



GesBIB



Impacto de la publicación científica CSIC

GesBIB: Impacto de la publicación científica CSIC

Publicaciones: afiliaciones, introducción

Información de las publicaciones



Información extraída de afiliaciones

Cómo se procesan afiliaciones en las fuentes

Cómo se procesan afiliaciones en GesBIB

Curado de datos en GesBIB

Árbol de entidades normalizadas

Información contenida en las afiliaciones



La afiliación que indican los autores en las publicaciones contiene mucha información valiosa que nos interesa extraer y procesar para incluirla en los distintos filtros, informes, indicadores asociados, etc.:

- Cuántos autores colaboran en la publicación
- A qué organizaciones pertenece cada autor
- En el caso del CSIC, a qué ICUs pertenece cada autor
- En base a la dirección postal de cada afiliación, qué países colaboran y cuántos de ellos son distintos
- Cuántas organizaciones diferentes colaboran, cuantas españolas o de otros países
- Cuantos autores españoles o extranjeros, cuantos autores CSIC o no CSIC

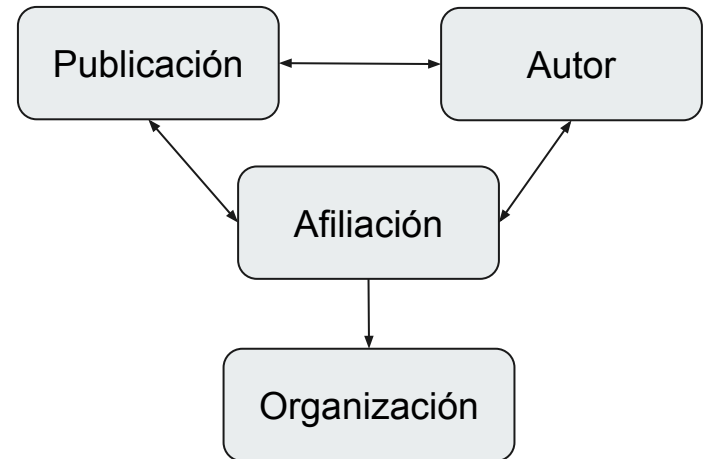
Cómo se procesan afiliaciones en las fuentes

No todas las fuentes de datos analizan las afiliaciones y autores de cada publicación, pero algunas como WoS o Scopus sí realizan un análisis y extracción de esta información. Otras fuentes, la mayoría, se limitan a extraer y normalizar de forma correcta los autores.

Para la extracción de afiliaciones, es necesario manejar tres tipos de datos que se relacionan entre sí:

- La publicación
- El autor
- La afiliación

Además, hay que tener en cuenta que cada afiliación puede hacer referencia a una o más Organizaciones.



Cómo se procesan afiliaciones en las fuentes



Debido a que no existe un estándar de normalización de estos datos ni un identificador persistente común a todas las fuentes, cada fuente de datos realiza su propia normalización y asigna sus propios identificadores:

- Autores: Researcher-ID en WoS, ScopusId en Scopus
 - ORCID es un identificador muy extendido y compartido entre fuentes, pero no todos los autores tienen uno.
- Organizaciones: Enhanced organization en WOS, Perfil institucional en Scopus
 - RoR y otras iniciativas intentan implantar un modelo similar a ORCID para organismos.
- Publicación: WOS ID y eID en Scopus
 - DOI es un estándar de facto con amplia adopción, pero no todas las publicaciones tienen uno.

Cómo se procesan afiliaciones en GesBIB



GesBIB gestiona información de distintas fuentes y la identifica de forma única, por lo que almacenamos para cada tipo de información sus identificadores en las fuentes:

- Autores: Researcher-IDs, ScopusIds, ORCIDs, ID de usuario CSIC.
- Organizaciones: Enhanced organization de WOS, Perfil institucional de Scopus, RoR e ID de ICU CSIC, Estrategias de búsqueda de cada ICU en WOS y Scopus.
- Publicación: WOS ID, eID, DOI, Handle de Digital.CSIC.

A menudo estos identificadores sufren cambios en sus fuentes de origen: se crean nuevos, se eliminan duplicados o erróneos.

Para poder mantener estos datos actualizados, GesBIB incorpora procesos que nos permiten sincronizar los datos, mediante APIs y plugins de conexión a las fuentes

Curado de datos en GesBIB



A partir de la información de afiliaciones proporcionada por cada fuente para una publicación, GesBIB realiza un proceso de análisis y extracción de datos de organizaciones:

1. En base a las estrategias de búsqueda de cada fuente, se analizan los ICUs CSIC que en teoría se encuentran referenciados en esa publicación
2. Después, para cada afiliación encontrada en la publicación, se analizan las organizaciones normalizadas que según la fuente corresponden a ese texto.
3. En caso de coincidencia, se asignan a la afiliación las organizaciones normalizadas
4. En caso de discrepancia, se inicia un proceso de revisión manual por personal de GesBIB
 - a. Si se trata de un error en la fuente, se notifica para su resolución.

Sobre los datos de la fuente se aplican las correcciones necesarias para que en GesBIB figure la información de forma correcta hasta que el proceso de rectificación en la fuente finaliza.

Árbol de Organizaciones

Cuando detectamos una organización normalizada en una afiliación, se establece una relación entre ambas. El análisis de todas las publicaciones CSIC da como resultado un árbol de organizaciones normalizadas.

Este es el ejemplo de una afiliación con texto “CSIC | IIQ”. Esta afiliación se relaciona con 3 organizaciones padre



Árbol de Organizaciones



El árbol de organizaciones tiene estas características:

- Cada nodo del árbol corresponde a una organización, del que podemos almacenar estos datos:
 - Nombre
 - Dirección, Ciudad, País
 - Página web
 - Vinculación CSIC: Si se trata de un instituto CSIC, instituto mixto, unidad asociada, departamento, etc.
- En el mismo árbol se encuentran organizaciones existentes en distintas fuentes de datos.
- Cada organización se etiqueta con las fuentes de datos a las que hace referencia
- Las relaciones entre nodos se establecen a partir de la información de afiliaciones o se han creado de forma manual
- Cada organización puede tener varios padres y varios hijos.

Árbol de Organizaciones



Para cada organización de este árbol guardamos la referencia a todas las publicaciones donde aparece, por lo que podemos, a partir de un conjunto de publicaciones, recrear el árbol de entidades que aparecen en ellas:

- ▼ [CSIC] Centro de Ciencias Humanas y Sociales (CCHS) (293/1)
 - ▼ [CSIC] Instituto de Historia (IH) (121/9)
 - [IH] Grupo de Arqueobiología (2/0)
 - [CSIC] Instituto de Lengua Literatura y Antropología (ILLA) (25/1)
- ▼ [CSIC] Instituto de Economía, Geografía y Demografía (IEGD) (31/0)
 - [IEGD] Laboratorio de Espectro-Radiometría y Teledetección Ambiental (SpecLab) (3/0)
- ▼ [CSIC] Instituto de Políticas y Bienes Públicos (IPP) (63/9)
 - [UA IPP-UGR] Grupo SCImago (4/0)
 - [CSIC] Instituto de Filosofía (IFS) (48/6)
 - [CSIC] Instituto de Lenguas y Culturas del Mediterráneo y Oriente Próximo (ILC) (21/0)
 - [CSIC] Instituto de Estudios Documentales sobre Ciencia y Tecnología (IEDCYT) (6/0)

Árbol de Organizaciones

Dentro de este árbol, algunos datos concretos se propagan desde los nodos superiores hacia los inferiores. Este es el caso de la información de pertenencia al CSIC.

