

Reportes estadísticos para repositorios digitales a partir de múltiples fuentes basado en el stack ELK

de Albuquerque Pablo C. (CIC / UNLP – Argentina)
Villarreal Gonzalo L. (CIC / UNLP – Argentina)
De Giusti Marisa R. (CIC / UNLP – Argentina).



Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional](#)



Introducción

Los Repositorios Institucionales (RI) tienen la necesidad de generar métricas de impacto

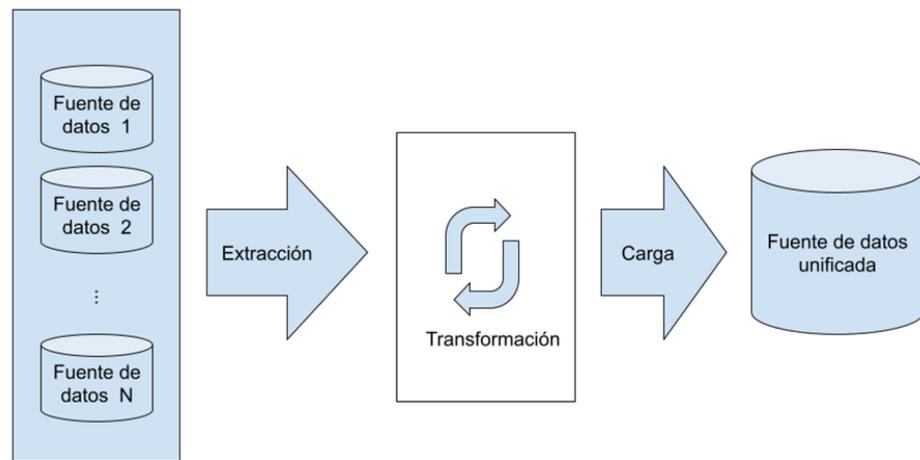
- comprender cómo son utilizados
- asistir a la toma de decisiones
- ofrecer servicios

Existen servicios de Web Analytics (WA)

- Limitaciones para integrar fuentes de datos
- No están diseñados para contemplar la estructura de un repositorio

Objetivos

- **Integrar y unificar** fuentes de datos
- Enriquecer los datos de un RI
- Asistir a la toma de decisiones
- Comprender cómo se utilizan los repositorios



Stack ELK

Elasticsearch

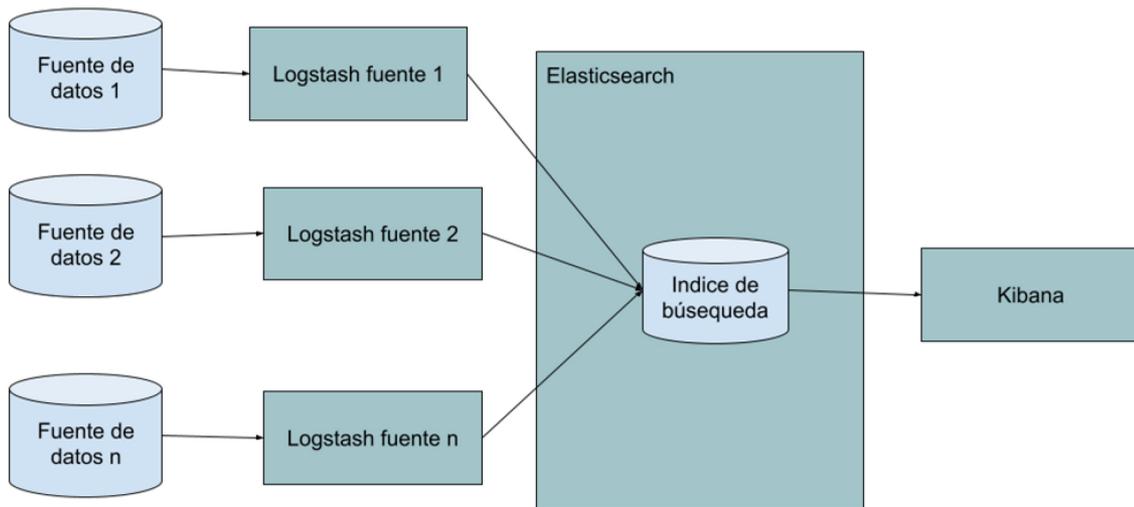
- Motor de búsqueda

Logstash

- Pipeline de procesamiento de datos

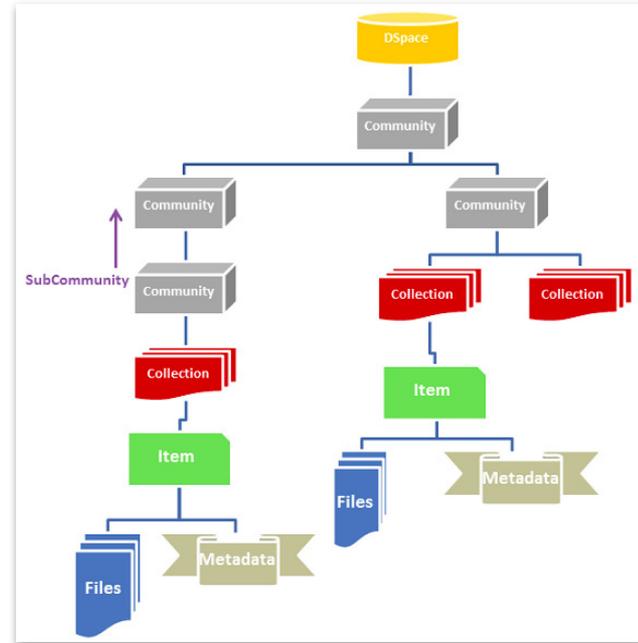
Kibana

- Herramienta de visualización de datos



Fuentes de datos para prototipo

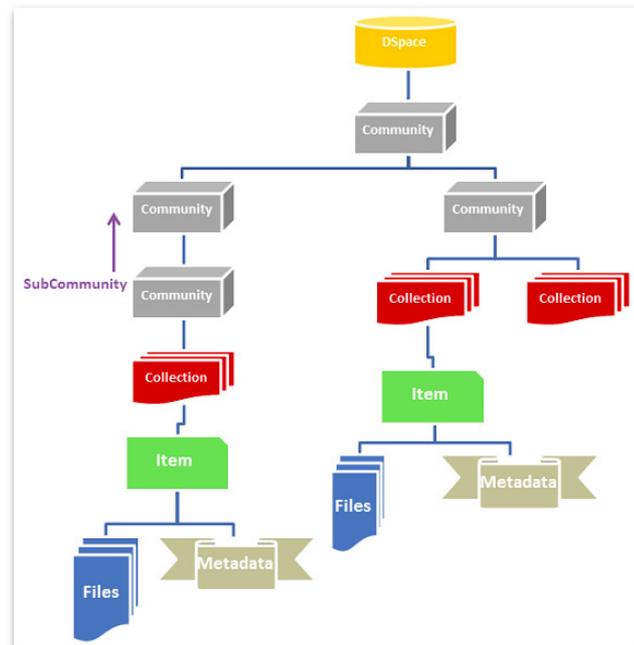
- Base de datos del repositorio CIC-DIGITAL



Functional Overview - DSpace 6.x Documentation - LYRISIS Wiki, 2018

Fuentes de datos para prototipo

- Base de datos del repositorio CIC-DIGITAL



Functional Overview - DSpace 6.x Documentation - LYRISIS Wiki, 2018

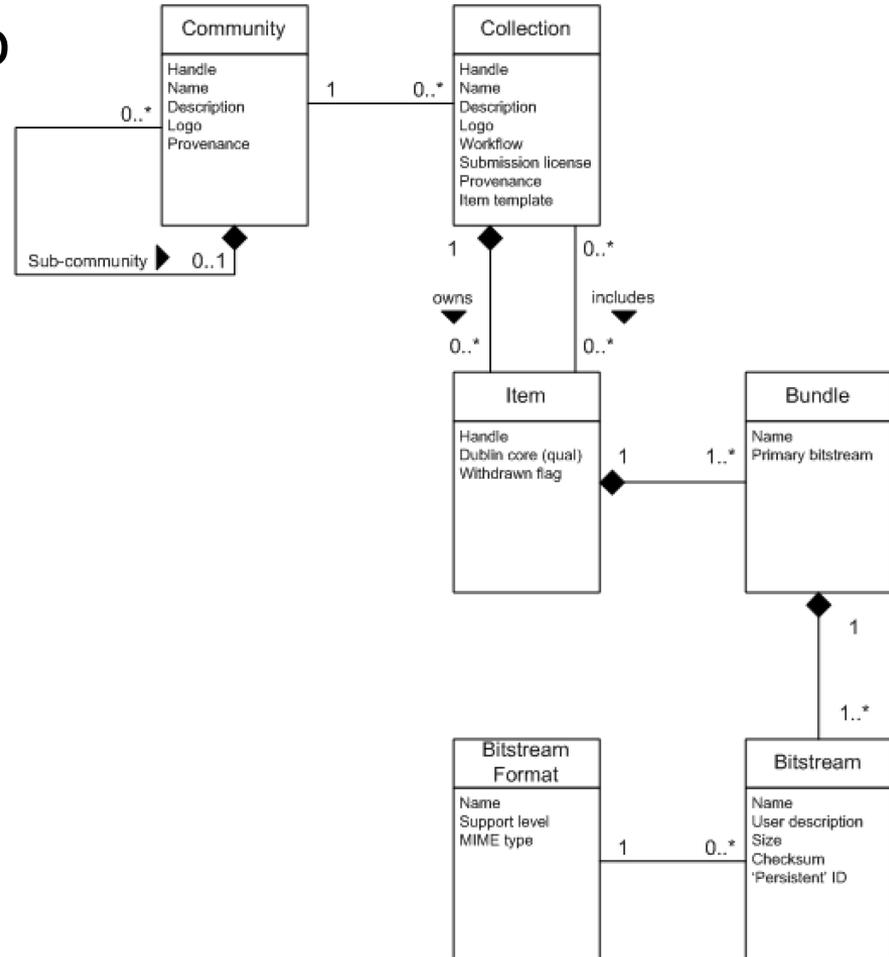
- Log de servidor Apache

```
186.0.176.252 - - [17/Jun/2021:00:00:11 -0300] "GET /bitstream/handle/10915/72076/Documento_completo.pdf?isAllowed=y&sequence=1 HTTP/1.1" 200 7470507 "http://sedici.unlp.edu.ar/handle/10915/72076" "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.101 Safari/537.36"
```

Fuente de datos del repositorio

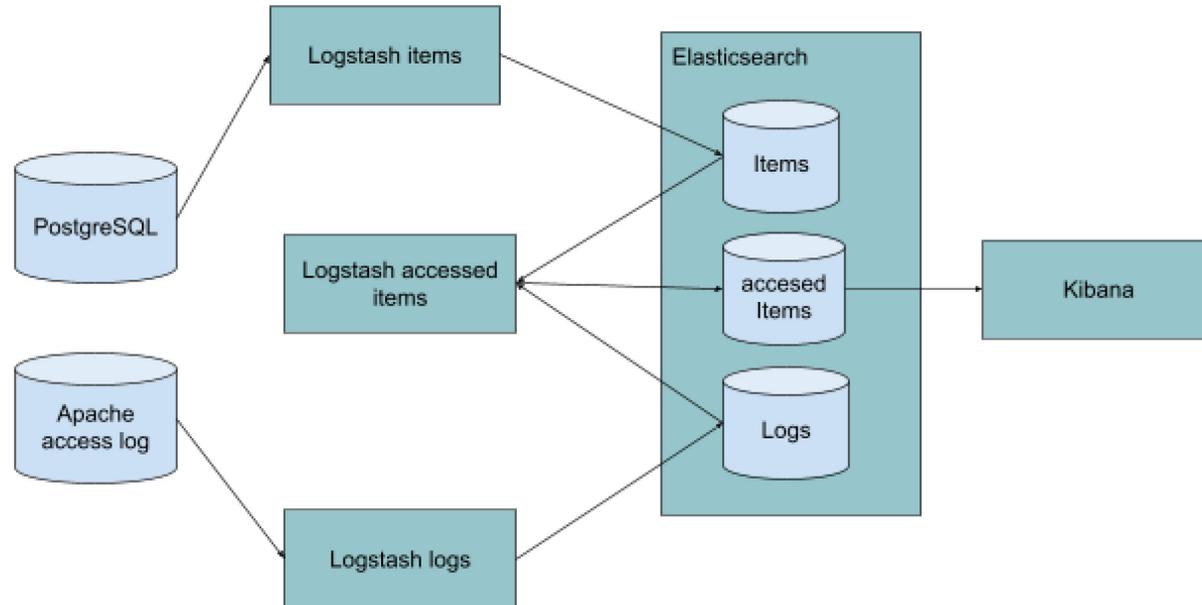
Ítems, comunidades y colecciones

- tiene asociado un Handle
- se asocian a una tabla de metadatos
 - Describen recursos
 - Almacenan información de eventos

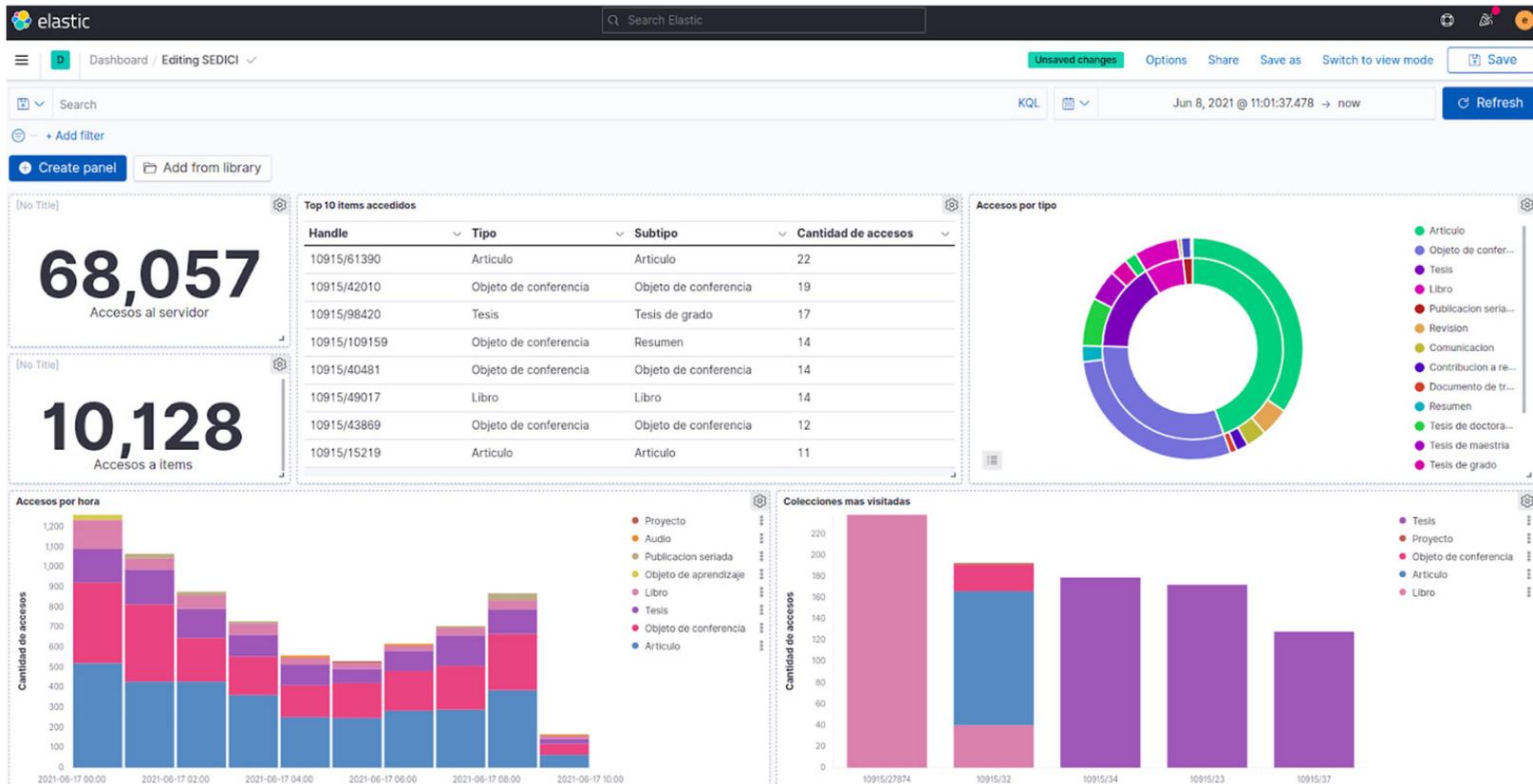


Prototipo ELK

Integrar fuentes de datos a partir del stack ELK



Prototipo ELK



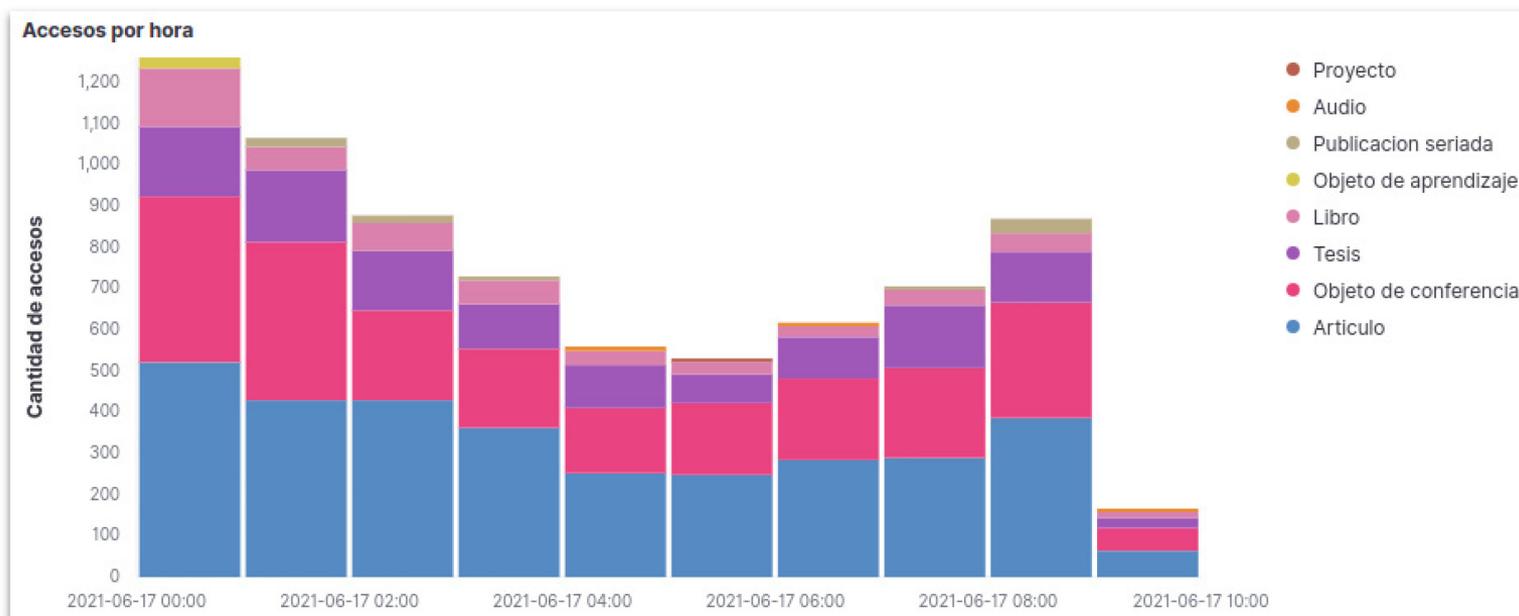
Prototipo ELK

Los 10 ítems más accedidos, junto con el tipo y subtipo de documento

Top 10 ítems accedidos 			
Handle	Tipo	Subtipo	Cantidad de accesos
10915/61390	Articulo	Articulo	22
10915/42010	Objeto de conferencia	Objeto de conferencia	19
10915/98420	Tesis	Tesis de grado	17
10915/109159	Objeto de conferencia	Resumen	14
10915/40481	Objeto de conferencia	Objeto de conferencia	14
10915/49017	Libro	Libro	14
10915/43869	Objeto de conferencia	Objeto de conferencia	12
10915/15219	Articulo	Articulo	11

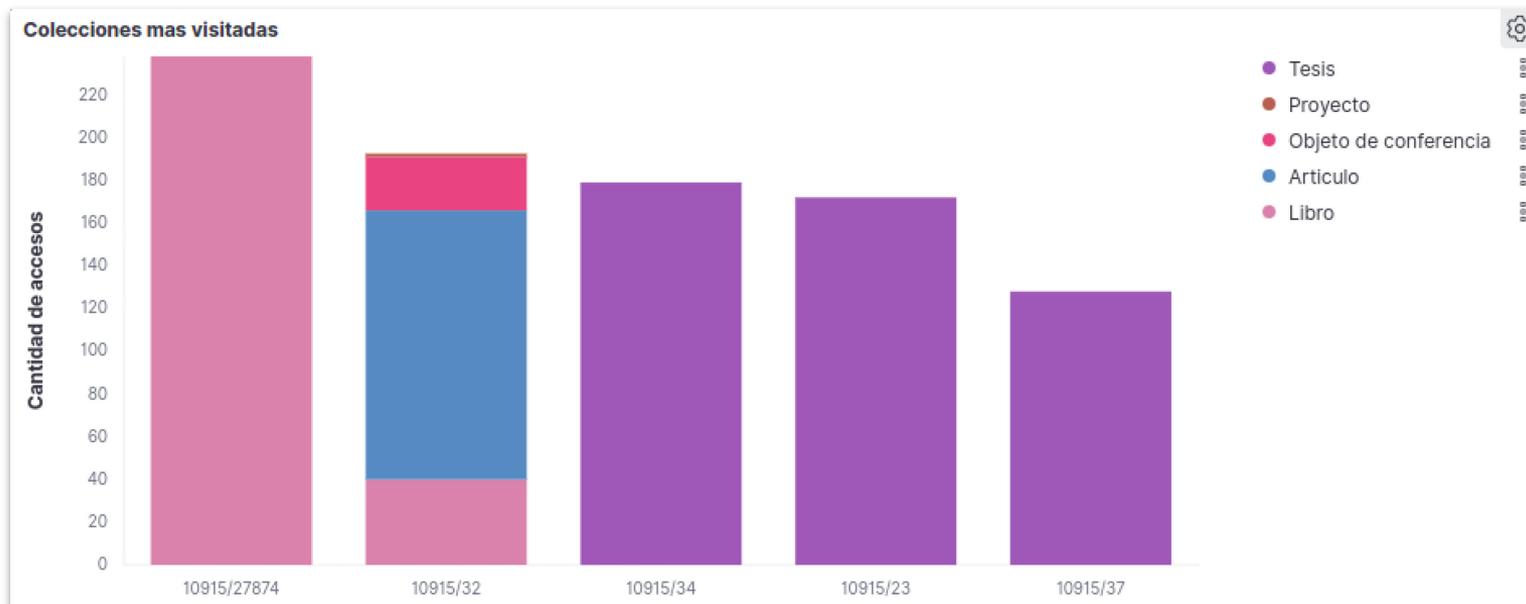
Prototipo ELK

Cantidad de accesos por hora y la distribución de tipos de documentos.



Prototipo ELK

Colecciones más visitadas, y distribución de tipos.



Conclusiones

El stack propuesto por ELK

- Facilita integración de múltiples fuentes de datos
 - Logs de acceso
 - Herramientas de Web Analytics (WA)
 - Bases de datos normalizadas
- Aumenta el control sobre los datos con respecto a otras herramientas de WA
- Permite comparar distintas fuentes de datos
- Permite tener una mejor perspectiva sobre lo que ocurre en un RI
- Implica un esfuerzo para normalizar los datos que otras herramientas ya resuelven
 - Ejemplo: Filtrado de bots

Muchas Gracias!

Pablo C. de Albuquerque - pablo@sedici.unlp.edu.ar
Gonzalo L. Villarreal - gonzalo@prebi.unlp.edu.ar
Marisa R. De Giusti - marisa.degiusti@sedici.unlp.edu.ar



Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional

