DATA-POWERED JOURNALISM

Tyler Dukes, WRAL.com, @mtdukes
eNCA // July 28, 2014 // Johannesburg

Google Drive

http://drive.google.com
An inexpensive alternative to Microsoft Office hosted in the cloud, Google Drive offers its own versions of Word, Excel, Powerpoint and more with much of the same functionality. Can also be embedded in stories.

Cometdocs

http://cometdocs.com
This free, Web-based file converter is useful for transforming PDFs into more useful formats, like text or Excel tables. Cometdocs Premium, a paid version, expands its limits on document sizes and includes optical character recognition for scanned PDFs.

DocumentCloud

http://documentcloud.org
This Web-based application offers a suite of tools to organize, annotate and analyze documents as well as share them with both colleagues and audiences. Documents are automatically processed with optical character recognition, allowing search and entity extraction to pull out things like names and dates.

Fusion Tables

http://tables.googlelabs.com
A beta project from Google, Fusion Tables allows users to mash up their data with other existing tables, like maps of political boundaries or poverty levels, using simple joins. Users can map the results and embed them in their websites quickly and easily.

Timeflow

http://reporterslab.org/timeflow
Timeflow is a visual tool for reporters looking to organize and analyze historical data on long-term stories. The downloadable java app makes it easier for reporters to tell in-depth narratives by tracking trends and making sense of temporal data.

Overview

https://www.overviewproject.org/
This Web-based application automatically sorts thousands of documents into topics and sub-topics. Its interface also links with DocumentCloud, allowing reporters to easily analyze large document sets.

Open Refine

http://openrefine.org/
OpenRefine (formally Google Refine), allows rapid cleaning of data with the combination of Excel-like formulas and text faceting/clustering. The downloadable application, which runs in a Web browser, groups similar words together based on multiple algorithms and allows users to quickly standardize names, businesses and other data. See Dan Nguyen's tutorial here.

Tabula PDF

http://tabula.nerdpower.org/
A free, downloadable, open-source application to convert tabular information locked inside PDFs into CSV data, which you can import easily into spreadsheets.

OTHER COOL STUFF

Mr. Data Converter
http://shancarter.github.io/mr-data-converter/
This open-source tool, built by Shan Carter, converts Excel data into one of several web-friendly structured formats, including HTML, JSON and XML.

Data Science Toolkit
http://www.datasciencetoolkit.org/
This toolkit features an entire suite of easy-to-use Web apps for doing all kinds of cool things to data, like converting PDFs to plain text and converting street addresses to coordinates. Also features an open API for more advanced users.

Jigsaw
http://www.cc.gatech.edu/gvu/ii/jigsaw/
An experimental tool born out of academia, this Java application helps users make sense of large collections of documents with the help of text analysis algorithms. It features a variety of different ways to look at the documents, from topic clustering to entity extraction.

FOIA Machine
http://www.foiamachine.org/
Online tool that helps you submit, organize and track public records requests at the state and federal level. Check out the team's session Friday at 9:30 a.m.

PLACES TO LEARN

Nieman Journalism Lab
http://niemanlab.org
An attempt to help journalism figure out its future in an Internet age. Often features new and emerging tools for journalists.

Source
http://source.mozillaopennews.org/en-US/
Source is a Knight-Mozilla OpenNews project designed to amplify the impact of journalism code and the community of developers, designers, journalists and editors who make it.

Knight Lab
http://knightlab.northwestern.edu/
A team of technologists, journalists, designers and educators working to advance news media innovation through exploration and experimentation. Often develop and release their own tools for journalists.

Data Driven Journalism
http://datadrivenjournalism.net
Dedicated to providing anyone interested in data driven journalism with a collection of learning resources, including relevant events, tools, tutorials, interviews and case studies.