It has moving parts!

Interactive visualisations in digital publications

James Cummings, Martin J. Hadley, Howard Noble
Academic IT Services, University of Oxford

james.cummings@it.ox.ac.uk
@jamescummings
Digital Scholarly Editions and Visualisations

- Digital Scholarly Editions sometimes include visualisations of the data they contain but more often than not this is limited to static visualisations.
- Most commonly page images, basic charts, timelines, social networks, and extracted tables of information.
- When they exist, they all use different technologies for processing and display.
- Often these are fixed visualisations: readers are unable to truly change the parameters or interact with the data.
- Usually these visualisations work only on the originating website, they can’t easily be embedded in articles reviewing them.
- Often the concentration is on the display/presentation of the editor’s viewpoint and conclusions rather than exposing the data for further query or ongoing analysis by readers.
Digital Data to be Extracted

- Purely digital editions sometimes present one or two visualisations to highlight the interests of those creating the editions;
- Many editors often think they have ‘only text’, not ‘data’; but the structures created by the edition contain more data than they think;
- Where feasible all digital editions should at least freely expose:
  - Full copies of underlying data available at a stable URL in interchange formats like TEI XML easily addressable at word level
  - The project-specific schema, TEI ODD Customisation file, or other machine-processable schema as part of project metadata and documentation
  - Extracted data on any named entities (people, places, organizations, works, etc.)
  - Individual and combined witnesses (in multi-witness editions), and data necessary for generation of stemmata codicum
  - Wordlists of distinct (at least) orthographic words and references to them
  - Exported lists of distinct abbreviations and expansions, and references to them
  - Pre-calculated statistics on any significant markup structures
- Data such as these should be considered low-hanging fruit, easy to produce, freely available, and enable secondary studies, comparison of editions, re-use of data
The Wandering Jew's Chronicle:
OR, The Old Historian. His Brief Declaration, Made in a mad passion, of each Coronation, That pass'd in this Nation, Since Williams's Invasion. For no great Occasion, But mere Recreation, To put off Vexation.
To the Tune of, Our Prince is welcome out of Spain.

1 When 2 William Duke 2 of Normandy 2
2 With all his Normans gallantly
3 This Kingdome did subside;
4 Full fifteen years of age I was,
5 And what I've since hath come to pass,
6 I can report for true.
7 I can remember since I went
8 From London to for Conquer Kent,
9 Where with a walking Wood,
10 The men of Kent compassed him,
11 And he for eye confirmd to them
12 King Edwards Laws for good.
13 Likewise, I William Rufus knew,
14 And saw the Arrow that him slow
15 By a forest side:
16 I well could tell if it live,
17 Or better tell you if he not,
18 Who next to him did ride.
19 First Henry 1, and Stephen knew,
20 Who no man here but I did view,
21 I saw them Crownd and dead:
22 I can remember well also,
23 The second Henry Royal show,
24 That day that he was dead.
25 I likewise was at Woodstock Bower,
26 And saw that sweet and famous flower,
27 Queen Eleanor so did spight;
28 I found the clew of thread again.
29 After that worse Knight was slain,
30 Was green, blew, red and white.
31 I saw King Richard in his shirt
32 Pull out a furious Lynne heart.
University of Oxford: Interactive Data Network

- A new service at the University of Oxford to provide support, training, and consultancy for interactive data visualisations
- Visualisations built on with a consistent set of standard technologies such as:
  - The R programming language
  - The Shiny web application framework
  - Open source packages and plugins
  - Javascript visualisation libraries
- All data and visualisation code is released openly
- Visualisations hosted on [shinyapps.io](http://shinyapps.io) and hosting provided to University of Oxford members as part of the service
- Data able to be read from many different repositories (institutional, Figshare, Zenodo, Google, Zotero)
- Workshops, training materials, how-tos, reading lists, and fully working examples provided for many different forms of data as case studies
- Not really ‘digital editions’ as we usually perceive them but DSEs should embed visualisations like this
Visualisation Showcase

This is a showcase of the interactive data visualizations built by members of the IDN as case studies for the Live Data Project, mostly built through 2016. Almost all of the examples below were built using R and Shiny. Details about the research behind the data, the data repositories hosting the data, and the researchers themselves are available by clicking on a visualisation of choice.

If you're interested in working with the IDN on a similar case study do get in touch, have a look at our templates and tutorials and read about the Oxford University funded Shinyapps.io subscription.

Family Policy Timelines Case Study
29 Sep 2016

Oxford Roman Economy Project Case Study
29 Sep 2016

Young Lives Case Study
29 Sep 2016

Online Labour Index Case Study
21 Sep 2016

Cultures of Knowledge Case Study
18 Sep 2016

Hepitopes Database Case Study
18 Sep 2016

German Migrant Letters
18 Feb 2016

Irish Surgeons in India
18 Feb 2016
IDN Modular Infrastructure

powered by OXFORD MOSAIC

Shiny

Drupal

iFrame

ORA

figshare

zenodo

plotly

Case Study Details

- **Dataset**: Irish surgeons in India
  - **Authors**: Name1, Name2
  - **Title**: Irish surgeons in India
  - **Abstract**: A brief summary of the study
  - **Keywords**: Keywords1, Keywords2

Case Study Overview

- A study on Irish surgeons in India presented at a conference.
- The study was presented in a visual format using tools such as Shiny and RStudio.
- The study was published in a peer-reviewed journal.
- The study's impact and significance in the field of medicine.
Cultures of Knowledge Case Study

- Cultures of Knowledge: Networking the Republic of Letters, 1550-1750
- CoK uses digital methods & interactive visualisations to reassemble & interpret the early modern correspondence networks
- Mellon-funded project running since 2009
- Created Early Modern Letters Online (EMLO) as a union catalogue of sixteenth-, seventeenth-, and eighteenth-century correspondence
- This case study visualised a subset of 1695 individuals correspondences from the EMLO dataset, stored in the FigShare repository
- Uses social network diagrams and datatables of extracted information
- This was used as part of an EU COST Network Workshop
- More: http://idn.web.ox.ac.uk/article/cultures-knowledge-case-study
Visualizations of the Prospographical Network of Samuel Hartlib

The graph below shows the whole network and how it is interconnected. People are represented in purple, organizations in green. The controls on the left allow you to search for a particular person or organization within the network and/or to select a date range. The last name of a person is shown below their node, positioning the cursor over the node displays their full name. The thickness of an edge (i.e. the line between two nodes) represents the number of connections between two people or organizations. The controls on the left allow you to highlight and/or exclude a specific relationship type.

Clicking on a node highlights the individuals or organizations directly connected to the selected individual, revealing a box showing the number and type of unique connections as well as a table with all events involving the selected individual. In the network graph, you can only select one individual at a time. You can also right-click on a node to find more information about it.

Here your cursor over the controls for more information.

### Columns to show

<table>
<thead>
<tr>
<th>Category</th>
<th>Event or Relationship Type</th>
<th>Primary Participant Name</th>
<th>Primary Participant Role</th>
<th>Secondary Participant Name</th>
<th>Secondary Participant Role</th>
<th>Date One Year</th>
<th>Date Two Year</th>
<th>Date Type</th>
<th>Location Type Ahead</th>
</tr>
</thead>
<tbody>
<tr>
<td>BasicData</td>
<td>FamilyRelationships</td>
<td>Hawk, Theodor</td>
<td>Son</td>
<td>Haas, Theodore, father of Theodore Haas (1695-1708)</td>
<td>Father</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BasicData</td>
<td>FamilyRelationships</td>
<td>Hawk, Theodor</td>
<td>Son</td>
<td>Teeswater, Marie</td>
<td>Mother</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>Study</td>
<td>Hawk, Theodor</td>
<td>Student</td>
<td>Heidelberg University</td>
<td>Administrator</td>
<td>1625</td>
<td>Before</td>
<td></td>
<td>Gladstone Hall, University of Oxford, Oxford, Oxfordshire, England</td>
</tr>
<tr>
<td>Education</td>
<td>Study</td>
<td>Hawk, Theodor</td>
<td>Student</td>
<td>Gloucester Hall, Oxford</td>
<td>Administrator</td>
<td>1625</td>
<td>1621</td>
<td>Then</td>
<td>Gladstone Hall, University of Oxford, Oxford, Oxfordshire, England</td>
</tr>
<tr>
<td>Education</td>
<td>Study</td>
<td>Hawk, Theodor</td>
<td>Student</td>
<td>Alien, Thomas</td>
<td>Teacher</td>
<td>1625</td>
<td>1621</td>
<td>Then</td>
<td>Gladstone Hall, University of Oxford, Oxford, Oxfordshire, England</td>
</tr>
<tr>
<td>Education</td>
<td>UniversityMobilization</td>
<td>Hawk, Theodor</td>
<td>Matriculant</td>
<td>Leiden University</td>
<td>Administrator</td>
<td>1630</td>
<td></td>
<td></td>
<td>Leiden, South Holland, Netherlands</td>
</tr>
<tr>
<td>SocialContacts</td>
<td>FirstContact</td>
<td>Hawk, Theodor</td>
<td>Acquaintance</td>
<td>Hartlib, Samuel</td>
<td>Acquaintance</td>
<td>1634</td>
<td>Before</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SocialContacts</td>
<td>Correspondence</td>
<td>Hawk, Theodor</td>
<td>Correspondent</td>
<td>Moreen, Marie</td>
<td>Correspondent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SocialContacts</td>
<td>Acquaintanceship</td>
<td>Hawk, Theodor</td>
<td>Acquaintance</td>
<td>Phill, John</td>
<td>Acquaintance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SocialContacts</td>
<td>Travel</td>
<td>Hawk, Theodor</td>
<td>Travel/Companion</td>
<td>Phill, John</td>
<td>Travel/Companion</td>
<td>1634</td>
<td>1638</td>
<td>Date</td>
<td></td>
</tr>
<tr>
<td>SocialContacts</td>
<td>Acquaintance</td>
<td>Hawk, Theodor</td>
<td>Acquaintance</td>
<td>Male, John (2)</td>
<td>Acquaintance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProfessionalActivity</td>
<td>Employment</td>
<td>Hawk, Theodor</td>
<td>Secretary</td>
<td>Crotch, Louis, Elazar Pellegrin</td>
<td>Employer</td>
<td>1644</td>
<td>1644</td>
<td>Date</td>
<td></td>
</tr>
<tr>
<td>BasicData</td>
<td>FamilyRelationships</td>
<td>Hawk, Theodor</td>
<td>Husband</td>
<td>Ceres, Elizabeth (d. 1683)</td>
<td>Wife</td>
<td>1665</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LearnActivity</td>
<td>MembershipOrLearnOrganization</td>
<td>Hawk, Theodor</td>
<td>Member</td>
<td>The Royal Society</td>
<td>Learn/Organization</td>
<td>1683</td>
<td>After</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LearnActivity</td>
<td>CreationOfWebActivity</td>
<td>Hawk, Theodor</td>
<td>Translator</td>
<td>Milos, John</td>
<td>Author</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SocialContacts</td>
<td>Acquaintanceship</td>
<td>Hawk, Theodor</td>
<td>Acquaintance</td>
<td>Milos, John</td>
<td>Acquaintance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SocialContacts</td>
<td>Acquaintanceship</td>
<td>Hawk, Theodor</td>
<td>Acquaintance</td>
<td>Horn, Robert</td>
<td>Acquaintance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BasicData</td>
<td>FamilyRelationships</td>
<td>Hawk, Theodor</td>
<td>Cousin</td>
<td>Spamer, Friedrich</td>
<td>Cousin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LearnActivity</td>
<td>CreationOfWebActivity</td>
<td>Hawk, Theodor</td>
<td>Translator</td>
<td>Dyke the Elder, Daniel</td>
<td>Author</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SocialContacts</td>
<td>Acquaintanceship</td>
<td>Hawk, Theodor</td>
<td>Acquaintance</td>
<td>Wehrle, Georg Rudolph</td>
<td>Acquaintance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SocialContacts</td>
<td>Acquaintanceship</td>
<td>Hawk, Theodor</td>
<td>Acquaintance</td>
<td>Romanek, Jan</td>
<td>Acquaintance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SocialContacts</td>
<td>Acquaintanceship</td>
<td>Hawk, Theodor</td>
<td>Acquaintance</td>
<td>Czernina, Johann Friedrich</td>
<td>Acquaintance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SocialContacts</td>
<td>Meeting</td>
<td>Hawk, Theodor</td>
<td>Participant</td>
<td>Christian V, King of Denmark and Norway</td>
<td>Participant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SocialContacts</td>
<td>Acquaintanceship</td>
<td>Hawk, Theodor</td>
<td>Acquaintance</td>
<td>Boyle, Robert</td>
<td>Acquaintance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LearnActivity</td>
<td>CreationOfWebActivity</td>
<td>Hawk, Theodor</td>
<td>Translator</td>
<td>Schier, Friedrich</td>
<td>Author</td>
<td>1623</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Showing 1 to 20 of 61 entries
German Migrant Letters Case Study

- Interactive visualisation of the correspondence network between German migrants in the 19th and 20th Century
- Tracking correspondence networks, specifically geolocation of the sender and recipient provides data for:
  - USA choropleth map
  - Letter routes (sender and destination addressee)
  - Migrant family locations in Germany
- Filter by date, toggle individual families
- Publication arising from and embedding the visualisation: “Writing home: how German immigrants found their place in the US”
  - Had thousands of tracked article reads in handful of months
  - Some readers citing the visualisations as what convinced them to contact the researcher to get the underlying data
  - Maps used as convincing aspects of a (now accepted) book proposal (and more visualisations will accompany the book)
- More: [http://idn.web.ox.ac.uk/article/german-migrant-letters](http://idn.web.ox.ac.uk/article/german-migrant-letters)
German Migrant Letters
18 February 2016  Felix Krausek (Politics and International Relations)

This case study is an interactive visualization of the correspondence network between German migrants in the 19th and 20th Century, primarily in the Americas. The case study comprises of three Story apps that provide an overview of how the migrant population dispersed over time using only cartograms.

Writing home: how German immigrants found their place in the US through the 19-20th Centuries, visualised with interactive maps

Click here to see the map...

Select ling Files
Status

- Include letters where source location unknown?
- Filter letters by date and
- Filter persons by age

Note that this map is a first version and is on its way to visualize the overall corpus. Any comments or suggestions you may have are very welcome. Please feel in touch with felix.krausek@politics.ox.ac.uk.

Case Study Details

- Department: Department of History
- Location: Felix Krausek (felix.krausek@politics.ox.ac.uk)
- Visitors: Felix Krausek
- Links to Visualizations:
  - https://github.com/irgoss/german_migrant_letters/tree/master/visualizations/plotmap
  - https://materialised.space/german_migrant_letters
  - https://materialised.space/german_migrant_letters/usa-chicago
- Where to visualize the map:
  - The Conversation
- Data Source: Felix Krausek
- Links to code:
  - https://github.com/irgoss/german-migrant-letters

Case Study Overview

Felix Krausek is a social history of the migration, he wanted to be able to show a number of things: Chicago, USA/Canada locations and letter routes.
The Oxford Roman Economy Project aims to study and curate the fundamentals of the Roman imperial project across all major economic activities.

This case study sought to investigate the cartographic and comparative visualisation capabilities of Shiny as a tool for exploring two subsets of the OxRep database - shipwrecks and mines.

You can filter maps and charts by date, on charts you can:
- Group by country / province / mine name
- Count by metals / mining techniques / number of mines
- Stack by percent / number of mines

Underlying OxREP databases are currently being reformulated but visualisations are planned on their databases of the Karanis Tax Rolls; Roman Mines; Olive Oil and Wine Presses; Shipwrecks; Stone Quarries; Water Technology, with more in the works.

More: [http://idn.web.ox.ac.uk/article/oxford-roman-economy-project-case-study](http://idn.web.ox.ac.uk/article/oxford-roman-economy-project-case-study)
Oxford Roman Economy Project Case Study

29 September 2016  Angelo Tredennick (Classics)  Nicholas Bay (Classics)

The Oxford Roman Economy Project is funded through the generosity of Hans-Lettner Dylla, primarily the Wrigley to study and map the fundamentals of the Roman imperial project across all major economic activities. This case study sought to investigate the cartographic and comparative visualization capabilities of Chiby as a tool for exploring two subsets of the Chiby database – shipwrecks and mints.

Case Study Details

<table>
<thead>
<tr>
<th>Department</th>
<th>Faculty of Classics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic Advisor</td>
<td>Angelo Tredennick</td>
</tr>
<tr>
<td>Nicholas Bay</td>
<td></td>
</tr>
<tr>
<td>Division</td>
<td>Humanities</td>
</tr>
<tr>
<td>Link to visualizations</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Where is visualization used?</td>
<td>Oxford University</td>
</tr>
<tr>
<td>Data Source</td>
<td>Oxford Database</td>
</tr>
<tr>
<td>Link to code</td>
<td></td>
</tr>
</tbody>
</table>

Case Study Overview

The aims were to add new interactive visualizations to their relevant database, the interactive maps and charts provided a novel way for users to explore their data.
Oxford Roman Economy Project Case Study

The Oxford Roman Economy Project is hosted through the generosity of Siena School Trustee, previously the AHRC, to study and curate the foundations of the Roman imperial project across its entire economic activities. This case study sought to investigate the geographic and computer visualisation capabilities of data as a tool for exploring two subsets of the Oxford database - shipwrecks and sites.

Oxford aims to make the vast database of Roman sites exploitable through maps & charts. This case study looks at data from shipwrecks & sites.

Case Study Details

- **Department**: Faculty of Classics
- **Academic**: Angela Traverso
- **Title**: Oxford Dataset
- **URL**: [https://www.oxforddataset.com](https://www.oxforddataset.com)
  - [Oxford Dataset](https://www.oxforddataset.com)
  - [Oxford Dataset](https://www.oxforddataset.com)
  - [Oxford Dataset](https://www.oxforddataset.com)

Case Study Overview

Young Lives wanted to add new interactive visualisations to their existing database. The interactive maps and charts provided a novel way for users to explore their data.
Oxford Roman Economy Project Case Study

29 September 2016  |  Angelo Tremontano (Classics)  |  Alastair Ray (Classics)

The Oxford Roman Economy Project is founded through the generosity ofRoyal Oxon. Their kind permission to work and create the fundamentals of the Roman Imperial project across all major economic activities. This user study will to investigate the cartographic and computational visualisation capabilities of data in a test, exploring two aspects of theFord database—shipments and taxes.

Case Study Details

- **Unit of Measure**: Tons
- **Time Period**: 47 AD
- **Group by**: Country
- **Map Types**:  
  - Shipwreck Map
  - Shipwreck Charts
  - Maps
  - Charts

Case Study Overview

Young Lives want to add nine interactive visualisations to their existing database, the interactive maps and charts provide a novel way for users to explore their data.
Oxford Roman Economy Project Case Study

The Oxford Roman Economy Project is funded through the generosity of Susan Lunt. Historically, the research has focused on the economic activity of Roman Britain. In recent years, the research has been expanded to explore the economic activity of the Roman Empire across all major economic activities. This case study aimed to investigate the cartographic and comparative visualization capabilities of the data for exploring pre-existing patterns in the Oxford database – especially in relation to

Oxford please to make the next database of Roman cities explorable through maps & charts. This case study looks at cities from different rich & ancient

<table>
<thead>
<tr>
<th>Case Study Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Department</strong>: Psychology and Economics</td>
</tr>
<tr>
<td><strong>Academics</strong>: University of Oxford</td>
</tr>
<tr>
<td><strong>Division</strong>: Digital Humanities</td>
</tr>
<tr>
<td><strong>Linking to Visualizations</strong>:</td>
</tr>
<tr>
<td><strong>Source</strong>: Oxford Database of Roman Cities (RDC)</td>
</tr>
<tr>
<td><strong>Developer</strong>: Oxford Mosaic</td>
</tr>
</tbody>
</table>

Case Study Overview

Yang Ling wanted to add more information to their database, the interactive maps and charts provided a novel way for users to explore their data.
Online Labour Index Case Study

- The Online Labour Index (OLI) is the first economic indicator that provides an online gig economy equivalent of conventional labour market statistics.
- It measures the utilization of online labour across countries and occupations by tracking the number of projects and tasks in real time.
- This case study sought to provide an interactive dashboard for tracking the change in the OLI overtime and broken down by occupation and worker location.
- It provides timeline graphs filterable by date, occupation, country, and a stacked chart able to be grouped by top 5 or 25 countries and occupation.
- As a result the data was used in a Guardian Newspaper article.
- More: [http://idn.web.ox.ac.uk/article/online-labour-index-case-study](http://idn.web.ox.ac.uk/article/online-labour-index-case-study)
Online Labour Index Case Study

21 September 2016

Viti Levuana’s (Oxford Internet Institute) Ottis Klass (Oxford Internet Institute)

The Online Labour Index (OLI) is the first economic indicator that provides an online gig economy equivalent of conventional labour market statistics. It measures the utilization of online labour across countries and occupations by tracking the number of projects and tasks in real time. This case study sought to provide an interactive dashboard for tracking the change in the OLI over time and broken down by occupation and worker location.

Online jobs in gig economy growing fast, finds the new Online Labour Index - visualised interactively over time, gig sector and worker location.

Case Study Details

Department: Department of History
Author(s): Viti Levuana
External Links:

- [Online Labour Index](https://www.oxfordinternetinstitute.org/online-labour-index/)

Case Study Overview

An online dashboard under heavy usage.
Releasing Code and Data Openly

- Where possible everything is released openly with CC (or similar) license:
  - R code for the Shiny app
  - Data underlying the visualisation
  - Templates for how to create this visualisation for yourself

- Who can display the visualisation is not controlled, this means that:
  - The academic can display it in their project’s webpages
  - And their department/college website (with no extra effort)
  - Reviewers can embed it in the review
  - National press can embed it in online articles

- Interactive visualisations help to bridge the data gap by giving human understandable views on the data that are able to be manipulated but also linking back to that data to enable those interested to find it

- Increasingly funding bodies are insisting the publications from their funded research publish the data that leads to the conclusion (and thus data journals are appearing and can be the data source for visualisations)
Oxford University Press: DataViz Project

- Joint project between Interactive Data Network and OUP
- Aim is to enable publication of interactive visualisations in OUP journal articles in a sustainable manner
- Paper as case study:
  - **Paper Title:** Comparison of genetic variants in matched samples using thesaurus annotation
  - **Authors:** Tomasz Konopka, Sebastian M.B. Nijman
  - **DOI:** doi.org/10.1093/bioinformatics/btv654
  - **Data Deposit:** https://dx.doi.org/10.6084/m9.figshare.4555441.v1 (all data within these visualisations is made available here)
  - **Visualisation Developer:** Martin John Hadley
  - **Motivation:** Calling changes in DNA, e.g. as a result of somatic events in cancer, requires analysis of multiple matched sequenced samples. Events in low-mappability regions of the human genome are difficult to encode in variant call files and have been under-reported as a result. However, they can be described accurately through thesaurus annotation—a technique that links multiple genomic loci together to explicate a single variant.

- Creating policies for issues such as how best to embed the visualisation, ongoing maintenance for these, and whether it should be permanently interactive (or only for a set number of years, replaced with static image later)
Where Does the Data Live?

For the Interactive Data Network we are exploring a wide variety of data sources:

- **Data Repositories**
  - DOI-providing data preservation archives; Rich metadata containers; APIs
  - ORA Data (UOxford Institutional Data Repository), Centre for Open Science, FigShare, Zenodo, etc.
  - Potential for embargoes where required

- **Hosted Databases**
  - Provide hosted databases with SQL-based query access over ODBC or similar
  - Institutional databases, cloud-hosted services such as Microsoft Azure, Amazon Web Services, Google Cloud and others
  - Full access control lists by various criteria (for those with access to write, or grab the raw data)

- **File Hosting Services**
  - Flat file hosting services ranging from sync-and-share services to static file repositories
  - Services such Dropbox, Google Drive, etc.
  - Potential limitation on size of dataset
  - Potentially controlled access depending on service
SNV called position
Mutation annotated position
x/y Estimate of AF [x/y] Estimate of AF forced by multi-sample analysis
Performance for calling somatic mutations from synthetic normal tumor sample pairs.

True positive rate (TPR) of various mutation detection approaches (note vertical axis does not start at 0 to emphasize the practically relevant TPR range). Dots represent synthetic tumor samples with different mutation allele frequencies (AF). The approaches are Muped, a local mutation calling at two mapping quality (MQ) thresholds, and thesaurus-assisted mutation calling at two mappability thresholds. Other methods are discussed in the Supplementary Text.

False discovery rates (FDR) for the same methods. The gray band indicates results with zero false positives and a small number of true positives, a performance regime that produces spuriously high FDRs based on Eq. (4).
Performance for calling de novo mutations in a synthetic family trio.

True positive rates (TPR) of various calling approaches. Dots represent actual calls at various thresholds for observed allelic frequency (AF); lines are simple interpolations. The approaches used to select candidates are local mutation calling at two mapping quality (MQ) thresholds, and the thesaurus assisted equivalents.

False discovery rates (FDR) for the same methods. The gray band indicates experiments with a small number of true positives and no false positives at all, for which Eq. (4) gives spuriously high FDR values.
Results on de novo mutation calling from a family trio from the Platinum dataset.

Family pedigree for sample NA12882. R1 and R2 denote two replicates of the child’s genome.

Concordance of de novo mutation calls in two replicates of sample NA12882 obtained using a high (MQ 15) mapping quality threshold. Axes display the allelic frequencies (AF) of candidate sites in the two replicates.

Concordance of de novo mutation calls identified through thioaur analysis of low mappability genomic regions (MQ 0) and not in the previous analysis. Allelic frequencies on the axes are averaged over all thioaur-linked genomic sites. (Thioaur-adjusted AF > 1 are plotted at unity.)

Analogous to panel the panel above, but with allelic frequencies on the axes computed using data from only one genomic locus per mutation candidate.
Long-term Preservation and Interactive Visualisations

- One of the problems with interactive visualisations is that they require infrastructure that needs to be maintained
  - The site hosting the project/edition needs to continue to be able to embed the visualisation
  - The visualisation engine still needs to be running
  - The data source still needs to be available
  - The various javascript and other libraries need to be updated (in cases of security updates)
- Most of these are slightly mitigated by institutional adoption but this is not a guarantee
- While the service runs it provides an unlimited account on ShinyApps.io
- Data should be deposited in well-known repositories for long-term preservation (we recommend our own institutional repository, FigShare, and Zenodo)
- For some publications:
  - Static image links to interactive visualization
  - Interactive visualization eventually replaced with static figure
Digital Scholarly Editions and Data

- We need to admit that our digital scholarly editions contain a lot of useful data other than the edited text.
- We need not only to expose the underlying data of our editions, but also adjunct derivable resources, and present these all in a visually communicative but interactive manner.
- We need to plan for obsolescence of all our digital tools and outputs in such a manner that the intellectual endeavour does not need to be repeated.
- We need to work collaboratively to meet these and similar challenges.

james.cummings@it.ox.ac.uk

@jamescummings