

The background of the slide is a monochromatic, blue-tinted image. It depicts a large, multi-masted sailing ship, likely a galleon, navigating through a turbulent sea with white-capped waves. In the lower-left foreground, a shark's dorsal fin is seen breaching the water's surface. The overall atmosphere is dramatic and historical.

Intercambios Oceánicos - Minería de datos de periódicos decimonónicos

Isabel Galina Russell

Miriam Peña

Ernesto Priani

Universidad Nacional Autónoma de México (UNAM)

III Congreso Internacional de la
Asociación Argentina de Humanidades Digitales

7 al 9 de noviembre 2018



Oceanic Exchanges

Tracing Global Information Networks In Historical Newspaper
Repositories, 1840-1914

Intercambios Oceánicos

Trazando redes de información global en repositorios de periódicos históricos,
1840-1914



Intercambios Oceánicos (OcEx)

- Proyecto de investigación internacional
- Humanidades digitales
 - Convocatoria de “Digging into data” – realizar minería de datos en colecciones de Humanidades
- Junio 2017 a mayo 2018 (24 meses)
- 6 países, 11 instituciones

Antecedentes - Viral texts project

- *¿Qué textos se reimprimían y por qué?*
- *¿Qué ideas – literarias, políticas, científicas, económicas, religiosas – destacaban?*
- *Herramientas computacionales lingüísticas*

(<https://viraltxts.org>)

- *Multi-linguismo*
- *Más datos, más colecciones*



Investigadores socios

- **Alemania:** Universität Stuttgart y Staatsbibliothek zu Berlin
- **Estados Unidos:** Northeastern University, North Carolina State University y University of Nebraska-Lincoln
- **Finlandia:** Turun Yliopisto
- **Holanda:** Universiteit Utrecht
- **México:** Universidad Nacional Autónoma de México
- **Reino Unido:** Loughborough University, University College London y University of Edinburgh
- 11 instituciones

Proveedores de datos

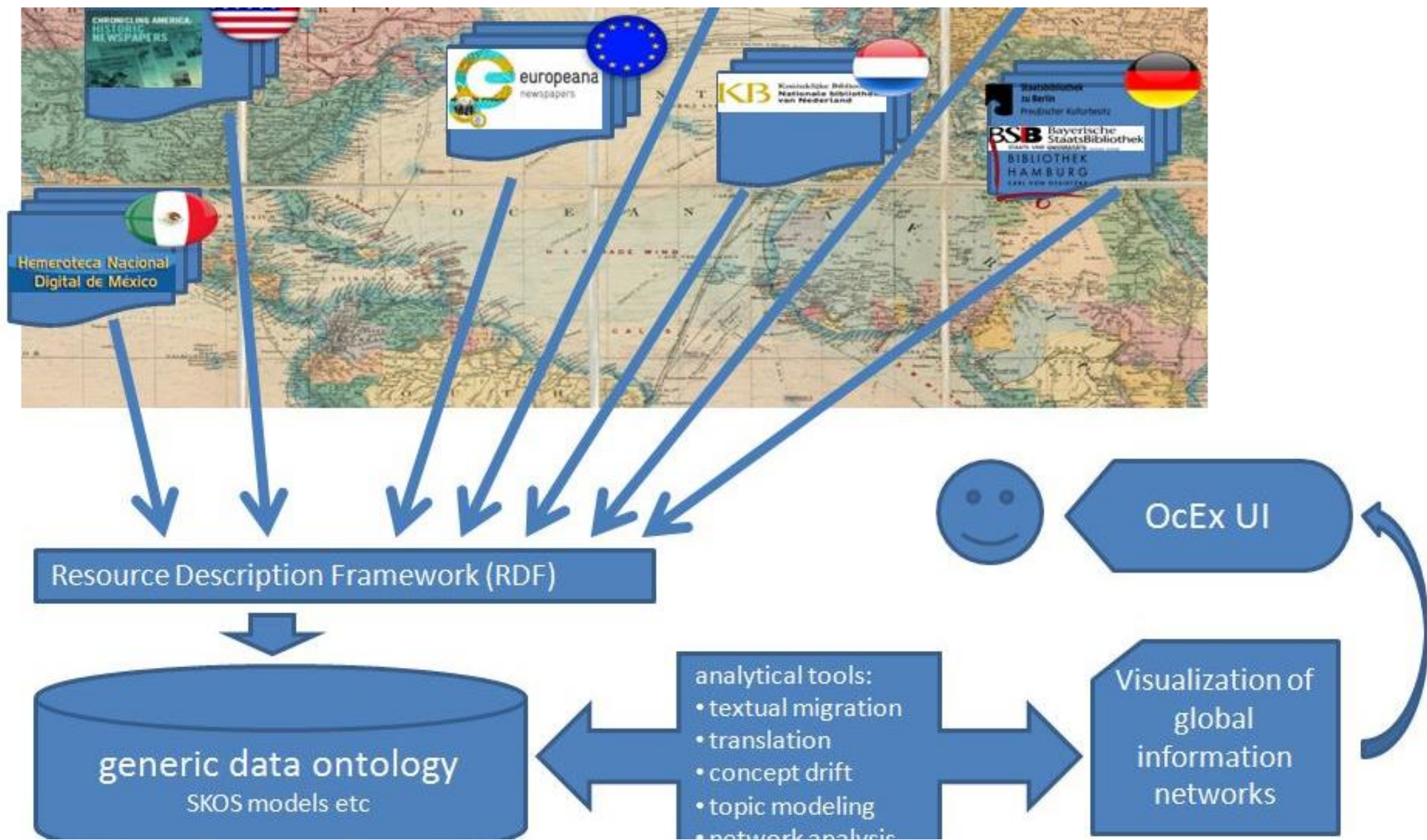
- Bavarian State Library
- Berlin State Library
- British Library
- Cengage
- Hamburg State Library
- National Library of Finland
- Biblioteca Nacional de México
- National Library of the Netherlands

90 millones de páginas digitalizadas de periódicos

Australia's Trove Newspapers, British Newspapers Archive, Chronicling America (U.S.), Europeana Newspapers, Hemeroteca Nacional Digital de México, National Library of Finland, National Library of the Netherlands, the National Library of Wales, y New Zealand's PapersPast

Objetivos

1. Analizar la reutilización de textos a través de estos archivos, utilizando modelos de lenguaje estadístico para detectar grupos de pasajes reimpresos;
2. Realizar modelos temáticos multilingües escalables de inferencia a través de este corpus para ayudar a discernir traducciones, temáticas compartidas y cambios de temáticas dentro y a través de diferentes idiomas, utilizando modelos temáticos plurilingüísticos (jerárquicos);
3. Construir clasificadores del diseño visual de los artículos periodísticos;
4. Crear un corpus de anotaciones sobre grupos de pasajes que registre las relaciones entre las distintas versiones de un texto;
5. Crear una ontología de géneros y formas de textos circulantes para apoyar esa anotación;
6. Realizar análisis de redes de los flujos agregados de información dentro y a través de las naciones, regiones y publicaciones;
7. Llevar a cabo un análisis de la dependencia de estos resultados a la disponibilidad del corpus en diferentes idiomas y niveles de acceso; y
8. Empacar el código, la información anotada (donde sea posible) y las ontologías para lanzamiento público.



Responder las preguntas:

- ¿Qué historias se difunden entre naciones y qué tan rápido?
- ¿Qué textos son traducidos y resuenan a través de distintos idiomas?
- ¿Cómo la migración de información textual facilita la circulación del conocimiento, ideas y conceptos, y cómo estas ideas son traducidas, transformadas y adaptadas conforme se mueven de un lugar del Atlántico al otro?

¿Cómo funciona en la práctica?

WP 1 - Coordinación general de los 6 países

WP 2 - Metadatos

WP 3 - Ontología

WP 4 - Migración textual y cultura viral

WP 5 - Migración conceptual

WP 6 - Resultados/lanzamiento de colecciones públicas.

WP2 y WP3 – Metadatos y ontologías

- Procesos de digitalización
- Calidad del OCR
- Tabla descriptiva metadatos HDNM
- Ontologías
- Historias orales

WP4 – Migración textual y cultural viral

- Creación de mini-corpus
- Eligieron:
 - Krakatoa (erupción del volcán), agosto de 1883
 - Caso Maine (Guerra España/Estados Unidos) , enero 1898
 - Batalla de Puebla Fusilamiento de Maximiliano (Invasión francesa), mayo 1862 y junio 1867
- Otros
 - Lajos Kossuth, Hungría (1802-1894),
 - Bobrikov, Finlandia
 - Cable transatlántico, 1858

Proteus

- Herramienta para la visualización de clusters y reimpressiones

The screenshot displays the Proteus web interface. At the top, there is a search bar containing 'Krakatoa' and a 'Search Clusters' button. The user is logged in as 'epriani' and can click 'LogOut' or a settings icon. On the left side, there are several filter panels: 'Refine by' (with 'Clusters' selected), 'Entity types' (set to 'none'), 'Document labels' (showing 'No labels yet for this document'), and 'Duplicate threshold' (set to 30). The main content area shows three search results for 'Krakatoa':

- Rank 1:** 33 reprints from 1883-08-31 to 1883-09-25 [ci68720321751] (view OCR). Published: 1883-08-31. The snippet includes text about volcanic activity in the straits of Sunda.
- Rank 2:** 7 reprints from 1883-09-01 to 1883-09-08 [ci214749093512] (view OCR). Published: 1883-09-01. The snippet describes the eruption of Mount Krakatoa and the resulting tidal waves.
- Rank 3:** 17 reprints from 1883-08-31 to 1883-09-05 [ci163209362803] (view OCR). Published: 1883-08-31.

Each result includes a 'subject:' field and a 'place:' field with a dropdown menu. The first result's 'place' dropdown is open, showing 'place'. Below each result is a checkbox labeled 'Show matching pages in this book...'.

WP5 – Migración conceptual

Estudiar conceptos a partir de las palabras con las que está asociado.

- La emergencia de EU, la declinación del Imperio Británico.
- Darwinismo, eugenesia, herencia,
- Derechos humanos, democracia, constitucionalista, libertad de expresión
- Tecnología
- Género, sufragistas, derecho al voto
- Migración, xenofobia

Retos y logros

- Organización y logística
 - Prioridades locales y globales
 - Participación internacional
- Incorporar a investigadores de otras áreas
- Recursos humanos
- Vacíos legales
- Empujar las fronteras
- Despertar la curiosidad

Equipo en México

- Ernesto Priani, profesor Filosofía, FFyL-UNAM
- Isabel Galina, investigadora, IIB-UNAM
- Miriam Peña, investigadora, IIB-UNAM
- Adán Lerma, Doctorando en Filosofía de la Ciencia, UNAM
- Rocío Castellanos, Doctorando en Historia, UAM
- Laura Domínguez, Doctora en Historia, UNAM
- Laura López, estudiante licenciatura Filosofía, FFyL-UNAM
- Juan Manuel Rodríguez Martínez, licenciatura Ciencias de la computación, UNAM
- Fabiola Delfín, Maestría en Humanidades Digitales, Claustro de Sor Juana

- HNDM (Depto. de Informática, Biblioteca Nacional)

- Ivan Meza Ruíz, Investigador Asociado IIMAS-UNAM
- Ximena Gutiérrez, Doctora en Ciencia e Ingeniería de la Computación, UNAM
- Jorge García Flores, Investigador, CNRS

¡Muchas gracias!

Intercambios Oceánicos / Oceanic Exchanges

oceanicexchanges.org

@OceanicEx

Zotero Library / Oceanic Exchanges

Github Oceanic Exchanges

Isabel Galina

igalina@unam.mx, [@igalina](#)

Ernesto Priani

ernestopriani@gmail.com, [@epriani](#)

Miriam Peña

miriampenapimentel@gmail.com, [@](#)

III Congreso Internacional de la
Asociación Argentina de Humanidades Digitales

7 al 9 de noviembre 2018