WOMAN POISONED HERSELF AND HER CHILDREN – *HARD NEWS* – **WOMAN** POISONED HERSELF AND HER ...

HE CURES ALL KINDS OF FEMALE WEAKNESS – *ADVERTISEMENT* – **HE** CURES ALL KINDS OF FEMALE ...

HE TRIED TO RAPE SARAH ROBINSON AND JANE KENDRICKS – *SHORT STORY* – **HE** TRIED TO RAPE ...

WHO GETS TO BE REMEMBERED AND HISTORICIZED BY WAYS OF – DIGITAL – RECORD CREATION?

For many German-Americans in the long nineteenth century, German-language newspapers were the primary source of both information and entertainment. So far, research on the German- American press has predominantly focused on the male editors, writers, or advertisers and their influence – and success stories – on U.S. politics and economy. These HIStories have entered schoolbooks and popular culture in the U.S. and Germany alike.

Beyond BUDWEISER

Budweiser's a friend of mine
The Mighty Workers of America
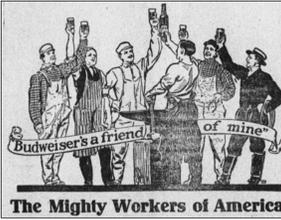
"No archive is innocent"
(Yale, 2015, p. 332).

Do digital newspaper archives offer ways to find HIDDEN/OTHER stories? Do they provide NEW ways to access and assess German-American newspaper literature to write HERstories of German Migration to the United States?
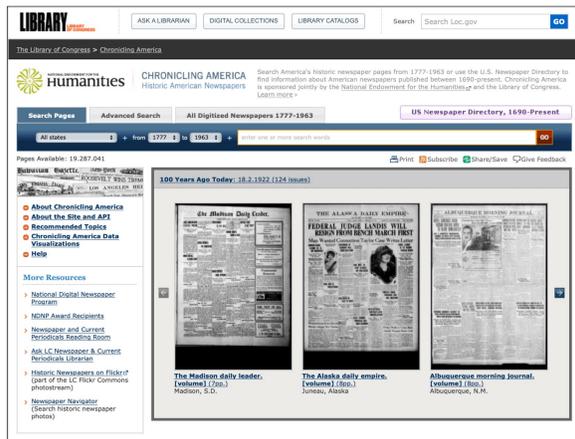

"... the traditional method of simple keyword searching is here to stay for the foreseeable future ..." (Hausdewell et al., 2020, p. 155).

"Data. A lot of data. Big Data? But few methods to access and assess it."
– Interfaces of digital newspaper archives such as *Chronicling America* offer no innovative (re-)search methods.

Researchers **Jana Keck** and **André Blessing** are creating a digital archive of popular newspaper literature using a human-in-the-loop approach.
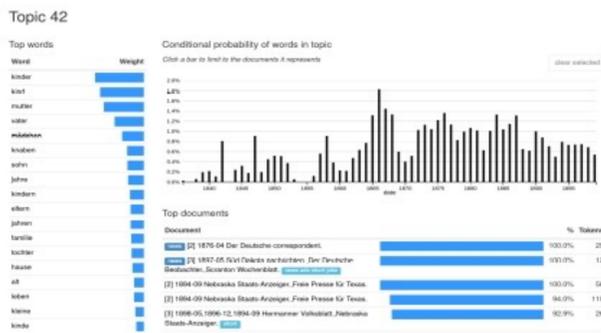
human in the loop corpus creation
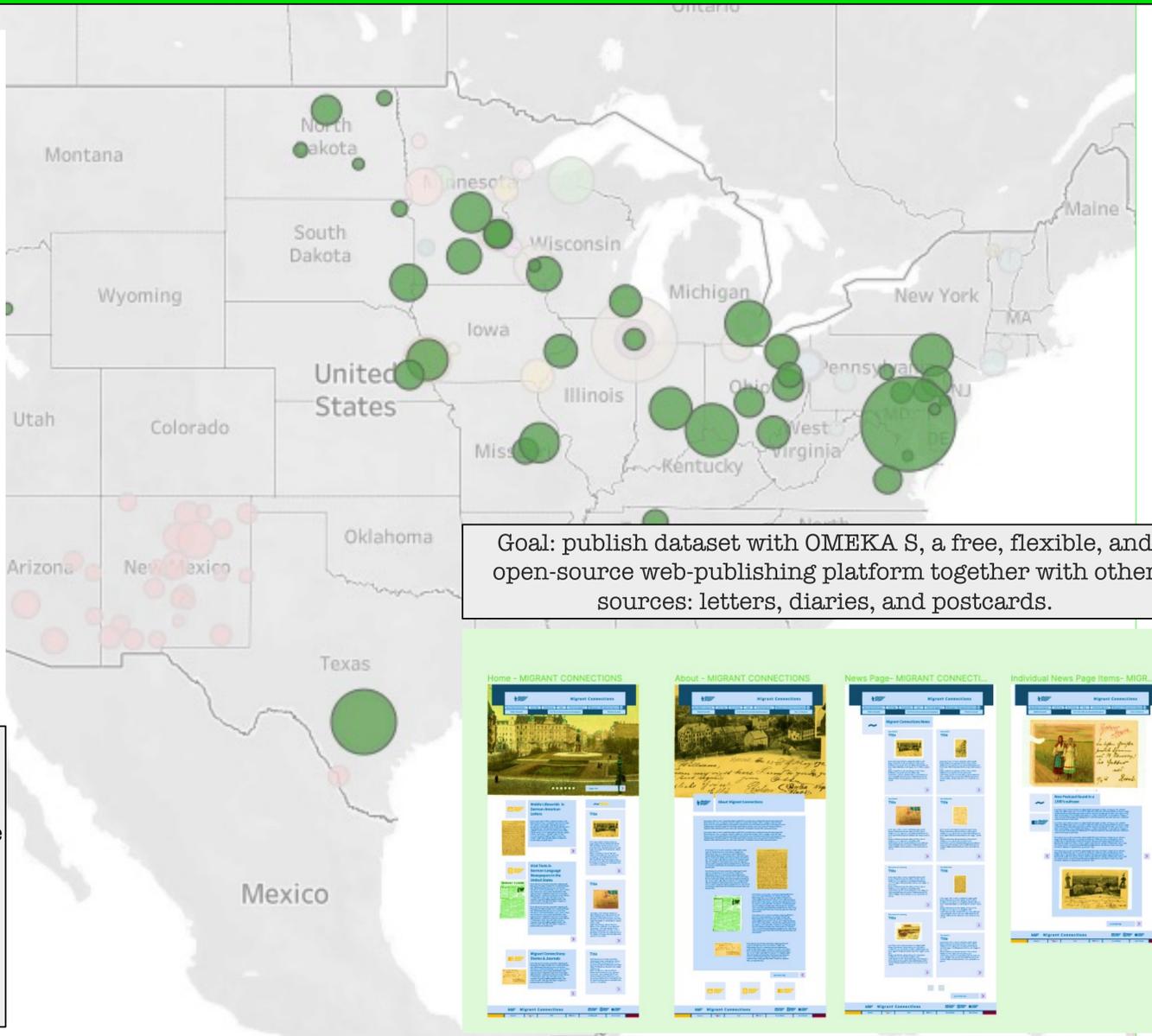
Goal: publish dataset with OMEKA S, a free, flexible, and open-source web-publishing platform together with other sources: letters, diaries, and postcards.

Corpus creation: texts from *Chronicling America* are not categorized into genres. To add another layer that will make an expanded search possible by adding genre as metadata, we are using unsupervised methods (*topic models*, *clustering*) and supervised classification methods enabled by manual annotations that were integrated into a web-based interface, an adaptation of the DFR-Browser, and extended by an annotation module.

Only "accessible" datasets can be passed on to future generations.
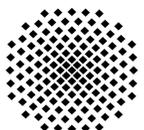
University of Stuttgart
Germany

Institut für Maschinelle Sprachverarbeitung

GHI **German Historical Institute Washington**