

CONFERENCIA INTERNACIONAL

BIREDIAL-ISTEC

18-20 de octubre de 2023

MONTEVIDEO • URUGUAY



Modelo dimensional para la medición de la producción académica

de Albuquerque, Pablo; Villarreal, Gonzalo Luján; De Giusti, Marisa Raquel

PREBI-SEDICI - UNLP

CESGI, Comisión de Investigaciones Científicas de la Provincia
de Buenos Aires

20 de Octubre 2023



Esta obra está bajo una [Licencia Creative Commons](https://creativecommons.org/licenses/by-nc-sa/4.0/)
Atribución-NoComercial-CompartirIgual 4.0 Internacional

Analizar la producción académica de una institución a lo largo del tiempo

¿Para qué?

Volumen de producción

Tipos de publicaciones

Investigadores

Redes de colaboración

Áreas de investigación

Tendencias temporales

etc

Desafíos para integrar la producción académica

- Diversidad de fuentes
- Calidad de los datos
- Desambiguación de autores
- Detección de duplicados
- Variaciones en instituciones y filiaciones
- Escalabilidad

Buscamos un modelo de datos que brinde una solución genérica

Modelo dimensional

Conjunto de técnicas utilizadas en el diseño de almacenes de datos (Data Warehouse)

Diseñado para

Proporcionar datos comprensibles para los usuarios *empresariales*.

Ofrecer un rendimiento de consulta rápido.

¿Cómo se estructuran los datos?

Hechos

Los eventos a analizar

Dimensiones

Atributos que dan contexto

Modelo dimensional - ¿Cómo diseñarlo?

Proceso de Diseño Dimensional de Cuatro Pasos (Kimball & Ross, 2013)

1. Seleccionar el proceso de negocio
2. Declarar la granularidad
3. Identificar las dimensiones
4. Identificar los hechos

¿Que proceso de negocio se modela?

Publicación de recursos académicos
generados por una institución científico-académica

Hechos

Los eventos a analizar

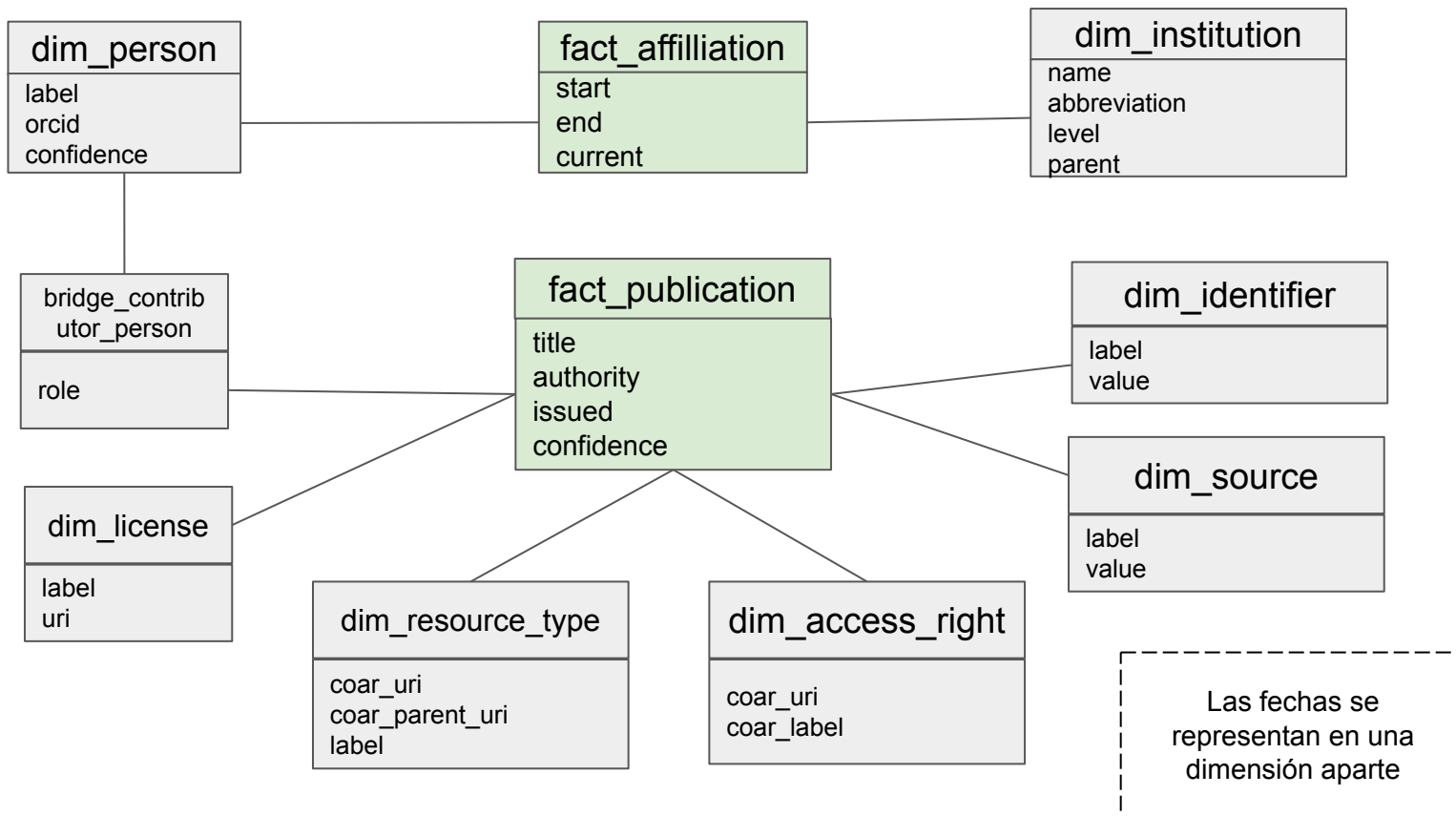
- Publicación de un recurso
- Filiación

Dimensiones

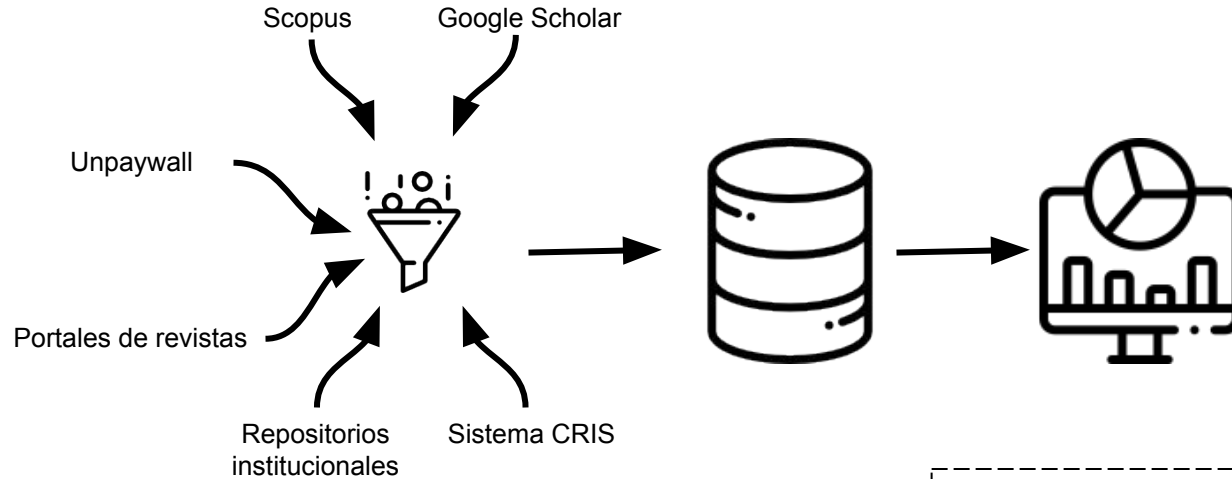
Atributos que dan contexto

- Institución
- Colaborador
- Disponibilidad o nivel acceso
- Licencia de uso
- Tipo de recurso

Modelo propuesto (simplificado)



Integración de datos - Procesos ETL



La normalización se realiza a través de tablas auxiliares que incorporan identificadores persistentes, vocabularios estandarizados, etc

Implementación en dos instituciones distintas

UNLP

- Universidad nacional pública
- Fundada en 1897
- Posee 17 facultades, 137 carreras de grado, 167 de posgrado
- + 150 centros I+D
- Integrada por alumnos, docentes, investigadores, personal administrativo

CIC

- Organismo científico provincial
- Fundado en 1956
- + 100 Centros I+D
- Integrada por becarios, pasantes, investigadores, personal de apoyo



Fuentes usadas

- Repositorios institucionales
- Vocabularios COAR
- RENAORG (SICYTAR)



Stack

- Almacenamiento
- Transformaciones
- Carga en BD
- Versionado de código
- Versionado de BD



Flyway



Trabajo a futuro

1. Implementar procesos ETL para incorporar más fuentes
2. Generación y monitoreo de métricas de rendimiento de procesos ETL
3. Incorporar hechos y dimensiones para poder medir el impacto de la producción
4. Implementación de modelo en entorno de testing y producción

CONFERENCIA INTERNACIONAL

BIREDIAL-ISTEC

18-20 de octubre de 2023

MONTEVIDEO • URUGUAY



¡Muchas gracias!



Esta obra está bajo una [Licencia Creative Commons](https://creativecommons.org/licenses/by-nc-sa/4.0/)
Atribución-NoComercial-CompartirIgual 4.0 Internacional

