

Sistemas para integrar bases de datos distribuidas, heterogéneas y autónomas

María Mora, INBio.
Costa Rica
mmora@inbio.ac.cr

Temas

- Introducción
- Estrategias de integración
- Sistemas consistentes de múltiples bases de datos
- Taxonomía de un sistema multi-base de datos
- ¿Qué es una red federada de bases de datos?
- Arquitecturas existentes de redes federadas
- Redes indexadas (GBIF, Red de especies y especímenes de IABIN)

Introducción

- Existe mucha información sobre biodiversidad en bases de datos administradas por instituciones alrededor del mundo.
- El integrar la información tendría un gran impacto en los procesos de investigación científica, educación y toma de decisiones en muchas disciplinas.
- Organizaciones con diferentes enfoques están realizando esfuerzos de integración de esta información, por ejemplo GBIF y IABIN.

Estrategias de integración

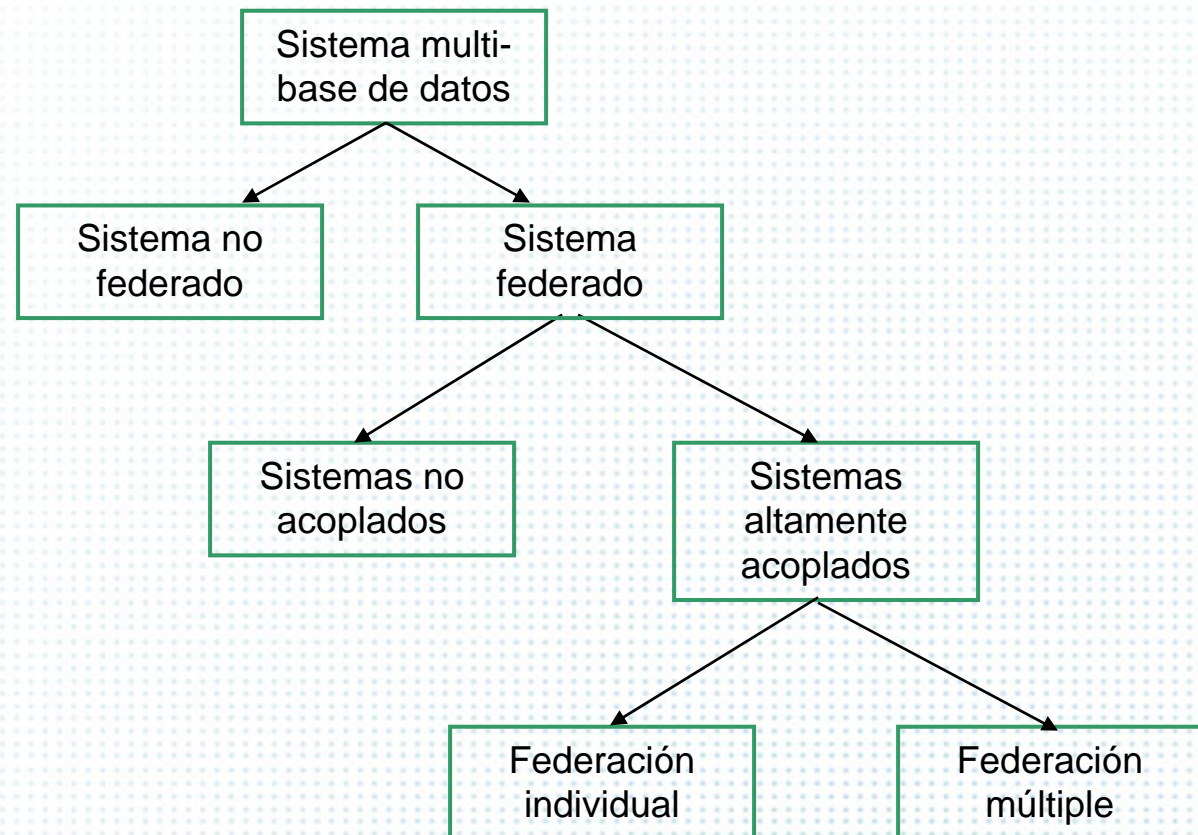
- **“Data warehouse”**
 - Datos centralizados.
 - Los dueños pueden mantener cierto control sobre los datos.
 - Un único nodo central (pueden existir replicación).
 - Más fácil de administrar.
- **Bases de datos distribuidas**
 - Se deben realizar cambios en los sistemas existentes para que realicen operaciones globales.
 - No existe autonomía en el manejo de las bases de datos.
 - Hay que reemplazar la base de datos centralizada por una distribuida.
- **Redes federadas**
 - Los datos permanecen en manos de los dueños.
 - Acuerdo común en un modelo de datos para el intercambio.
 - Las consultas pueden tomar mucho tiempo a medida que aumenta el número de nodos de la red.
- **Redes indexadas**
 - Los datos permanecen en manos de los dueños.
 - Acuerdo común en un modelo de datos para el intercambio.
 - Los elementos más utilizados en las consultas son almacenados en un índice.

Sistemas consistentes de múltiples bases de datos

Pueden ser caracterizados por tres dimensiones:

- **Distribución:** Datos distribuidos, las bases de datos existen antes de formarse la federación.
- **Heterogeneidad:** Debido al uso de diferentes motores de base de datos, a diferentes formas de modelar los datos, a diferencias en la semántica de los datos.
- **Autonomía:** En el diseño, en la comunicación, en la ejecución y en la asociación.

Taxonomía de un sistema multi-base de datos



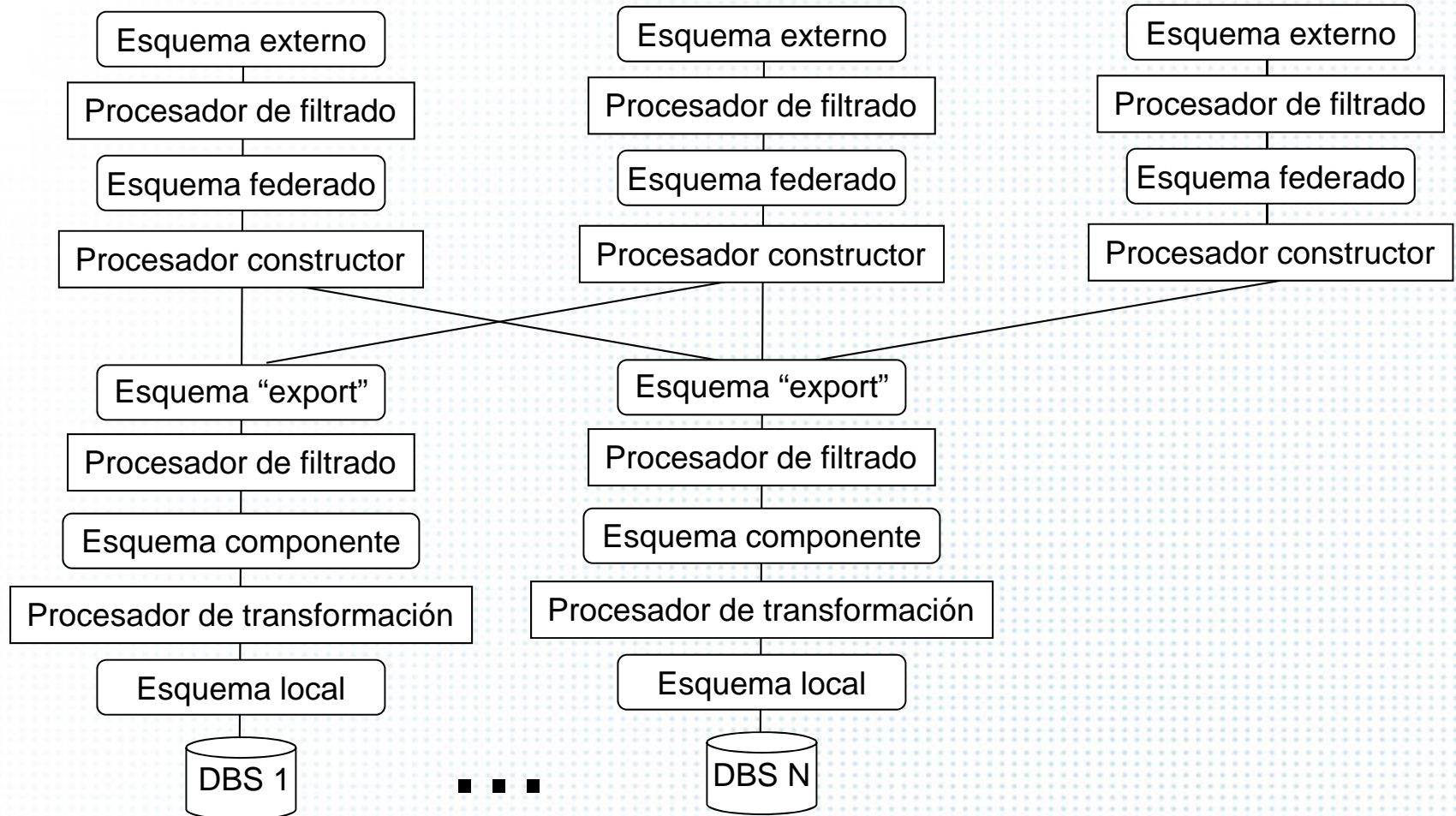
Clasificación basada en la dimensión de autonomía.

Sheth and Larson: **Federated database systems for managing distributed, heterogeneous and autonomous databases**. ACM Computing Survey, Vol. 22, N° 3, Set. 1990.

¿Qué es un sistema federado de bases de datos?

- Es una colección de sistemas de base (DBS) de datos autónomos pero que colaboran con la federación.
- Los DBS componentes continúan con su operación normal.
- La cantidad de integración depende de las necesidades de los usuarios de la federación y de la disposición de los administradores de compartir información.
- No hay un control centralizado en la arquitectura de los sistemas federados.
- Es un estado intermedio entre la no integración y la integración total.
- Permiten realizar operaciones locales y globales.
- En algunos casos la federación puede estar formada de DBS heterogéneas.

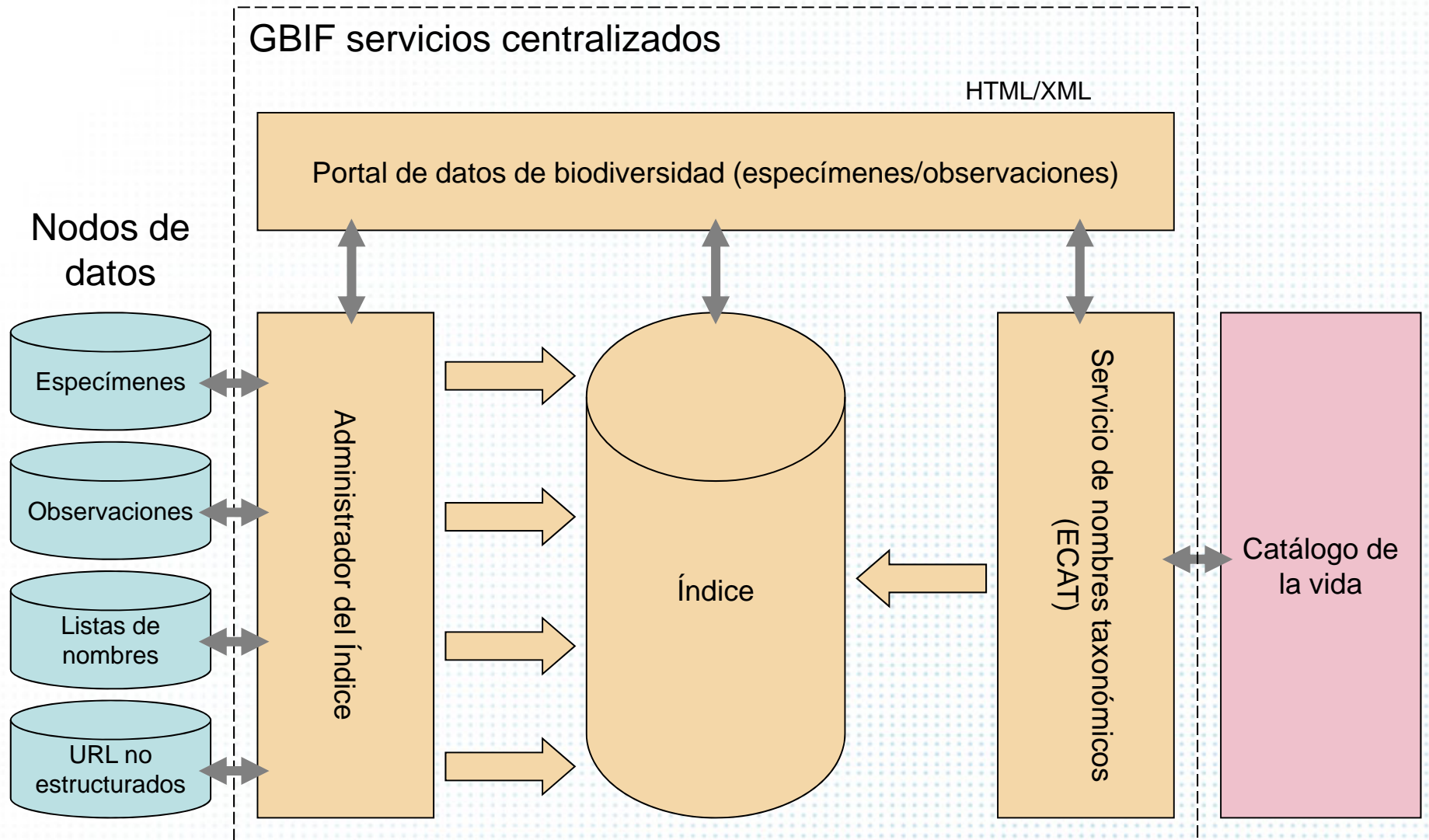
Arquitecturas existentes de redes federadas



Redes indexadas (GBIF, Red de especies y especímenes de IABIN)

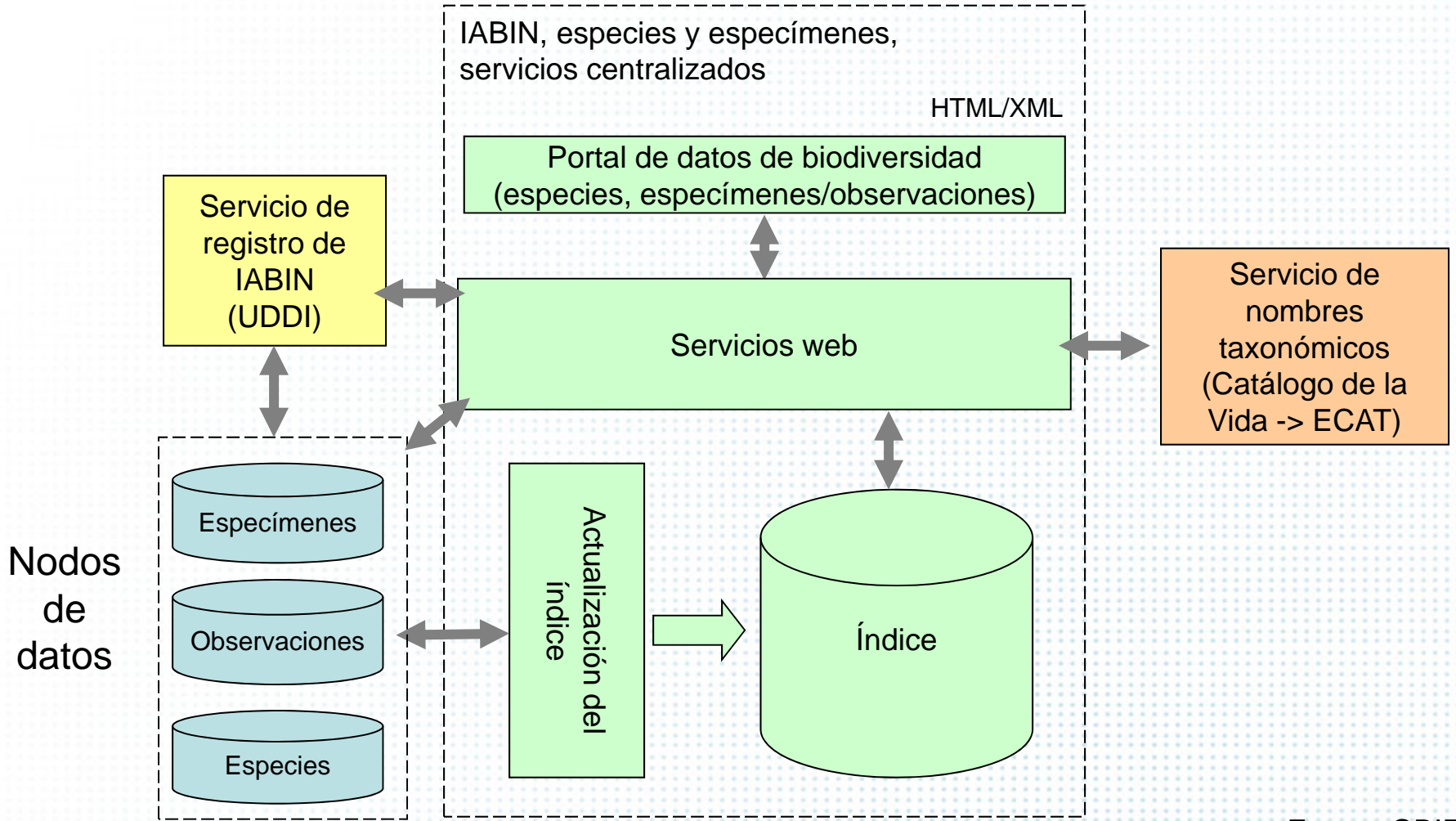
- Es un sistema de bases de datos federadas. Utiliza una base de datos local, que actúa como caché o índice, para hacer las búsquedas de forma más eficientes.
- GBIF utiliza un índice centralizado que se consulta para descubrir en cuáles proveedores de datos se encuentra la información solicitada por el usuario.

Redes indexadas (GBIF)



Fuente GBIF

Redes indexadas (Red de Especies y Especímenes de IABIN)



Fuente GBIF

Referencias

- Red Interamericana de Información sobre Biodiversidad (IABIN). **Plan de implementación del proyecto**. USA, 16 de agosto de 2004. 144pp.
<http://www.iabin.net/spanish/index.shtml>
- Hobern, Donald. **GBIF Biodiversity Data Architecture**. Agosto 2003.
<http://www.gbif.org>
- Sheth and Larson. **Federated database systems for managing distributed, heterogeneous and autonomous databases**. ACM Computing Survey, Vol. 22, N° 3, Set. 1990.

Preguntas

¡Muchas Gracias!