



Relación proyección de tres entregables - Línea Cienciometría INM

- P1: Documento con la metodología para el análisis cienciométrico...
- P2. Documento con la identificación de fuentes de datos disponibles...
- P3. Plan de trabajo para la implementación y operación de la metodología...

Entregable 1. Metodología

Introducción

La cienciometría es una disciplina que analiza la producción científica mediante métodos cuantitativos, con el propósito de evaluar el impacto, la evolución y las dinámicas de la ciencia y la tecnología. Para ello, estudia diversas fuentes como publicaciones científicas, patentes, proceedings (actas de congresos), convocatorias, agendas de investigación y otros insumos que permiten comprender cómo se genera, organiza y difunde el conocimiento.

A través de indicadores y técnicas de análisis como redes de colaboración, coautoría, coocurrencia de términos, citas, estructuras temáticas o análisis de contenidos, la cienciometría ofrece una visión integral de la actividad científica. Esto permite identificar patrones, dinámicas de crecimiento, áreas estratégicas y vacíos de conocimiento, y proporciona evidencia para fortalecer la toma de decisiones en ciencia, tecnología e innovación.

En particular, el análisis de publicaciones y patentes facilita la detección de señales tempranas (temas emergentes) y megatendencias (líneas consolidadas de investigación), así como la comprensión de los modos de producción científica basados en la colaboración entre instituciones, países y disciplinas. De esta manera, la cienciometría no solo revela el pasado y el presente del avance científico, sino que también contribuye a proyectar trayectorias tecnológicas y anticipar escenarios futuros, reduciendo la incertidumbre en la formulación de políticas públicas, estrategias institucionales y decisiones de inversión en I+D.

Metodología para la realización de análisis cienciométricos

A continuación, se detalla el procedimiento para la construcción de análisis cienciométricos con el que se pueda analizar el panorama de investigación de un campo a partir de las tendencias temáticas.

En el marco del Modelo Integrado de Planeación y Gestión (MIPG) la Gestión del Conocimiento y la Innovación es la séptima dimensión MIPG que parte de la premisa de que el conocimiento es el activo más importante de una entidad pública. No se trata de guardar archivos ni de almacenar procesos y procedimientos con buenas prácticas y lecciones aprendidas, sino de un proceso dinámico de acciones orientadas a generar, identificar, capturar, valorar, transferir y preservar el conocimiento (tanto tácito como explícito) para mejorar la gestión y facilitar la innovación.

La dinámica de la gestión del conocimiento tiene como base, para que las cuatro dimensiones operen de forma armónica, la designación de un líder o gestor del conocimiento, cuya responsabilidad es coordinar las actividades que permitan tanto la gestión del conocimiento en la entidad como su relación con las demás dimensiones del MIPG. Este líder o gestor del conocimiento tiene la responsabilidad de convocar expertos, sintetizar la información aportada en los talleres y presentar el informe final.

Para efectos del primer entregable, definido en términos de una “metodología para el análisis cienciométrico de la producción científica del INM”, se presentan los procedimientos necesarios para establecer un panorama de la creación intelectual del INM que pueda integrar un sistema de gestión del conocimiento en el marco del MIPG. Sin embargo, dado el bajo número de resultados en las principales bases de datos, presentamos primero las consultas generales para el INM y luego el procedimiento para crear un panorama cienciométrico de las tendencias de investigación a nivel global.

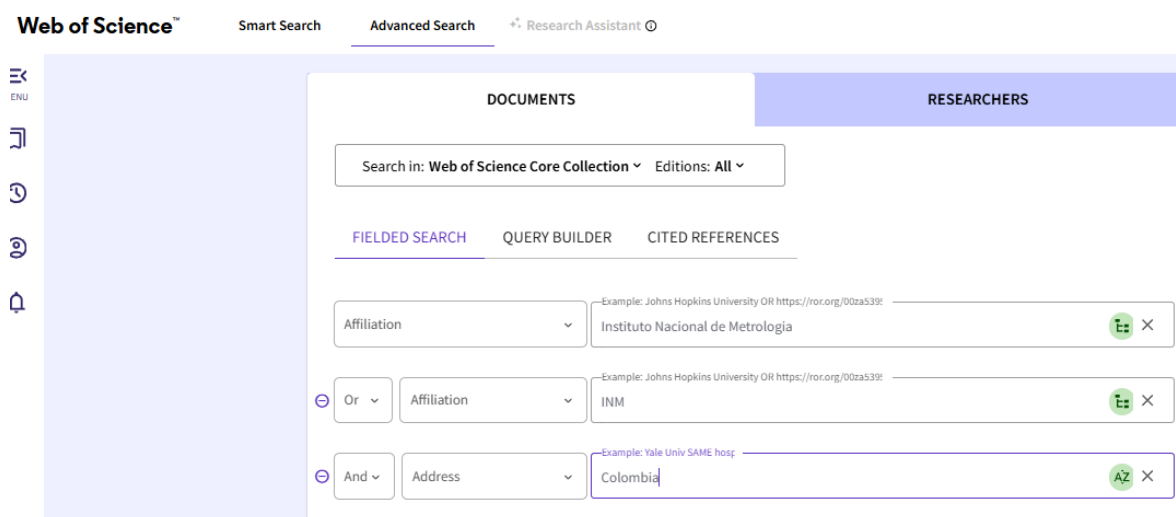
Fase 1. Identificación de la producción científica y tecnológica del INM

1. Definición de la consulta en las bases de datos.

Se propone, si bien en el segundo entregable se presenta una exposición de las bases de datos, utilizar cuatro fuentes: dos de acceso por suscripción (Web of Science y Scopus) y dos de acceso abierto (OpenAlex y Lens).

Hacer una lista de las formas en que se pudo codificar el INM y filtrar por país: Colombia. Ejemplo. La producción de la entidad puede estar por el nombre completo, por la sigla o por abreviaturas de la entidad, por ejemplo, Ins Nal Met.

A continuación, se presentan las búsquedas sobre la producción intelectual del INM en documentos científicos y patentes.



Web of Science™ Smart Search **Advanced Search** Research Assistant

DOCUMENTS **RESEARCHERS**

Search in: **Web of Science Core Collection** Editions: **All**

FIELD SEARCH QUERY BUILDER CITED REFERENCES

Affiliation **Instituto Nacional de Metrologia** Example: Johns Hopkins University OR https://ror.org/00za539 **×**

Or Affiliation **INM** Example: Johns Hopkins University OR https://ror.org/00za539 **×**

And Address **Colombia** Example: Yale Univ SAME hosp **×**



Scopus

[Search](#) [Sources](#) [SciVal](#) [?](#) [🔔](#)

Start exploring

[Documents](#) [Authors](#) [Researcher Discovery](#) [Organizations](#) [Search tips](#)

Search within Affiliation	Search documents Instituto Nacional de Metrologia	×	🗑️
OR			
Search within Affiliation	Search documents INM	×	🗑️
AND			
Search within Affiliation country	Search documents Colombia	×	🗑️

Es fundamental incluir el país y no buscar en Scopus por Organization ya que solo sale registrado el INM de Brasil.



Scopus

Instituto Nacional de Metrologia


Santa Alexandrina St, 416, [Rio de Janeiro, RJ, Brazil](#) [📞 60069940](#)


2.736
Documents

849
Authors

[🔔 Set document alert](#)

[✍️ Give feedback](#)




Works ▾

Data Version 2 is live! [Learn more here.](#)

Unsaved search ▾

1
Institution
is
Instituto Nacional de Metrología de Colombia
X

+

Works
Stats

Evaluation of pesticide residues in honey from different geographic re-
gions of Colombia

46 results


English - EN ▾
Our Apps ▾
Pricing
About ▾
Work Ap

167,632,408 Patents (94,698,163 Simple families)
Explore Science, Te

FILTERS
Date Range
Flags
Jurisdiction
Applicants
Inventors
Owners
Agents & Attorneys
Legal Status
Document Type
Cited Works
Biologicals
Classifications
Document Family
Query Tools

New Patent Search
Patents (167,632,408) = All Docs
Filters: No filters applied

Patent Records
167,632,408
Patent Citations
493,595,078
Cites Patents
55,248,659
Citi
6.

Structured Search
Query Text Editor
Profiles

Field
Predicate: AND OR

Owner Name
Instituto Nacional de Metrologia
+ -

Owner Name
INM
+ -

Applicant Country
Colombia
+ -

2. Panorama cientímetro sobre campos de la ciencia y la tecnología

La formulación de un estudio cientímetro inicia con la definición del tema concreto a estudiar. Para ello, se convoca a los integrantes de un grupo de investigación o laboratorio, así como a reconocidos expertos externos a la organización que puedan aportar ideas para la identificación y selección de las palabras clave que delimitan el tema de interés. Un taller de post-it en el que cada experto proponga sus palabras clave es recomendable.

El líder deberá sintetizar las palabras clave y, con ellas, realizar una búsqueda inicial de revisiones de la literatura y de manuales (handbooks) para darse un panorama del tema. Luego convoca a una segunda sesión en la que expone los resultados de su análisis a los expertos, quienes, en una discusión abierta, aportan ideas para definir conjuntamente el conjunto de palabras que delimitan el tema de investigación y el periodo de análisis. El taller finaliza con la definición de un tesoro, esto es, un conjunto de palabras clave de sustantivos con sus respectivos sinónimos. Esto implica que el taller finaliza con una actividad en la que cada término está asociado a su conjunto de palabras similares y abreviaturas. Por ejemplo, artificial intelligence tiene la abreviatura AI así como una serie de conceptos asociados: Machine Learning (ML), Computational Intelligence (CI), Deep Learning (DL), Neural Networks (NN), Convolutional Neural Networks (CNN), Computer Vision (CV) y Natural Language Processing (NLP). Finalmente, el líder debe incluir en la discusión del taller el rango de tiempo para la elaboración de la información, buscando un consenso sobre los años en que se realizarán las consultas a las bases de datos.

3. Estrategia de búsqueda

La búsqueda general en bases de datos se basa en la organización de palabras clave, denominada técnicamente tesoro, y en operadores booleanos. Los términos del taller de expertos deben estar en inglés (idioma universal de la ciencia) e incluir los posibles sinónimos. La ecuación de búsqueda tiene las siguientes opciones:

- AND: para combinar conceptos (ej: Energy AND Storage).
- OR: para sinónimos (ej: Battery OR Accumulator).
- " ": Para frases exactas.

También es necesario:

- Aplicar filtros: tipo de documento y área temática. Se sugiere realizar un análisis independiente de artículos, revisiones de la literatura y memorias de eventos (proceedings). Asimismo, es importante, de acuerdo con el objetivo de la investigación, filtrar las áreas de la ciencia que son de interés para la investigación. Por ejemplo, en IA donde salen resultados en todos los campos del conocimiento, si el enfoque es técnico sería apropiado excluir las publicaciones en el campo de las ciencias sociales.
- El ejercicio finaliza con la exportación de los resultados en formatos compatibles para el software de analítica de contenidos que se expone a continuación en el punto 4 (TXT, CSV, RIS, BibTeX).

El líder deberá determinar las fuentes de consulta en función de la disponibilidad de acceso a bases de datos científico-tecnológicas. Este apartado se detalla en el segundo informe.

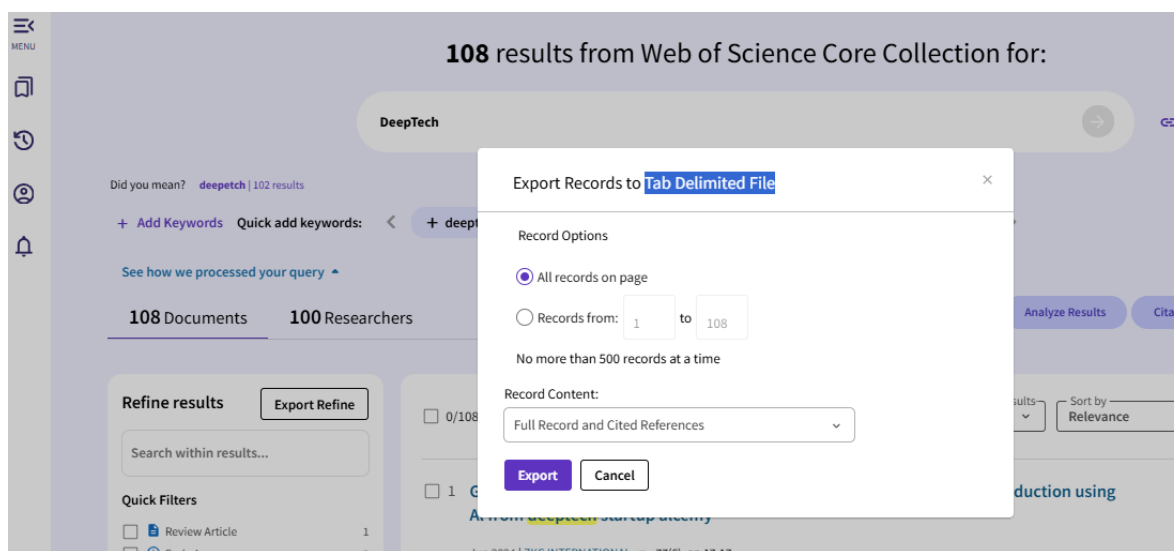
- Acceso abierto: Dimensions, Lens.org, OpenAlex.
- Suscripción: Web of Science, Scopus.

4. Creación de Mapas con VOSviewer

VOSviewer es una herramienta especializada en la construcción y visualización de mapas cuantitativos de acceso abierto. Fue desarrollada por el Centro de Estudios de Ciencia y Tecnología (CWTS) de la Universidad de Leiden y se utiliza ampliamente en análisis cuantitativos para representar las relaciones entre publicaciones científicas, autores, instituciones, palabras clave y citas. Para los intereses del INM, se presenta el procedimiento para la creación de mapas de coocurrencia de palabras, mediante el cual se identifican los temas emergentes y consolidados de la investigación en un campo determinado.

4.1. Preparación de Datos

- Se realiza la consulta con las palabras clave resultantes del proceso anterior y, luego de aplicar los filtros seleccionados, se procede a descargar los archivos.
- Exporta los resultados de la búsqueda en las bases de datos (Web of Science, Scopus, OpenAlex, Lens etc.).
- Formatos compatibles: .txt, .csv o archivos bibliográficos como .ris.





Avanciencia
ASOCIACIÓN COLOMBIANA PARA EL AVANCE DE LA CIENCIA

Scopus

Search within results

Save search

Set search alert

Search documents *
deeptech

File types

- CSV
- RIS
- BibTeX
- Plain text

Reference managers

- Mendeley
- Refworks (RIS)
- Zotero (RIS)
- EndNote (RIS)

Documents

19 documents

Refine search

Search within results

Filters

Document title

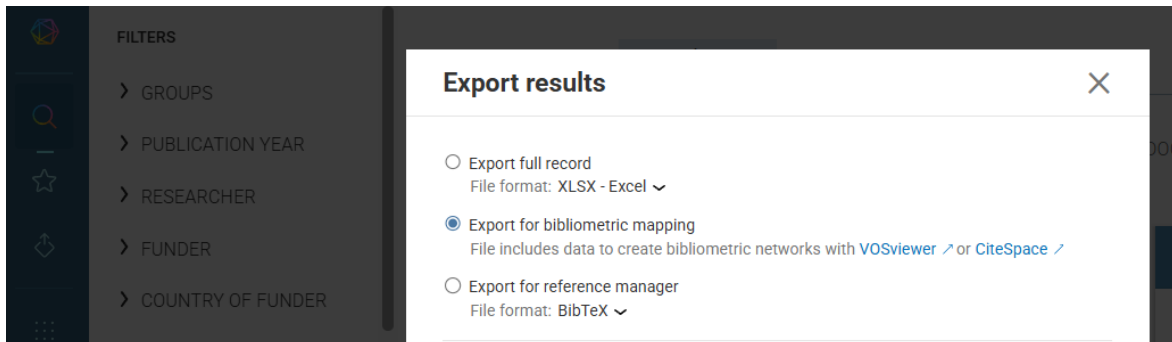
Autl

1

Cosmos 1.0: a multidimensional map of the emerging technology frontier

Gor

Grif



FILTERS

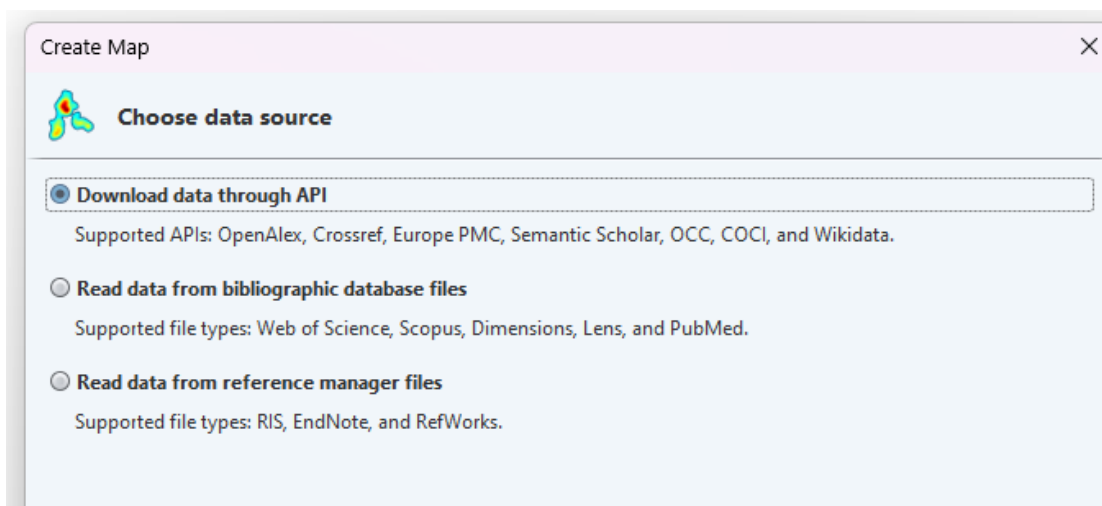
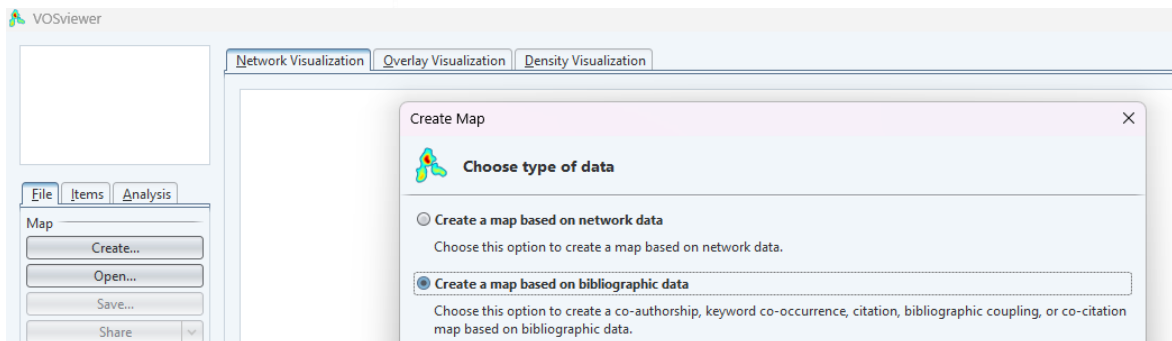
- GROUPS
- PUBLICATION YEAR
- RESEARCHER
- FUNDER
- COUNTRY OF FUNDER

Export results

- Export full record
File format: XLSX - Excel
- Export for bibliometric mapping**
File includes data to create bibliometric networks with [VOSviewer](#) or [CiteSpace](#)
- Export for reference manager
File format: BibTeX

4.2. Abrir VOSviewer

- Descargar e instalar VOSviewer desde <https://www.vosviewer.com>.
- Ejecutar el programa y seleccionar Create: Create a map based on bibliographic data y luego Read data from bibliographic database files.
- Carga tus archivos (Scopus, WoS, Dimensions, RIS).
- Unidad de análisis: Co-ocurrence y All keywords.
- Método de conteo: Full counting.
- Ajustar parámetros: número mínimo de ocurrencias, buscar una cantidad mínimo de 200 y máximo de 800 palabras.



Create Map



Choose type of analysis and counting method

Type of analysis: ?

- ☐ Co-authorship
- ☒ Co-occurrence
- ☐ Citation
- ☐ Bibliographic coupling
- ☐ Co-citation

Unit of analysis:

- ☒ All keywords
- ☐ Author keywords
- ☐ KeyWords Plus

Counting method: ?

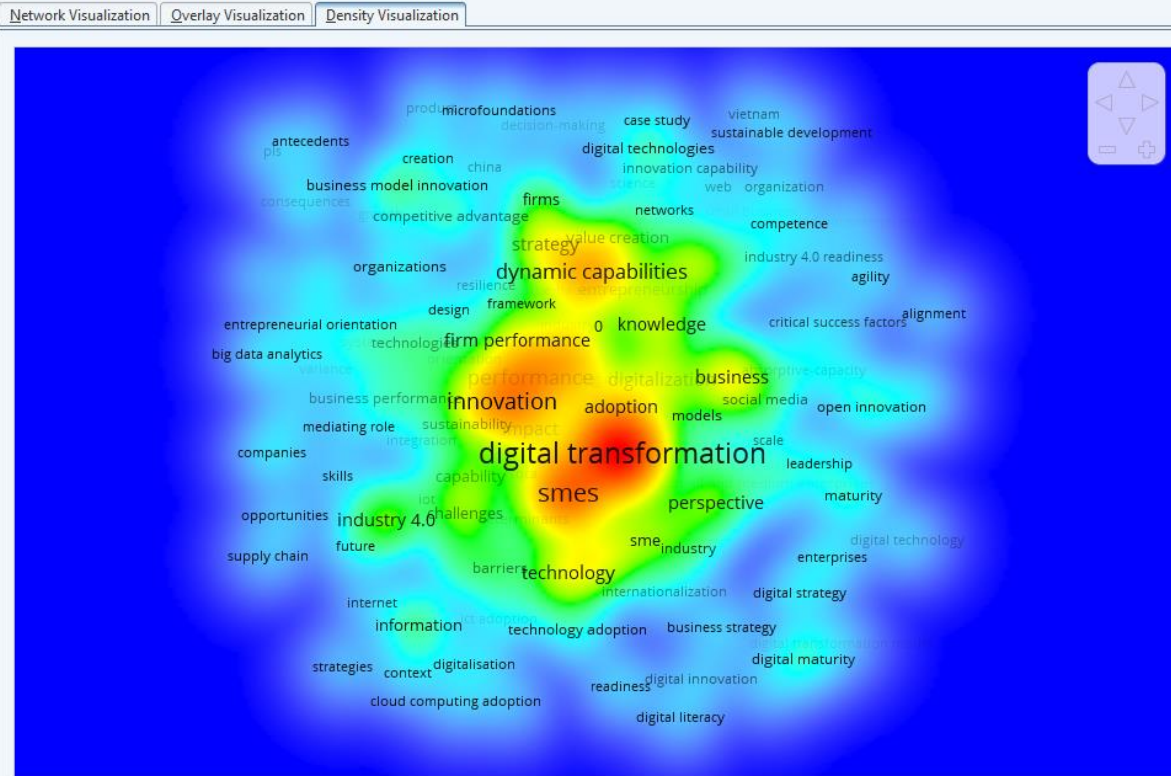
- ☒ Full counting
- ☐ Fractional counting

4.3. Análisis Clave

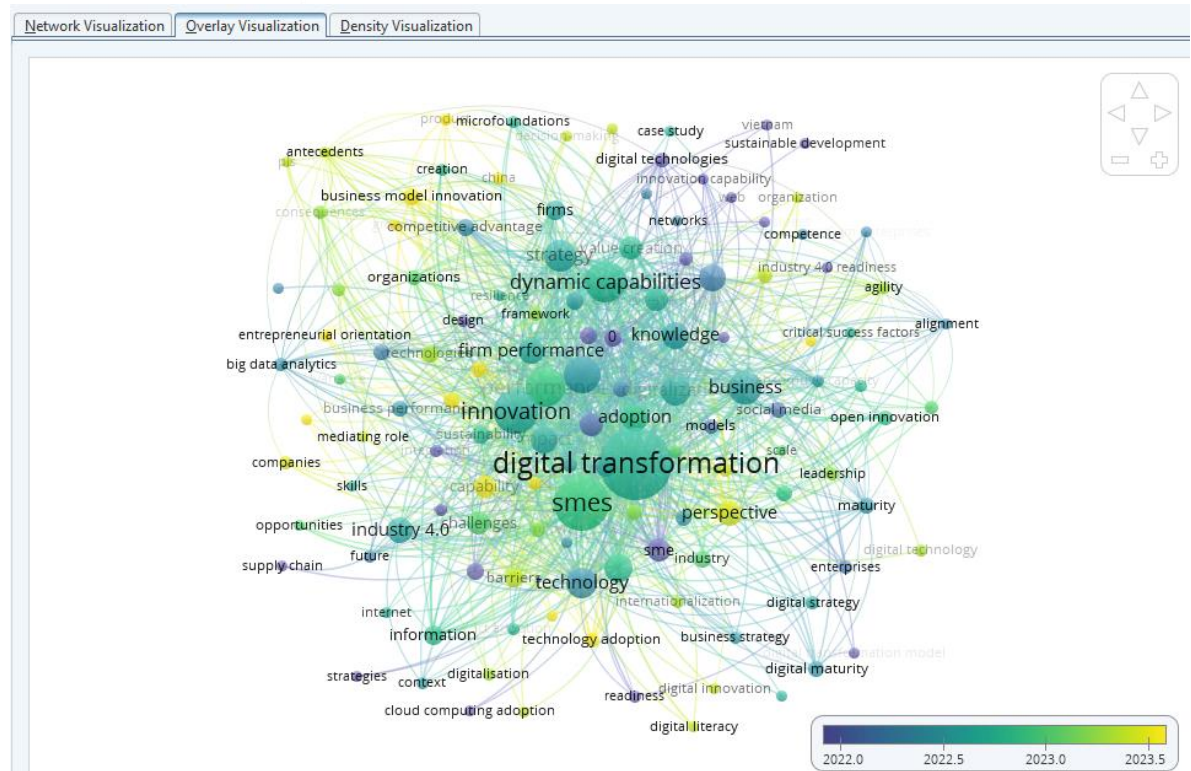
Para las tendencias, en el análisis de coocurrencia de palabras clave se identifican grupos de palabras por colores y los enlaces entre ellas para definir etiquetas y darles sentido a los temas de investigación. Si "Blockchain" aparece junto a "Logistics" con una conexión directa y con el mismo color, hay una tendencia de aplicación ahí.

- Clústeres (Colores): Agrupaciones temáticas.
- Nodos (Círculos): Representan términos. Su tamaño indica la frecuencia (peso). Cuanto más grande, más investigado es el tema.
- Enlaces (Líneas): Representan la relación. Su grosor indica la fuerza del vínculo (cuántas veces aparecen juntos).
- Colores: evolución temporal o agrupamiento temático.

- T: 317 4288010 M: info@avanciencia.org W: www.avanciencia.org**
Dir: Cra 16 # 31A - 16 - Bogotá Colombia



- Mapa evolutivo (Overlay Visualization)
 - Azul oscuro: temas antiguos/consolidados.
 - Amarillo/Verde claro: temas recientes/emergentes.



Construcción de escenarios

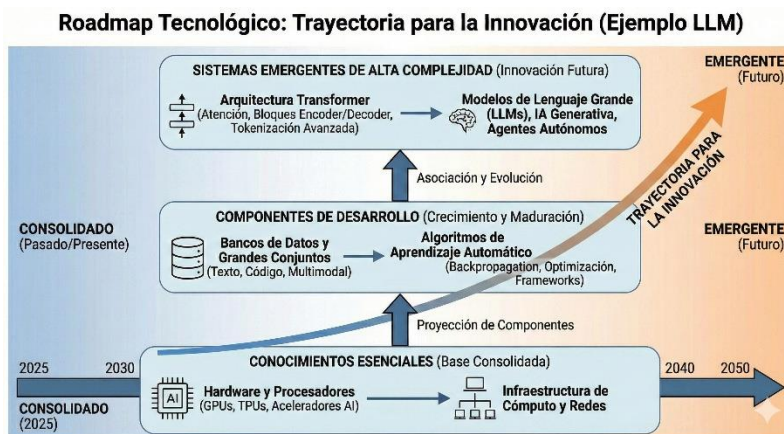
El análisis de contenido de publicaciones y patentes nos permite detectar señales débiles (temas emergentes) y megatendencias (temas consolidados). Al analizar el pasado y el presente de la producción intelectual, podemos proyectar trayectorias tecnológicas y reducir la incertidumbre en la toma de decisiones estratégicas.

De los mapas de VosViewer se necesita:

- Tópicos dominantes (clústeres grandes).
- Tópicos emergentes (clústeres pequeños pero conectados).
- Vacíos o brechas (áreas con poca conexión) que representan oportunidades de innovación.
- Evolución temporal de temas
- Identificar clústeres con alta y baja conexión interdisciplinaria.

Trazar una línea de tiempo en la que se ubiquen los temas de consolidados a emergentes, buscando asociaciones que permitan definir una trayectoria para la innovación. En la base, identificar los conocimientos esenciales para el desarrollo de alta complejidad, que se proyectan a partir de sus componentes. Por ejemplo, para un sistema de large language modeling LLM la base necesaria es el hardware y los procesadores, indispensables para los desarrollos. Luego se proyectan los bancos de

datos y sobre ellos los algoritmos de entrenamiento para el aprendizaje de máquina y finalmente se proyecta la arquitectura del transformer con los tipos de tokens y los bloques encoder y decoder.



Segundo entregable:

Introducción

Las bases de datos de artículos científicos y patentes son repositorios organizados que recopilan información sobre publicaciones académicas y desarrollos tecnológicos. Las de artículos incluyen metadatos (autores, título, resumen, palabras clave), referencias y, en muchos casos, el texto completo de investigaciones publicadas en revistas científicas. Las de patentes, por su parte, contienen información sobre invenciones registradas, incluyendo descripciones técnicas, reivindicaciones y datos legales. Estas bases son esenciales para los estudios cuantitativos porque permiten analizar la producción científica, las tendencias tecnológicas y la relación entre la investigación y el desarrollo industrial.

Las bases de datos se organizan como sistemas de indexación y resumen para facilitar la búsqueda de información a partir de la selección de revistas por su calidad y la normalización de la información y asignación de descriptores (palabras clave, códigos temáticos) para que las publicaciones puedan



recuperarse de manera eficiente. El resumen proporciona una síntesis del contenido del documento a través del título, el resumen y las palabras clave, lo que permite evaluar rápidamente el documento. Estos sistemas son fundamentales para la minería de datos porque garantizan la coherencia en la clasificación y la interoperabilidad entre diferentes plataformas.

Las bases de datos se dividen en dos grandes categorías: por suscripción y de acceso abierto. Las primeras requieren pago o licencias institucionales para acceder a sus contenidos completos y herramientas avanzadas; suelen ofrecer cobertura exhaustiva, alta calidad en la indexación y métricas robustas. Las segundas son gratuitas y permiten el acceso libre a los datos, aunque pueden tener limitaciones en cobertura, actualización y funcionalidades analíticas. Ambas son complementarias: las de suscripción aportan profundidad y confiabilidad, mientras que las abiertas favorecen la democratización del conocimiento.

Bases de datos por suscripción

El Web of Science (WoS)

Es una plataforma propiedad de Clarivate que actúa como una base de datos bibliográfica multidisciplinaria y un índice de citas. Fue creada en 1997 como evolución del Science Citation Index, desarrollado inicialmente por el Institute for Scientific Information (ISI). Hoy en día, el Core Collection de WoS incluye más de 79 millones de registros (y más de 171 millones en toda la plataforma) en 256 disciplinas que abarcan artículos, actas de congresos, reseñas, editoriales y más, desde 1900 hasta la actualidad. WoS es considerado estándar para realizar análisis cuantitativos, evaluaciones de la investigación y estudios comparativos a nivel nacional o institucional. Las referencias citadas permiten rastrear el impacto y la evolución científica, mapear disciplinas y detectar áreas emergentes.

WoS ofrece diez índices importantes dentro del Core Collection, que cubren disciplinas como ciencias naturales (Science Citation Index Expanded), ciencias sociales (Social Sciences Citation Index), artes y humanidades (Arts & Humanities Citation Index), y fuentes emergentes (Emerging Sources Citation Index), además de índices especializados en congresos, libros y química. También tiene el sistema Derwent Innovation Index especializada en patentes y organizada para el análisis tecnológico y vigilancia competitiva. Cada entrada incluye referencias citadas, lo que permite rastrear la evolución de las ideas, medir el impacto científico y realizar estudios bibliométricos con herramientas robustas. La plataforma se distingue por un riguroso proceso de selección y curación editorial: un equipo de Clarivate evalúa títulos con criterios de objetividad, relevancia y formato técnico, y luego indexa metadatos estandarizados (DOI, autores, afiliaciones, resumen) a partir de PDFs y XML garantizando la calidad y precisión de la base de datos.



Funcionalidades:

- Búsqueda de referencias citadas para identificar cómo y dónde se ha citado un trabajo.
- Perfiles enriquecidos de investigadores.
- Métricas consolidadas (Journal Impact Factor, h-index, conteo de citas, análisis de redes).
- Mapeo de áreas temáticas consolidadas y emergentes.

Ventajas

- Calidad y confiabilidad: metadatos curados y normalizados por expertos. Bibliografía depurada.
- Cobertura histórica sólida: desde principios del siglo XX.
- Herramientas avanzadas: análisis de redes de citas, métricas, tendencias.
- Independencia editorial: sin influencia de los editores de las publicaciones.

Limitaciones

- Cobertura sesgada: favorece disciplinas de ciencia y tecnología, y revistas en inglés, con menor visibilidad en humanidades, idiomas locales y ciencias sociales.
- Acceso restringido: requiere suscripción institucional costosa.
- Restricciones para minería de datos: limitaciones en acceso masivo incluso con licencia activa.

Scopus

Scopus es una base de datos de indexación y resumen creada por Elsevier en 2004. Contiene más de 100 millones de registros, incluyendo 71 millones de documentos con citas desde 1970, unos 6,6 millones anteriores a esa fecha, y más de 44 millones de patentes. Su cobertura abarca 23,000 revistas, 9,8 millones de actas de congresos, 210,000 libros y material "Articles-in-Press". El contenido de Scopus es seleccionado por un comité independiente, el Content Selection & Advisory Board (CSAB), y se actualiza de forma constante. Proporciona metadatos estandarizados como resúmenes, palabras clave, afiliaciones y referencias citadas, además de perfiles detallados de autores e instituciones, generados mediante algoritmos y revisión manual. También ofrece herramientas avanzadas para investigar, como la monitorización de tendencias, alertas y análisis de redes de citas. Gracias a su cobertura, limpieza de datos y herramientas, Scopus es ampliamente utilizada para evaluaciones de investigación, ciencia de políticas, rankings institucionales, análisis de redes de colaboración y estudios de tendencias.



Funcionalidades

- Cobertura amplia y multidisciplinaria: ciencia, tecnología, medicina, ciencias sociales y humanidades.
- Tipos de documentos: revistas, conferencias, libros, preprints, patentes.

Métricas integradas:

- CiteScore (journals)
- SJR (Scimago Journal Rank)
- SNIP (impacto normalizado por campo)

Indicadores de citas por artículo, h-index, Field-weighted Citation Impact, además de PlumX Metrics para uso e impacto digital agrupando datos en cinco categorías principales:

- Uso: Descargas, vistas, clics en documentos.
- Capturas: Guardado en gestores de referencias (Mendeley), marcadores, favoritos.
- Menciones: Apariciones en blogs, Wikipedia, noticias.
- Redes sociales: Tweets, posts en Facebook, LinkedIn.
- Citas: Métricas tradicionales en artículos y patentes.

Ventajas

- Alta cobertura y datos multidisciplinarios.
- Base normalizada y confiable por expertos, con procesos de calidad y limpieza y curación de datos.
- Perfiles robustos de autores e instituciones, útiles para análisis de redes y tendencias.
- Métricas ricas y variadas para medir impacto en diferentes niveles.

Limitaciones

- Sesgo hacia revistas en inglés y campos STEM, cobertura menos fuerte en humanidades y ciencias sociales.
- Es de una editorial (Elsevier) lo que puede presentar sesgos a las revistas de esta casa.
- Requiere suscripción institucional.
- Potenciales errores en métricas y dependencia del proceso de selección de la CSAB.

Dimensions Plus

Dimensions Plus es la edición comercial y avanzada de la base de datos Dimensions, desarrollada por Digital Science desde 2018. Nació para ofrecer no solo un índice de publicaciones y citas, sino una plataforma integral que conecta proyectos financiados, resultados académicos, patentes, ensayos clínicos y documentos de políticas públicas, brindando una visión holística del ciclo de investigación. Dimensions Plus incluye funciones extendidas como mayor capacidad de exportación



de datos en formatos CSV o XLSX, herramientas para mapas bibliométricos (como integración con VOSviewer y CiteSpace), acceso a Altmetric a nivel de referencia individual, y a textos completos, tanto de acceso abierto como vía suscripción institucional. Contiene más de 159 millones de publicaciones, 170 millones de patentes, 7,9 millones de proyectos financiados, 938 000 ensayos clínicos.

Funcionalidades

- Visualización y análisis: mapas, gráficos, aplicaciones preconfiguradas e integración de IA para resumen de tendencias
- Exportación avanzada: datos descargables en formatos estructurados para análisis interno o con herramientas externas
- APIs y BigQuery: acceso programático y análisis a gran escala mediante Google BigQuery.
- Integración institucional: Single Sign-On, sincronización con repositorios, ORCID, ReadCube y acceso a textos suscritos.

Ventajas

- Cobertura multidimensional y riqueza de datos vinculados entre financiaciones, publicaciones y patentes.
- Cobertura superior a la de Scopus y WoS en términos de número de revistas y diversidad regional.
- Capacidad de exportación y mapeo robustos, esenciales para estudios bibliométricos avanzados.
- Herramientas de IA que facilitan análisis de tendencias, resúmenes automáticos y descubrimiento de conexiones.
- Integración fluida con flujos de trabajo de investigación y gestión institucional.

Desventajas

- Acceso restringido: requiere suscripción institucional o licencia Plus/Analytics, lo cual implica un costo elevado.
- Curva de aprendizaje: la riqueza de funciones hace que se necesite entrenamiento para aprovechar al máximo la plataforma.
- Cobertura aún limitada en algunas disciplinas en ciencias sociales y en campos emergentes de la ciencia básica.



- Costos de infraestructura: el análisis a gran escala vía BigQuery e integración institucional pueden requerir recursos adicionales.

Bases de datos de acceso abierto

OpenAlex

El 31 de diciembre de 2021 Microsoft cerró su plataforma Microsoft Academic Graph (MAG). OpenAlex surgió como una propuesta de OurResearch —una organización sin ánimo de lucro dedicada a herramientas de ciencia abierta— para crear una base bibliográfica completamente abierta y universal, tomando la posta del MAG. OpenAlex se lanzó en enero de 2022, combinando datos del MAG e incorporando una API gratuita. A marzo de 2024 ya contenía datos sobre más de 209 millones de obras, 13 millones de autores, 109 mil instituciones y 65 mil conceptos de Wikidata, con más de mil millones de conexiones generadas. OpenAlex agrega, normaliza y enriquece metadatos de múltiples fuentes: MAG, Crossref, ORCID, ROR, Unpaywall, DOAJ, PubMed, PMC, ISSN International Centre, arXiv, Zenodo, etc. Su código es open-source, y los datos se publican bajo una licencia CC0, lo que permite uso, redistribución y reutilización sin restricciones. Más de 250 millones de registros de obras (artículos, libros, preprints, tesis, datasets, etc.) y cobertura de 256 disciplinas, con especial atención a trabajos no anglófonos y del Sur Global.

Funcionalidades

Incluye relaciones interconectadas entre:

- Works: publicaciones académicas
- Authors: perfiles vinculados a ORCID (Open Researcher and Contributor ID) es un identificador único y persistente para investigadores que permite distinguir de manera inequívoca a cada persona en el ecosistema académico y científico.
- Institutions: alineados con ROR (Research Organization Registry) que es un identificador global único para instituciones de investigación.
- Sources: revistas, repositorios.
- Topics/Concepts: más de 65 000 conceptos de Wikidata clasificados jerárquicamente

Ventajas

- Cobertura global y multidisciplinaria



- Incluye artículos, libros, preprints, datasets y más, con especial atención a la diversidad geográfica y lingüística.
- Integración con estándares abiertos
- Compatible con ORCID, ROR, DOI, Unpaywall (una herramienta y base de datos que permite acceder legalmente a versiones en acceso abierto de artículos científicos), lo que facilita la interoperabilidad.
- Transparencia y licencia CC0 con datos reutilizables sin restricciones, fomentando la ciencia abierta.
- Actualización frecuente, comunidad activa y soporte institucional

Desventajas

- Calidad variable de metadatos. Al depender de múltiples fuentes, puede haber inconsistencias o registros incompletos.
- Menos herramientas analíticas integradas. No ofrece dashboards avanzados como Scopus o Web of Science; requiere desarrollo externo.
- Cobertura desigual en ciertas áreas. Algunas revistas o disciplinas pueden estar menos representadas.
- Dependencia de identificadores externos. Si un DOI o ORCID no está bien registrado, la información puede faltar.
- Curva de aprendizaje para uso avanzado. Para aprovechar la API y snapshots se necesita conocimiento técnico.
- Sin métricas propietarias.
- No incluye indicadores como CiteScore o Journal Impact Factor, lo que limita comparaciones tradicionales.

Lens

Lens.org, también conocido como The Lens, es una plataforma gratuita y abierta que integra bases de datos de patentes y de literatura académica. Sus orígenes se remontan a 1998 con el proyecto Patent Lens, liderado por Cambia, para mejorar la transparencia del sistema global de patentes. En 2017 se amplió con la incorporación de búsqueda académica, integrando miles de millones de registros provenientes de Microsoft Academic, PubMed, Crossref y posteriormente OpenAlex y Unpaywall, creando una única interfaz pública que permite analizar la interrelación entre investigación científica y desarrollos tecnológicos. Actualmente alberga más de 272 millones de trabajos académicos, 155 millones de patentes y 495 millones de secuencias biológicas, facilitando una visión completa del flujo de conocimiento entre academia e innovación.

Funcionalidades



- Búsqueda avanzada de patentes (más de 95-100 jurisdicciones) con operadores booleanos, clasificadores y filtros.
- Búsqueda y análisis de literatura académica, incluyendo enlaces con patentes citantes.
- Visualización de redes y análisis de influencia entre artículos y patentes (PatCite, In4M).
- Herramienta PatSeq para explorar secuencias biológicas contenidas en patentes.
- Creación y gestión de colecciones/carpetas para agrupar, anotar y compartir patentes y literatura.
- Perfiles Lens, que integran obras académicas y patentes asociadas a ORCID.
- Generación de reportes dinámicos y reutilizables, combinando consultas y colecciones con análisis.
- APIs y descargas masivas, tanto para datos de patentes como de literatura académica y secuencias.

Ventajas

- Integración de datos académicos y de patentes.
- Cobertura extremadamente amplia: centenares de millones de registros y secuencias.
- Herramientas avanzadas de análisis como redes de influencia, secuencias biológicas, perfiles y reportes estructurados.
- API robusta y posibilidad de descarga masiva para integraciones personalizadas o análisis bibliométricos.
- Transparencia total: los datos son rastreables, con metadatos que preservan su origen y modificaciones.

Desventajas

- Curva de aprendizaje: para explotar APIs, análisis de redes y PatSeq se requieren conocimientos técnicos.
- Limitaciones en visualización: aunque ofrece visualizaciones básicas, no dispone de paneles analíticos tan pulidos como herramientas comerciales (ej. Derwent, Clarivate).
- Variabilidad en calidad de metadatos: al integrarse de múltiples fuentes, algunos registros pueden contener errores o estar incompletos.
- Dependencia de fuentes externas: si una base (como PubMed o Crossref) no actualiza correctamente, puede haber lag o ausencias de datos.
- Sin métricas propietarias estandarizadas, como factor de impacto o CiteScore, limitando comparaciones con sistemas tradicionales.

Crossref

Crossref es una infraestructura digital sin ánimo de lucro fundada en 1999 por un grupo de editores académicos para implementar identificadores DOI persistentes y así facilitar el enlace entre referencias en publicaciones científicas. Oficialmente incorporada en enero de 2000 como



Publishers International Linking Association (PILA), su sistema fue lanzado en junio del mismo año. Hoy agrupa a unos 23 000 miembros de más de 160 países, incluyendo editoriales, universidades, bibliotecas, repositorios y financiadores. Maneja metadata de más de 150 millones de registros, propiciando más de 2 000 millones de consultas mensuales y cerca de 1 100 millones de resoluciones de DOI al mes.

Funcionalidades

- Registro y asignación de DOIs a artículos, libros, capítulos, datos, preprints, conferencias, etc., mediante metadatos estandarizados.
- API REST abierta y servicio de búsqueda para consultar metadatos disponibles públicamente.
- Metadata Plus: versión avanzada con mayor acceso a APIs (OAI-PMH, REST) y formatos ampliados.
- Cited-by: herramienta que permite descubrir cuántas veces ha sido citado un objeto por otros miembros.
- Crossmark: informa sobre actualizaciones, correcciones o retractaciones vinculadas a un DOI.
- Similarity Check: sistema antiplagio basado en Crossref y Turnitin, con acceso a artículos y web.
- Register of Funders: base de datos de financiadores, útil para rastrear investigaciones patrocinadas y reportar transparencia.
- Event Data: recoge interacciones online (blogs, redes) sobre contenidos, proporcionando contexto y uso.

Ventajas

- Infraestructura abierta y gratuita hasta cierto nivel, fomentando acceso amplio a metadata académica.
- Cobertura global y diversidad disciplinaria, con metadatos de revistas, libros, conferencias, datos y más desde 150 M de registros.
- Interoperabilidad y persistencia gracias al uso de DOI para referencias confiables y enlazables.
- Servicios complementarios (Cited-by, Crossmark, Similarity Check) que agregan control de calidad, seguimiento de impacto y seguridad mediante plagio.
- Alta escalabilidad, con miles de miembros y miles de millones de interacciones

Desventajas

- Metadatos incompletos en muchas publicaciones: cerca de un tercio carecen de referencias; solo dos tercios incluyen resúmenes u ORCID.
- Sin acceso al texto completo, solo enlaces; depende de repositorios externos o editoriales.



- Servicios avanzados de pago (Metadata Plus, Similarity Check) pueden resultar costosos para instituciones pequeñas.
- Orientado a metadata, no a análisis ni visualización avanzada; requiere integrar herramientas externas.

TERCER ENTREGABLE

Plan de trabajo para la implementación.

La creación de capacidades para la cienciometría requiere la realización de ejercicios prácticos a partir de talleres, en los que se pueda adquirir familiaridad con las bases de datos y las herramientas de visualización, así como con la elaboración de resultados en mapas que marquen la ruta de las trayectorias científico-tecnológicas para facilitar la toma de decisiones.

Se sugiere que el INM, en el marco de la siguiente propuesta, pueda diseñar e implementar un sistema de gestión del conocimiento cuyo eje central sean los estudios cienciométricos de inteligencia tecnológica. En lo que compete a este entregable, a continuación, se detallan los pasos para la implementación del entregable 1 que consiste en la metodología de análisis cienciométrico.

Enero

1. Asignación de un funcionario como gestor de conocimiento.
2. Exposición de los componentes esenciales de un documento científico por tipología y de una patente, identificando los metadatos a ser procesados en los estudios cienciométricos por parte de Avanciencia.
3. Realización de un taller aplicado con la metodología por parte de Avanciencia para la apropiación del gestor de conocimiento y los líderes de los grupos y laboratorios del INM.

Febrero

4. Presentación de la metodología a los investigadores del INM por parte del gestor de conocimiento para la realización de pruebas.
5. Realización de talleres de búsquedas en bases de datos por parte del gestor de conocimiento, quien convocará a los investigadores para el entrenamiento en la

formulación de consultas. Se propone priorizar las bases de datos de acceso abierto si no se tiene acceso a las bases de pago por suscripción.

6. Realización de un taller para la definición del conjunto de palabras clave a ser usadas en las consultas de las bases de datos. Este taller se puede realizar para la metrología en general como campo de estudio y hacer un zoom en metrología cuántica como campo emergente de desmaterialización y universalización democrática de la medida.
7. Realización de un taller para la realización de consultas en bases de datos de documentos científicos y otro taller para las consultas de patentes, con el fin de descargar los archivos en el formato indicado para el VosViewer
8. El gestor de conocimiento realizará talleres para la construcción de mapas de co-ocurrencia de palabras en VosViewer con los archivos descargados para dejar disponibles los mapas de redes de palabras, de líneas de tiempo y mapas de calor.

Marzo

9. Talleres para la elaboración de mapas de trayectorias y rutas científico-tecnológicas. El gestor de conocimiento realizará ejercicios para combinar ambos mapas en una propuesta gráfica que responda a una pregunta específica.
10. El gestor de conocimiento realizará talleres para los grupos y laboratorios del INM que culminen en un ejercicio de inteligencia tecnológica.