Digital Data Repositories in Historical Research

Adjunct Professor Asko Nivala
TIAS / Department of Cultural History, University of Turku

TURUN YLIOPISTO

1. COMHIS: Text Reuse in Finnish Newspapers and Journals, 1771–1920 (PI: Hannu Salmi)
2. Gale Cengage NCCO modules (Commercial license)
3. Romantic Cartographies: Lived and Imagined Space in English and German Romantic Texts, 1790–1840 (Postdoctoral Project in Turku Institute for Advanced Studies)

TURUN YLIOPISTO

Computational History and the Transformation of Public Discourse in Finland, 1640–1910

- Academy of Finland, 2016-2019
- a consortium together with the National Library of Finland, University of Helsinki, University of Turku
- Consortium PI: Hannu Salmi (University of Turku)
- Subproject PIs: Kimmo Kettunen (National Library of Finland), Tapio Salakoski (University of Turku), Mikko Tolonen (University of Helsinki)
- Our WP in Turku works with the text reuse in Finnish press

TURUN YLIOPISTO

XML Data dump from National Library of Finland

- 1771–1874 (15 GB), 1875–1920 (371 GB)
- XML METS/ALTO files, OCR (Abbyy)
- Swedish, Finnish (and some German, Russian, English)
- CC-BY license
- T. Pääkkönen, J. Kervinen, A. Nivala, K. Kettunen, E. Mäkelä: Exporting Finnish Digitized Historical Newspaper Contents for Offline Use. *D-Lib Magazine* vol. 22, no. 7 2016, DOI: 10.1045/july2016-paakkonen.

TURUN YLIOPISTO

BLAST

- Abbyy OCR is very noisy for Fraktur typeset
- BLAST is developed in bioinformatics for detecting and aligning similar passages in very noisy data
- A. Vesanto, A. Nivala, H. Rantala, T. Salakoski, H. Salmi & Filip Ginter, ’Applying BLAST to Text Reuse Detection in Finnish Newspapers and Journals, 1771–1910’, Proceedings of the 21st Nordic Conference of Computational Linguistics. Gothenburg, Sweden, 23–24 May 2017 (Linköping 2017), 54–58, http://www.ep.liu.se/ecp/133/010/ecp17133010.pdf

TURUN YLIOPISTO

http://www.comhis.fi

- Another problem: our research group recognised 13,8 million clusters of textual reuse consisting of 61 million passages
- How to browse this data? Normal databases do not scale well > we used Apache Solr (Lucene)
- We have published research articles about our data, but we also decided to open our web interface for other researchers and general audience

TURUN YLIOPISTO

Gale Cengage NCCO modules

University of Turku has access to three commercial NCCO modules

1. European Literature, 1790-1840: The Corvey Collection
2. Mapping the World: Maps and Travel Literature
3. Science, Technology, and Medicine, part I

If you want access, contact: asko dot nivala at utu dot fi

TURUN YLIOPISTO

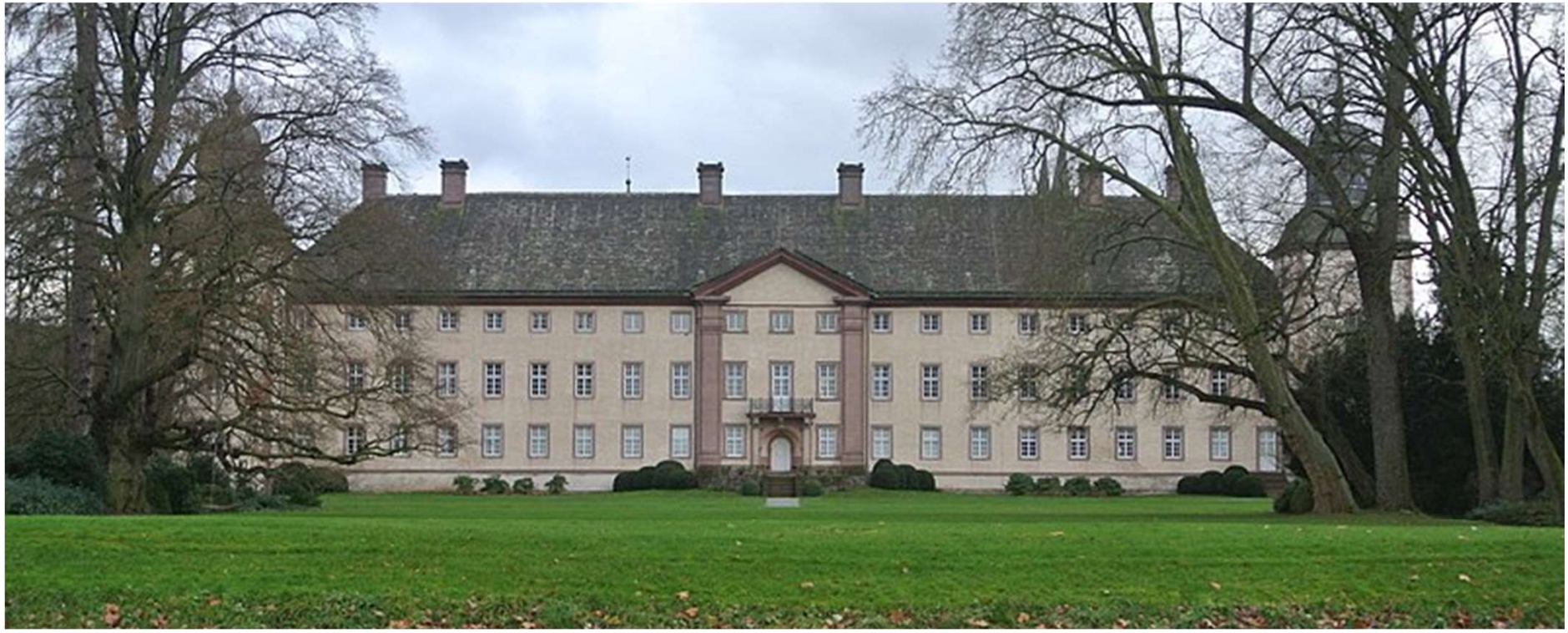
The Corvey Library

- consists of more than 72,000 volumes
- was among the most important bibliographical findings of the 1970s, at least for the study of Romanticism (Rainer Schöwerling)
- provides outstanding data for book history
- ”in certain peak years – more than 90% of the total fiction output from the British Isles is represented in this forgotten library of the eastern Westphalian provinces” (Werner Huber)

TURUN YLIOPISTO

Schloss Corvey

Höxter, North Rhine-Westphalia, Germany



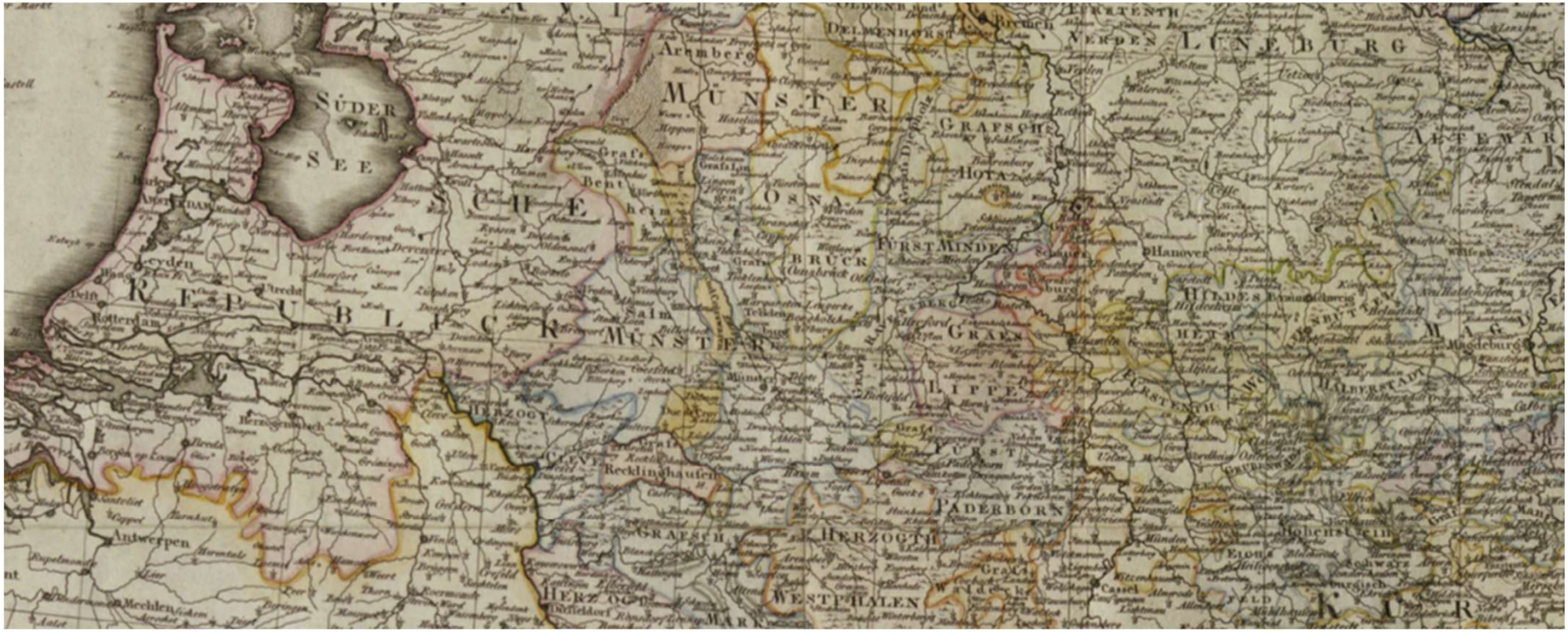
TURUN YLIOPISTO

NCCO: The Corvey Collection 1790–1840

- NCCO The Corvey Collection includes the full-text of more than 9,500 English, French and German titles
- excellent metadata compared with other collections like Archive.org, Google Books or British Library 19th literature
- full-text has been produced with Abbyy OCR (optical character recognition) > not working great with German gothic typeset
- the NCCO collection includes the original image files

TURUN YLIOPISTO

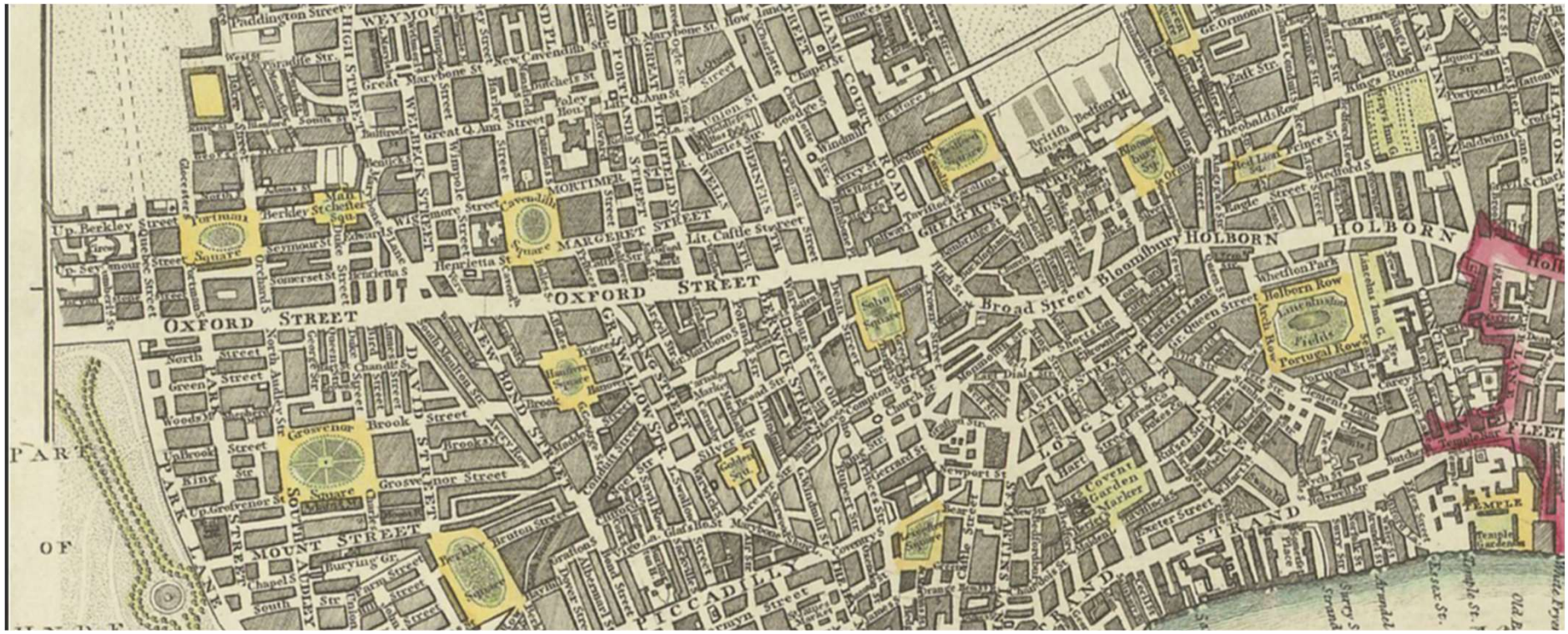
NCCO Mapping the World: Maps and Travel Literature



”General-Karte von Deutschland, Der Batavischen Und Helvetischen Republik, Ober-Und Mittel-Italien, und Dem Östlichen Theil Der Französischen Republik; In Zwey Sectionen. Entworfen von D. F. Sotzmann. 1063. (24.)." *British Library: 19th Century European Sheet Maps*, Primary Source Media, 1803. Nineteenth Century Collections Online, http://tinyurl.galegroup.com/tinyurl/6cjn73. Accessed 5 June 2018.

TURUN YLIOPISTO

A detail of a map of London (1797)



Darton, William and Harvey, Josiah. ”A New Pocket Plan of London, Westminster and Southwark with All the Adjacent Buildings. Also a Correct Lift of Upwards of 300 Hackney Coach Fares. 1797. Maps Crace Port. 5.181:181.” *Crace Collection of Maps of London*, Primary Source Media, 1797. Nineteenth Century Collections Online, http://tinyurl.galegroup.com/tinyurl/6cjHt2. Accessed 5 June 2018.

TURUN YLIOPISTO

Romantic Cartographies: Lived and Imagined Space in English and German Romantic Texts, 1790–1840

- Turku Institute for Advanced Studies 2017–2019
- Spatial analysis of English and German Romanticism, mainly novels and travelogues
- Digital humanities project using software, e.g. named entity recognition, geoparsing
- Focusing on the references to urban cities, natural sites and historical places
- Destabilisation of centre–periphery distinction

TURUN YLIOPISTO


TURUN
YLIOPISTO