la normalización y el uso adecuado de los nombres de los autores podría formar parte de la discusión que proponen y abordan los manuales metodológicos, sean ellos las propias normas de normalización o los reputados libros de metodología científica en las más diversas áreas del conocimiento y para los diferentes niveles de académicos, investigadores y profesionales. Dado que estos manuales proporcionan orientación sobre las prácticas de investigación, desde, por ejemplo, el diseño de los proyectos, hasta los detalles sobre la investigación cuantitativa y cualitativa, sería beneficioso que estas obras, tan consultadas y utilizadas como referencia por muchos investigadores, también trataran de temas como el control de autoridades.

3.2 Uso de la tecnología para desambiguar autorías e instituciones

Los avances en las herramientas tecnológicas aportan soluciones y agilidad en procesos que fuera del entorno digital exigirían tiempo y esfuerzo a los profesionales en actividades como la construcción de control de autoridades, listas de materias (índices), nombres de instituciones, etc.

El término *Author Name Disambiguation* (AND) o (Desambiguación del Nombre del Autor) se encuentra en la literatura científica y aborda técnicas o discusiones relativas a la aplicación de soluciones tecnológicas que, con la colaboración de profesionales y académicos de las Ciencias de la Información y la Informática, buscan resolver o minimizar los problemas de ambigüedad en la autoría.

Smalheiser y Torvik (2009) señalan cuatro retos para esta desambiguación: un investigador puede publicar con diferentes nombres a lo largo de su carrera académica; puede haber investigadores con nombres idénticos; metadatos incompletos por parte de, por ejemplo, las revistas científicas o las bases de datos, que pueden almacenar los nombres de los autores sin otra información como su nacionalidad, títulos académicos y otros datos identificativos; y el creciente número de publicaciones no solo con autoría múltiple, sino también con múltiples áreas e instituciones implicadas. Estos factores hacen reflexionar sobre la importancia de las técnicas y procesos que optimizan la desambiguación aliados a las políticas establecidas por las fuentes de información científica, como las revistas científicas o las bases de datos, para que los metadatos de los registros estén indexados con una riqueza de información suficiente para permitir la desambiguación realizada por las herramientas tecnológicas.

La cantidad de información que se recupera en las búsquedas realizadas en un entorno web, independientemente de su asertividad y relevancia, hace que el proceso de tratamiento informativo o desambiguación de la autoría sea cada vez menos práctico de realizar manualmente. Para Veve (2009), en cambio, la desambiguación automática de los nombres nunca podría ser realizada solo por soluciones tecnológicas, y que siempre sería necesario el ojo de un humano para completar el proceso con éxito. Para Elliott (2010), un enfoque híbrido, que trabaja con soluciones tecnológicas combinadas con la verificación humana, parece ser una excelente alternativa; sin embargo, es necesario que cada caso abordado tenga en cuenta sus particularidades.

En la literatura científica, es posible encontrar varios estudios que abordan técnicas o metodologías para realizar la desambiguación de nombres de autores, ya sea en formato de revisión, es decir, enumerando las alternativas disponibles en el momento en que se desarrollaron los estudios, o análisis centrados en fuentes de información específicas, como los métodos y esfuerzos de las bases de datos científicas. En el presente trabajo se mencionarán algunos de los principales estudios y algunas de las técnicas que abordaron.

Smalheiser y Torvik (2009), autores de varios estudios sobre la desambiguación del nombre de los autores, comentan que parte de los enfoques de investigación sobre el tema comparten líneas generales con el *machine learning*, con el objetivo de agrupar o clasificar los trabajos según sus autores, aplicable a la mayoría de las bases de datos bibliográficas existentes.

Una técnica también comentada en la revisión de Smalheiser y Torvik (2009) es el *stylometry*, que es un proceso utilizado para diferenciar a los autores según sus estilos de escritura, incluyendo el análisis y la comparación de la estructura de la oración, la gramática, el contexto de la publicación (*post* del blog, artículo académico, etc.), entre otros aspectos.

Smalheiser y Torvik (2009) presentan en su revisión un proyecto propio, denominado *Author-ity Project*, que tuvo como primer paso la desambiguación de la autoría en la base de datos Medline, destacando el nivel de dificultad que presenta esta fuente, que cuenta con publicaciones científicas de una misma gran área de conocimiento, que es la de las Ciencias de la Salud, y pueden existir autores que tengan los mismos nombres, publiquen sobre los mismos temas y tengan las mismas afiliaciones.

El *Author-ity Project* tiene un modelo estadístico que se basa en una comparación vectorial que incorpora seis ítems de los registros de Medline, las palabras del título, el nombre de la revista, los nombres de los coautores, los descriptores de *Medical Subject Headings* (MeSH), el idioma y las palabras de afiliación, además de comprobar la estructura del nombre del autor (cuáles son las iniciales del nombre, el sufijo, etc.); en un segundo momento, también se incorporó como ítem el correo electrónico de los autores (Smalheiser y Torvik, 2009).

Un proyecto mencionado por Elliott (2010) es el de la British Library's Name Project, que fue una respuesta al creciente número de repositorios institucionales en el Reino Unido, y su diseño corresponde a la creación de un número único, o identificador único (ID) que vincularía diferentes formas de nombres de autores.

Específicamente en el caso de las instituciones, la desambiguación por nombres de instituciones, o desambiguación del nombre de la institución (IND), puede reunir diferentes formas del nombre de una institución en la misma entidad institucional, haciendo la normalización de los nombres institucionales para optimizar la precisión de la recuperación de los registros en las bases de datos (Huang *et al.*, 2020). Huang *et al.* (2020, p. 390) afirman que, los estudios relevantes sobre el tema pueden dividirse en tres categorías:

- a) **La IND basada en similitud de palabras:** comprueba el grado de similitud entre dos *strings*, normalmente mediante un análisis porcentual; las métricas y algoritmos más utilizados son la similitud del coseno, la distancia euclidiana, la distancia de Hamming y la distancia de edición;

b) **La IND basada en reglas y estadísticas:** combina un modelo de reglas o *rule model* con técnicas de identificación de similitudes de palabras mediante un algoritmo;

c) **La IND basada en recurso genérico de clasificación:** utiliza técnicas de aprendizaje automático, considerando que cada institución está representada por un perfil que utiliza características generales independientes de la lexicalización.

Las técnicas utilizadas para desambiguar los nombres de las instituciones, así como en el caso de los nombres de los autores, utilizan tecnología y conocimiento que provienen de toda la informática general, requiriendo profundos conocimientos técnicos para su aplicabilidad.

Los métodos manuales o automáticos desarrollados y utilizados para desambiguar las autorías personales o institucionales en los diferentes sistemas o incluso en los catálogos de las bibliotecas, además de ser imprescindibles en la recuperación de la información bibliográfica por parte de sus potenciales usuarios, son necesarios para optimizar los procesos evaluativos a los que se someten los investigadores e instituciones de la comunidad científica. Las evaluaciones, especialmente en su perspectiva cuantitativa, implican la recuperación de información bibliográfica por sus nombres en diferentes fuentes de información y, si hay problemas en la recuperación, hay un impacto directo en los resultados de la evaluación.

3.3 Impacto en los indicadores métricos científicos

La evaluación de la ciencia es un tema que plantea varias discusiones, especialmente en relación con el uso adecuado y responsable de las métricas, que suelen ser más cuantitativas que cualitativas, lo que genera una afirmación pertinente que debe ser considerada por los evaluadores.

La evaluación es intrínseca a la actividad científica y para muchas de estas actividades se utilizan fuentes de información bibliográfica para recoger parte de los insumos que serán evaluados, con el fin de definir diversas cuestiones como el otorgamiento de becas por parte de las agencias de fomento, la contratación o el despido de investigadores, entre otras. Esta recuperación informativa está directamente relacionada con los nombres de los autores y las instituciones y el impacto de las posibles irregularidades en su representación.

El proceso de evaluación de la investigación científica debe llevarse a cabo según diferentes niveles de medición del impacto:

[...] es importante entender que el impacto puede -y debe- medirse a diferentes niveles. Podemos hablar del impacto de un artículo, que es diferente del impacto de la revista en la que se publicó ese artículo, así como del autor que lo escribió, o de la institución a la que pertenece ese autor, porque cada una de estas instancias tiene sus propias medidas de impacto independientes (Nascimento, 2017, p. 28).

Strotmann y Zhao (2012) afirman que los estudios basados en la autoría constituyen uno de los aspectos más importantes de la bibliometría, aunque su mayor inconveniente es la ambigüedad. Así, las evaluaciones de la investigación científica por parte de los autores o las instituciones necesitan necesariamente recuperar sus producciones y la forma más utilizada de hacerlo sería por su nombre. De esta manera, estas necesidades de búsqueda se alinean con las necesidades informativas de los procesos evaluativos en ciencia y la discusión sobre la recuperación efectiva de la información, considerando las ambigüedades de los nombres que se pueden encontrar en las fuentes de información bibliográfica.

Boudry y Durand-Barthez (2020) reiteran que la correcta identificación de los investigadores y su producción científica beneficia a todas las partes implicadas en la investigación. Sin embargo, encontrar toda la producción científica de un investigador o de una institución en el contexto actual de proliferación de publicaciones en revistas *online* es como buscar una aguja en un pajar (Jinha, 2010).

Sin embargo, aunque el problema de la ambigüedad del nombre del autor es bien conocido y se han propuesto y explorado varios métodos para la AND, [...] la práctica común en los estudios bibliométricos basados en el autor ha sido agrupar a todos los autores con el mismo apellido e inicial del nombre, intercambiando una alta recuperación por una menor precisión a la hora de identificar las obras del autor para su análisis. Este uso todavía predominante del enfoque AND simplificado probablemente se debe en gran medida a la forma en que los nombres de los autores han sido referenciados durante mucho tiempo en las publicaciones periódicas e indexados en las bases de datos bibliográficas, especialmente para las referencias citadas (Strotmann y Zhao, 2012, p. 1822, traducción nuestra).

El papel de las bases de datos bibliográficas en el entorno digital destaca por su potencial para reunir, a través de sus plataformas, grandes cantidades de publicaciones científicas que servirán de insumo para el desarrollo científico en las próximas décadas. De este modo, la calidad de las bases de datos es un factor primordial para garantizar la eficacia del

proceso de comunicación científica y sus errores pueden afectar directamente a la identificación, selección, extracción, clasificación, ordenación y tabulación de los datos, que pueden utilizarse para realizar estudios o evaluaciones de la ciencia (Ruiz-Pérez; López-Cózar y Jiménez-Contreras, 2002). En este sentido, la recuperación efectiva de los nombres de los autores y de las instituciones incide directamente en el proceso de evaluación de la ciencia, que utiliza indicadores cuantitativos, bibliométricos y, más recientemente, altmétricos.

Según Sugimoto y Larivière (2018, p. 50, traducción nuestra), "la autoría es el corazón de los indicadores bibliométricos: son los autores cuyos nombres e instituciones aparecen en un documento, los que se miden". Tal y como se ha comentado en relación con la desambiguación de nombres de autores realizada por técnicas y métodos que implican tecnología, es importante reflexionar que, para la realización de los indicadores utilizados en las evaluaciones científicas, se debe considerar también la necesidad de obtener información adicional sobre el autor, más allá de la publicación.

Uno de los usos de la investigación bibliométrica es extraer información que no se presenta explícitamente en las bases de datos tradicionales, como inferir patrones de nuevas tecnologías, el grado de novedad y originalidad en las patentes, o identificar el género del autor, la edad académica o las trayectorias migratorias de las publicaciones de investigación. Exponer la información implícita puede aumentar el valor de los datos y su poder explicativo. El nombre del autor representa un dato del que se puede deducir más informaciones (Karaulova, Gök y Shapira, 2019, p. 488, traducción nuestra).

Por su parte, las métricas alternativas, que también se utiliza en los procesos evaluativos en ciencia, pueden definirse como "indicadores de la comunicación científica que ayudan a comprender cómo se ven y utilizan los resultados de la investigación en entornos *online*, complementando el análisis tradicional basado en el recuento de citas" (Nascimento, 2017, p. 56).

Sin embargo, la altmetría es todavía poco utilizada en varios países, incluido Brasil, y no es totalmente comprendida por los investigadores. Esta falta de comprensión, junto con la ausencia de interacción y perfiles en las redes académicas, hace que el indicador altmétrico no se considere a menudo para integrar las evaluaciones de la investigación científica. Como las menciones en la web deben hacerse siempre a través del identificador persistente, esta norma no siempre se conoce y muchas menciones no son contabilizadas por las herramientas que agregan este tipo de métricas, como Altmetric y PlumX.

[...] la falta de normalización y calidad de los metadatos de las publicaciones nacionales impide el correcto reconocimiento de los elementos individuales de la producción académica y el seguimiento de las citas recibidas en línea. La baja participación de la comunidad científica brasileña en la web y en los entornos en línea es otro factor que inhibe la adopción de métricas alternativas (Silva y Gouveia, 2020, p. 15).

En este sentido, las fuentes de información, grupos editoriales o independientes, pero de alguna manera vinculados a la comunidad científica, y los investigadores, han creado espacios digitales para que se pueda añadir información sobre los investigadores y sus investigaciones más allá de las publicaciones, ya sea en forma de sus metadatos o del texto completo. En estos entornos, es posible vislumbrar una interacción entre la comunidad académica y, en algunos de ellos, recoger toda la trayectoria académica de un investigador, independientemente de su aportación y de las fuentes en las que se haya realizado, sin restricciones.

3.4 Autenticidad y autoridad de contenidos en ambiente digital

Atribuir correctamente la autoría de los contenidos en los entornos digitales se convierte en un reto, sobre todo si se tiene en cuenta la velocidad de difusión de los contenidos por parte de las distintas plataformas. La información científica, asimismo, se comparte a gran escala en entornos digitales, entre miembros de la Academia.

Sin embargo, el intercambio desenfrenado de contenidos sin autenticidad probada, sean de carácter científico o no, acaban contribuyendo a las llamadas *fake news*, dándoles aún más fuerza y poder en su difusión. Estas noticias falsas o contenidos de sospechosa autenticidad, consiguen difundir calumnias, difamaciones y desinformación que contribuyen a la desinformación de la sociedad a gran escala.

La información científica también es víctima de las *fake news* y muchos contenidos se comparten en las redes sociales como si tuvieran validez científica o como si fueran dichos por supuestos investigadores vinculados a universidades o instituciones de investigación nacionales o internacionales de renombre.

Para ayudar en el proceso de identificación correcta de los autores de los contenidos científicos, Gasparyan *et al.* (2016) enumeran una serie de opciones que pretenden identificar, validar y generar indicadores de calidad científica, considerando el impacto que la ambigüedad o la falta de precisión en los nombres de los autores causan en la recuperación informativa.

Escrito con precisión / nombre, medio y apellido abreviados
Afiliaciones vinculadas
Contactos postales y de <i>e-mail</i>
Inscripción
Fotos en artículos o en plataformas institucionales o individuales
Notas biográficas en portales académicos, institucionales o sociales, incluyendo páginas en la Wikipedia
Videos con resúmenes de artículos
<i>Open Researcher and Contributor ID (ORCID)</i>
Registros académicos en bases de datos bibliográficas o plataformas (PubMed, Scopus, Web of Science y Google Scholar)
Perfil en redes académicas y corporativas (ResearchGate y LinkedIn)
Perfil en redes sociales (Facebook, Twitter)
Estilo de escritura académica

Cuadro II. Opciones para identificar y evaluar autores científicos. Fuente: adaptado de Gasparyan *et al.*, 2016, p. 172, traducción nuestra.

El cuadro II muestra algunas opciones utilizadas en los medios digitales para desambiguar la autoría considerando nuevas plataformas creadas con este objetivo. Sin embargo, es importante recalcar cuánto pueden aportar los estándares de catalogación y ser aplicados también en un entorno digital, pues establecen pautas directamente relacionadas con el formato de la descripción de los autores en los materiales bibliográficos. En este sentido, es fundamental destacar, por encima de todo, la RDA.

RDA fue diseñado para su uso en nuevas estructuras de bases de datos en la web y puede adoptarse para hacer que los datos sean visibles y utilizables en un entorno digital, siempre con el objetivo de facilitar la interoperabilidad de los datos manteniendo la normalización como un aspecto esencial (Oliver, 2011). La interoperabilidad de los datos es una preocupación legítima que deben tener los responsables de construir los sistemas de información, sobre todo si se considera que, una vez que los datos están correctamente disponibles, otros sistemas similares pueden aprovecharlos y cargarlos en su banco. En la práctica, por ejemplo, se puede aplicar en relación con el control de autoridad: si en el momento de la presentación de un artículo científico el autor proporciona su ORCID, estos datos pueden ser interoperables para que la base de datos también pueda leerlos y almacenarlos.

La correcta identificación de los investigadores y de sus investigaciones tiene un impacto directo en la autenticidad y la autoridad de los contenidos científicos en el entorno digital, cruzando varias cuestiones que remiten al establecimiento de directrices por parte de los editores y las revistas, así como a otras iniciativas que orientan a los investigadores para una conducta adecuada, no solo en relación con la forma de comunicar sus nombres y afiliaciones, sino también en relación con la ética vinculada al proceso.

4 CONSIDERACIONES FINALES

Para que la recuperación de la información científica sea precisa y veraz, es necesario que los datos pasen por procesos de organización y representación adecuados, de modo que proporcionen resultados relevantes para las necesidades de información del investigador. En este sentido, el control de las autoridades personales y organizativas se realiza a través de metodologías y normas establecidas a nivel nacional e internacional por el área de Biblioteconomía, Documentación y Ciencia de la Información. Otro esfuerzo por contribuir a la estandarización y recuperación de la información se encuentra en los formatos bibliográficos estandarizados, aunque existen diferentes patrones de uso según áreas específicas de conocimiento, que proponen la estandarización para ayudar en el proceso de intercambio de información, objetivo intrínseco de su concepto. Sin embargo, muchos estándares bibliográficos no son tan rígidos, permitiendo adaptaciones por parte de sus usuarios, y algunos vacíos que dejan pueden no permitir una recuperación efectiva de la información, especialmente en un entorno digital.

Las publicaciones científicas tienen la opción de ser recuperadas a través de sus autores y de sus respectivas instituciones de origen, instituciones de fomento o instituciones en las que se haya realizado la investigación. Para la presente investigación, se indagó las acciones e iniciativas para controlar la información con el fin de resaltar su papel fundamental en la sistematización para tener una recuperación y acceso relevante.

El control de autoridad adoptado se realiza a partir de principios y reglas orientados a normalizar los puntos de acceso en los catálogos bibliográficos, incluso antes de la aparición y consolidación de las fuentes de información en el entorno

web. En este contexto, la visibilidad de la investigación implica visibilidad de sus autores y sus instituciones, dándoles crédito por la realización y propiciando el contacto entre pares y el establecimiento de posibles asociaciones.

Sin embargo, son muchas las personas que intervienen en el proceso de publicación de la ciencia y, por lo tanto, el estándar utilizado no es siempre el mismo. La publicación, antes de estar disponible para su consulta y uso por parte de la comunidad académica, pasa por sus autores, revisores y editores y, finalmente, se indexa en las fuentes de información. En este sentido, es evidente que un estándar establecido a lo largo de este flujo de trabajo podría minimizar las variaciones de representación de los nombres de los investigadores y de sus instituciones, aunque, aun así, podrían surgir variaciones debidas a errores humanos, como errores ortográficos o simplemente de digitación.

Ante tantas implicaciones de la inadecuada representación de los investigadores y sus instituciones, potenciadas por la cantidad de información disponible en el entorno digital, surgen iniciativas tecnológicamente nativas con el objetivo de ayudar en los procesos de desambiguación e identificación de nombres, despertando varios proyectos y metodologías discutidas en la literatura y aplicadas en diversas fuentes informativas, involucrando codificación, lenguajes de programación y uso de algoritmos. Son alternativas que buscan consolidar lo que ya está disponible en el entorno digital y no realizan la inserción de nombres fuera de la norma.

Como se ha observado, el entorno digital ha permitido un gran acceso y una difusión exponencial de la información integral y/o de los metadatos disponibles. Esta información no siempre es veraz o completa, lo que provoca confusión en relación con lo que se expone. Desde la perspectiva particular de esta investigación, se exploraron iniciativas en relación con el control de la autoridad personal e institucional, entrando en las cuestiones de las normas bibliográficas y los manuales metodológicos, mostrando que las tecnologías para la desambiguación pueden ser una alternativa para dicha actividad. Para que se sigan desarrollando estándares apropiados y nuevas tecnologías que contribuyan a este proceso, es imperativo que la comunidad científica desarrolle nuevas investigaciones y que estén ampliamente disponibles para dar continuidad a la discusión sobre el tema.

La información no estandarizada causa problemas en cuanto a la recuperación correcta impactando negativamente en la métrica/evaluación científica que puede causar no solo resultados incorrectos, sino también el no estímulo por parte de las agencias de financiación. Así, es evidente que la disponibilidad de los documentos científicos debe pasar obligatoriamente por el correcto tratamiento y la normalización documental para transmitir su autenticidad.

5 BIBLIOGRAFÍA

- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *ABNT NBR 10520: Citações em documentos – Apresentação*. Rio de Janeiro: ABNT, 2002.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *ABNT NBR 12225: Lombada – Apresentação*. Rio de Janeiro: ABNT, 2004.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *ABNT NBR 6022: Artigo em publicação periódica técnica e/ou científica – Apresentação*. Rio de Janeiro: ABNT, 2018a.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *ABNT NBR 6023: Referências – Elaboração*. Rio de Janeiro: ABNT, 2018b.
- ASSUMPÇÃO, F.S.; SANTOS, P.L.V.A. da C. y ZAFALON, Z.R. O controle de autoridade no domínio bibliográfico: os catálogos digitais. *Biblios*, 2017a, vol. 68, p. 21-33, [en línea]. Disponible en: <http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1562-47302017000300002> [Consulta: 30 de julio de 2021]
- ASSUMPÇÃO, F.S.; SANTOS, P.L.V.A. da C. y ZAFALON, Z.R. O controle de autoridade no domínio bibliográfico: os catálogos em livros e em fichas. *Biblios*, 2017b, vol. 67, p. 84-98, [en línea]. Disponible en: <http://dx.doi.org/10.5195/biblios.2017.341>.
- BELLINI, A. de B. *et al.* Normalização documental: o caminho de acesso à produção científica. In: *SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS...* 2006, Salvador. Anais [...] Salvador: UFBA, 2006, [en línea]. Disponible en: <<http://repositorio.febab.org.br/items/show/5428>> [Consulta: 20 de junio de 2021]
- BOUDRY, C. y DURAND-BARTHEZ, M. Use of author identifier services (ORCID, ResearcherID) and academic social networks (Academia.edu, ResearchGate) by the researchers of the University of Caen Normandy (France): a case study. *Plos One*, 2020, vol. 15, n° 9, e0238583, [en línea]. Disponible en: <https://doi.org/10.1371/journal.pone.0238583>.
- BOURNE, C.P. Frequency and impact of spelling errors in bibliographic data bases. *Information Processing & Management*, 1977, vol. 13, n° 1, p. 1-12.
- BRANDT, S. Authority files for microcomputer databases. *Special Libraries*, 1988, vol. 79, n° 4, p 296-301.

- CLACK, D.H. Authority control and linked bibliographic databases. *Cataloging & Classification Quarterly*, 1988, vol. 8, nº 3-4, p. 35-46.
- CRESPO, I.M. y RODRIGUES, A.V.F. Normas técnicas e comunicação científica. *Revista Digital de Biblioteconomia e Ciência da Informação*, 2011, vol. 9, nº 1, p. 36-55, [en línea]. Disponible en: <https://doi.org/10.20396/rdbci.v9i1.1918>.
- CUTTER, C.A. *Rules for a dictionary catalog*. 4th ed. Washington: Government Printing Office, 1904.
- DIAS, M.M.K. Normas técnicas. In: CAMPELLO, B.S.; CENDÓN, B.V.; KREMER, J.M. (Orgs). *Fontes de informação para pesquisadores e profissionais*. Belo Horizonte: Ed. UFMG, 2000. p. 137-152.
- ELLIOTT, S. Survey of author name disambiguation: 2004 to 2010. *Library Philosophy and Practice*, 2010, vol. 473, s. n., p. 1-11, [en línea]. Disponible en: <https://digitalcommons.unl.edu/libphilprac/473/> [Consulta: 13 de junio de 2021]
- FACHIN, O. *Fundamentos de metodologia*. 3. ed. São Paulo: Saraiva, 2001.
- FEDERAÇÃO BRASILEIRA DE ASSOCIAÇÕES DE BIBLIOTECÁRIOS – FEBAB. *Código de Catalogação Anglo-Americano*. 2. ed. rev. 2002. Preparado sob a direção do Joint Steering Committee for Revision of AACR. São Paulo: FEBAB: Imprensa Oficial do Estado de São Paulo, 2004.
- FONSECA, E.N. da. *Problemas de comunicação da informação científica*. São Paulo: Thesaurus Editora, 1973.
- FOX, R. From strings to things. *Digital Libraries Perspective*, 2016, vol. 32, nº 1, p. 2-6, [en línea]. Disponible en: <https://doi.org/10.1108/DLP-10-2015-0020>.
- GASPARYAN, A.Y.; YESSIRKEPOV, M.; GERASIMOV, A.N.; KOSTYUKOVA, E.I. y KITAS, G.D. Scientific author names: errors, corrections, and identity profiles. *Biochemia Medica*, 2016, vol. 26, nº 2, p. 169-173, [en línea]. Disponible en: <http://dx.doi.org/10.11613/BM.2016.017>.
- HAAK, L.L.; MEADOWS, A. y BROWN, J. Using ORCID, DOI, and Other open identifiers in research evaluation. *Frontiers in Research Metrics and Analysis*, 2018, vol. 3, nº 2, [en línea]. Disponible en: <https://doi.org/10.3389/frma.2018.00028>.
- HERRERO PASCUAL, C. El control de autoridades. *Anales de Documentación*, 1999, nº 2, p. 121-136, [en línea]. Disponible en: <https://revistas.um.es/analesdoc/article/view/2621> [Consulta: 30 de mayo de 2021]
- HOPPEN, N.H.F.; MAGNUS, A.P.M. y VANZ, S.A. de S. Referências de atos normativos brasileiros na comunicação científica internacional: estilos Vancouver e APA. *Prisma.com*, 2018, vol. 26, p. 41-58, [en línea]. Disponible en: <https://brapci.inf.br/index.php/res/v/73975> [Consulta: 20 de junio de 2021]
- HUANG, Y.; LI, J.; SUN, T. y XIAN, G. Institution information specification and correlation based on institutional PIDs and IND tool. *Scientometrics*, 2020, vol. 122, p. 381-396, [en línea]. Disponible en: <https://doi.org/10.1007/s11192-019-03268-9>.
- IFLA WORKING GROUP ON FLANAR. *Functional Requirements for Authority Data: a conceptual model*. Den Haag: IFLA, 2013, [en línea]. Disponible en: https://www.ifla.org/files/assets/cataloguing/frad/frad_2013.pdf [Consulta: 12 de junio de 2021]
- INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. *Declaração dos Princípios Internacionais de Catalogação (PIC)*. Holanda, 2016. Tradução de Marcelo Votto Teixeira, [en línea]. Disponible en: https://www.ifla.org/files/assets/cataloguing/icp/icp_2016-pt.pdf [Consulta: 10 de junio de 2021]
- INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. *Name of persons*. [s. l.]: IFLA, 2020, [en línea]. Disponible en: <https://www.ifla.org/node/4953#:~:text=According%20to%20the%20International%20Cataloguing,conventions%20of%20the%20country%20and> [Consulta: 10 de junio de 2021]
- JINHA, A.E. Article 50 million: an estimate of the number of scholarly articles in existence. *Learned Publishing*, 2010, [s. l.], vol. 23, p. 258-263, [en línea]. Disponible en: <https://doi.org/10.1087/20100308>.
- KARAULOVA, M.; GÖK, A. y SHAPIRA, P. Identifying Author Heritage Using Surname Data: An Application for Russian Surnames. *Journal of the Association for Information Science and Technology*, 2019, vol. 70, nº 5, p. 488-498, [en línea]. Disponible en: <https://dx.doi.org/10.1002/Fasi.24104>.
- MEADOWS, A.J. *A comunicação científica*. Brasília (DF): Briquet de Lemos, 1999.
- MEY, E.S.A. y SILVEIRA, N.C. *Catalogação no plural*. Brasília, DF: Briquet de Lemos, 2009.
- MUGNAINI, R.; DIGIAMPIETRI, L.A.; OLIVEIRA, L.C. de y FERREIRA, S.M.S.P. Normalização de nomes de autores em fontes de informação institucionais: proposta de um método automático de verificação de erros. Em *Questão*, 2012, vol. 18, nº especial, p. 263-279, [en línea]. Disponible en: <https://seer.ufrgs.br/EmQuestao/article/view/33265> [Consulta: 12 de junio de 2021]
- NASCIMENTO, A.G. do. *Altmetria para bibliotecários: guia prático de métricas alternativas para avaliação da produção científica*. São Paulo: Scorecci, 2017.
- OLIVER, C. *Introdução à RDA: um guia básico*. Brasília, DF: Briquet de Lemos, 2011.
- PATRIAS, K. y WENDLING, D. (ed.). *Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers*. 2a ed. Bethesda (MD): National Library of Medicine: 2007, [en línea]. Disponible en: <https://www.ncbi.nlm.nih.gov/books/NBK7256/> [Consulta: 20 de junio de 2021]

- RODRIGUES, M.E.F.; LIMA, M.H.T. de F. y GARCIA, M.J. de O. A normalização no contexto da comunicação científica. *Perspectivas em Ciência da Informação*, 1998, vol. 3, nº 2, p. 147-156, [en línea]. Disponible en: <<https://brapci.inf.br/index.php/res/v/33139>> [Consulta: 17 de junio de 2021]
- ROSADO, K.M.L. y DIAS, C. da C. Controle de autoridade de nomes de pessoas: uma proposta a partir da análise da Biblioteca Digital do Ministério Público Federal. *Múltiplos Olhares em Ciência da Informação*, 2019, [s. l.], vol. 9, nº 2, p. 1-19.
- ROSADO, K.M.L. y DIAS, C. da C. Representação e recuperação de nomes de pessoas em catálogos de autoridades. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, 2020, vol. 25, p. 1-25, [en línea]. Disponible en: <https://doi.org/10.5007/1518-2924.2020.e70349>.
- RUIZ-PÉREZ, R.; LÓPEZ-CÓZAR, E.D. y JIMÉNEZ-CONTRERAS, E. Spanish personal name variations in national and international biomedical databases: implications for information retrieval and bibliometric studies. *Journal of the Medical Library Association*, 2002, vol. 90, nº 4, p. 411-430, [en línea]. Disponible en: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC128958/>> [Consulta: 27 de junio de 2021]
- SANDBERG, J. y JIN, Q. How should catalogers provide authority control for journal article authors? Name identifiers in the linked data world. *Cataloguing & Classification Quarterly*, 2016, vol. 54, nº 8, p. 537-552, [en línea]. Disponible en: <https://doi.org/10.1080/01639374.2016.1238429>.
- SEVERINO, A.J. *Metodologia do trabalho científico*. São Paulo: Cortez, 2007.
- SILVA, J.A.D. da y GOUVEIA, F.C. Uso de identificadores persistentes e research blogging por blogues científicos em língua portuguesa e sua cobertura pela altmetric. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, 2020, vol. 25, p. 1-19, [en línea]. Disponible en: <https://doi.org/10.5007/1518-2924.2020.e71277>.
- SMALHEISER, N.R. y TORVIK, V.I. Author name disambiguation. *Annual Review of Information Science and Technology*, 2009, vol. 43, p. 287-313, [en línea]. Disponible en: <http://abel.lis.illinois.edu/tutorial/ARIST_preprint.pdf> [Consulta: 13 de junio de 2021]
- STROTMANN, A. y ZHAO, D. Author name disambiguation: what difference does it make in author-based citation analysis. *Journal of the American Society for Information Science and Technology*, 2012, vol. 63, nº 9, p. 1820-1833, [en línea]. Disponible en: <https://doi.org/10.1002/asi.22695>.
- SUGIMOTO, C.R. y LARIVIÈRE, V. *Measuring research: what everyone needs to know*. New York: Oxford University Press, 2018.
- VEVE, M. Supporting name authority control in XML metadata: a practical approach at the University of Tennessee. *Library Resources & Technical Services*, 2009, vol. 53, p. 41-52, [en línea]. Disponible en: <https://doi.org/10.5860/lrts.53n1.41>.