

In-workshop Notes for the Planning Workshop

Ethical Visualization in the Age of Big Data: Contemporary Cultural Implications of Pre-Twentieth-Century French Texts

When: 9–10 January, 2020

Where: Lincoln Hall, University of Nevada, Reno, U.S.A.

[Full briefing materials](#)

Session 1: Scholarly Framing

Visual analysis

KM - “easy” way to start examining images - Pixplot <https://dhlabs.yale.edu/projects/pixplot/>

Giles Bergel (Oxford, <http://seebibyte.org/>) work with visual analysis

<http://vgg.gitlab.io/seebibyte-blog/digital-humanities/2017/11/24/Matching-printers-woodblocks-to-their-impressions-with-VISE.html>

Computer vision people

Thomas Smits - <https://lab.kb.nl/person/thomas-smits>

Melvin Wevers - <http://www.melvinwevers.nl/>

Lauren Tilton, University of Richmond - <https://rhetoric.richmond.edu/faculty/ltilton/>

Taylor Arnold, University of Richmond - <https://math.richmond.edu/faculty/tarnold2>

Frederico Freitas, NCSU - https://history.ncsu.edu/people/faculty_staff/fsantos2

(for similar kinds of images in periodicals, modern photos)

Fred Harris, UNR - <https://www.unr.edu/cse/people/fred-harris>

Logical home for our data: the BnF

KM - Person to talk to at BnF about data - Emmanuelle Bermès (and new BnF Data Labs that is launching soon)

CG - in France, [Ortolang](#) is the place to store corpora that are used by Natural Language Processing people. It is the French node of the European [CLARIN](#) repository.

Ethical consideration follow-up links

How to do reproducible research ethically - a collaborative, distributed labor project with a home at the Turing: <https://the-turing-way.netlify.com/introduction/introduction> (e.g. in terms of thinking about what we need to do to make data useable after individual projects)

On describing your data: see Gebru et al. (2019), “Datasheets for Datasets” for a good list of guidelines for the description of new data: <https://arxiv.org/pdf/1803.09010.pdf>

On identifying problematic descriptions and slurs of people: see research on identifying ‘abusive language online’ in NLP, for example: <https://www.aclweb.org/anthology/W18-5117.pdf>

Session 2: Metadata Structure

BnF InterMarc metadata: <https://catalogue.bnf.fr/ark:/12148/cb34446531j.intermarc>

Hans Sloane catalogues TEI project -

https://research.britishmuseum.org/research/research_projects/all_current_projects/enlightenment_architectures.aspx

Session 3: Natural Language Processing (nlp) Workflow

Event-based viz environment developed for humanities projects: Nodegoat -

<https://nodegoat.net/>

Edinburgh Geoparser - <https://www.ltg.ed.ac.uk/software/geoparser/>

(demo: <http://jekyll.inf.ed.ac.uk/geoparser/>)

The Getty Thesaurus of Place Names - <https://www.getty.edu/research/tools/vocabularies/tgn/>

Turing Institute uses Wikipedia to disambiguate place names

Ethics in NLP coursework example: http://faculty.washington.edu/ebender/2017_575/index.html

Session 4: Domain Adaptation (for early modern French)

BiLSTM Neural Networks (Bidirectional Long Short-Term Memory)

Annotation GUI similar to brat (Inception, but here called WebAnno):

<https://discovery.clarin.eu/?entityID=https%3A%2F%2Fwebanno.sfs.uni-tuebingen.de&return=https%3A%2F%2Fwebanno.sfs.uni-tuebingen.de%2FShibboleth.sso%2FLogin%3FSAMLDS%3D1%26target%3Dss%253A017c5b0ec462266d5159d43f28c26060b0fe6214d049ac8dcb34f00f42db043e>

(if you sign up for a CLARIN.eu website account, you can set up annotation exercise without belonging to an EU institution) <https://inception-project.github.io/>

BRAT: <https://brat.nlplab.org>

‘Inception’. <https://inception-project.github.io/>

New CLEF shared task on historical NER for newspapers, including French (though data looks to be Swiss newspapers?). May be worth signing up for the shared task to get access to the training/evaluation data. <https://impresso.github.io/CLEF-HIPE-2020/datasets.html>

Session 5: Ethical Data Visualization

<http://hdlab.stanford.edu/projects/writing-rights/>
https://opus.lib.uts.edu.au/bitstream/10453/37041/1/0143_SMT_Vol13_P07_Kasunic-Sweetapple_v1%20%282%29.pdf
<https://www.uts.edu.au/about/faculty-design-architecture-and-building/staff-showcase/writing-rights>

Ethical Visualization Workflow - <https://github.com/kathep/ethics>

Doris Kosminsky's lab - <https://labvis.eba.ufrj.br/>

Related project: <http://www.atilf.fr/spip.php?rubrique88>
and
https://digitalkoine.github.io/atlas_nazifascist_chronology/

Image analysis, related work: Lyneise Williams (UNC Chapel Hill) describe her work on black bodies and image reproduction. It is amazing. She is examine how reproduction in 20th and then microfilming and other processes strip away the advances of rotogravure.

<https://veracollaborative.com/>
https://dcic.umd.edu/fall-2018-dcic-speaker-series/fall_2018_dcic_speaker_series_lyneise-williams/

Deaths on slave ships, visualized in beads -

<https://labvis.eba.ufrj.br/projetos/a-dispersao-de-africanos-escravizados-pelo-mundo-atlantico/>

Session 6: Mapping and Interface Design Usability

From session 3 to discuss here: how to refer to old geographic location names in the context of newer location names.

<http://revolt.axismaps.com/map/>
<https://apps2.cer-rec.gc.ca/pipeline-incidents/>
<https://northeastern.theopenscholar.com/elizabeth-maddock-dillon/publications/design-remapping-colonial-archive>

<https://corpus-synodaliu.com/index.html>

[Interesting: they ask for a password to get access to the dataset? They include a sentence: "Please contact Rowan Dorin (dorin@stanford.edu) to request access to the database."]

<http://encyclopedia.uchicago.edu/>

Session 7: Adapting and Integrating Existing Open Source Projects

Centroid-based toponym resolution -- the earliest work is from the Perseus Project: Smith and Crane (2001), "Disambiguating Geographic Names in a Historical Digital Library"

https://link.springer.com/chapter/10.1007/3-540-44796-2_12

Talk to Eric Steiner at Spatial History Lab - thinks a lot about systems of collaboration and one-off tool use - <https://digitalhumanities.stanford.edu/users/erik-steiner>

Radical Collaboration model from Living with Machines:

<https://livingwithmachines.ac.uk/how-we-collaborate/>

Session 8: Input and Output (i/o), Ongoing Maintenance & Preservation

Notes go here

Glossary

BnF - [The Bibliothèque nationale de France](#)

TEI - [Text Encoding Initiative](#)

BiLSTM - Bidirectional Long Short-Term Memory

Visualization Examples

<http://apps2.cer-rec.gc.ca/pipeline-incidents/>

Pedro Cruz Visualizing Empires Decline <https://vimeo.com/11506746>