



- Datos
  - Datos registrados, que tienen un significado implícito, sobre fenómenos del mundo real
  - Por lo general representan valores (números, caracteres) variables (cualitativa o cuantitativa)
  - Se utiliza para transmitir, almacenar y deducir información

- metaDatos

- Facilitan la comprensión de las relaciones y la utilidad de la información de datos

título

Central do Brasil  
Walter Sales  
Fernanda Montenegro

director

actriz principal

**dato**

Valor sin un significado explícito

**información**

Significado asociado o deducido de un conjunto de datos y asociaciones entre ellos

**conocimiento**

Información adicional extraída de los datos o de expertos del dominio de aplicación



**conocimiento**

**información**

**dato**



dato

8.848

*La altura del monte Everest es 8.848m*

información

conocimiento

*Monte Everest es la montaña más alta de la superficie de la Tierra y mide 8.848m*

Proceso de **información**  
para generar **conocimiento**

**conocimiento**

**información**

**dato**

Procesar los **datos** para  
generar **información**

Es interés común de las organizaciones públicas y privadas

# ¿Cuáles son las fuentes de datos?





- Datos
  - Los datos pueden ser almacenados en
    - Fuentes de datos privadas - disponibles en las organizaciones
    - Fuentes de datos públicas - disponibles en la Web

- Datos
  - Pueden tener formatos diferentes
    - Datos estructurados (por ejemplo, bases de datos relacionales)
    - Datos semi-estructurados (por ejemplo, documentos XML)
    - Datos no estructurados (por ejemplo, documentos de texto)

- Datos estructurados
  - Los datos de los sistemas transaccionales
  - Los datos almacenados en bases de datos relacionales
  - Tienen una estructura fija y bien definida (esquema de base de datos)
    - Esquema predefinido
    - Todos los datos de acuerdo con el esquema

Tabla de Autores

Tabla de Géneros

| <u>cod</u> | título                 | cod_ autor | ano  | cod_género |
|------------|------------------------|------------|------|------------|
| L01        | Mar Muerto             | autor_1    | 1936 | género_1   |
| L02        | El Camino hacia el Mar | autor_1    | 1938 | género_2   |
| L03        | El mundo de la Paz     | autor_1    | 1951 | género_3   |
| L04        | El Quince              | autor_2    | 1930 | género_1   |
| L05        | Tieta de Agreste       | autor_1    | 1977 | género_1   |

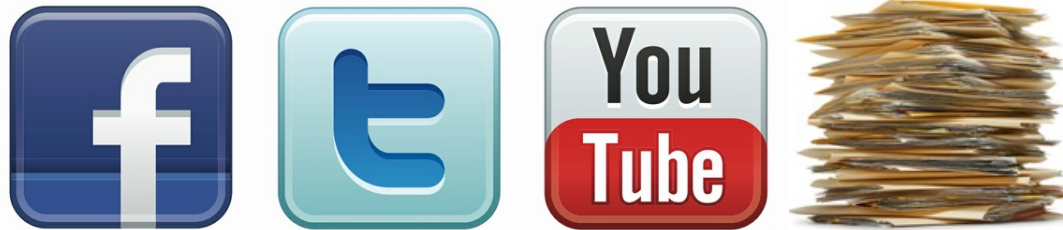
Tabla relacional - Libros

- Datos semi-estructurados
  - Ausencia de una estructura regular, o la estructura puede evolucionar impredecible
  - Los datos pueden estar incompletos
  - Estructura irregular (datos heterogéneos)
  - Los tipos son sólo indicativos
  - La estructura puede ser implícita
  - Los datos en la Web

```
<?xml version="1.0" encoding="utf-8"?>
<livraria>
  <livro id="L01" ano="1936">
    <autor> Jorge Amado </autor>
    <titulo>Mar Morto</titulo>
  </livro>
  <livro id="L04" ano="1930">
    <autor>
      <nome>Rachel</nome >
      <sobrenome>de Queiroz</sobrenome >
    </autor>
    <titulo>O Quinze</titulo>
    <genero> Romance </genero>
  </livro>
</livraria>
```



- **Datos no-estructurados**
  - **Falta** de estructura
  - Los datos que no han sido "tratado" o modelo
  - Los datos almacenados en archivos o documentos

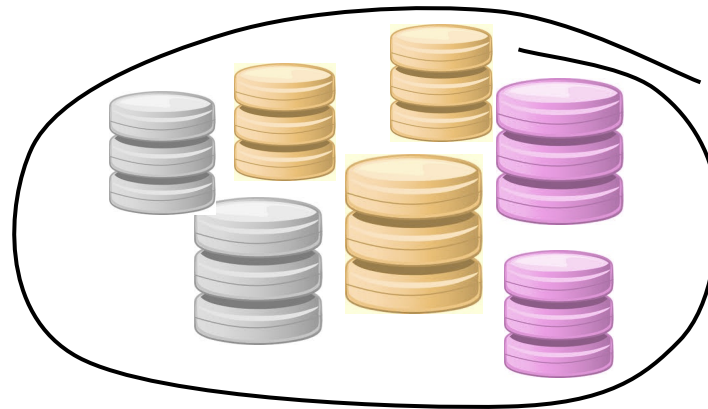




- Integración dos Datos
  - Presentación de una visión uniforme y coherente de los datos
  - Identificación de los datos complementarios y redundantes
  - Resolución de inconsistencias

# ¿Cómo integrar?

¿Cómo ofrecer una visión global de datos distribuidos a través de fuentes de datos **heterogéneas** y **autónomas**?



resumen de los  
datos

## Tipos de heterogeneidad

Diferentes  
estructuras

**Sintáctico**

Modelos  
diferentes

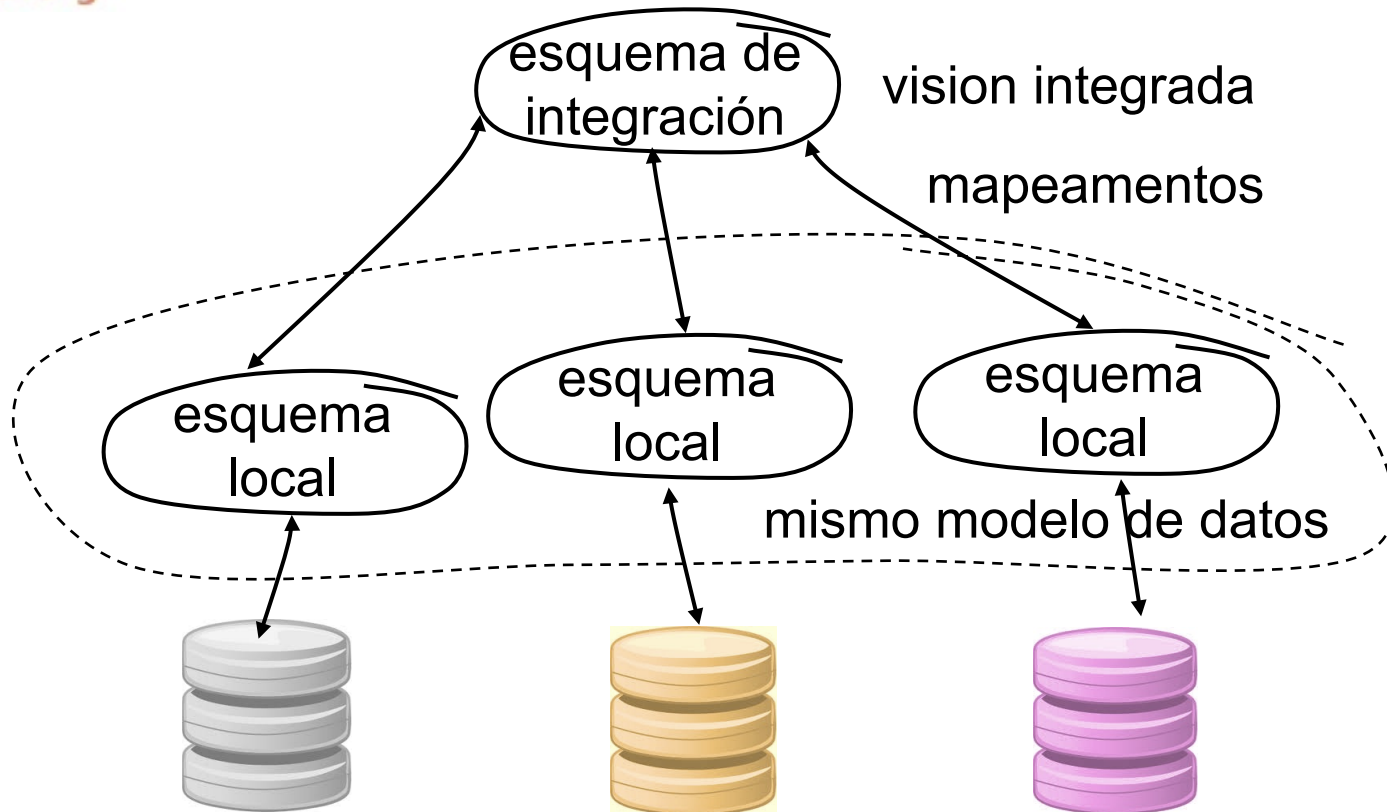
**Estructural**

Nombres  
diferentes

**Terminología**

Significados  
diferentes

**Semántica**

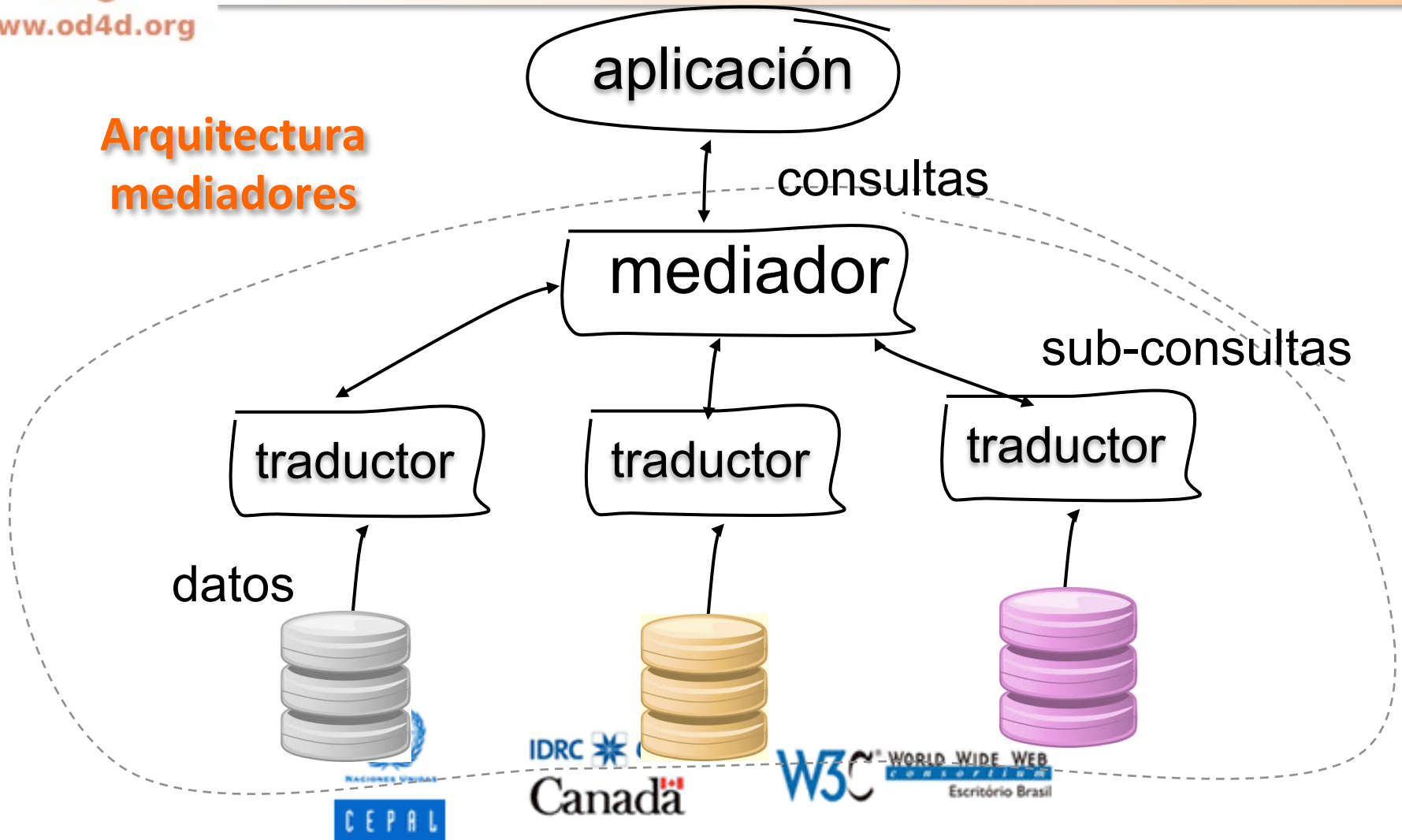




- Arquitecturas de Integración
  - Mediadores
  - Datawarehouse
  - P2P
  - Enfoque pay-as-you-go



## Arquitectura mediadores



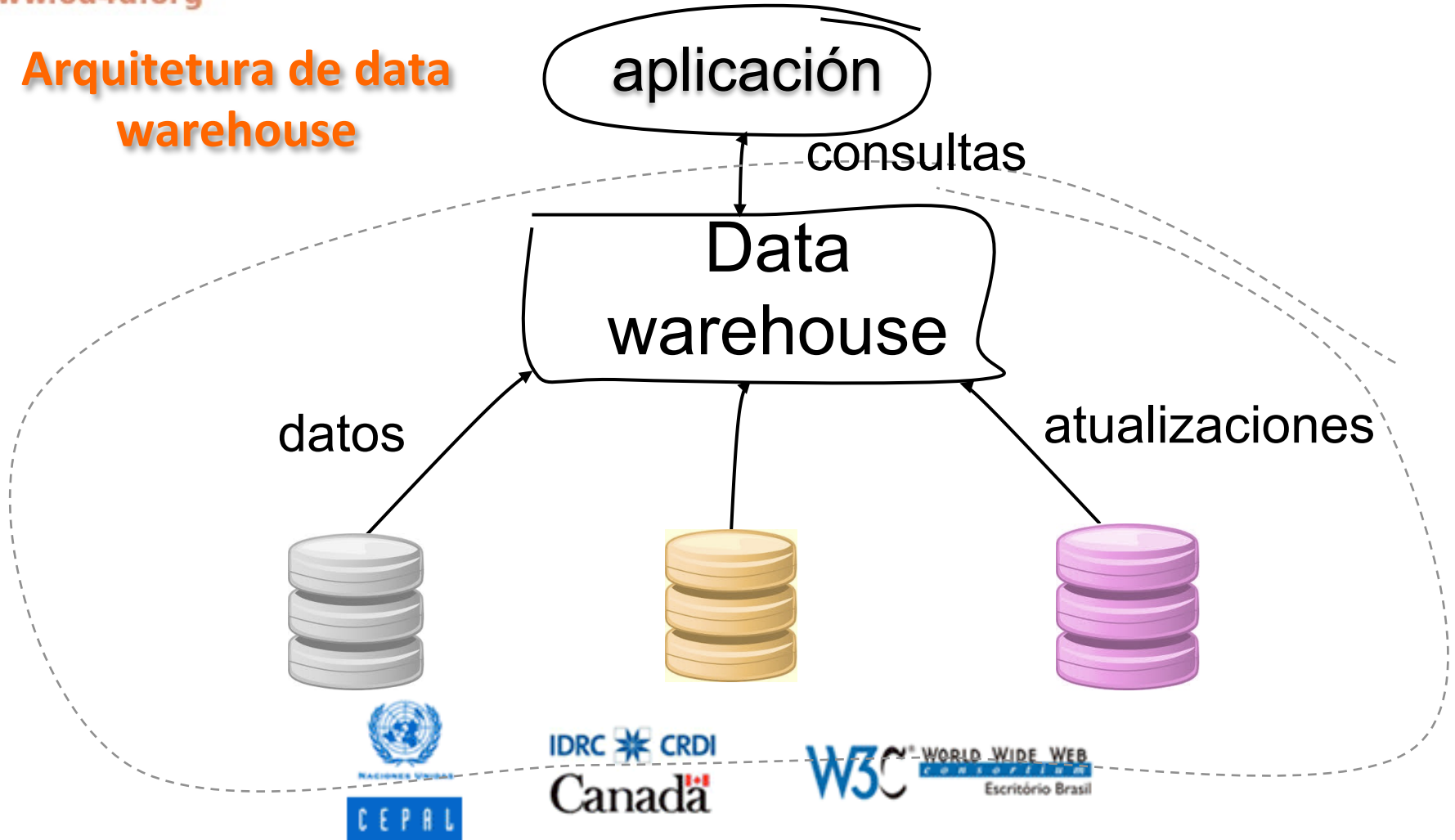


www.od4d.org

# Integración de Datos

Bernadette Lóscio/CIn-UFPE

## Arquitectura de data warehouse



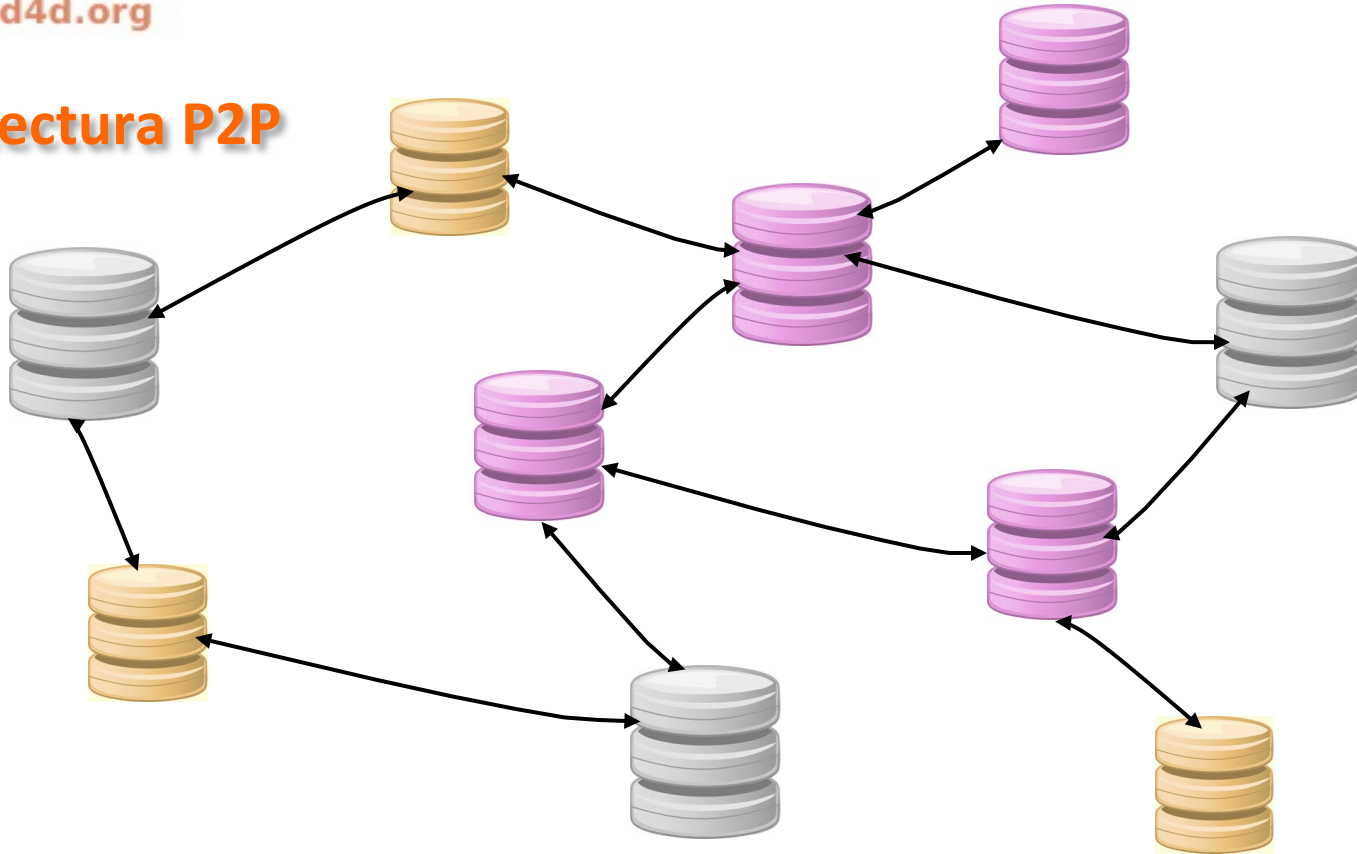


www.od4d.org

# Integración de Datos

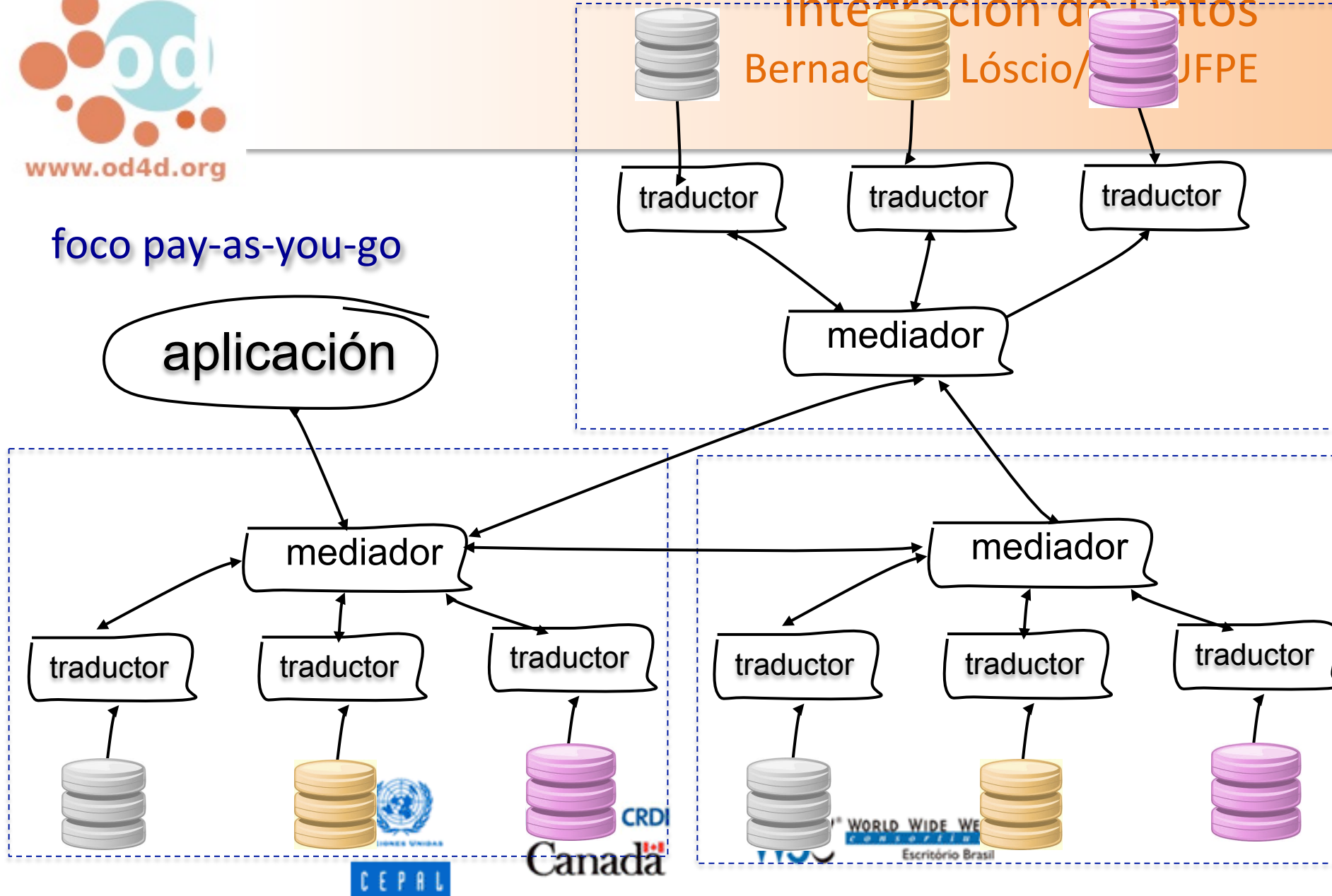
Bernadette Lóscio/CIn-UFPE

## Arquitectura P2P



foco pay-as-you-go

# Integración de Datos Bernardo Lóscio / UFPE



- Arquitecturas de integración
  - La elección de la arquitectura depende de algunos factores:
    - La cantidad de fuentes de datos para ser integrado
    - La frecuencia de actualización fuentes
    - La infraestructura de comunicación
    - ...S



- Arquitecturas de integración
  - La elección de la arquitectura responde a algunas preguntas:
    - ¿La integración de datos se materializará o virtual?
    - ¿Se utiliza un esquema de integración único o múltiples esquemas?
    - ¿Puede definir asignaciones entre las fuentes de datos o simplemente entre las fuentes y el esquema de integración?



- Interoperabilidad

- Una solución de integración de datos debe proporcionar:
  - Interoperabilidad sintáctica: la adopción de un modelo de datos común
  - Interoperabilidad estructural: las asignaciones definen
  - Interoperabilidad semántica: el uso de vocabularios

La interoperabilidad es la capacidad de un sistema (o no) para comunicar de forma homogénea (o tan cerca de ella) con otro sistema (similar o no)

# Web Semántica y Linked Data

- Proporcionan soluciones y tecnologías apropiadas para resolver el problema de la integración de datos
- **Modelo de datos flexible** para la representación de datos en la Web
- **Ontologías** ayudar a resolver el problema de la **heterogeneidad semántica**

# Conclusiones