

Módulo de capacitación para la recolección y el análisis de indicadores de producto de las actividades de ciencia y tecnología

Autor: Anna María Prat



Este módulo fue elaborado en el marco del proyecto "Fortalecimiento del sistema de información sobre la red interamericana de ciencia, tecnología e innovación", dentro de su componente "Capacitación y asistencia técnica para la mejora en la recolección y análisis de indicadores de ciencia, tecnología e innovación en países de América Latina y el Caribe", bajo la coordinación de Mario Albornoz. La autora de este módulo fue Anna María Prat.



PRESENTACIÓN	7
1.INTRODUCCIÓN	8
2.INDICADORES BIBLIOMÉTRICOS	12
2.1 Algunas leyes y conceptos	13
2-1-1 Ley de Lotka	13
2.1.2 Ley de distribución de Bradford	13
2.1.3 Ley de Zipf	14
2.1.4 Derek de Solla Price. Crecimiento y contemporaneidad de la ciencia	14
2.1.5 Eugene Garfield y los índices de citas	15
2.2 Fuentes y análisis de datos	15
2.2.1 Web of Science	16
2.2.2 SCOPUS	17
2.2.3 PASCAL y FRANCIS	18
2.2.4 Google Académico	18
2.2.5 SciELO	19
2.2.6 Bases de datos especializadas	19
2.2.7 Bases de datos latinoamericanas	20
2.1.8 Redalyc y Latindex	20
2.2.9 Bases de datos nacionales complementarias	20
2.2.10 Acceso a datos comparativos generales. SciMago y RICYT	21
2.3 Definiciones y metodología	22
2.3.1 Metodología general	22
2.3.2 Indicadores de actividad (indicadores cuantitativos)	24
2.3.3 Indicadores de calidad (Indicadores cualitativos)	26
3 INDICADORES DE PATENTES	30
3.1 Procedimientos de solicitudes de patentes	31
3.2 Uso de patentes como fuente de indicadores	32
3.2.1 Bases de datos	32
3.2.2 Información disponible	32
3.2.3 Tipos de medición	33
3.2.4 Información cualitativa	34
4 WEBOMETRIA O CIBERMETRIA	35
5 MINERÍA DE DATOS, MINERÍA DE TEXTOS, DESCUBRIMIENTO DE CONOCIMIENTO	36
BIBLIOGRAFIA	39

» PRESENTACIÓN

El presente módulo de capacitación ha sido elaborado en el marco del proyecto "Fortalecimiento del sistema de información sobre la red interamericana de ciencia, tecnología e innovación", inscripto dentro del Diálogo Regional de Política en ciencia y tecnología impulsado por el Banco Interamericano de Desarrollo.

La intención de las actividades de capacitación impulsadas desde el proyecto es mejorar la cobertura de indicadores de ciencia, tecnología e innovación de los países de la región mediante la creación o la consolidación de sus capacidades. Se apunta así a que estos países, y en especial aquellos de menor desarrollo relativo, puedan ampliar el conjunto de indicadores que producen y sean capaces de sostener este esfuerzo en el tiempo.

La experiencia de América Latina y el Caribe muestra que la escasez de personal capacitado en la construcción de indicadores actúa como un obstáculo para la generación de información confiable acerca de las actividades científicas, tecnológicas y de innovación. Tal información, con la adecuada calidad, debería ser la base sobre la cual se tomen decisiones políticas en esos ámbitos. Las tareas de capacitación pretenden, en tal sentido, contribuir a la generación de una masa crítica de profesionales entrenados para relevar información y construir indicadores confiables, los cuales resultan indispensables para precisar las variables sobre las que habrá de operar la política a ser implementada.

Se presenta a continuación el módulo de capacitación sobre indicadores del producto de la investigación, enfocado especialmente en realizar un recorrido por los diferentes sistemas para evaluar este aspecto, para lo cual brinda un panorama de la construcción de indicadores bibliométricos y de patentes, así como de las llamadas "webometría" y "cibernetría" y de nuevas herramientas para el estudio de este campo, tales como la minería de datos. El material recoge la experiencia en esta materia desarrollada en el seno de la Red Iberoamericana e Interamericana de Indicadores de Ciencia y Tecnología (RICYT). El módulo está destinado principalmente a quienes, dentro de instituciones como los consejos de investigación de los países de América Latina y el Caribe, tienen la necesidad de medir la actividad científica desarrollada con los recursos asignados por los estados. Los indicadores de producto cobran su mayor sentido en el marco de un conjunto de indicadores destinados a mostrar, evaluar y orientar la política científica y los instrumentos desarrollados para fomentar la investigación, el desarrollo y la innovación a nivel nacional.

Así, el presente módulo pretende ser un aporte para la mejora de las capacidades metodológicas y técnicas de los organismos de las áreas de ciencia, tecnología e innovación en la región y así contribuir, en la medida de lo posible, a la construcción de capacidades que contribuyan al desarrollo integral de los países de América Latina y el Caribe.

01. Introducción

Las actividades de investigación apuntan a obtener como resultado la creación de nuevo conocimiento o de nuevos procesos, productos y servicios. A medida que los presupuestos destinados a la investigación, desarrollo e innovación (I+D+i) crecen en cada uno de los países, se hace necesario conocer, medir y evaluar lo producido con los recursos invertidos.

El producto (output) de la actividad de investigación adquiere diferentes formas y varía según cada disciplina. En general éstas se pueden agrupar en:

- **Comunicación de resultados**, tradicionalmente a través de publicaciones, especialmente artículos científicos, monografías, presentaciones a congresos especializados, etc. En la actualidad han surgido otras formas de comunicar conocimiento científico, las cuales plantean nuevos desafíos tanto para su medición como para su evaluación.
- **Protección de la propiedad de los nuevos procesos o productos**, especialmente a través de patentes u otros resguardos, relacionados con la posibilidad de aplicación y transferencia del conocimiento.
- **Otros productos**, tales como software diseños industriales, material multimedia y muchos otros, según la disciplina en la que se originan.

El producto de la actividad de investigación se puede medir y generalmente es cuantificable. Es por ello que se pueden construir indicadores específicos para conocerlo y evaluarlo.

Sin embargo, junto a estos productos se originan, además, lo que se ha denominado "efectos" (outcomes) de la actividad de I+D+i, que varían sobremanera de un área especializada a otra y de un sector de la economía a otro, tales como:

- Producción de graduados de alta calidad
- Innovaciones tecnológicas con un gran impacto en la economía o en el bienestar social
- Ampliación de la capacidad de servicios de consultoría e investigación contratada
- Relaciones internacionales con potencialidad de intercambio, participación en redes, acceso a resultados y facilidades de investigación de alto costo, etc.

Estos efectos adquieren gran importancia en el momento de evaluar las potencialidades de instituciones y grupos de investigación y son difícilmente cuantificables. En general, no se miden en forma sistemática al momento de generar indicadores de producto de la actividad nacional. Pueden, sin embargo, llegar a ser tanto o más importantes que los resultados cuantificables y responder más directamente a políticas nacionales o sectoriales de I+D+i. En algunas áreas como salud, producción de alimentos o preservación del medio ambiente, entre otras, esto resulta evidente y es un aspecto a tener en consideración al momento de evaluar resultados.

Antes de iniciar el trabajo de medición de la actividad científica es útil recordar algunos conceptos generales:

- Cada indicador que se genera y usa es la medida de algo que tiene sentido sólo con relación a otras variables.
- Los indicadores pueden mostrar una situación dada o la realidad en un momento determinado, y su evolución en el tiempo permiten evaluar políticas, acciones o instrumentos.

- Detrás de cada indicador existe una cantidad de información recopilada con métodos y normas aceptados internacionalmente, que avalan su veracidad y compatibilidad.
- Previamente a la propuesta de uso de un nuevo indicador es necesario tener una definición clara de su objetivo, la forma de obtención de la información necesaria y la confiabilidad de los datos que lo sustentan.
- La producción de cada indicador tiene un costo, requiere de información confiable, de recursos humanos capacitados para llevar a cabo el trabajo y, muchas veces, de infraestructuras tecnológicas adecuadas.

En el caso de la producción científica, los indicadores que se generan son generalmente de dos tipos:

- indicadores de la actividad (indicadores cuantitativos), medidos, por número de publicaciones, número de patentes u otros productos;
- indicadores de la calidad de lo producido (indicadores cualitativos). medidos a través del análisis de su impacto y uso. Es importante saber no sólo cuánto se ha producido sino también cuán "bueno" es lo producido.

Si bien es relativamente fácil contar el número de publicaciones o patentes y otros productos generados por un grupo de investigación, institución o país, resulta más difícil y costoso analizar la calidad de lo producido. Sin embargo, indicadores de este tipo son los que se requieren al momento de evaluar, y no sólo medir, la actividad científica en diferentes niveles y en distintos momentos de su quehacer.

B-C. Björk, del Departamento de Gestión y Organización de la Escuela Sueca de Negocios y Administración de Empresas, ha propuesto un modelo gráfico del proceso de comunicación científica, con el objeto de ayudar a las discusiones y el análisis de los factores que intervienen en el trabajo científico. Se trata de un conjunto treinta y tres gráficos en los que se analizan cada uno de los procesos, personas y factores económicos que intervienen en el trabajo científico e ilustran los diferentes niveles en los que se requiere contar con indicadores de producto, tanto para medir la actividad como evaluarla y, sobre todo, el tipo de indicadores que deberían generarse. Se presenta a continuación el primer gráfico de la serie (Gráfico 1).

En la etapa "Financiar I+D", tanto a nivel nacional como institucional, se requiere información acerca de la cantidad y calidad de conocimiento y productos generados por el solicitante, tanto para el caso de proyectos financiados por el sector público como para los financiados por el sector privado. Se evalúa la solicitud y la capacidad de llevar a cabo la investigación, en base en los antecedentes previos del investigador o grupo.

Se ha dicho que la comunicación científica es la culminación de un proceso de financiamiento, investigación, discusión y evaluación por pares y, al mismo tiempo, el inicio de un proceso de premios y sanciones, de búsqueda de visibilidad y de permanencia. Las formas de comunicación pueden variar de una disciplina a otra, pero, en general, toda la comunidad científica se encuentra sometida a estas normas.

En palabras de Van Raan (2004): "podemos definir el avance científico como el incremento sustantivo de nuestro conocimiento acerca de 'todo'. En líneas generales diferenciamos el conocimiento básico (comprender) del conocimiento aplicable (uso). Este conocimiento puede ser 'tácito' (artesanía, habilidad, el saber hacer) o codificado (archivado y públicamente accesible)"

Los científicos han comunicado o publicado sus resultados en forma relativamente ordenada desde el siglo XVII. Es sin duda a partir de la aparición de las primeras revistas científicas, el *Journal des Scavants* y las *Philosophical Transactions of the Royal Society*, que se impone el concepto de publicar dichos resultados en revistas de carácter internacional, con la previa evaluación por pares. Intercambiar conocimiento e información es parte sustantiva del quehacer científico. Se publica para someter los resultados a juicio de las comunidades especializadas, asegurar la prioridad del descubrimiento, reclamar el derecho a la propiedad intelectual, acumular conocimiento para usos futuros. El objetivo de la comunicación científica es el registro, evaluación, diseminación y acumulación del conocimiento, hechos y percepciones humanas. Sin duda las publicaciones no son los únicos elementos importantes del proceso de medir los resultados de la ciencia, pero son una parte importante del intercambio de información. Se estima que, en la actualidad, cada año se agregan cerca de 1.000.000 de nuevas publicaciones al acervo mundial del conocimiento.

A partir de estos hechos, surgen a comienzos del siglo XX la bibliometría y la cientometría, vinculadas con los estudios de sociología de la ciencia. La bibliometría y la cientometría son disciplinas relacionadas, si bien poseen características y métodos propios. Fue Derek de Solla Price, en 1963, en su clásico *Little Science, Big Science* -traducido al español como *La Ciencia de la Ciencia*-, quien analizó el sistema de comunicación científica, su crecimiento exponencial y su efecto en las comunidades científicas y su comportamiento, así como los efectos de ese crecimiento en la sociedad contemporánea. Esta obra ha sido considerada como el primer trabajo sistemático para usar la publicación científica como elemento de medida.

Antes de cerrar esta introducción conviene puntualizar algunos conceptos básicos. Lo que interesa medir con los indicadores de producto es:

- **Producción** científica, medida por el número de publicaciones individuales o de un grupo de investigación y/o el número de patentes.
- **Productividad**: número de publicaciones o patentes según la capacidad de investigación del grupo (por ejemplo, número de jornadas completas equivalentes -JCE- dedicadas a investigación).
- **Eficiencia**: número de publicaciones o patentes por recursos invertidos.
- **Impacto**: número de veces que un trabajo es citado en revistas de corriente principal, o por patentes. Si bien el impacto no es sinónimo de calidad, es una de las medidas utilizadas para medirla.

La evaluación del producto tiene grados crecientes de complejidad, según el tamaño del grupo a evaluar o medir. Los indicadores bibliométricos tienen mayor relevancia al medir y comparar la actividad de investigación a nivel nacional, institucional o por disciplinas o áreas de desarrollo y presentan mayores problemas al ser aplicados a grupos de investigación o investigadores individuales. En estos casos es recomendable complementar las técnicas bibliométricas puras con otro tipo de evaluaciones tales como evaluación por pares, encuestas y otras metodologías.

Asimismo, es necesario tener en consideración que las formas de comunicación en ciencia varían sobremanera de una disciplina a otra. Tal como ya dijo Kuhn en 1962, cada comunidad científica de una disciplina dada sigue su propio paradigma, que se refleja en su forma de organización, lenguaje y comunicación de resultados. Si bien los artículos científicos son el medio más utilizado para comunicar resultados de investigación en las ciencias básicas y algunas aplicadas, no es necesariamente el sistema de comunicación usado en forma natural en otras áreas del conocimiento. Este es el caso de las ciencias sociales o las ingenierías. Es importante este hecho al hacer análisis o producir indicadores por disciplina, área del conocimiento o incluso de instituciones y personas.

02. Indicadores bibliométricos

La bibliometría se ha constituido, en los últimos años, en una disciplina madura, aceptada en forma generalizada, junto con la evaluación por pares, como una forma de describir, analizar y evaluar la actividad de la comunidad científica.

Fue Allan Pritchard en 1969 quien acuñó el término "bibliometría", definida como "la aplicación de métodos matemáticos y estadísticos a libros y otros medios de comunicación". El mismo año, Nalinov y Mulchenko hablaron por primera vez de "cientometría", definiéndola como "la aplicación de aquellos métodos cuantitativos que se refieren al análisis de la ciencia vista como un proceso de información". En 1994, Glänzel y Schoepflin definieron la bibliometría más ampliamente como "aquella disciplina que abarca todos los aspectos cuantitativos y modelos de la comunicación científica, almacenamiento, disseminación y recuperación de información científica".

Son muchas las aplicaciones de la bibliometría y sus métodos, y se utilizan en diversas áreas del conocimiento, entre las que se pueden mencionar especialmente:

- **Historia de la ciencia.** Contribuye al estudio del desarrollo de disciplinas científicas a partir de los resultados obtenidos por los investigadores. Permite trazar correlaciones entre disciplinas y estudiar su influencia en el desarrollo de cada una de ellas, etc.
- **Sociología de la ciencia.** Entrega información acerca de las comunidades científicas, su estructura en una sociedad determinada, sus motivaciones, comportamiento, organización, creación de redes, influencia entre comunidades, relaciones institucionales, etc.
- **Bibliotecología y gestión de información.** Es útil para determinar las colecciones núcleo de una biblioteca institucional o de una disciplina dada. Se utiliza también para el manejo de colecciones (determinación de cantidad de periódicos que se requieren para cubrir el 50%, el 80% o el 90% de un área científica), la generación de bases de datos etc.
- **Política científica.** Permite producir indicadores que sirvan en la medición de la producción, productividad y calidad científica, necesaria para evaluar y orientar programas nacionales de C+T+i. Éste es actualmente el campo de aplicación más importante, donde se destacan los estudios y análisis comparativos del desempeño científico en todos sus niveles.

La variedad de aplicaciones ha producido diferentes enfoques y metodologías debido a los diversos puntos de vista y objetivos de cada una de ellas. Es necesario contemplar este aspecto al generar indicadores bibliométricos con una metodología dada, a fin de evitar distorsiones que afecten el análisis de los resultados.

2.1 ALGUNAS LEYES Y CONCEPTOS

El análisis estadístico de la literatura científica empezó mucho antes de que se acuñara el término bibliometría.

2.1.1 LEY DE LOTKA

En 1926, Alfred Lotka publicó en el *Journal of the Washington Academy of Sciences* su trabajo acerca de la frecuencia de distribución de la productividad científica. Su propósito era "determinar, de ser posible, qué hombres de diferente calibre contribuyen al progreso de la ciencia". Para su estudio se basó en el primer índice decenal de *Chemical Abstracts* y en el *Gesichtstafeln der Physik*. En su estudio, conocido hoy como la Ley de Lotka, determinó que con independencia de la disciplina científica, el número de autores que publican n trabajos, es inversamente proporcional a n^2 y que la proporción de todos los autores que hacen una sola contribución es cerca del 60%.

$A(n)=A(1)/n^2$, donde

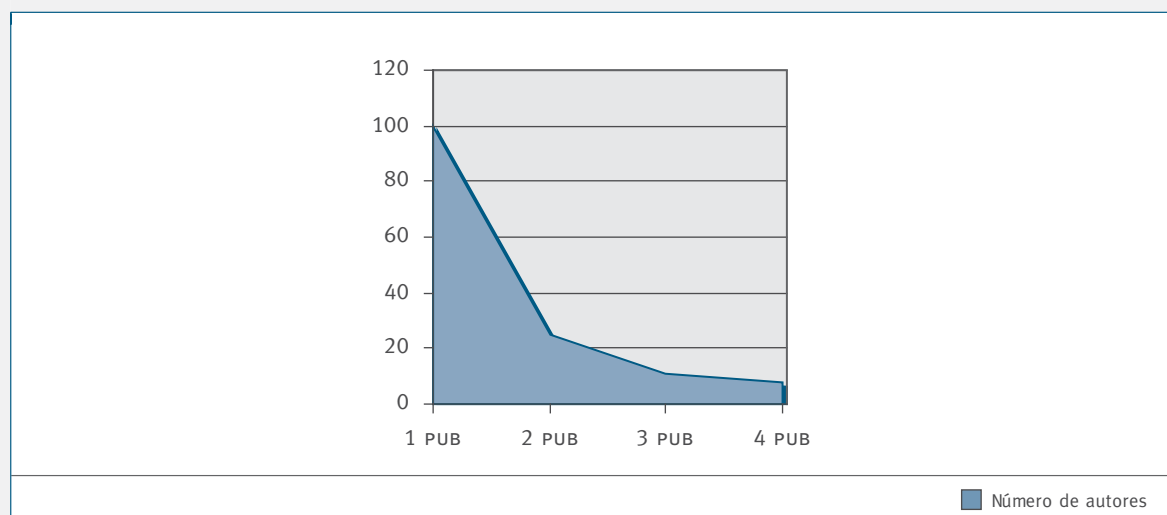
$A(n)$ es el número de autores con n trabajos,

$A(1)$, es el número de autores con 1 trabajo y

n^2 el número de trabajos al cuadrado.

El Gráfico 2 muestra la representación gráfica de la Ley de Lotka, que sirve hasta el día de hoy para determinar cuáles son los autores más productivos en una disciplina o institución dada, los grupos más eficientes, etc.

>> Gráfico 2 La representación de la Ley de Lotka

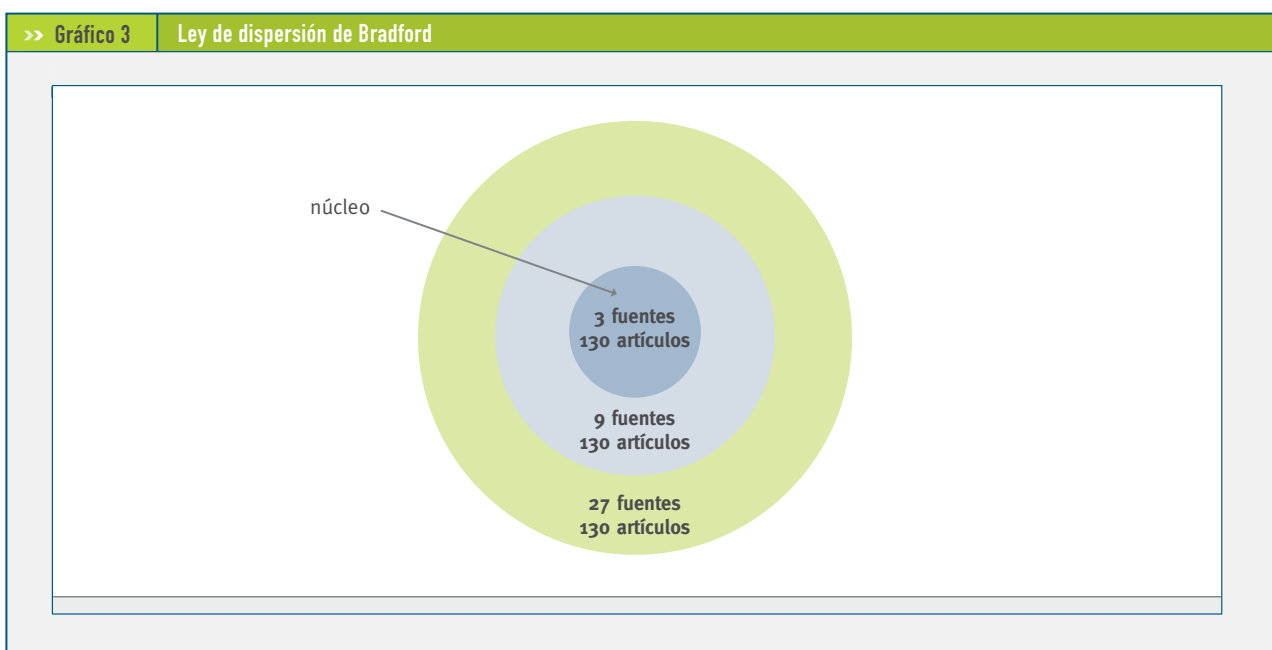


2.1.2 LEY DE DISTRIBUCIÓN DE BRADFORD

Pocos años después, Samuel C. Bradford, publicó su estudio acerca de la distribución de artículos en revistas científicas, publicado primero en la revista *Engineering* en 1934 y en su libro *Documentation* en 1948. El objetivo del estudio era determinar un método para identificar las revistas más productivas en un tema dado y tratar de poner orden en lo que denominó el "caos documental". Encontró que "si las revistas científicas se disponen en orden decreciente de productividad de

artículos en un tema determinado, se puede distinguir un núcleo de revistas más específicamente dedicado al tema y diversos grupos o zonas que contienen el mismo número de artículos que el núcleo 1: $n_1: n_2 \dots$ Esta afirmación se conoce como la "ley de la dispersión de Bradford". El Gráfico 3 ilustra esta dispersión.

Este concepto ha sido ampliamente utilizado para determinar colecciones núcleo en bibliotecas y para la cobertura de bases de datos. Más adelante, Eugene Garfield amplió este concepto, en el sentido que las publicaciones incorporadas en la zona 3 en muchos casos incluyen material de interés no sólo para la disciplina en estudio, sino también para otras disciplinas. Esto permite determinar un cierto número de publicaciones máximo que asegure no sólo la cobertura de diferentes disciplinas sino también áreas multidisciplinarias.



2.1.3 LEY DE ZIPF

A su vez, George K. Zipf formuló en 1949 una importante ley de la bibliometría y de la lingüística, que derivó de diversos estudios de frecuencia de palabras en un texto dado. De ahí derivó su "Principio del menor esfuerzo": "el principio del menor esfuerzo significa (...) que una persona tratará de solucionar su problema de tal manera de minimizar el trabajo total que debe gastar en solucionar tanto su problema inmediato como sus probables problemas futuros" (Zipf, 1949). Este concepto ha sido ampliamente utilizado en estudios lingüísticos (de hecho la primera aplicación que hizo Zipf fue el análisis de ocurrencias de términos en el *Ulises* de James Joyce) y tiene en la actualidad aplicación en minería de textos, especialmente en la generación de "clusters" de documentos similares.

2.1.4 DEREK DE SOLLA PRICE. CRECIMIENTO Y CONTEMPORANEIDAD DE LA CIENCIA

Como ya se ha mencionado, fue Derek de Solla Price quien sentó las bases de las técnicas modernas de evaluación y estudio de la comunicación científica, sus flujos y el crecimiento y envejecimiento del conocimiento. A partir de sus trabajos surgieron numerosos estudios y se desarrollaron modelos utilizados hasta el día de hoy. De Solla Price postuló que el crecimiento de la información científica es superior a otros fenómenos sociales, pero muy similar al de otros fenómenos obser-

vables en las ciencias naturales. El crecimiento de la literatura científica es exponencial, o sea que la información existente se duplica cada 10 a 15 años, dependiendo del área del conocimiento de que se trate. En este crecimiento, identificó cuatro etapas: la etapa 1, de los precursores, la etapa 2, del crecimiento exponencial, la etapa 3, del crecimiento lineal y la etapa 4, del colapso del campo científico. Este concepto del crecimiento exponencial de la literatura científica se complementa con el concepto de la contemporaneidad de la ciencia. Ello significa que el número de científicos vivos también aumenta de tal manera que el número de científicos en la actualidad constituye casi el total de todos los que han existido en el pasado más los actuales. Sin embargo, como un crecimiento de este tipo no puede mantenerse hasta el infinito, debe existir un límite de saturación.

En un trabajo publicado en 1965, De Solla Price estudió las redes de artículos científicos, basado en los análisis que se podían llevar a cabo con el Science Citation Index. De Solla Price, así como otros sociólogos de la ciencia, como Thomas Merton, Jonathan Cole y especialmente Francis Narin en su trabajo *Evaluative Bibliometrics*, entre muchos otros, comenzaron a utilizar la información producida por la recopilación de publicaciones y sus citas en forma sistemática, para sus estudios y análisis del crecimiento y comportamiento de los grupos y las disciplinas científicas.

2.1.5 EUGENE GARFIELD Y LOS ÍNDICES DE CITAS

En 1955, Eugene Garfield, en un artículo publicado en la revista *Science*, desarrolló el concepto del uso de las citas como un elemento importante en el análisis de la comunicación científica. Poco tiempo después se creó el Institute for Scientific Information (ISI) y comenzó a publicarse el Science Citation Index, que constituye una herramienta nueva para la recuperación de información. En poco tiempo, se vio su potencialidad para realizar estudios de diversa índole, especialmente en lo que se ha denominado el "factor de impacto", basado en la cantidad de veces en que un artículo es citado por otros y también para evaluar las revistas de acuerdo con el número de citas que han recibido en un año determinado. A partir de los trabajos de Garfield surgieron una serie de otros estudios, basados en la potencialidad del uso de las citaciones para llevar a cabo análisis cada vez más especializados. Más adelante se discutirán varias aplicaciones concretas.

También se debe a Garfield la ampliación de la ley de dispersión de Bradford para calcular el número de publicaciones necesarias para cubrir la totalidad de la información relevante, y no sólo la de una disciplina; ello se denominó "ley de la concentración de Garfield". En ella se señala que para cualquier campo de la ciencia los artículos se concentran en las mismas revistas multidisciplinarias de alto impacto o de "corriente principal", ya que las revistas "núcleo" de una disciplina son también parte de la "cola del cometa" de otra disciplina. Garfield calculó que con alrededor de 1.000 revistas de corriente principal estaría cubierto el 95% de la literatura citada.

2.2 FUENTES Y ANÁLISIS DE DATOS

Las publicaciones, a pesar de no ser el único producto del trabajo científico, son consideradas como la unidad básica para medir dicho producto, especialmente los artículos publicados en revistas de carácter internacional, con comité editorial compuesto por connotados especialistas en el tema.

Ya se ha comentado, pero vale la pena insistir, que no en todas las áreas del conocimiento los investigadores se comunican a través de artículos científicos, pero dado que este es el medio más común para medir y evaluar a la comunidad científica y académica, y producir indicadores comparables a nivel internacional, hay que reconocer que, en cierta medida, se presiona a los investigadores a utilizar este sistema de comunicación. En muchas áreas es tanto o más importante el libro (o la monografía), los informes de progreso, las presentaciones a congresos, etc., y sin embargo muchos estudios, especialmente a nivel nacional o institucional, no suelen considerar estas otras formas de comunicación al momen-

to de medir la actividad total.

Es importante llamar la atención sobre esta situación cuando se apliquen indicadores bibliométricos en la medición y evaluación de ciertas áreas del conocimiento, o de departamentos e institutos de una universidad o, asimismo, a investigadores individuales.

En todo caso, y a nivel internacional, son los artículos científicos los que se utilizan como unidad de medida al generar indicadores para medir la productividad en forma comparable.

Estos artículos son fuentes importantes de información, con las que se pueden llevar a cabo una serie de estudios y mediciones. Cada artículo contiene datos tales como: nombre del o de los autores, institución de pertenencia (filiación), direcciones institucionales, título de la revista (que no sólo aporta información acerca del tema, sino que también su importancia y prestigio), conceptos, a través de palabras claves, referencias (citas), etc. El conjunto de estos datos, debidamente organizados en bases de datos, constituyen las fuentes básicas de información para la construcción de indicadores y la realización de análisis y evaluaciones.

Uno de los principales objetivos de los estudios bibliométricos es generar indicadores consistentes y comparables. Para lo cual es necesario contar con bases de datos lo suficientemente amplias y con información internacional que registren la totalidad de los datos requeridos para llevar a cabo estudios bibliométricos, lo que no siempre ocurre con las bases de datos e índices especializados, cuyo único objetivo es la recuperación de información.

Para los fines de los análisis bibliométricos es importante que para cada artículo se consignen los nombres de todos los autores participantes en el trabajo y la filiación completa de cada uno de ellos. Es indispensable, además, trabajar con las citas recibidas, para poder conocer la influencia que el trabajo ha tenido sobre el trabajo de otros.

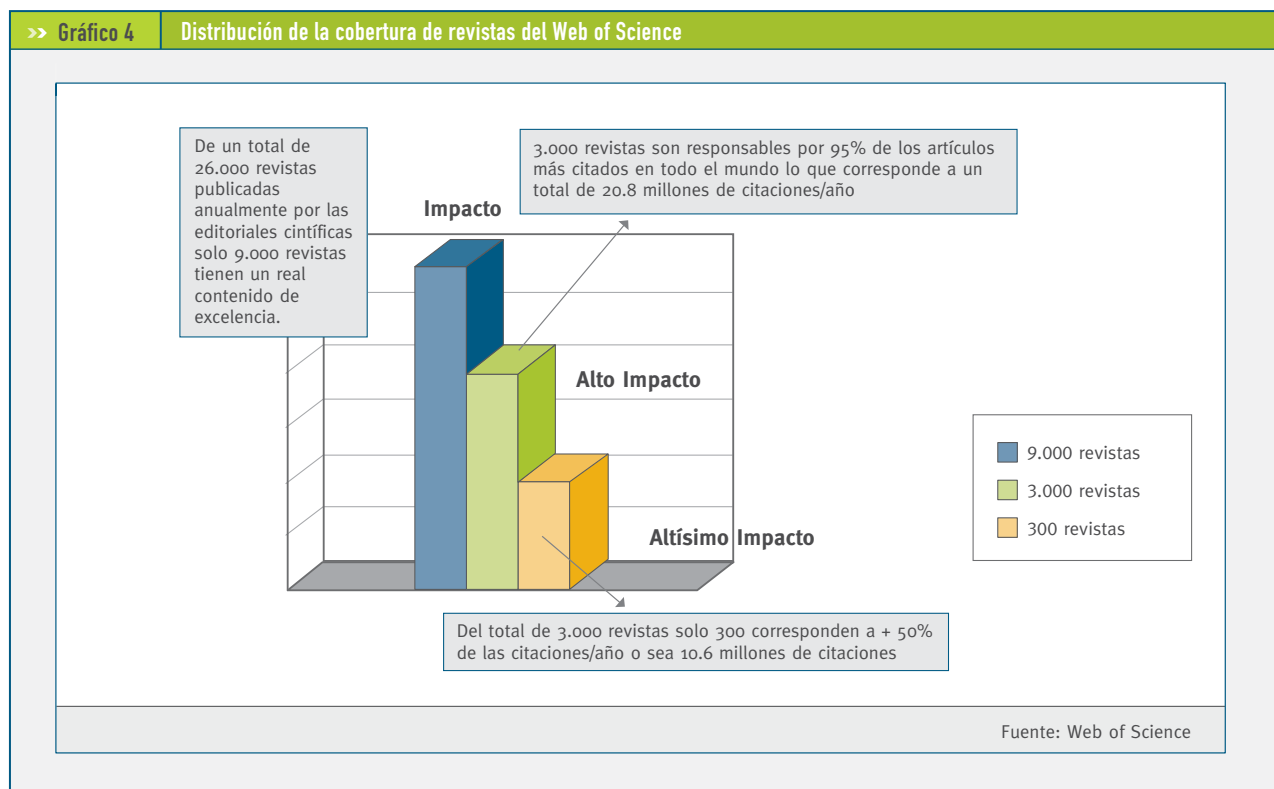
2.2.1 WEB OF SCIENCE

Durante muchos años, la única base de datos que cumplía con los requisitos indispensables era el Science Citation Index y los demás índices de citas (ciencias sociales y arte y humanidades) de ISI (actualmente Thomson Reuters). Sin embargo, desde hace unos pocos años, la empresa Elsevier ha creado la base de datos Scopus, que permite llevar a cabo los mismos estudios y análisis, pero todavía, por razones de comparación internacional, los datos de los CI (Citation Indexes) y del Web of Science siguen siendo los más utilizados.

Siempre se ha discutido la validez de estos índices para medir la producción de Latinoamérica, debido a la mayor cobertura que ellos brindan de las revistas en idioma inglés y de origen anglosajón, a pesar de que en estos últimos años ha habido una apertura mayor a las publicaciones en otros idiomas y procedentes de otras regiones.

La cobertura actual del Web of Science y los CI de Thomson Reuters es de 9.000 títulos de revistas, divididas en: Science Citation Index Expanded (6.100 títulos), Social Science Citation Index (1.790 títulos) y Arts & Humanities Citation Index (1.125 títulos). Según la información proporcionada, de estos 9.000 títulos, 300 revistas acumulan más del 50% de las citaciones/año, o sea 10,6 millones de citas (ver Gráfico 4). Los criterios de selección de las revistas se basan en la aplicación de la Ley de Bradford y su ampliación por Garfield. Además de su evaluación formal, se observan, durante un cierto

período de tiempo, las citas recibidas y se mide su factor de impacto. Cada año se sacan de la base de datos un cierto número de títulos y se incorporan nuevos. El Gráfico 4 muestra la distribución de la cobertura de revistas del Web of Science.



En forma complementaria, Thomson Reuters publica anualmente el *Journal Citation Reports* (JCR), en el que se indica para cada revista el factor de impacto (FI) en un año determinado. El FI se calcula sobre la base del número total de citas recibidas por los trabajos publicados en la revista en los últimos dos años, dividido por el número de trabajos publicados. Este factor de impacto, por lo tanto, varía de un año a otro (situación que debe ser recordada según el uso que se quiera hacer de dicho indicador); el factor de impacto es ampliamente utilizado no sólo para evaluar la revista en sí, sino también los trabajos que en ella se publican. La aceptación de un trabajo en una revista de alto impacto asegura el reconocimiento de la calidad del mismo, pero no necesariamente el impacto real del artículo propiamente tal.

2.2.2 SCOPUS

La base de datos Scopus, de Elsevier, cubre también la mayor parte de las áreas del conocimiento, especialmente agricultura, biología, química, geología, economía, negocios, ingeniería, salud, ciencias de la vida, matemáticas, física, psicología y ciencias sociales. Scopus proporciona información completa para cada artículo, lo que permite llevar a cabo estudios bibliométricos de alta calidad, y también incorpora enlaces a los artículos originales.

La base Scopus tiene una amplia cobertura de revistas, que actualmente supera los 16.000 títulos, un poco más de la mitad de las revistas académicas y científicas con comité editorial que se publican en el mundo; asimismo, incorpora un mayor registro de revistas latinoamericanas y procedentes de países no anglosajones. Incluye información con enlaces a textos

completos desde 1996 hasta la fecha.

Además de revistas, Scopus incluye monografías seriadas, actas de congresos, y publicaciones comerciales. A través del programa Scirus revisa repositorios locales y páginas web con información sustantiva.

El sistema de evaluación de las revistas se lleva a cabo a través de grupos de investigadores y bibliotecarios, pertenecientes a diversas disciplinas y provenientes de diferentes regiones, que sugieren la inclusión de los títulos más relevantes en cada área o región geográfica. Se ha desarrollado para la base una interfaz muy amistosa, que permite llevar a cabo diferentes análisis, especialmente el índice h, tan importante hoy para evaluar personas, instituciones o revistas.

Estas dos bases de datos son prácticamente las únicas que cubren todas las áreas del conocimiento, y cuentan con la totalidad de los metadatos requeridos para llevar a cabo análisis bibliométricos detallados.

2.2.3 PASCAL Y FRANCIS

Las bases de datos PASCAL (sigla de "Programme Appliqué à la Sélection et à la Compilation Automatique de la Littérature") cubre las ciencias exactas, naturales y aplicadas, mientras que la base FRANCIS (sigla de "Fichier de Recherches Automatisées sur les Nouveautés, la Communication et l'Information en Sciences Sociales et Humaines") abarca las ciencias sociales y las humanidades; ambas poseen datos desde 1984 hasta el presente. Son producidas por el Institut de l'Information Scientifique et Technique (INIST) del CNRS de Francia y continúan el Bulletin Signalétique del CNRS, publicado en papel por el mismo INIST desde los años sesenta.

Las bases abarcan la totalidad de las áreas del conocimiento y constituyen un importante índice con resúmenes. Sólo en los últimos años han incluido las filiaciones de los coautores, si bien no registran información sobre citas, lo que reduce bastante su utilidad como herramienta para estudios bibliométricos más avanzados. Las bases son especialmente importantes para conocer las publicaciones de origen europeo.

PASCAL y FRANCIS pueden ser consultadas para búsquedas por materias, en francés, inglés o español, basadas en un tesoro trilingüe. Actualmente entregan información de más de 4.000.000 de artículos, actas de conferencias, tesis de doctorado e informes científicos.

2.2.4 GOOGLE ACADÉMICO

En estos últimos años Google ha producido el Google Académico, desde el cual pueden hacerse búsquedas de trabajos que han sido publicados o se encuentran accesibles a través de Internet. Incluye bases de datos, colecciones de revistas electrónicas, repositorios institucionales y nacionales, bibliotecas y otros archivos de literatura científica o académica.

El resultado de búsqueda presenta un conjunto de trabajos académicos, con revisión por pares, ordenados por número de citas (bajo la opción "todos los artículos") o por fecha de publicación (bajo la opción "artículos recientes"). Puede incluir múltiples versiones de un mismo artículo, relacionados entre sí (por ejemplo, el documento previo a la impresión, una presentación a una conferencia, el artículo de una revista y el publicado en una antología). Al agrupar estos artículos se puede visualizar con mayor precisión el sistema de comunicación científica y las distintas versiones de una publicación efectuada como resultado de una misma investigación.

El resultado de búsqueda contiene información bibliográfica como: título, nombres de los autores y datos de la publicación, número de citas y la posibilidad de visualizarlas, además de enlaces a bases de datos y bibliotecas que incluyen el artículo completo o lo tienen en sus colecciones. También existe la posibilidad de realizar búsquedas por los principales co-autores, posibilitando así el análisis de las redes y la cooperación científica.

Dada su estructura, se recomienda a quien haya de utilizar Google Académico como fuente de datos tener cuidado de no contabilizar el mismo artículo más de una vez, ya que, como se ha dicho, el producto incluye todas las versiones de un mismo trabajo, sus posibles ubicaciones y las citas al mismo trabajo.

Esta es una herramienta de búsqueda cada vez mejor estructurada y con mayor información y posibilidades de búsqueda y análisis, lo que lo convierte en una fuente de datos de importancia, además de la ventaja, para todos los países, de ser de acceso abierto.

2.2.5 SciELO

SciELO (sigla de Scientific Electronic Library Online) es una hemeroteca científica creada en Brasil en 1997. De la red SciELO participan hoy catorce países de Iberoamérica a través de colecciones nacionales, seleccionadas y evaluadas con criterios de calidad comunes. Constituye una de las fuentes que empieza a tener información suficiente para llevar a cabo estudios y análisis de la producción regional.

Actualmente, SciELO cubre más de 550 revistas iberoamericanas, con una colección de más de 180.000 artículos y más de 3.500.000 de citas. Permite el acceso por diferentes criterios de búsqueda, ya sea en cada una de las colecciones de cada país participante, o bien a la colección completa de la hemeroteca través del Portal SciELO (<http://www.scielo.org/php/index.php?lang=es>). SciELO ha contribuido a dar visibilidad y accesibilidad a las revistas científicas latinoamericanas y a aumentar su impacto a nivel internacional. Por este motivo se ha constituido en un sistema válido para medir y evaluar la producción nacional de los países que la conforman.

Dede sus inicios, SciELO consideró la posibilidad de realizar estudios bibliométricos con la información recopilada y procesada. Para cada artículo, además de los metadatos necesarios para estudios bibliométricos, se detallan cada una de las citas y se llevan a cabo los enlaces necesarios para su utilización en estudios y análisis de impacto.

Al cubrir cada vez un mayor número de países y disciplinas, empieza a tener sentido la producción de indicadores complejos basados en la base de datos común. Actualmente, en algunos países, tales como Brasil, Chile y Argentina, se pueden llevar a cabo mediciones básicas complementarias a las llevadas a cabo con el Web of Science y Scopus.

2.2.6 BASES DE DATOS ESPECIALIZADAS

Existen, además de las ya nombradas, una gran cantidad de bases de datos especializadas, diseñadas originalmente para el análisis y la recuperación de información. Estas bases son útiles, a pesar de sus limitaciones, tales como no incluir la filiación de los coautores y no proporcionar datos acerca del uso de los artículos a través de las citas obtenidas. Sin embargo resulta interesante consultar tales bases para realizar estudios disciplinarios, en forma complementaria a las bases de datos más generales. Las más reconocidas y utilizadas son:

- Medline, en el campo de la biomedicina, considerado a su vez como la "corriente principal" en su campo,

- Chemical Abstracts, en la química,
- INSPEC, en la física y sus disciplinas afines,
- Compendex, en la ingeniería y
- CAB, en el campo de las ciencias agropecuarias.

2.2.7 BASES DE DATOS LATINOAMERICANAS

Entre las bases de datos regionales, todas ellas disponibles en Internet mediante el sistema de acceso abierto, se cuenta especialmente con las siguientes:

- CLASE: base de datos en línea de la Dirección General de Bibliotecas de la UNAM (México), que recoge más de 160.000 registros de documentos publicados en más de 1.200 revistas latinoamericanas especializadas en ciencias sociales y humanidades desde 1979. Se actualiza diariamente.
- PERIÓDICA: también producida por la Dirección General de Bibliotecas de la UNAM (México), recoge más de 160.000 registros bibliográficos de documentos publicados en cerca de 1.300 revistas latinoamericanas especializadas en ciencia y tecnología desde 1979 a la fecha. Se actualiza diariamente.
- LILACS (Literatura Latinoamericana y del Caribe en Ciencias de la Salud): recoge registros de la literatura científica y técnica en salud, producida por autores latinoamericanos y del Caribe y publicada en los países de la región, a partir de 1982. En los últimos años se complementa la información con la filiación de todos los autores y coautores, lo que facilita la preparación de muchos estudios en el área.

2.2.8 REDALYC Y LATINDEX

Cabe destacar, además, otros dos programas regionales, REDALYC (sigla de la Red de Revistas Científicas de América Latina y el Caribe, España y Portugal) y LATINDEX (sigla del Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal), que ofrecen diferentes tipos de información utilizable para estudios regionales.

REDALYC es un proyecto impulsado por la Universidad Autónoma de Estado de México (UAEM), con el objetivo de contribuir a la difusión de la actividad científica editorial que se produce en y sobre Iberoamérica, especialmente en el área de las ciencias sociales. Se trata de una hemeroteca de acceso abierto, que incluye los textos completos de más de 500 revistas en formato pdf y está disponible en Internet en <http://redalyc.uaemex.mx/> en la modalidad de acceso abierto. El programa se lanzó en el año 2002.

LATINDEX, por su parte, se inició en 1997 y es producto de la cooperación de una red de instituciones a nivel nacional para reunir y diseminar información bibliográfica sobre las publicaciones científicas seriadas producidas en la región. La información que recopila esta base se refiere a la revista propiamente dicha, y no cubre los artículos que en ella se publican. Hasta ahora existen tres tipos de productos: el directorio, que es una lista de revistas académicas y de investigación de la región; el catálogo, que ofrece una selección de dichas revistas de acuerdo con criterios de calidad editorial aprobados; y el Sistema, que actualmente incluye más de 3.000 títulos. La base ofrece también enlaces a revistas electrónicas que permiten acceder a textos completos de artículos.

2.2.9 BASES DE DATOS NACIONALES COMPLEMENTARIAS

Las bases de datos Web of Science y Scopus tienen un alto costo y por consiguiente no son accesibles a todos los países.

En algunos casos, esas bases han sido suscritas por un consorcio o una entidad nacional, pero hay muchos países que no cuentan con ninguna información para llevar a cabo estudios de producto. Resalta así la importancia de contar con bases de datos nacionales que contengan la información científica generada en el país (a la espera de la creación de repositorios nacionales o institucionales) a fin de facilitar, al menos, la medición de lo producido por las diferentes instituciones y en las diferentes áreas.

Al generar estas bases de datos es necesario definir, desde el inicio, la totalidad de los metadatos que deben incluirse y que se requieren para llevar a cabo estudios bibliométricos básicos:

- Nombres de los autores y coautores, normalizados para facilitar su identificación y evitar alcances de nombres.
- Filiación de los autores y coautores: instituciones a las que pertenecen, al nivel de detalle institucional al que interese llegar. Es indispensable la normalización de los nombres de las instituciones, para recuperar datos al momento de iniciar estudios.
- Referencia bibliográfica completa: título de la revista (también normalizado), año de publicación, volumen, número páginas, o editorial en el caso de monografías, etc.
- Palabras claves: para facilitar la búsqueda y recuperación de información disciplinaria, generalmente incluidos en la publicación misma.
- Clasificación de disciplinas: de acuerdo a alguna convención internacional.

Si no existe una base de datos nacional, es muy probable que existan bases de datos institucionales que puedan consolidarse para formar una sola fuente de información acerca de la producción científica nacional. En algunos casos pueden ser de mucho interés otros tipos de fuentes, como por ejemplo las tesis de postgrado, que es importante incluir en una base de datos nacional para efectuar estudios locales. La mayoría de los países de la región cuentan en la actualidad con bases de datos de tesis en formato de acceso abierto, y en muchas instituciones se crean repositorios de material académico de acceso abierto en Internet.

Cabe destacar que al recurrir a bases de datos internacionales para llevar a cabo estudios bibliométricos, la información que se utiliza ha sido previamente evaluada, lo que permite la comparación de resultados entre países. Sin embargo, si se trata de empezar a conocer la producción del país en todas sus áreas puede ser útil el uso de las bases de datos nacionales, siempre y cuando en su preparación exista un mínimo control de calidad de la información que en ellas se incorpora.

2.2.10 ACCESO A DATOS COMPARATIVOS GENERALES. SCIMAGO Y RICYT

Para indicadores generales comparativos a nivel internacional o regional, en este momento, se dispone de información por país y disciplina, en forma abierta libremente accesible a través de Internet en diferentes sitios, tales como el de SCImago y el de RICYT.

SCImago es un grupo de investigación con sede principal en la Universidad de Granada (España). El grupo ha trabajado durante varios años construyendo el Atlas de la Ciencia, del que se ha publicado uno correspondiente a España. En Internet se hallan disponibles los atlas de Argentina, Brasil, Chile, Colombia, Cuba, España, México, Portugal, Venezuela y Perú. El Atlas de la Ciencia está basado en información de las bases de datos del Web of Knowledge de Thomson Reuters, y está disponible en <http://www.atlasofscience.net/>

El grupo también mantiene el portal SCImago Journal and Country Rank (SJR), basado en información de la base de datos Scopus. En dicho portal se analiza y ordena la información bibliométrica por país y por revista, se hacen análisis y comparaciones de la producción por países y revistas y se ofrecen gráficos para visualizar las disciplinas y áreas del conocimiento con mayor relevancia en cada uno de los países. Está disponible en <http://www.scimagojr.com/>. En colaboración con Universia, SCImago produce el Ranking Iberoamericano de Instituciones de Investigación (conocido como RI3), disponible en la dirección <http://investigacion.universia.net/isi/isi.html>.

La RICYT (sigla de la Red Iberoamericana e Interamericana de Indicadores de Ciencia y Tecnología), por su parte, publica anualmente indicadores regionales, disponibles en el sitio web <http://www.ricyt.org>. La publicación incluye un capítulo de indicadores de producto, en el que se dedica una sección a los indicadores de patentes y los bibliométricos. Sobre estos últimos se entrega información acerca de número de publicaciones por país, utilizando bases de datos tales como Science Citation Index Search, PASCAL, INSPEC, COMPENDEX, Chemical Abstracts, Medline, CAB, ICYT, la base de datos del Consejo Superior de Investigaciones Científicas de España, y bases de datos regionales (CLASE, PERIODICA y LILACS).

Esta información se complementa con tablas que relacionan el número de publicaciones por habitantes, el PIB, el gasto nacional en I+D y por cada 100 investigadores. La información se encuentra abierta en Internet en la dirección <http://www.ricyt.org/interior/interior.asp?Nivel1=1&Nivel2=2&Idioma=>

2.3 DEFINICIONES Y METODOLOGÍA

Como se ha dicho anteriormente, el producto de la actividad científica se puede medir y por lo general cuantificar. Pero también se requiere la medición de la calidad de esta producción científica. Para estos efectos es necesario generar indicadores de producción, los cuales pueden ser de dos tipos: indicadores de actividad (cuantitativos) e indicadores de calidad (indicadores cualitativos.)

2.3.1 METODOLOGÍA GENERAL

Al iniciar cualquier trabajo de medición o evaluación del trabajo científico utilizando métodos bibliométricos es importante tener en cuenta diversas consideraciones. La primera de ellas es que si bien las técnicas usadas para medir productividad (estudios cuantitativos) pueden ser bastante seguras, en el momento de evaluar el trabajo científico (estudios cualitativos) la bibliometría debe ser solamente una herramienta complementaria a otras formas de evaluación, tales como la evaluación por pares. A pesar de que los estudios bibliométricos dan una visión bastante concreta de la realidad, no deben tomarse en forma aislada como una verdad absoluta.

La segunda consideración es que es importante no perder de vista que la bibliometría es muy eficiente para medir y evaluar ciencia básica, pero es mucho menos precisa al evaluar las ciencias aplicadas o las ingenierías, y menos aún las ciencias sociales o las humanidades.

Otro aspecto que influye en los resultados es la amplitud de los datos utilizados y el nivel de análisis que se va a llevar a cabo. Cuanto más general sea el estudio (nivel nacional o institucional) mejores serán los resultados. A nivel individual, la información y los análisis serán mucho menos eficientes.

Al utilizar la mayoría de las bases de datos generales o especializadas para análisis cuantitativos o cualitativos es necesari-

rio revisar previamente cuán exacta es la información que se extraerá de ellas. La mayoría de las bases de datos se construyen con los datos proporcionados por el autor de los trabajos y la revista en la que se publican. Esto significa que puede haber muchas formas de identificar instituciones y personas. Si bien muchas de las bases de datos han hecho esfuerzos para normalizar lo más posible el nombre de las instituciones, al menos al primer nivel, todavía falta mucho camino por recorrer.

Es imprescindible, una vez identificado el conjunto de datos sobre los cuales elaborar los estudios, llevar a cabo un trabajo bastante detallado de normalización manual de los datos. Esto es válido tanto en el caso de nombres de las instituciones, en sus diferentes niveles, como en el caso de autores. Generalmente los nombres de los autores aparecen en dichas bases tal como firmó el autor, o simplemente con su apellido y las iniciales de su nombre. Hay autores que han firmado de diferentes maneras a lo largo de su carrera y otros que tienen apellidos muy comunes y producen una gran ambigüedad al momento de hacerse una búsqueda por su nombre. Este es un factor que debe tomarse en consideración al utilizar cualquier base de datos existente. También vale la pena, al hacer búsquedas para evaluación de investigadores individuales, una vez reconocidas todas las formas en que puede aparecer el nombre, asegurarse de que el nombre fue ingresado correctamente en la base de datos, sin sufrir fallas de digitación.

Es importante insistir en la necesidad de comparar siempre entre conjuntos similares; como dice la frase, "las peras con las peras y las manzanas con las manzanas". Esto es especialmente válido al comparar la producción en diferentes áreas del conocimiento. Ya se ha dicho que cada disciplina tiene sus formas propias de comunicar resultados y de citar trabajos previos. No es lo mismo publicar y citar en matemáticas que hacerlo en biología molecular. Por lo tanto no es válido hacer un ordenamiento de investigadores de acuerdo con su productividad sin separarlos previamente por áreas disciplinarias.

Tampoco es válido comparar sin mayor análisis la producción y las citas de un investigador veterano, con muchos años de experiencia, con un estudiante de postgrado, aunque sea en la misma disciplina. Aunque esto sea bastante de sentido común, es importante tener siempre en cuenta la necesidad de comparar entre conjuntos de similares características.

Un tema que se ha discutido mucho a nivel internacional es cómo asignar a cada autor la proporción del trabajo en el que ha participado como coautor. Salvo en el caso de trabajos en ciertas áreas del conocimiento en que se sobrepasan los treinta coautores por artículo, es razonable asignar a cada autor la totalidad del trabajo, ya que es muy difícil distinguir qué parte del mismo se le debe a cada uno. Se supone que todos los coautores de un trabajo científico deberían poder responder del contenido del mismo. Si bien se han detectado algunos fraudes (por ejemplo, los casos de personas que se colocan siempre como autores de trabajos de sus equipos en los que nunca han participado), también es posible detectarlo, dado que su producción media puede superar en mucho la media de la disciplina. En la actualidad, cuando lo que se busca es formar redes de trabajo y colaboración entre personas de diferentes disciplinas para solucionar un problema dado, resulta contradictorio castigar a estos autores no dándoles el crédito por la autoría de un trabajo científico en el que participaron activamente. En todo caso, es un tema a decidir al iniciar el análisis de la producción, especialmente en el caso de investigadores individuales. Esto mismo es válido en el caso de estudios por institución, como se indica más adelante.

Se puede trabajar con dos tipos de mediciones simples:

1. Mediciones absolutas. Conteo de, entre otros:

- Número de artículos en una base de datos dada
- Número de artículos por año

- Número total de citas
- Combinaciones entre sí o con otros indicadores complementarios

2. Mediciones relativas, entre otras:

- Citas por artículo comparada con citas por artículo en una disciplina dada
- Citas vs. citas esperadas por artículo
- Porcentaje de artículos citados vs. artículos no citados, comparado con el promedio del área
- Rango dentro de un área disciplinaria, entre grupos de pares por número de artículos, citas o citas por artículo

En el Web of Science, la disciplina que se asigna a un artículo corresponde a la de la revista en que el artículo fue publicado. Se ha discutido mucho la validez de esta clasificación, pero dado el número de artículos incorporados diariamente a la base parece difícil otra aproximación. En todo caso sirve para normalizar las mediciones por disciplina, y a pesar de las críticas es usado por la mayoría de los organismos que producen indicadores de producción.

Es recomendable, especialmente al utilizar datos bibliométricos para evaluar personas o grupos de investigación, llevar a cabo diversas mediciones complementarias, para asegurarse del real impacto del investigador y sus trabajos. Hay que tener siempre en cuenta que las mediciones bibliométricas tienen muchas sutilezas que las hacen no siempre exactas al momento de evaluar individuos. Más vale asegurarse de la veracidad y aplicabilidad de un dato, llevando a cabo diversas mediciones, como por ejemplo el número total de citas y el número de citas esperadas, comparado con la disciplina y con el tipo de trabajo (artículo, revisión etc.).

En todo caso, después de cualquier análisis es importante preguntarse siempre si los resultados son razonables. Es bueno revisar los estudios cada vez que un resultado aparece muy fuera de lo esperado. Esto es siempre válido, pero lo es especialmente si se trata de evaluar personas o grupos de investigación.

2.3.2 INDICADORES DE ACTIVIDAD (INDICADORES CUANTITATIVOS)

En general estos indicadores son resultado del conteo de publicaciones. Se trata de números absolutos que requieren ninguna o poca elaboración estadística. Es posible agruparlos en:

Número de publicaciones académicas: da cuenta del número de publicaciones a nivel nacional, institucional o en una disciplina determinada. Sirve para fines comparativos.

Previamente a la generación de este indicador es importante definir qué se entiende por publicación científica, ya que las bases de datos descritas en la sección anterior registran una variedad de fuentes y de tipos de publicaciones. Sin embargo, sólo son aceptadas como contribuciones científicas nuevas, para estos efectos, los artículos (que incluyen también las presentaciones a congresos en texto completo), las notas y las revisiones. En algunas áreas, como por ejemplo en la física, también se incluyen las cartas.

Tanto en el Web of Science como en SCOPUS es posible discriminar el tipo de publicación a incluir en el momento de efec-

tuar el conteo. El número de publicaciones por sí mismo puede servir para comparar la producción científica entre países o instituciones, pero es conveniente relacionar también dicho indicador con otros indicadores generales, tales como el número de investigadores (como medida de productividad) o el gasto en investigación (como medida de eficiencia). Esto permite comparar en mejor forma países o instituciones con mayor o menor capacidad científica.

Número de publicaciones por institución: se ha discutido mucho acerca de cómo contabilizar las publicaciones cuando participan en ellas más de una institución, ya sea nacional o internacional. En general, las publicaciones se contabilizan por cada una de las instituciones de cada autor o coautor participante. Si participan varios autores de una misma institución se consideran una sola vez. Si participan autores de más de una institución se cuenta una vez en cada una de las instituciones. En el momento de hacer el conteo a nivel nacional se cuenta como una sola publicación. Esto significa que la suma de las contribuciones por institución será mayor que la suma de las contribuciones a nivel nacional.

Este sistema es más usado que el de asignar un porcentaje de publicación a cada institución participante, de acuerdo con el número de instituciones de los coautores. Este último sistema es en general poco usado ya que es sumamente difícil determinar cuál es la contribución real de cada una de las instituciones participantes en el momento de asignar a cada una de ellas un valor específico, sobre todo al manejar una gran cantidad de publicaciones. Un caso especial lo constituyen algunas áreas en las que participan un número exagerado de coautores e instituciones. Tal es el caso de ciertos ensayos clínicos de nuevos medicamentos, donde pueden participar más de cincuenta instituciones de diferentes países, o en alguna de las áreas de la física de alta energía, entre otras. En este caso, la recomendación es no tomar en cuenta estos trabajos, ya que difícilmente uno de los participantes puede declararse responsable intelectual de todo o parte del trabajo.

En todo caso, asignar un porcentaje de cada trabajo a las instituciones participantes aparece como un castigo a aquellas que tienden a colaborar más con otras. Esto entraría en contradicción con las políticas nacionales para el fomento de colaboraciones interinstitucionales a nivel nacional y de esfuerzos conjuntos de investigación. En cualquier caso, es una decisión a tomar en el momento de iniciar el conteo de las publicaciones de una institución.

Una medición interesante es calcular qué proporción de la información producida en el país es producida por cada una de las instituciones nacionales.

Publicaciones por área geográfica: lo mismo sucede en el momento de contar las publicaciones a nivel nacional o por regiones administrativas (estados, provincias, regiones, etc.) y la colaboración internacional.

Se ha discutido mucho si un trabajo con coautores de diferentes países debe contarse como una contribución en cada uno, o debe calcularse en forma proporcional. Para fines comparativos entre países es preferible asignar una publicación a cada país en el que figure al menos un coautor de ese país.

Lo mismo sucede en el momento de contar las publicaciones por regiones geográficas dentro de un país. Por ejemplo, si en un trabajo participan dos autores de la Región de Valparaíso y uno de Santiago de Chile, se asigna una publicación a Valparaíso y una a Santiago, y obviamente se cuenta una sola cuando se hace el recuento a nivel nacional.

En la actualidad es importante también generar indicadores de colaboración internacional, tanto en lo que hace al número

de colaboraciones por país, como al porcentaje de colaboración internacional sobre el total de la producción científica en un período de tiempo determinado. Este es un índice importante para medir la magnitud de la colaboración internacional y detectar redes de investigación y sus vinculaciones e influencias.

Publicaciones por disciplina: la producción media de artículos por persona o por grupos de investigación varía considerablemente de una disciplina a otra. Este es un hecho que debe tomarse en consideración cada vez que se llevan a cabo estudios o análisis por disciplina.

Los documentos en el Web of Science se clasifican por disciplina de acuerdo con una grilla que comprende un número de disciplinas generales subdivididas en especialidades, cada una de ellas mutuamente exclusiva. Esto permite clasificar las revistas en una sola disciplina. Los contenidos de toda la revista se asignan a esa disciplina dada. Esta clasificación es usada por la mayoría de los organismos nacionales de producción de indicadores. Sin embargo, si bien esto facilita la producción de indicadores por disciplinas y su comparación a nivel internacional, no siempre es suficiente para realizar análisis más finos a nivel local y, sobre todo, produce distorsiones al medir la producción individual.

Un investigador puede considerarse a sí mismo como perteneciente a una disciplina pero las revistas en las que publica pueden pertenecer a otras áreas. Esto puede producir problemas para comparar el número de artículos de una persona respecto de una disciplina dada. Este es un elemento a tomar en consideración según el tipo de análisis que se quiera llevar a cabo.

Publicaciones por sector económico: puede utilizarse la clasificación de sectores económicos definidos por la OCDE en el Manual Frascati, que comprende cuatro sectores: empresas (incluidas las públicas y privadas), enseñanza superior, administración (incluye a todos los organismos de la administración pública salvo empresas públicas y educación superior), e instituciones privadas sin fines de lucro. Sin embargo, en muchos países se suelen definir para fines de análisis de producto algunos sectores en forma independiente, como por ejemplo los hospitales, el sector defensa, etc.

La información necesaria para este tipo de mediciones, se obtiene del campo institucional y su asignación a un solo sector económico.

En resumen: dependiendo del tipo de estudio o análisis pueden llevarse a cabo análisis bastante simples de producción medida de acuerdo con diferentes criterios, y compararlos con diferentes parámetros a nivel nacional o con estudios similares a nivel internacional. Es necesario informar en cada ocasión qué base de datos se ha utilizado, los tipos de publicaciones incorporadas, como se calcularon los trabajos por autor o por institución, los años en que se llevó a cabo la medición y la clasificación de disciplinas utilizada.

2.3.3 INDICADORES DE CALIDAD (INDICADORES CUALITATIVOS)

Número de citas: las citas son una medida de la influencia de un artículo o de su autor en la comunidad científica. Se trata de un parámetro socio-epistemológico complejo, que probablemente implica un factor de calidad, pero este factor no es ni equivalente ni está inequívocamente correlacionado con la calidad científica (Seglen, 1992).

Se ha discutido mucho acerca de la validez de las citaciones para evaluar un trabajo científico. Las razones por las que se cita un autor o un trabajo son muy variadas; pueden ser valederas o simplemente responder a otras motivaciones: presti-

gio del autor o de la institución a la que pertenece, capacidad de la revista de atraer a ciertos autores, redes de colaboración e influencia, etc. Sin embargo, siguen siendo una unidad válida para llevar a cabo una serie de estudios y análisis, siempre que se tenga claro que las citas no son, necesariamente, un indicador de calidad.

Algunos problemas detectados incluyen, entre otros, las citas negativas, o sea que un autor o un trabajo pueden ser citados por que se trata de un problema polémico o porque contiene errores flagrantes. Asimismo, el factor idioma influye en el momento de citar un trabajo; en muchas disciplinas, por caso, es necesario escribir en inglés, la "lengua franca" de la ciencia, para ser leído y citado. Otro problema es el de los "círculos de amigos" que citan en forma cruzada sus trabajos. Las autocitas o referencias de un autor a sus propios trabajos también son consideradas como uno de los aspectos a tomar en cuenta en el momento de utilizar las citas como medio de evaluación. Es natural y legítimo que un autor cite trabajos suyos previos que fueron considerados para llevar a cabo el actual trabajo. Sin embargo, las autocitas pueden ser la forma en que un autor apunte a aumentar su impacto. En general, los evaluadores que revisan los artículos antes de su publicación toman en cuenta este tipo de citas y se produce así un cierto control. Como en otros casos, el número de autocitas depende también de las disciplinas en las que trabaja el autor: en algunas se produce un mayor número de autocitas que en otras. Es importante, al analizar personas, considerar cada caso en forma individual, en especial en los que se salen de la media. Si un autor se autocita en forma exagerada podría tratarse de que es el único o uno de los pocos que trabaja en un tema muy especializado. En todo caso, es posible separar las autocitas del total de citas recibidas y entregar ambos datos por separado.

El factor tiempo debe también ser considerado para medir o evaluar individuos. Muchas veces es necesario utilizar espacios de tiempo diferentes para contar el número de artículos y el número de citas de una persona. El tiempo necesario para que un trabajo sea citado varía mucho de una disciplina a otra, pero también un trabajo más antiguo tiene muchas más posibilidades de haber sido citado alguna vez.

Se dice que más del 50% de los artículos incluidos en el Web of Science no son citados dentro de los cinco años posteriores a su publicación, pero esto no significa que no vayan a ser citados nunca. De hecho se ha estudiado que sólo un 10% de los artículos no son citados después de diez años de su publicación.

El número total de citas de un artículo se complementa a menudo con el dato del número esperado de citas, calculado por la cantidad de citas promedio que obtuvieron los artículos publicados en esa misma revista el mismo año de publicación. Es una medida relativa que permite poner cada artículo en un cierto contexto.

Índice-h: este índice fue sugerido por el físico Jorge E. Hirsch, que lo utilizó como una manera de determinar la calidad relativa de los físicos teóricos. Sirve para cuantificar la productividad científica real y el impacto científico aparente de un investigador. El índice se basa en el conjunto de los trabajos más citados de un investigador y el número de citas que han recibido. El índice también puede aplicarse a grupos de investigación de una misma disciplina. Para calcularlo, Hirsch dice: "Un científico tiene índice h si el h de sus N_p trabajos recibe al menos h citas cada uno, y los otros $(N_p - h)$ trabajos tienen como máximo h citas cada uno".

En otras palabras, un científico tiene índice h si ha publicado h trabajos con al menos h citas cada uno. Por ejemplo, un investigador tendrá un valor h de 75, si el trabajo número 76 de su lista de publicaciones, ordenadas por número de citas, ha sido citado 75 veces o menos, pero cuyo trabajo número 75 ha sido citado 75 o más veces. Ese investigador tendrá 75 trabajos con al menos 75 citas en cada uno de ellos.

Así, el índice h es el balance entre el número de publicaciones y las citas que ha recibido. El índice fue diseñado para medir eficazmente la calidad del investigador, diferenciando a aquellos investigadores con gran influencia en el mundo científico de aquellos que simplemente publican muchos trabajos. Tanto el Web of Science como Scopus calculan actualmente el índice h. El índice h da una idea mucho más exacta de la calidad de los trabajos de un investigador que el mero recuento de citas o del número de publicaciones.

Índice de inmediatez: es una medida de la actualidad e impacto de un trabajo publicado. Es calculado, cada año, por Thomson Reuters en el Journal Citation Reports, sobre la base del número de veces que un artículo publicado en un año X es citado durante el mismo año, en revistas indizadas, dividido por el número de artículos, revisiones, actas y notas publicados durante ese año. Este índice varía mucho de una disciplina a otra y debe tomarse con mucho cuidado al momento de ser utilizado para medir a individuos. Puede medir no sólo la calidad e impacto de la investigación, sino también las reacciones negativas.

Factor de impacto: mide las citaciones recibidas por las revistas en ciencia y ciencias sociales. Se utiliza como una forma de caracterizar la importancia de una revista en su área, y fue concebido como una manera de evaluar las revistas a ser incluidas en las bases de datos de los Citation Index, hoy Web of Science. Es, por lo tanto, una medida para evaluar la revista y no los artículos que ella incluye. Su fórmula es la siguiente:

$$\text{Factor de impacto} = \frac{\text{Números de artículos publicados por una revista en los últimos 2 años}}{\text{Números de citas de artículos en el mismo período}}$$

El factor de impacto ha sido muy mal utilizado para evaluar artículos específicos e incluso autores. Debe tenerse especial cuidado al aplicarlo, ya que es una medida ambigua, que varía año a año y que fue concebida para fines muy diferentes a aquellos para los que se la utiliza en la actualidad. Cada año se publica el *Journal Citation Reports* con la información completa de las revistas incluidas en el Web of Science.

SCImago Journal Rank (SJR): en estos últimos años el grupo SCImago ha desarrollado este índice, que se encuentra disponible en su portal Journal and Country Rank (<http://www.scimagojr.com/>). Este nuevo índice ordena las revistas de acuerdo con su grado de influencia. Su cálculo se basa en la transferencia del prestigio de una revista a otra, a través de las referencias que una revista hace al resto de las revistas y a sí misma. La información acerca de cómo se calcula el SJR se encuentra detallada en el mencionado portal. El índice se calcula sobre la base de la información contenida en la base de datos Scopus de Elsevier y se encuentra libremente disponible en Internet.

El índice permite llevar a cabo búsquedas por área temática, categoría temática y país. Para cada revista entrega información acerca de variables tales como SJR, índice h de la revista, número total de documentos en períodos de un año o de tres años, número total de referencias, total de referencias en los últimos tres años, número de citas por documento y número de referencias por documento. Para cada revista, además, se puede solicitar su análisis con información más detallada, incluyendo el porcentaje de colaboración internacional de los artículos publicados, complementado con los gráficos correspondientes. También pueden llevarse a cabo comparaciones entre revistas.

Finalmente, en el Cuadro 1 se presenta un resumen de los pasos necesarios para iniciar la preparación de indicadores bibliométricos.

>> Cuadro 1	Pasos previos a la preparación de indicadores bibliométricos
1. Seleccione la base de datos	
Seleccione la base de datos según el tipo de indicadores a producir. De ser posible, descargue los documentos con información pertinente a una base de datos local (por ejemplo, todo lo publicado o patentado por autores residentes en el país).	
2. Limpie la base de datos	
Es conveniente tomar en cuenta sólo las publicaciones consideradas contribuciones originales: artículos, presentaciones a congresos publicados en extenso, artículos de revisión, notas. Elimine lo demás.	
3. Normalice los nombres de las instituciones	
Las bases de datos publican los nombres de las instituciones de muy diversas maneras. Decida hasta qué nivel institucional requiere normalizar.	
4. Normalice, en lo posible, nombres de autores	
Si va a trabajar en la evaluación de investigadores individuales, debe normalizar sus nombres previamente. Las bases de datos, generalmente, sólo indican el apellido del autor y las iniciales de su nombre. Puede haber diferentes formas en que el autor firmó y alcances de nombres. Compruebe si se trata del autor que busca, comparando la institución y la disciplina en la que trabaja.	
5. Decida la clasificación disciplinaria que va a utilizar	
Si va a utilizar una clasificación de disciplinas similar a la aplicada en la producción de otros indicadores, haga una tabla de equivalencias entre la clasificación propuesta por la base de datos y la clasificación a usar.	
6. Decida cómo va a considerar los trabajos en colaboración	
Tanto en el caso de medición de trabajos por institución o como en el de trabajos por autor, decida previamente si los va a contar como un trabajo para cada una de las instituciones o autores participantes o va a asignar un porcentaje del trabajo a cada una de ellas. En el caso de trabajos con más de treinta autores, deberá decidir si los incluye o los elimina.	
7. Decida cómo va a medir el impacto	
Número total de citas, número total de citas vs. citas esperadas, índice h, dependiendo de si mide el impacto institucional, el impacto de una disciplina, o el impacto de los trabajos de una persona.	
8. Decida cómo va a tratar las autocitas	
Defina si se van a incluir como válidas, si se las va a eliminar de la medición o si se determinará caso a caso, según la disciplina.	
9. Prevea el tipo de indicadores compuestos que va a producir	
Defina el tipo de cruces de variables que va a tener que llevar a cabo. Asegúrese que la información que necesita está disponible (por ejemplo, jornadas completas equivalentes dedicadas a I+D, número de habitantes, sectores económicos, etc.).	
10.	
Si va a utilizar la información bibliométrica para evaluación de personas o grupos de investigación tenga mucho cuidado. Cada disciplina es diferente. Estos indicadores son sólo indicativos. Deben complementarse con evaluaciones por pares u otros métodos.	

03. Indicadores de patentes

Una patente es un instrumento legal que le entrega un monopolio temporal a un inventor a cambio de la publicación detallada del invento. Por lo tanto le permite al inventor protegerse y beneficiarse de su invento y a la sociedad ganar a través de la amplia diseminación del conocimiento producido al crear el invento. (Nesta y Patel (2004).

El término "patente" deriva del latín "patens" que significaba "estar abierto" (abrir algo para su inspección pública). Posteriormente fue utilizado en Inglaterra en el siglo XVII con el nombre de "Letter Patent", que eran decretos reales que le otorgaban derechos exclusivos a determinadas personas. La primera patente dada en Europa de las que tenemos noticia, data de 1421 y fue otorgada en Florencia.

Por lo tanto, de acuerdo con la definición original de la palabra, una de las finalidades de la legislación sobre patentes es la de inducir al inventor a hacer público su invento para el avance del conocimiento y la sociedad, a cambio de detentar el derecho de exclusividad durante un período limitado de tiempo. El uso temporal de lo inventado es un derecho que otorga el Estado; los procedimientos para obtener este derecho, los requisitos para patentar y el período por el cual se le otorgan esos derechos exclusivos varían de país en país, de acuerdo con la legislación nacional y los acuerdos internacionales vigentes suscritos por el país e incorporados a su legislación. Las patentes se otorgan a productos nuevos susceptibles de ser explotados industrialmente por el solicitante (el inventor o su representante).

Una patente no concede el derecho a usar el invento, sino más bien a excluir a otros de fabricar, usar, aplicar, vender, ofrecer a la venta o importar el producto patentado por un período determinado, el cual generalmente es de veinte años.

La información contenida en una patente es de acceso público, de modo que cualquier persona puede acceder a ella, pero sólo el titular la puede utilizar con fines comerciales. El concepto en el que se basa el sistema consiste que al proteger el invento de ser copiado, con lo cual el Estado apunta a fomentar la invención. La creación está protegida durante un plazo de tiempo determinado, permitiendo al inventor ser el único que pueda venderlo o explotarlo. De esta forma, su beneficio es más rentable y puede recuperar los recursos invertidos en la investigación. Las patentes son un vehículo de difusión efectiva y rápida de las innovaciones y nuevas tecnologías, y facilitan el acceso al nuevo conocimiento generado.

Cada país tiene su propia legislación de propiedad industrial, que fija la forma y las condiciones para otorgar una patente y el alcance de la protección concedida; ello significa que el proceso de otorgamiento de patentes y de derechos sobre ellas es de carácter nacional. En algunos países se permite llevar a cabo investigación sobre la información contenida en las patentes, siempre y cuando los resultados no sean comercializados; en otros, en cambio, no se permite el uso de esa información como fuente de nuevas investigaciones o creación de nuevo conocimiento.

Los países han creado oficinas de patentes, como responsables de la tramitación de solicitudes y de su otorgamiento. En el caso de existir demandas acerca de la propiedad de una patente, éstas son resueltas por los tribunales nacionales respectivos. Patentar un producto en un país no lo protege necesariamente en los demás países. Por consiguiente, es importante patentar los productos de interés tecnológico y comercial en los países industriales donde pueda haber una mayor posibilidad de aplicación, comercialización y uso.

Actualmente existe la tendencia para armonizar las leyes nacionales de propiedad intelectual, a través de tratados internacionales, como el Acuerdo sobre los Aspectos de los Derechos de Propiedad Intelectual relacionados con el Comercio (ADPIC, conocido en inglés como TRIPS) de la Organización Mundial del Comercio (OMC). En él se establecen una serie de principios básicos sobre la propiedad industrial y la propiedad intelectual, dirigidos a armonizar y concordar la legislación entre los países firmantes en relación al comercio mundial.

El Tratado de Cooperación en materia de patentes, conocido como PCT, que ha sido suscrito por más de 130 países, crea un procedimiento único para solicitar patentes que estén protegidas en todos los países miembros. Este procedimiento permite iniciar una solicitud con carácter internacional, pasar el proceso de examen e inscribir la patente, una vez otorgada por el país de origen, en los demás países en los que se desee proteger el producto.

3.1 PROCEDIMIENTOS DE SOLICITUDES DE PATENTES

Si bien todas las legislaciones nacionales varían, existen procedimientos comunes que es importante conocer en el momento de trabajar con patentes como fuente de información para la producción de indicadores.

En la terminología legal, "inventor" significa la persona que produce una invención, mientras que "beneficiario" define al propietario individual o institucional de la invención.

La solicitud de patente es un formulario en el que debe entregarse información sobre el beneficiario y el inventor o inventores, junto con una descripción técnica detallada en la que se estipule en qué medida y en qué áreas tecnológicas la invención representa una innovación. Para esto se puede hacer referencia a otras patentes y/o a la literatura científica que avalen la novedad de la invención. Si fuera necesario, la solicitud aportará también dibujos, gráficos, diagramas u otros anexos. La información debe ser lo suficientemente detallada como para permitir evaluar la originalidad y novedad de la invención.

Para obtener una patente, una invención debe ser ante todo nueva, y no haber sido divulgada en ninguna forma. Debe ser también "inventiva o creativa", o sea, sobrepasar las destrezas y conocimientos de un profesional especialista en el problema.

Una vez completada la solicitud, la oficina respectiva le asigna uno o más examinadores, especialistas en el área y expertos en legislación de patentes. Para asegurar la novedad e inventiva, el o los examinadores deben comparar la solicitud con el estado del arte publicado al momento de presentar la solicitud. Esto significa la revisión no sólo de otras patentes (otorgadas o no) sino también de otro material ya publicado sobre el tema, tales como publicaciones científicas, libros, informes técnicos, tesis etc. Sobre la base de ésta búsqueda, el o los examinadores entregan un informe a la oficina de patentes de cada país. Si el informe acepta que el invento es novedoso se otorga la patente; en caso contrario se le permite al solicitante discutir con el o los examinadores y hacer las modificaciones a la solicitud que permitan demostrar la novedad. El procedimiento puede tener variaciones según las legislaciones locales, pero este es, en general, el más aceptado.

Cabe recordar que la patente otorga protección sólo en el país en el que se concede, que puede no necesariamente ser el país de residencia del solicitante. Una vez otorgada la patente debe iniciarse el trámite de solicitud de patentes en los demás países donde interese protegerlo. Actualmente, este trámite se facilita mediante una solicitud a la Organización Mundial de Propiedad Intelectual (OMPI) y su tratado de protección en materia de patentes (el PCT).

El proceso de obtención de una patente es generalmente caro. Se incurre en gastos de contratación de un abogado especialista en patentes, de inscripción, examen, costos de búsqueda de información, costos de traducción en el caso de solicitudes en el extranjero, tarifas de los agentes de patentes extranjeros o abogados especializados en el extranjero, etc. En varios países de la región existen subsidios especiales para fomentar la solicitud de patentes, tanto en el país como en el exterior; sin embargo, lamentablemente, todavía es escaso el número de resultados de investigación e innovaciones patentados en o por los países de América Latina y el Caribe.

Es necesario recalcar que el proceso de obtención de una patente es largo. En general, para asegurar la prioridad de un invento se toma en consideración la fecha en la que se llevó a cabo la inscripción de la solicitud en alguna oficina de patentes. En muchos casos, esta solicitud se hace pública al cabo de periodos que van de un año a dieciocho meses. La protección definitiva a través de la concesión de una patente, tanto en el país como en las principales oficinas de patentes del mundo, puede tomar algunos años. No obstante, el producto se considera protegido desde el momento de su inscripción.

3.2 USO DE PATENTES COMO FUENTE DE INDICADORES

Al igual que en el caso de los indicadores bibliométricos de publicaciones científicas, es importante tener presente que existen variaciones importantes en cuanto a la importancia del uso de patentes de un sector a otro. Hay estudios recientes que muestran que el uso de patentes en sectores como el automotriz es muy bajo; sin embargo, las patentes son de mucha importancia para el sector farmacéutico. El sector productor de software prefiere el uso de propiedad intelectual (derechos de autor) más que el uso de patentes. Una vez más, es importante comparar entre sectores similares en el momento de llevar a cabo estudios en que se utilice información acerca de patentes. También es necesario reiterar que existen grandes variaciones en los procedimientos y criterios para la obtención de patentes de un país a otro.

3.2.1 BASES DE DATOS

Al iniciar el trabajo de preparación de indicadores de patentes es importante, en primer lugar, utilizar la información proporcionada por la oficina de patentes del propio país, además de las bases de datos internacionales más importantes. Entre estas últimas se encuentran la de la Oficina de Patentes de los Estados Unidos (USPTO, en <http://patft.uspto.gov/>) y la de la Oficina de Patentes Europea (disponible en <http://www.epo.org/patents.html>). Ambas contienen suficiente información, con el detalle necesario para llevar a cabo búsquedas de patentes otorgadas a personas no residentes, un dato importante para el análisis a nivel nacional, ya que, como se ha dicho, la tendencia es patentar no sólo en el país de origen del inventor, sino en los principales países o regiones en los que pueda comercializarse el producto.

Además de las base de datos mencionadas, de acceso abierto, existen bases de datos comerciales que aportan información para realizar estudios más específicos. Entre ellas cabe mencionar Derwent World Patents Index (DWPI), considerada la base internacional más completa. Incluye en la actualidad información acerca de más de 35.000.000 de patentes procedentes de más de cuarenta oficinas de patentes del mundo y de todos los sectores tecnológicos. Esta base permite llevar a cabo estudios de citas.

3.2.2 INFORMACIÓN DISPONIBLE

Al igual que en el caso de los artículos científicos, una patente es un documento que contiene en sí una gran cantidad de información que puede ser utilizada para producir indicadores o llevar a cabo estudios más detallados sobre el avance de una tecnología dada. La información incluye:

- Número de registro de la patente

- Estado de la patente (si es una solicitud o una patente concedida)
- Nombre y dirección del inventor o inventores
- Nombre y dirección del beneficiario o beneficiarios de la patente
- Código que señala si se trata de un ciudadano, un residente o un no residente en el país
- Fecha de la solicitud
- Clasificación técnica de la patente
- Referencias a otras patentes y a la literatura científica, tanto incluida en la patente misma, como aportada por los examinadores

En cada patente se asigna un código de clasificación a cada uno de sus componentes, por lo tanto una patente puede contener varios códigos que pueden ser asignados a uno o diferentes sectores tecnológicos. Para estudios por sector tecnológico existen diferentes clasificaciones, mediante las cuales se puede agregar la información que entrega la clasificación técnica de la patente. Es importante decidir, antes de iniciar un análisis por sectores tecnológicos, qué agrupación o clasificación utilizar

3.2.3 TIPOS DE MEDICIÓN

Los tipos de medición que se llevan a cabo, a nivel macro, para fines meramente comparativos, incluyen:

Número de patentes solicitadas y concedidas a nivel nacional, separadas por patentes solicitadas y concedidas a residentes y no residentes. Este indicador proporciona información acerca del número de inventos llevados a cabo en el país y el grado de protección que ciertos productos extranjeros tienen a nivel nacional.

Número de patentes solicitadas y concedidas en oficinas de patentes del extranjero, por parte de inventores residentes en el país. Este indicador proporciona información acerca de la actividad de innovación tecnológica que se ha estimado necesario proteger a nivel internacional.

Número de patentes solicitadas y concedidas en el país y en el extranjero por residentes en el país, por sector tecnológico. A partir de 1975, las principales oficinas de patentes del mundo utilizan una clasificación única de sectores tecnológicos para organizar la información. Con esta información es posible llevar a cabo estudios más detallados acerca de sectores tecnológicos en los que se trabaja más intensamente en el país.

Número de patentes solicitadas y concedidas en el país y en el extranjero por residentes en el país, por sector económico y por empresas. Este es un indicador importante para conocer qué sectores de la economía están llevando a cabo mayores esfuerzos en I+D con potencial aplicación tecnológica. También es interesante para conocer la colaboración entre sectores (educación superior y empresas, por ejemplo) en este campo, así como detectar las empresas con mayor grado de innovación tecnológica.

Colaboración internacional en la producción de patentes. Es interesante para medir el grado de colaboración internacional en el campo de la innovación tecnológica y la participación de residentes en el país en redes internacionales de desarrollo

de nuevas tecnologías. Se puede medir utilizando la información proporcionada en las direcciones de los inventores.

3.2.4 INFORMACIÓN CUALITATIVA

El análisis de citas, en el caso de las patentes, puede aportar información interesante para realizar estudios, análisis y evaluación de la innovación tecnológica a nivel nacional. Narin, Breitzman y Thomas (2004) definen algunos índices importantes para el trabajo de análisis con citas de patentes:

Índice corriente de impacto (Current Impact Index, CII). Muestra el impacto de las patentes de una empresa en el desarrollo tecnológico más reciente. Es una medida de la frecuencia con que las patentes de los últimos cinco años de una empresa dada son citadas en las patentes concedidas o registradas en el último año. Como éste es un índice que se calcula cada año, y siempre sobre las patentes concedidas en los últimos cinco años, cuando el número de patentes de una empresa decae esto se refleja en el índice CII del año.

Enlace científico (Science Linkage SL). Mide cuánta tecnología utilizada o aplicada en una empresa se basa en investigación científica de punta. Se calcula sobre la base del número promedio de referencias en las patentes de una empresa a trabajos científicos, que no son a su vez patentes. Se presume que las empresas cuyas patentes citan una gran cantidad de artículos científicos trabajan con resultados de investigación científica más nueva.

Ciclo de tiempo tecnológico (Technology Cycle Time, TCT). En general las empresas que innovan con mayor rapidez tienden a ser más exitosas en el desarrollo de productos que las empresas que se basan en tecnologías más antiguas. Este indicador se calcula sobre la antigüedad media de las patentes citadas en las patentes de una empresa. El hecho de citar patentes más antiguas puede ser un indicio de que la empresa está utilizando tecnologías más antiguas. El promedio de TCT es de aproximadamente tres a cuatro años en sectores industriales en rápido desarrollo, como en el caso de la industria electrónica, y de hasta quince años en sectores que cambian más lentamente, como por ejemplo el de la construcción de barcos.

Estos indicadores pueden ser aplicados también para realizar análisis a nivel nacional. Pueden ser especialmente interesantes los estudios de la influencia de la investigación en I+D producida en el país, en el desarrollo de nuevas tecnologías y su apropiación por parte de las empresas tanto nacionales como extranjeras.

Conocer cuántos trabajos científicos originados en el país son citados por patentes internacionales, en qué sectores tecnológicos y procedentes de qué áreas del conocimiento y sectores institucionales, constituye otro dato de importancia.

Otro indicador interesante es el de co-actividad. En aquellas áreas en que se produce un mayor intercambio entre ciencia y tecnología, los investigadores tienden a participar activamente en ambos tipos de trabajos. Esta co-actividad puede tomar diferentes formas, tales como patentar a través de la universidad, de socios en la empresa o en forma personal. Existen diversos estudios acerca de este doble rol de un investigador que es a la vez un inventor y los diversos niveles de relación que se establecen entre la academia y la empresa. Un investigador puede ser el autor tanto de trabajos de investigación como de patentes. Un análisis de este tipo de trabajos se encuentra en el artículo de Bassecouard y Zin (2004).

La OECD publicó en 1994 un Manual de Patentes, que se encuentra en revisión, y puede ser utilizado para fines metodológicos.

04. Webometría o cibermetría

Desde mediados de los años noventa se están llevando a cabo estudios acerca de la naturaleza del World Wide Web. Para ello se aplican modernas metodologías infométricas al estudio de contenidos, estructura de los enlaces y motores de búsqueda. Este tipo de estudios ha sido denominado webometría o cibermetría, y se considera parte del conjunto de disciplinas del que forman parte la bibliometría, cientometría e infometría.

El término webometría fue acuñado por Almind e Ingwersen en 1997 y se lo ha definido como "el estudio de los aspectos cuantitativos de la construcción y uso de recursos de información, estructuras y tecnologías en la Web conducentes a análisis de tipo bibliométrico e infométricos". La definición toma en cuenta tanto la construcción como el uso de la Web. De ella se desprenden cuatro áreas especializadas de investigación webométrica:

- Análisis de contenido de páginas Web
- Análisis de las estructuras de los enlaces Web
- Análisis del uso de la Web (incluyendo los archivos log del comportamiento de búsqueda y navegación de los usuarios)
- Análisis de la tecnología Web (incluyendo la evaluación del comportamiento de los motores de búsqueda)

Son de interés para el caso de los indicadores de producto, los trabajos que se han desarrollado para la descripción y evaluación de los contenidos en Internet, la obtención de estadísticas sobre el impacto de la red en distintas actividades de interés científico-técnico, económico y social y el análisis de los patrones de comunicación científica a través de la Web.

A partir del año 2004, el Laboratorio de Cibermetría del CINDOC-CSIC de España publica el Ranking Web de las universidades del mundo, que es actualizado dos veces al año y cubre más de 16.000 instituciones de educación superior de todo mundo. En este ranking se usa el concepto de "presencia Web", que según dicho laboratorio "mide la actividad y visibilidad de las instituciones y es un buen indicador del impacto y prestigio de las universidades. La posición en el ranking resume el rendimiento global de la universidad, aporta información para aquellos candidatos a estudiantes o académicos y refleja el compromiso para con la diseminación del conocimiento científico".

Para hacer el ranking se toman en consideración diversos parámetros, entre ellos las páginas con contenido académico (notas de cursos, conjuntos de datos, material gráfico, software y muchos otros), las publicaciones científicas en diferentes grados de edición, el número de enlaces a la página, el número de visitas, etc.

En forma complementaria se publica el Ranking Web de centros de investigación del mundo. El directorio incluye en la actualidad más de 7.000 centros de investigación. En forma separada se ofrece un ranking de los principales consejos de investigación del mundo, entre ellos el CNRS de Francia, el CONICET de Argentina, el Consejo Superior de Investigación Científica de España (CSIC), el National Institute of Health (NIH) de Estados Unidos, los Institutos Max Planck de Alemania, etc. A partir de 2008 se publica también el Ranking Web de los repositorios del mundo. Estos listados pueden ser consultados en la página Web especializada del Laboratorio de Cibermetría: http://www.webometrics.info/index_es.html.

05. Minería de datos, minería de textos, descubrimiento de conocimiento

En los últimos años, con la explosión de información de todo tipo en formato electrónico, se han desarrollado técnicas para descubrir datos, información y sus relaciones, que permiten su mayor y mejor recuperación, comprensión y uso. Estas técnicas facilitan el descubrimiento, captura, tratamiento, análisis y visualización de la información. Se basan en una combinación de técnicas avanzadas de estadística, inteligencia artificial, sistemas de información y visualización, para la obtención de conocimiento a partir de bases de datos de grandes dimensiones.

A diferencia de la recuperación de información tradicional, estas técnicas no sólo recuperan documentos concretos, sino que también descubren relaciones entre documentos afines, los jerarquizan y generan informes, mapas o representaciones gráficas con los datos obtenidos, que permiten la comprensión fácil de la información recuperada.

Estas técnicas están siendo ampliamente usadas en la empresa, en marketing, en la detección de fraudes, en las finanzas, en industrias manufactureras y en el trabajo científico (como por ejemplo en la búsqueda de relaciones en bancos de genes). Se las utiliza, asimismo, para detectar a tiempo problemas que afectan a la salud pública, para predicciones climáticas, para el tratamiento de millones de imágenes astronómicas en los grandes observatorios, etc.

La toma de decisiones de política científica también ha recurrido al apoyo de estas técnicas, ya que posibilitan analizar en mayor profundidad el estado de la I+D de un país o de una disciplina dada, detectar líneas emergentes de investigación, relaciones entre ciencia y su aplicación en la producción de bienes y servicios, redes de colaboración para enfrentar nuevos problemas científicos emergentes, etc.

La minería de textos está siendo utilizada cada vez con mayor éxito, con base ya no sólo en datos numéricos, sino en el descubrimiento de nuevo conocimiento a partir del lenguaje utilizado en los textos mismos. Existen en la actualidad numerosas herramientas para el análisis y procesamiento en línea que pueden llevar a cabo en forma interactiva estas búsquedas y la visualización de sus resultados. En general, estas herramientas basan sus búsquedas en los metadatos que cada documento electrónico posee.

En 1996, Fayad brindó una de las definiciones más citadas de lo que se entiende por Knowledge Discovery and Data Mining (KDD): "El proceso no trivial de identificar estructuras de datos válidas, novedosas, potencialmente útiles y finalmente 'comprensibles'". Es decir, se trata de descubrir el significado que poseen los datos almacenados en grandes bancos, explorar y analizar las bases de datos disponibles para ayudar a la toma de decisiones, además de facilitar la extracción de la información existente en los textos y crear sistemas inteligentes capaces de entenderlos. Fayad propone los siguientes pasos para trabajar con cualquier sistema de minería de textos:

- Definición del objetivo del estudio. Su alcance, conocimiento previo del tema.
- Selección de las fuentes. El proceso de selección de las fuentes a ser explotadas implica el conocimiento de las fuentes disponibles, conocimiento del tema y una comprensión de los objetivos del trabajo.
- Limpieza de las fuentes y los datos. Esto implica eliminar los datos que pueden hacer "ruido" o ser redundantes, decidir estrategias para completar datos inexistentes en algunos campos y decidir cómo tratar variaciones de situaciones previsibles en el tiempo.
- Buscar las características útiles para representar los datos, de acuerdo con los objetivos del trabajo. En esta etapa se

pueden definir las variables efectivamente útiles, etc.

- Comparar los objetivos del primer punto. Objetivos del trabajo con un tipo específico de minería de datos: resumen, clasificación, regresión, clasificación, clustering, etc.
- Análisis exploratorio y selección de los modelos e hipótesis, selección de los algoritmos y metodología ser usada.
- Minería de datos. Buscar estructuras de interés, como clasificaciones, árboles, regresiones o clusterings.
- Interpretación de las estructuras encontradas. Visualización de dichas estructuras.
- Finalmente, uso del conocimiento producido, revisando su posible validez, su posible conflicto con conocimiento previamente aceptado, etc.

La ventaja del uso de estas tecnologías es la posibilidad de utilizar diferentes bases de datos en forma simultánea y complementaria, permitiendo una selección de fuentes y extracción de contenidos más apropiada para cada estudio o análisis. En el caso de los estudios bibliométricos es interesante la posibilidad de utilizar no sólo base de datos de artículos científicos, sino también de otros tipos de materiales de importancia para el estudio, tales como informes técnicos, normas, patentes, imágenes y muchos otros.

En la actualidad existen muchos programas para llevar a cabo este tipo de análisis, algunos de los cuales han sido utilizados en la región para estudios bibliométricos; entre ellos está Collexis, utilizado en Brasil por Bireme, y Stanalyst, empleado en un proyecto conjunto entre Bireme, CAICYT de Argentina y CONICYT de Chile.

Algunas aplicaciones de interés regional son los Atlas de la Ciencia elaborados por SCImago para algunos países de la región, basados en la información del Web of Science.

Estas técnicas de minería de datos, minería de textos y descubrimiento de conocimiento ofrecen en la actualidad amplias posibilidades para emprender estudios del estado de los sistemas de I+D, su producción e impacto, con mucha más precisión que la lograda hasta este momento. Esta es una muy interesante línea de investigación a ser iniciada en muchos de los países de la región.

AGUILLO, I. (2003): *Cibernetría. Introducción teórico-práctica a una disciplina emergente*, Madrid, CINDOC-CSIC (disponible en <http://internetlab.cindoc.csic.es/cursos/cibernetria.pdf>).

BASSECOULARD, E. y ZITT, M. (2004): "Patents and publications", en: *Handbook of quantitative Science and Technology Reseach. The use of Publication and Patent Statistics in Studies of S&T Systems*, Kluwer Academic Publishers.

BJÖRK, B.-C. (2007): "A model of scientific communication as a global distributed information system", *Information Research*, 12(2) paper 307 (disponible en <http://InformationR.net/ir/12-2/paper307.html>).

CRONIN, B. (2005): *The Hand of Science. Academic writing and its rewards*, London, Scarecrow Press.

COURTIAL, J. P. (1990): *Introduction a la Scientométrie. De la bibliométrie a la veille technologique*, Paris, Anthropos.

DERNIS, H., GUELLEC, D. y VAN POTTELSBERGHE, B. (2001): "Using patent counts for cross-country comparisons of technology output", *STI Review* No. 27.

FAYAD, U., PIATETSKY-SHAPIO, G. y SMYTH, P. (1996): "From Data Mining to Knowledge Discovery in databases", *AI Magazine*, Fall, pp. 37-54 (disponible en: <https://www.aaai.org/aitopics/assets/PDF/AIMag17-03-2-article.pdf>).

FAYAD, U., PIATETSKY-SHAPIO, G. y SMYTH, P. (1996): "From Data Mining to knowledge discovery: an overview", en: *Advances in knowledge discovery and data mining*, AAAI/MIT Press.

GARFIELD, E. (1955): "Citation indexes for science - new dimension in documentation through association of ideas", *Science* 122(3159), pp. 108-111.

GILBERT, K. (2004): "Técnicas híbridas de inteligencia artificial y estadística para el descubrimiento de conocimiento y minería de datos" (disponible en: <http://www.lsi.us.es/redmidas/Capitulos/LMD12.pdf>).

GLÄNZEL, W. y SCHOEPLIN, U. (1994): "A Stochastic Model for the Ageing Analyses of Scientific Literature", *Scientometrics*, 30 (1), pp. 49-64.

GLÄNZEL, W. (2003): "Bibliometrics as a research field. A course on theory and application of bibliometric indicators. Course Handouts" (disponible en: http://www.norslis.net/2004/Bib_Module_KUL.pdf).

HIRSCH, J. E. (2005): "An index to quantify an individual's scientific research output", *Proc.Nat.Acad.Sci.*, vol. 102, 16569 (disponible en [arXiv:physics/0508025v5](https://arxiv.org/abs/physics/0508025v5)).

INGWERSEN, P. y BJÖRNEBORN, L. (2004): "Methodological issues of Webometric Studies", en: *Handbook of quantitative Science and Technology Research. The use of Publication and Patent Statistics in Studies of S&T Systems*, Dordrecht, Kluwer Academic Publishers.

KING, D. W., TENOPIR, C. y CLARKE, M. (2006): "Measuring Total Readings of Journal Articles", *D-Lib Magazine*, 12(10) (disponible en: <http://www.dlib.org/dlib/october06/king/10king.html>).

KUHN, T. (1962): *The Structure of Scientific Revolutions*, Chicago, University of Chicago Press. (ed. española editada por el Fondo de Cultura Económica, México).

LEWISON, G. (2005): "Beyond SCI citations: New ways to evaluate research", *Current Science*, 89(9) (disponible en: <http://www.ias.ac.in/currsci/nov102005/1524.pdf>).

- LEYDESDORFF, L. (2001): *The challenges of scientometrics. The development, measurement and self-organization of Scientific Communications*, Universal publishers.
- LOSIEWICZ, P., OARD, D. y KOSTOFF, R. (2000): "Textual data mining to support Science and Technology management", *Journal of Intelligent Information Systems*, 15, pp. 99-119.
- MOED, H. F., GLÄNZEL, W. y SCHMOCH, U. (2004): *Handbook of Quantitative Science and Technology Research. The use of publication and patent statistics in studies of S&T Systems*, Dordrecht, Kluwer Academic Publishers.
- NALINOV, V. y MULCHENKO, E. M. (1969): *Nauchnometria*, Moscú, Isd. Nauka.
- NARIN, F. (1976): *Evaluative Bibliometrics. The use of Publication and Citation Analysis in the Evaluation of Scientific Activity*, Cherry Hill, NJ, Computer Horizons (disponible en: http://www.conceptsymbols.com/narin/narin_1975_eval-bibliometrics_images.pdf).
- NARIN, F., BREITZMAN, A. y THOMAS, P. (2004): "Using Patent citation indicators to manage a stock portfolio", en: *Handbook of quantitative Science and Technology Research. The use of Publication and Patent Statistics in Studies of S&T Systems*, Dordrecht, Kluwer Academic Publishers.
- NESTA, L. y PATEL, P. (2004): "National Patterns of Technology accumulation: use of Patent statistics", en: *Handbook of quantitative Science and Technology Research. The use of Publication and Patent Statistics in Studies of S&T Systems*, Dordrecht, Kluwer Academic Publishers.
- OCDE (1994): *The Measurement of Scientific and Technological Activities. Using Patent data as Science and technology Indicators. Patent Manual 1994*, Paris, OCDE (disponible en: <http://www.oecd.org/dataoecd/33/62/2095942.pdf>).
- OKUBO, Y. (1997): "Bibliometric indicators and analysis of research systems. Methods and examples", OCDE, STI Working Papers 1997/1. <http://ideas.repec.org/p/oec/stiaaa/1997-1-en.html#download>
- PENDLEBURY, D. (2008): *White Paper. Using bibliometrics in evaluating research*, Philadelphia, Thomson Reuters Scientific.
- PRICE, Derek De Solla (1965): "Networks of Scientific Papers", *Science*, 149 (3683), pp. 510-515.
- PRICE, Derek De Solla (1963): *Little Science, Big Science*, Columbia University Press (ed. en español como *Hacia una ciencia de la ciencia*, Barcelona, Ariel, 1973).
- SANCHO, R. (2002): "Indicadores de los sistemas de ciencia, tecnología e innovación", *Economía industrial*, Nº 343 (disponible en: <http://www.mityc.es/NR/rdonlyres/82F8051A-742D-4F3D-8149-4C2B76A8DoA6/o/097SANCHO.pdf>).
- SEGLEN, P. (1992): "The Skewness of Science", *JASIS* 43(9), pp. 628-638.
- SPINAK, E. (1996): *Diccionario enciclopédico de bibliometría, cienciometria e informetria*, Caracas, UNESCO.
- VAN RAAN, F. J. (2004): "Measuring Science", en: *Handbook of quantitative Science and Technology Research. The use of Publication and Patent Statistics in Studies of S&T Systems*, Dordrecht, Kluwer Academic Publishers.
- VARGAS QUESADA, B. y MOYA-OBREGÓN, F. (2007): *Visualizing the structure of science*, New York, Springer.
- WEBB, C., DERNIS, H., HARHOFF, D. y HOISL, K. (2005): "Analysing European and international patent citations: a set of EPO patent database building blocks", OCDE. STI Working Paper 2005/9 (disponible en: <http://www.oecd.org/dataoecd/18/17/35520805.pdf>).

