Wikidata in the curriculum for Data Science for Design MSc

Wikidata assignment; developing data literacy and data visualization skills.

Date Added: 2nd July, 2018.

What was done

Wikimedian in Residence, Ewan McAndrew, supported Assistant Professor Benjamin Bach (Lecturer in Design Informatics and Visualisation) and Research Associate Dr. Joe Corneli (Centre for Intelligent Systems and their Applications) integrate a Wikidata assignment into the Data Science for Design MSc - Semester One curriculum. This course provides an introduction to programming, used in support of the development of data science techniques, to give a practical facility in manipulating, analysing, visualising and contextualising data. The outcome was 3219 items of data on accused witches in Scotland created in Wikidata (spanning 1563 to 1736). We also now have data on 2356 individuals involved in trying these accused witches. Finally we have 3210 witch trials themselves and two data visualization videos created by the two groups of students working on the project. Thirteen datasets were pitched at the entire cohort of forty-five postgraduate students by a variety of researchers and students were invited to work on these projects in groups of three. Two groups volunteered to surface the Survey of Scottish Witchcraft database and were supported over the course of five weeks to analyse, model and process the data from the Survey of Scottish Witchcraft into Wikipedia's sister project, Wikidata.

The Data Fair

At the Data Science for Design MSc's "Data Fair" on 26th October 2017, "problem holders" from across the university shared their data-oriented problems with 45 masters students in Design Informatics. Approximately 13 datasets were pitched for the students to choose from to work on in groups of three. Ewan McAndrew, Wikimedian in Residence at the University of Edinburgh, presented a project in a 5 minute pitch to import the university's Survey of Scottish Witchcraft database into Wikipedia's sister project, Wikidata, a central hub for structured, machine-readable, linked open data.

This Survery of Scottish Witchcraft database began life in the 1990s before being realised in 2001-2003. It had as its aim to collect, collate and record all known information about accused witches and witchcraft belief in early modern Scotland (from 1563 to 1736) in a Microsoft Access database and to create a web-based user interface for the database. Since 2003, the data has remained static in the Access database and so students at the 2018 Data Fair were invited to consider what could be done if the data were exported into Wikidata, given multilingual labels and linked to other datasets? Beyond this, what new insights & visualisations of the data could be achieved?

Two groups of three students volunteered to take part in the project and arranged with Ewan McAndrew to attend two follow-up workshops to discuss how to divide the task of analysing, processing and importing the data.

Overview

Summary

Data Science for Design MSc students volunteered on a 5 week project to surface the data from the Survey of Scottish Witchcraft database into Wikipedia's sister project, Wikidata. The outcome was 3219 items of data on accused witches in Scotland (spanning 1563 to 1736) created in Wikidata. There is also now have data on the 2356 individuals involved in trying these accused witches. Finally there is data on the 3210 witch trials themselves and two data visualization videos created by the two groups of students working on the project.

Integrating Wikidata into coursework allows students and staff to develop their understanding of, and engagement with, issues such as: data completeness; data ethics; digital provenance; data analysis; data processing; as well as making practical use of a raft of tools and data visualisations.

Context
- **Level:** Postgraduate
- **Discipline/Subjects:** Edinburgh College of Art; Data Science for Design MSc
- **Supported by:** Learning, Teaching & Web Services.

Contact Details
- Assistant Professor Benjamin Bach (Lecturer in Design Informatics and Visualisation)
- Research Associate Dr. Joe Corneli (Centre for Intelligent Systems and their Applications)
- Ewan McAndrew, The University of Edinburgh's Wikimedian-in-Residence.

Themes
- Curriculum Development
- Learning Communities

Two workshops to analyse, process and import the data

A similar methodology to managing Wikipedia assignments was employed; making the transition from managing a Wikipedia assignment to managing a Wikidata assignment an easy one.

On 8 November 2017, the two groups of students underwent a 1.5 hour practical induction on working with Wikidata and third party applications such as Histropedia ("the timeline of everything") before being introduced to the Access database. Afterwards, the students spent an hour discussing collaboratively how best to divide the task of analysing and exporting the data before deciding one group would work on (1) importing records for the 3,219 accused witches while the other group would work on (2) the import of the witch trial records and (3) the people associated with these trials (lairds, judges, ministers, prosecutors, witnesses etc).

The groups researched and submitted their data models by email for review by Ewan McAndrew, Wikimedian in Residence.

Once their models had been checked and agreed upon via email, the students were ready to process the data from the Access database into a format Wikidata could import (making use of the handy Wikidata plug-in on Google Spreadsheets).

The processing of the data took place in a 2 hour workshop on 22 November 2017. Upon completion of the import, the students could then choose how to visualise this newly added data in a number of ways; such as maps, timelines, graphs, bubble charts. Data visualisation videos and more.

The students finished their project by showcasing their data visualisation videos at the end of project presentation day on the 30th of November 2017.

Here are two data visualisation videos they produced:

- https://media.ed.ac.uk/media/Introduction+to+Scottish+Witchcraft/1_kt5nbmou
- https://media.ed.ac.uk/media/The+trial+of+witchcraft-data+visualisation/1_v4116vif

Motivation and aims

There were several areas of interest: course leaders on the Data Science for Design MSc were keen for the students to work with 'real world' datasets in order to give them practical experience ahead of their dissertation projects.

> *"A common critique of data science classes is that examples are static and student group work is embedded in an 'artificial' and 'academic' context. We look at how we can make teaching data science classes more relevant to real-world problems. Student engagement with real problems—and not just 'real-world data sets'—has the potential to stimulate learning, exchange, and serendipity on all sides, and on different levels: noticing unexpected things in the data, developing surprising skills, finding new ways to communicate, and, lastly, in the development of new strategies for teaching, learning and practice."*
> – Towards Open-World Scenarios: Teaching the Social Side of Data Science by Dave Murray Rust, Joe Corneli and Benjamin Bach.

The learning outcomes were that on completion of this course, the student would be able to:

1. Program: Identify and deploy strategies for writing, understanding and managing computer programs using Python and version control
2. Data: The ability to wrangle, analyse, learn from and visualise a range of data, in a way that demonstrates its relevance to particular contexts of enquiry
3. Communicate: Communicate around socially relevant issues, supported by the use of multiple data sources and appropriate analysis
4. Professionalism: Working in collaborative, interdisciplinary teams to a high professional standard.

Furthermore, the pressing need for implementing data literacy in the curriculum to produce a workforce equipped with the data skills necessary to meet the needs of Scotland's growing digital economy presents a massive opportunity for educators, researchers, data scientists and repository managers alike. Therefore, the aim was also that this project would aid the students' understanding of data literacy through the practical application of working with a real-world dataset and help shed new light on a little understood period of Scottish history. This, in turn, would help fuel discoveries by dint of surfacing this data and linking it with other related datasets across the UK, across Europe and beyond. As the Survey of Scottish Witchcraft's website states itself *"Our list of people involved in the prosecution of witchcraft suspects can now be used as the basis for further inquiry and research."*

Successes and lessons learnt

Over the course of a month, six postgraduate students were motivated to volunteer to surface a much-loved repository of information, the Survey of Scottish Witchcraft database, into Wikidata. This was a project that the students felt proud to take part in and found "*very meaningful*".

This demonstrated that the Data Fair model of providing a variety of research data projects for students to gain 'real world' experience of analyzing, processing, importing and visualizing data can be employed in the curriculum to good effect in the case of working with Wikidata. This proved beneficial for the students in terms of improving their understanding of data literacy and for the surfacing of the data itself in order to enable further insights and research. Feedback from the students was extremely positive (see below) although improving the documentation available for Wikidata work is an area for future development.

Scalability and transferability

The success of this project indicates how this type of project, with students practically engaging with data, lends itself to incorporation into the curriculum and can be potentially scaled up to provide a successful learning experience for larger cohorts. Requirements for staffing include tutors to support the teaching of data processing & data visualisation skills and a Wikimedian to provide Wikidata training. Venues should be equipped with learning technology for training purposes and should support students with personal computers.
This type of coursework can be applied in a number of subjects interested in the developing of data skills and knowledge transaction.
Wikidata has a particular need, & an ongoing commitment to, improving its coverage of notable datasets; be it biographical, biomedical, geographical, taxonomical, or bibliographical in nature.

Assessment

Coursework counted for 100% of the Data Science for Design MSc programme. assessed through two individual reports (1000 words each, 30% each), a group report (2000 words, 30%) and an ongoing professional collaboration component (10%) assessed through contributions to code and data repositories, and online group discussion.

This *Data Fair* group project involved the submission of an Individual Report (addressing Learning Outcomes 1 and 2) and involves an open ended analysis of student-selected data
. The Group Report addressed Learning Outcome 3 and involved analysis and reporting on a particular problem holders social situation in a small team. The purpose of this exercise was to have the students think about an audience and an engaging story.

Students had two weeks to submit the Individual Report analyzing the datasets and the final 3 weeks beyond that to come up with some engaging form of presenting the data (group work). This end-piece was to communicate insights from or around the data to a specific audience. This was to be shaped by what the students considered made most sense in terms of the data.

Some suggested outputs were:

1. An interactive web site
2. Interactive visualizations on a website
3. Data comics (http://datacomics.net)
4. Infographic
5. Data video
6. A standard scientific presentation / report.

The students in this assignment opted for data visualization videos and these were presented on 30 November 2017 and showcased with the other projects on this website: http://aviz.fr/~bbach/datafair/

Feedback
Formative feedback was provided verbally during weekly tutorials.
Written formative feedback was provided on initial submission of the individual report outputs.
Summative feedback was provided following the submission of all outputs (individual reports and group work) at the end of the course.

Feedback from students at the conclusion of the project included:

"After we analysed the data, we found we learned the stories of the witches and we learned about European culture especially in the witchhunts."

"We had wanted to do a happy project but finally we learned much more about these cultures so it was very meaningful for us."

"In my opinion, it's quite useful to put learning practice into the real world so that we can see the outcome and feel proud of ourselves… we learned a lot."

"Thank you for inviting us and appreciating our video. It's an unforgettable experience in my life. Thank you so much."

Links

- Blog article https://blog.wikimedia.org.uk/2018/03/data-on-the-history-of-scottish-witch-trials-added-to-wikidata/

- Course catalogue: http://www.drps.ed.ac.uk/17-18/dpt/cxdesi11100.htm

- Survey of Scottish Witchcraft database: http://www.shca.ed.ac.uk/Research/witches/

- Outcomes of the project: https://thinking.is.ed.ac.uk/wikidata-workshop/real-world-datasets-the-survey-of-scottish-witchcraft/

- Interview with the students: https://commons.wikimedia.org/wiki/File:Wikidata_in_the_Classroom_-_interview_with_Data_Science_for_Design_MSc_students.webm