**Oceanic Exchanges: Tracing Global Information Networks In Historical Newspaper Repositories, 1840-1914**

Poster at DHN 2018 conference

Abstract

Presenters: Hannu Salmi, Asko Nivala, Mila Oiva, Otto Latva – Cultural History, University of Turku

*Oceanic Exchanges: Tracing Global Information Networks in Historical Newspaper Repositories, 1840-1914* (OcEx) is a Digging into Data – Transatlantic Platform funded international and interdisciplinary project with a focus on studying spreading of news globally in the nineteenth century newspapers. The project combines digitized newspapers from Europe, US, Mexico, Australia, New Zealand, and the British and Dutch colonies of that time all over the world.

The project examines patterns of information flow, spread of text reuse, and global conceptual changes across national, cultural and linguistic boundaries in the nineteenth century newspapers. The project links the different newspaper corpora, scattered into different national libraries and collections using various kinds of metadata and printed in several languages, into one whole.

The research group members come from Finland, the US, the Netherlands, Germany, Mexico, and UK. OcEx’s participating institutions are Loughborough University, Northeastern University, North Carolina State University, Universität Stuttgart, Universidad Nacional Autónoma de México, University College London, University of Nebraska-Lincoln, University of Turku, and Utrecht University. The project’s 90 million newspaper pages come from Australia's Trove Newspapers, the British Newspapers Archive, Chronicling America (US), Europeana Newspapers, Hemeroteca Nacional Digital de México, National Library of Finland, National Library of the Netherlands (KB), the National Library of Wales, New Zealand’s PapersPast, and a strategic collaboration with Cengage Publishing, one of the leading commercial custodians of digitized newspapers.

**Objectives**

Our team will hone computational tools, some developed in prior research by project partners and novel ones, into a suite of openly available tools, data, and analyses that trace a broad range of language-related phenomena (including text reuse, translational shifts, and discursive changes). Analysing such parameters enables us to characterize “reception cultures,” “dissemination cultures,” and “reference cultures” in terms of asymmetrical flow patterns, or to analyse the relationships between reporting targeted at immigrant communities and their surrounding host countries.

OcEx will leverage existing relationships and agreements between its teams and data providers to connect disparate digital newspaper collections, opening new questions about historical globalism and modeling consortial approaches to transnational newspaper research. OcEx will take up challenging questions of historical information flow, including:

1. Which stories spread between nations and how quickly?

2. Which texts were translated and resonated across languages?

3. How did textual copying (reprinting) operate internationally compared to conceptual copying (idea spread)?

4. How did the migration of texts facilitate the circulation of knowledge, ideas, and concepts, and how were these ideas transformed as they moved from one Atlantic context to another?

5. How did geopolitical realities (e.g. economic integration, technology, migration, geopolitical power) influence the directionality of these transnational exchanges?

6. How does reporting in immigrant and ethnic communities differ from reporting in surrounding host countries?

7. Does the national organization of digitized newspaper archives artificially foreclose globally-oriented research questions and outcomes?

**Methodology**

OcEx will develop a semantic interoperable knowledge structure, or ontology, for expressing thematic and textual connections among historical newspaper archives. Even with standards in place, digitization projects pursue differing approaches that pose challenges to integration or particular levels of analysis. In most, for instance, generic identification of items within newspapers has not been pursued. In order to build an ontology, this project will build on knowledge acquired by participating academic partners, such as the project TimeCapsule at Utrecht University, as well as analytical software that has been tested and used by team members, such as viral text analysis. The members of the Finnish team have already worked on text reuse in Finnish newspapers and journals (the COMHIS project), but OcEx offers the possibility to expand this analysis into the study of transnational and transcontinental flows of information. OcEx does not aim to create a totalizing research infrastructure but rather to expose the conditions by which researchers can work across collections, helping guide similar projects in future seeking to bridge national collections. This ontology will be established through comparative investigations of phenomena illustrating textual links: reprinting and topic dissemination. We have divided the tasks into six work packages:

**WP1: Management**

➢ create an international network of researchers to discuss issues of using and accessing newspaper repository data and combine expertise toward better development and management of such data;

➢ assemble a project advisory board, consisting of representatives of public and private data custodians and other critical stakeholders.

**WP2: Assessment of Data and Metadata**

➢ investigate and develop classifier models of the visual features of newspaper content and genres;

➢ create a corpus of annotations on clusters/passages that records relationships among textual versions.

**WP3: Creating a Networked Ontology for Research**

➢ create an ontology of genres, forms, and elements of texts to support that annotation;

➢ select and develop best practices based on available technology (semantic web standard RDF, linked data, SKOS, XML markup standards such as TEI).

**WP4: Textual Migration and Viral Texts**

➢ analyze text reuse across archives using statistical language models to detect clusters of reprinted passages;

➢ perform analyses of aggregate information flows within and across countries, regions, and publications;

➢ develop adaptive visualization methods for results.

**WP5: Conceptual Migration and Translation Shifts**

➢ perform scalable multilingual topic model inference across corpora to discern translations, shared topics, topic shifts, and concept drift within and across languages, using distributional analysis and (hierarchical) polylingual topic models;

➢ analyze migration and translation of ideas over regional and linguistic borders;

➢ develop adaptive visualization methods for the results.

**WP6: Tools of Delivery/Dissemination**

➢ validation of test results in scholarly contexts/test sessions at academic institutions;

➢  conduct analysis of the sensitivity of results to the availability of corpora in different languages and levels of access;

➢  share findings (data structures/availability/compatibility, user experiences) with institutional partners;

➢  package code, annotated data (where possible), and ontology for public release.