

# Data Visualization in the Digital Humanities: a case study of the University of Sussex's Humanities Lab visualization project for reuse and engagement

This case study looks at the data visualization work undertaken by David Banks and colleagues from the University of Sussex to better engage with historical datasets and to allow researchers to dive into datasets deeper and more meaningfully.

## About David Banks

David Banks is a Full Stack Web Developer and Data Visualisation Research Fellow at the University of Sussex. David works within the Sussex Humanities lab which specializes in Digital Humanities. Research done within the lab is cross-disciplinary across the humanities, social sciences, and informatics.

David works on a number of different projects inside the lab as a developer creating web applications for data use and reuse. Most of David's work is in visualizing research data. "My job is to take something like a dataset from social sciences or something text-based from history and create a novel visualization from it," said David. "Because these projects happen over a long span of time, we're designing our interfaces in an iterative dialogue with humanists."

The rationale for creating a visualization is different for each project. For some projects, David and his colleagues are looking to share the impact of the research either more widely with the public or with the academic community. In these cases, the goal is to make the results of the research and the data itself more visible and interpretable to people who may not know much about it.

One project David has worked on is [The Planet Bethlehem Archive](#), an open access digital resource that documents Bethlehem's history and culture created in collaboration with filmmaker Leila Sansour, historian Jacob Norris and the University of Sussex. David worked with Research Data & Digital Preservation Technologist Adam Harwood and Research Fellow in Digital Humanities Dr Ben Jackson, both from the University of Sussex, to develop a program for researchers in numerous projects to upload research data to [sussex.figshare.com](https://sussex.figshare.com). Based on feedback Adam and Ben received from other researchers, alongside collaboration and testing with Research Assistant in Digital Archiving Freja Howat-Maxted, David developed a desktop application using Figshare's API (Application Programming Interface) for researchers to be able to upload their large files and associated metadata together. The application can be [found here](#).

Researchers have been using David's application to upload images and metadata to [The Planet Bethlehem Archive's Figshare page](#). In the case of these researchers, the application allowed them to upload the files and metadata stored in an Excel spreadsheet, meaning each item created is automatically populated with the corresponding metadata held in the spreadsheet, saving their researchers time by not having to manually add the metadata to the item page. The next stage of development will be for researchers to upload interview transcripts to Figshare, many of which have sensitive information in them.

David and his colleagues also build tools for researchers themselves to use during active research for interrogation and interpretation of large-scale data sets. David works in the [Text Analytics Group](#) at the University of Sussex where they use natural language processing techniques to allow researchers to address research questions arising from large text corpora, questions that would be intractable without the use of computational methods.

One of David's next projects is to develop a semantic querying tool using the Old Bailey Proceedings dataset<sup>1</sup>. These are a text history of all of the trials that occurred at the Old Bailey from the 1600s to the early 1900s. This tool would allow someone to run machine learning classifiers over text datasets, annotate certain parts of them, then there will be a visual interface for querying the annotations.

*“This will allow users to query, for example, proceedings of sailors arrested at night, or other semantic criteria the user has in mind.” said David. “This information doesn’t come from atomic metadata stored on the set: rather, users train active-learning classifiers to enable the construction of queries like these.”*



Visualizing phrase usage in the Old Bailey dataset.

Tim Hitchcock, Robert Shoemaker, Clive Emsley, Sharon Howard and Jamie McLaughlin, et al., The Old Bailey Proceedings Online, 1674-1913 ([www.oldbaileyonline.org](http://www.oldbaileyonline.org), version 7.0, 24 March 2012).

For more information on The Planet Bethlehem Archive, visit:

[sussex.ac.uk/history/bethlehem/archive](http://sussex.ac.uk/history/bethlehem/archive)

To see data from the archive, visit:

[sussex.figshare.com/The\\_Planet\\_Bethlehem\\_Archive](http://sussex.figshare.com/The_Planet_Bethlehem_Archive)