Visualising data stories together: Reflections on data journalism education from the Bournemouth University Datalabs Project

Anna Feigenbaum, Einar Thorsen, Daniel Weissmann and Ozlem Demirkol. Bournemouth University

Introduction

More and more journalists are turning to data-driven storytelling techniques as a way to both interpret and communicate news using data. To date, discussions regarding the expanding field of data journalism, and the associated novelty of data storytelling tools and practices, have primarily taken place in industry forums and blogs, rather than in academia.

The earliest forms of data journalism go back to the late 18th and early 19th century reporting on domestic trade, education figures, and mortality rates (Rogers, 2013). With the rise of the computer in the 1950s, data driven journalism or Computer-Assisted Reporting (CAR) as it was referred to, emerged. CAR mainly focused on statistics in answering journalistic question (Lewis, 2015). It wasn't until the 1980s and 1990s that data driven journalism made its way into the newsroom after Pulitzer Prize winning stories paved the way for a wider use and the acceptance of data journalistic techniques (Coddington, 2015). Today, innovation in data journalism predominantly takes place in the newsrooms of the Guardian, the BBC, the New York Times and the Washington Post. Due to the costs of doing high quality, innovative data journalism, these well-funded newsrooms continue to lead the way.

Offering a useful summary of recent developments in data journalism education, Hewett (2016) argues that higher education institutions have been slow to incorporate digital training and the rise of data journalism into their curricula. While industry publications are calling for these skills, as outlined in reports by Nieman Lab, Journalism.co.uk and PBS MediaShift, the formal adoption of such pedagogy remains limited in both the US and the UK. Filling the gap in professional knowledge bases, projects including the Data Journalism Handbook (http://datajournalismhandbook.org/1.0/en/) and the Data Driven Journalism website (http://datadrivenjournalism.net/) offer open educational resources for self-directed learning (Hewett 2016).

Yet these online tools alone cannot, and should not, take the place of enhancing the journalism education offered in Higher Education. Addressing the "constraints and influences faced by journalism education in teaching data journalism" Hewett (2016) summarises the particular interdisciplinary challenge that data journalism faces:

The development of data journalism education may be impacted more profoundly and/or immediately than other areas of journalism education because of its interaction with other fields – notably statistics, computing, data science and visualization. For 'mainstream' journalism education, these may be an optional extra; for data journalism, they relate to its core functions (p.3).

As news rooms become increasingly interdisciplinary, so too must journalism education. Research in major newsrooms conducted by Weber and Rall (2012) emphasised this significance of knowledge exchange and skill sharing across disciplines as journalists, designers and programmers collaborate.

In order to embrace collaboration and begin to address the interdisciplinary challenges facing data journalism education, in January 2015 we established the Bournemouth University Datalabs project. The aim of the project was to develop a sustainable, iterative model for co-creating a curriculum on data storytelling

for journalism education. To accomplish this aim, we brought together a multidisciplinary, cross-faculty team of researchers and students that worked in collaboration with journalists and external stakeholders. We combined expertise from journalism studies, media and communications, geography, computer science and data analytics to identify and problem-solve key issues in data storytelling, particularly around questions of data gathering, analysis and visualisation.

Using an action research approach (Greenwood and Levin 2007), we designed a programme of activities that could work across existing knowledges to co-create "the ground for new learning for all participants" (p.107). The activities were organised around key problems in the field as they are encountered through the data storytelling process: data gathering and cleaning, analysis and visualisation, including geographical and social network mapping. In the workshop space, as well as afterwards through survey responses and recorded team reflections, knowledge generation arose out of participants' "conscious attempts to solve practical problems" (p.108). This active research approach enabled journalism students to work alongside researchers, professional journalists and NGOs to explore the possibilities and challenges associated with the communicative power of visualisations, maps and related interactive digital media for telling data stories on sensitive subjects of societal importance, such as policing, crisis, migration and human rights.

In this article we briefly reflect on current educational and professional challenges facing the emergent area of data storytelling and offer a pragmatic, reflective analysis of our Datalabs project. Covering the opportunities and obstacles our participatory workshop model provides, we share some key insights that arose out of our collaborations toward building interdisciplinary journalism education. We conclude by exploring how this model can be utilised by other universities and organisations in order to encourage the growth of data journalism within journalism education and training.

The Challenges of Data Storytelling in Journalism

With the recent growth of open data and big data, it has become more important for journalists to understand and access the datasets made available by governments and organisations in order to use this newfound commitment to transparency as a way of holding power to account. Likewise, Freedom of Information laws have created new opportunities for journalists to access data (Bowles et al 2014). However, datasets are often large, messy or complex, making it difficult to interpret and analyse in order to find the stories that matter. This is one of the key challenges that face the growing field of data journalism. Paul Bradshaw (2015) discusses how the rise of such novel practices, including "networked and interdisciplinary modes of collaboration; global publication; an increased reliance on visual communication; and new forms of user-driven storytelling using interactivity and/or personalisation" bring up new conflicts between ethical principles (p.203). Proactive principles of seeking truthful and accurate information must be balanced with restraining principles related to privacy, protecting sources and accountability.

Echoing some of these concerns, with big data particularly in mind, Boyd and Crawford (2012) stress the importance of signalling that big data is not always better data. Issues of accuracy and claims to objectivity can be misleading. Indeed, journalists' fraught relationship with statistics is both longstanding and well documented, and reflects "a vast misunderstanding, underestimation and ignorance of the nature and the role of statistics in daily news work" (Nguyen and Lugo-Ocando, 2016 p. 4). Despite this, statistical analysis rarely forms part of any journalism programmes, neither in Higher Education nor within professional training courses. That is, ensuring journalists move beyond use of statistics as a rhetorical device (McConway, 2016), to developing "a permanent determination to question data and a basic level of statistical reasoning" (Nguyen and Lugo-Ocando, 2016 p. 5) in new and innovative ways (Hewett, 2016).

Likewise, in their comparative assessment of data journalism education across six European countries, Splendore et al (2016) concluded that future journalism training in Higher Education should provide knowledge and skill acquisition around issues of transparency, accountability and journalism ethics. This new terrain of data journalism creates new challenges for journalism educators who must make extra effort to balance how we teach on issues of transparency, accuracy and stylised storytelling in data-driven reporting and visualisation.

Challenges to Gathering Data

While open data initiatives have led to a proliferation of what kinds of stories people can tell with data,

sensitive issues often have no straightforward data source. Where documents are available, they are often scattered across agencies and organisations. Moreover, data on sensitive topics is often kept hidden, deemed too confidential to be made open. This 'uneven transparency' raises important questions about the duty to document (Larsen and Walby 2012), particularly in regard to security issues where obtaining health and human rights information on vulnerable populations (for example prisoners, detainees, or those living in conflict zones) becomes difficult, trumped by a greater interest in keeping data protected for proclaimed 'national interests'.

In other cases, data is simply not recorded or not recorded in any structured way. Efforts in recent years to aggregate police killings in the United States are one prominent example of how data can be effectively crowd-sourced. For this project information was mined from public media and police reports, then verified through online and offline sources by journalists on the Guardian's 'The Counted' project, in efforts to create an interactive dataset (Gray et al 2016). While incomplete and likely dotted with inaccuracies, in such cases some data is better than no data (or no disclosed data). The participatory process of collecting this data, and then curating it through visualisation in a way that was both emotional and user-friendly, engaged a diversity of readers, including campaign groups like Black Lives Matters, and public health researchers at Harvard, who called for police killings to be deemed a public epidemic. Such citizen generated and civil society data can be used as an advocacy tool, putting pressure on organisations and governments to establish better public data protocols, and at times, shape new practices and policies (Gray et al 2016).

Challenges to Visualising Data

Employing visual narratives, interactives and maps can help vitalise complex data sets. Research has found that visualising data can help audiences process complex information and enables them to make comparisons, while personalisation aids engagement and connection with the issue (Green and Myers 2010). Research in this emergent area has pointed to the importance of bringing narrative and rhetorical skills into data visualisation practices (Hullman and Diakopolous 2011), as well as to the need for data visualisations to be well contextualised and situated in broader articles (Kosara and MacKinley 2013). Delineating types of visualisations and testing their success on readers, Segel and Heer (2010) have called for more research into reader engagement, noting the promise of eye-tracking studies and monitoring user interaction with data visualisations.

Recent work by Helen Kennedy and the Seeing Data team (visualisingdata.com) is paving the way toward audience studies of data visualisation. Combining focus groups that used Talking Mats with visualisation diaries, the Seeing Data team found that statistical and language literacies, confidence, available time, previously held opinions and the source of visualisations all effected how people see data (Kennedy et al 2015). Echoing previous research (Segel and Heer 2010, Hullman and Diakopoulos 2011), the Seeing Data team highlighted the importance for data visualisation design to convey emotion, trust and well thought out user-experience (Kirk et al 2016). Such research on reception studies can usefully inform data journalism pedagogy and professional practice.

Just as research on data visualisations remains sparse, there is also a gap in the literature on cartography in the media. The use of static and interactive maps in journalism is proliferating, yet research on cartography remains almost exclusively within the field of geography. Little contemporary work has been published examining connections between cartographic practice and their connection to media audiences or journalism education (Churchill and Stege 2006, Demirci and Zeliha 2015). This means that while maps are proliferating in the news, there is currently a lack of journalistic education on how power inequalities can get embedded and reproduced through these visualisations. As Doug Specht (2015) writes, "there has been much research into the modes of knowledge construction through geographical and cartographic artefacts, and much of this points to classical linear constructions of knowledge by experts that is then imposed upon the other" (Specht 2015, p. 236). To understand maps from a critical perspective it is important to learn about how spatial knowledge and geographic information is codified and symbolised (visually represented). Yet the historic links between cartography, colonisation, military invasion and Western Imperialism rarely comprise part of journalism or data journalism education. Incorporating critical cartography into journalism education is a crucial component of enhancing data storytelling and visualisation practice in data journalism.

In relation to digital cartography, online platforms like Google Maps have transformed ways in which people relate to visual representations of geographic space. From Google's Street View to ArcGIS's StoryMaps

platform, people are engaging their geographical imaginations to produce personalised visual representations of spaces that matter to them. Likewise, people use RSS feeds and APIs to dynamically draw information from web and social media data sources to create real-time map visualisations and mash-ups (Crampton 2009). Such interactive visualisations are often populated with user-generated or crowd-sourced content, engaging the activity of citizen scientists and citizen journalists. For example, maps that allow a reader to drill down into how issues affect their local area, tap into individuals' curiosity and motivate them to engage more deeply with a story (Wilson 2011). These practices open up new opportunities for civic authorship and challenge the idea of cartography as fixed, objective and authoritative (Dodge and Kitchen 2013). Yet, without journalism education that incorporates critical cartography, all this digital map-making runs the risk of reproducing unequal and marginalising power relationships.

This increased emphasis on mapping and visualising data has brought with it the need to pay more attention to the importance of understanding design principles (Segel and Heer 2010). Storytelling with data involves implementing design techniques and expressive practices in order to give data context and meaning in ways that connect to different audiences. Alongside basic principles of graphic design, iconography, colour theory and user experience design must be brought together (Kirk 2012). Researchers have found that as with advertising or the presentation of statistics more broadly, data visualisation has also been found to enhance persuasion, particularly when viewers do not hold strong initial attitudes on a subject (Pandey et al 2014). This is an important area of consideration for journalists.

In contrast to more traditional areas where data visualisations are used, such as economics or scientific visualisation where graphics are intended for highly trained audiences (Gershon et al 2001), in news journalism visualisations are targeted at wider publics. This means that visuals must communicate complex information to everyday audiences that often do not have specialist expertise in these areas. News audiences come with differing types and levels of personal skills, education, and tastes that journalists need to accommodate in their production of data visualisations. For these reasons, Hullman and Diakopoulos (2011) call for an increased responsibility for designers to consider how visualisation choices can manipulate user interpretations.

Challenges to Skill Development

In addition to the specific challenges of gathering and visualising data, like the introduction of any new skills in a newsroom, visualising data requires time and resources. While we have seen a rise in the number of free and user-friendly software tools, handbooks and online resources in recent years, financial and time pressures facing both newsrooms and the education sector make it difficult to integrate these new techniques and tools into everyday practices and routines (Hewett 2016, Splendore et al 2015). Moreover, unless a news organisation is already aware of new technologies in the field of data journalism, they are unlikely to be able to access this software and training material in the first place. Traditional structures for journalism training and funding often prohibit exposure to or adoption of new technologies and practices.

The Bournemouth University Datalabs Project

Addressing this emergent field of data storytelling and its importance to the future of journalism education, the aims of the Bournemouth University Datalabs project were to:

Bring together a multidisciplinary, cross-faculty team of media, data science and GIS researchers and students to work in collaboration with journalists and digital designers.

Establish a sustainable training model for data literacy, data-driven research and data storytelling.

Co-create resources and output targeted at journalists to maximise impact.

Our team at Bournemouth University worked with our NGOs, journalists and digital designer stakeholder partners to run hands-on data aggregation, visualisation and digital storytelling workshops designed specifically for addressing civic and humanitarian issues. The BU team comprised of:

Dr Anna Feigenbaum, PI (Senior Lecturer in Digital Storytelling)

Dr Einar Thorsen, CI (Principal Lecturer in Journalism and Communication)

Dr Phillipa Gillingham, CI (Senior Lecturer in Biogeography)

Dr Shelley Thompson (Senior Lecturer in Corporate and Marketing Communications)

Dr Dan Jackson (Principal Lecturer in Corporate and Marketing Communications)

Dr Nathan Farrell (Lecturer in Communication and Media)

Dr Hamid Bouchachia (Associate Professor, School of Design, Engineering and Computing)

Dr Brad Gyori (Senior Lecturer in Digital Storytelling)

Oz Demirkol (Senior Research Assistant)

Daniel Weissmann (Senior Research Assistant)

Laura McKenna (Research Assistant)

Alongside the interdisciplinary team at BU we worked with:

CVG Design - a New York based Graphic Designer specialising in community art as urban intervention.

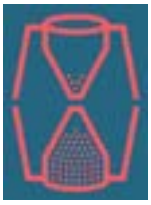
Jay Cassano - an independent journalist covering the intersection of technology and politics.

Julio Molina Montenegro - an award-winning filmmaker with over ten years of international experience in the fields of documentary, sound recording, post-production and media technology.

Our Partner Organisations

Coordinated by John Horne, a doctoral candidate at Birmingham University and organiser at the NGO Bahrain Watch, our diverse stakeholder partners came from journalism organisations, digital design teams and human rights NGOs.

RECLAIM BRIXTON	**BAGGAGE**	Brixton based community initiative monitoring neighbourhood gentrification and housing struggles.
BAHRAIN WATCH	**Bahrain Watch**	An independent organisation that seeks to promote democracy, equality and social justice in Bahrain, through evidence-based research investigations and advocacy.
THE BUREAU OF INVESTIGATIVE JOURNALISM	**Bureau of Investigative Journalism**	Research, investigations, reporting and analysis which is of public benefit by undertaking in depth research into the governance of public, private and third sector organisations and their influence.
CAMPAIGN AGAINST ARMS TRADE	**Campaign Against Arms Trade**	UK-based organisation working to end the international arms trade.

Corporate Watch	**Corporate Watch**	Independent research group, investigating the social and environmental impacts of corporations and corporate power.
CORRUPTION WATCH	**Corruption Watch**	An investigative organisation that details and exposes instances of corruption and their subjective and objective impact on democracy, human rights and development across the world in order to precipitate strong action against it.
IRIN	**IRIN News**	Independent news agency delivering unique, authoritative and independent reporting from the frontlines of crises to inspire and mobilise a more effective humanitarian response.
	**Minute Works**	A multidisciplinary graphic design studio specialising in sustainable communications.
	**Occupy Design**	UK collective set up as part of the Occupy movement in London in 2011. Organises workshops and the production of visuals for, and with, a number of grassroots social movements. Crisis graphics for crisis times!
Omega Research Foundation	**Omega Research Foundation**	Non-profit research foundation providing rigorous, objective, evidence-based research on the manufacture, trade in, use of, military, security and police (MSP) technologies.
reported.ly	**Reported.ly**	A global team of journalists covering stories of international importance through social media and citizen networks with a focus on social movements, civil rights, conflict and human rights.

SMALL MEDIA	**Small Media**	An action lab helping the free flow of information and creative expression in closed societies, with training, technology and research initiatives that focus in Iran
TACTICAL TECHNOLOGY COLLECTIVE	**Tactical Technology Collective**	Non-profit organisation, working since 2003 to advance the use of information and digital technologies by advocates and activists worldwide.
Tear Gas Research Connection	**Tear Gas Research Connection**	Research project bringing together existing knowledge around tear gas and the impacts less lethals have on people and their lived environments. Working with international researchers, NGOs, journalists and tactical technologists, we contribute to news reports, public debate and policy-making on the safety and social impacts of tear gas technologies.

Original Datalabs Workshops

As a team, the first set of workshops we hosted were three 'Datalabs' on-site at Bournemouth University. Each Datalab ran for two days, consisting of an interactive masterclass and a discussion-based workshop on Day 1, followed by a guided computer lab session on the afternoon of Day 1 and into Day 2. During these lab sessions, students, staff and external stakeholders worked on live datasets using free and open source software for data analysis and storytelling.

Data Scraping and Cleaning – Monday March 16 and Tuesday March 17, 2015

This masterclass and workshop introduced participants to tools and techniques for scraping data from the web. We went over how you format and organise data in spreadsheets. Discussing basic methods for data storytelling, the workshop provided training in how to find data stories within datasets.

Datasets: TheyWorkForYou.com data on MP profiles

Tools: import.io & openRefine

Data Visualisation with Mapping – Monday April 27 and Tuesday April 28, 2015

This masterclass and workshop introduced participants to a variety of mapping techniques and technologies. Offering a beginner lesson in the principles of GIS, participants learned how to work with base maps, data layers and symbolizing to present spatial representations of data. On Day 2, participants were introduced to social network mapping, using Twitter APIs to map the use of hashtags.

Datasets: UK Parliamentary Constituencies

Tools: QGIS & CartoDB

Digital Storytelling with Data – *Thursday May 14 and Friday May 15, 2015*

In this masterclass and workshop session we introduced participants to the emerging interdisciplinary field of data storytelling. We explored different forms of digital narratives and discussed the principles behind storytelling with data for diverse audiences. In the workshop session we reflected on the ethical questions that come with turning numbers in narratives. Following this we delivered a hands-on lab session digging for stories in a data set that could be visualised using the tools we had introduced at the first two Datalab sessions.

Datasets: MP travel to MENA region

Tools: import.io, openRefine, CartoDB

Subsequent Datalabs Activities

Drawing on our experiences with our initial Bournemouth-based Datalabs workshops, we then hosted a series of events both at Bournemouth University and in London. These events were designed around the project aims, bringing together a wider network of partners from the local and national community.

Mapping for Justice

2 June 2015, with Richmond, American University in London

In this public showcase and discussion, we explored how investigative reporting on a wide variety of social issues, from drone strikes to council house evictions, can use mapping and GIS techniques to tell data-driven stories. From oral history to Twitter data-mining, our featured mapmakers use a range of methods to make social issues visible. We discussed how - whether made with paper and pen, or on open source digital platforms - maps can reach broad audiences and engage with people's geographic imaginations. Attended by 80 people, this public engagement event was co-hosted with Richmond the American International University in London.

Turning Numbers into Narratives: An introduction to digital storytelling with data

14 May 2015, Interdisciplinary Research Week, Bournemouth University

In this masterclass, we introduced participants to the emerging interdisciplinary field of data storytelling. Drawing on a range of contemporary examples from documentary journalism, we presented a variety of digital narrative structures that can be used to tell stories with data. Guest speaker Malachy Brown from reported.ly, who use social media to report on human rights and social justice around the world, discussed how journalists can use social media tools to conduct investigative research and find news beyond the mainstream. With over 50 participants from across all faculties of the university, as well as outside, our event drew together insights from the fields of communications, journalism, geography, psychology and data science.

Data Storytelling Workshop

12 November 2015, ESRC Festival of Social Science, Boscombe

In this masterclass and workshop, we introduced participants to the data collection tools and methods of data storytelling. Using a dataset from migrant crisis case, we discussed the differences of good and bad data, inaccuracies in data recording and other problems journalists face while working with data. After presentations on narrative structures and character creation, drawing from the same example, participants were given different migrant crisis actors and asked to work in teams to write a story. With the participation of 25 professionals from NGOs, academia and local councils, we wrapped up by discussing the challenges and strategies in pulling together data from Web.

BU Datalabs Training - Policing and Human Rights

13 January 2016, Centre for Excellence in Media Practice, Bournemouth University

In this training workshop, we introduced undergraduate and postgraduate journalism and politics students to the ways of finding relevant and accurate data on policing and human rights through the use of advanced Google searching. We then provided training in how to navigate published FOI requests and their provided documentation for producing data-driven stories. At the workshop we also covered some of the pitfalls of data collection in sensitive areas, such as policing and human rights, in relation to how the reliability of data and challenges of working with unstructured vs structured data.

BU Datalabs Hackday - Policing and Human Rights

9 February 2016, Journalism School, Bournemouth University

At this hackday students who attended the training session put their new skills to work, producing a report commissioned by the Council of Europe. During the hackday we investigated reports on member states of the Council of Europe between 2006 and 2016, digging into human rights investigations on the excessive use of tear gas on peaceful and civilian protesters. Our final report detailed key findings regarding human rights concerns, deaths and injuries related to tear gas, and the lack of adequate and transparent data recorded on police use of force. Recommendations made in this report, with our stakeholder partner Omega Research Foundation, were subsequently adopted by the Council of Europe.

Creating the Datalabs Format

We designed our datalabs activities to adopt the principles of a hackerspace and of an open-space learning environment. Combining the collaborative, problem-based dimension of a hackerspace with a lab-based learning format, our workshops offer participants hands-on experience and insight into applied use of data storytelling tools and techniques. After the workshops we solicited feedback from participants using survey monkey. Facilitators held a de-briefing meeting where oral feedback on the events was recorded through minute taking. We also followed up with our stakeholder partners, receiving updates on their uptake and use of data journalism techniques and technologies in their storytelling practices.

Datalab as Hackerspace

Traditional hackerspaces were first established in Germany in the late 1980s as “places in the community where local programmers can collectively meet, work, and share infrastructure” (Borland 2007). Based around a community of programmers with an interest in developing their computer skills, these hackerspaces were relatively homogenous, drawing together people with similar backgrounds and experiences (Cavalanti 2013). This shared work culture and background among participants created a high creative output as the programmers could work creatively with a community of complementary skill sets, mind-sets and languages, to co-create and learn from each other. With the hackerspace in mind, we first brought our partners and participants together to present the tools we would be using in the hackerspace format the following day.

However, unlike a hackerspace, where participants typically have a similar proficiency and experience with digital communications technologies, our Datalabs aimed to accommodate a broader range of abilities and backgrounds. With participants coming from journalism, geography and data science, skillsets and vocabularies were highly varied. This meant utilising scaffolded tool training (Hmelo-Silver et al., 2007), in which we took participants on a step-by-step guided lesson in how to use data gathering and visualisation tools at the beginning of every workshop.

Challenges & Opportunities

McLoughlin and Lee (2008) argue that during interdisciplinary workshops: “learning focused on knowledge creation and networking, offers the potential for transformational shifts in teaching and learning practices, whereby learners can access peers, experts, the wider community, and digital media in ways that enable reflective, self-directed learning” (2008 p19). This aspect of a transformational shift in learning practices was evident as we had to guide our partners and stakeholders not only on the same technical skill level in relation to computer programing and computer literacy, but also to a shared understanding of the language and vocabulary used across fields of digital media, GIS and data science. Even seemingly basic terms are prescribed different meanings across these approaches to working with data. For our partners and stakeholders to communicate effectively within our interdisciplinary workshop environment, shared definitions and understandings were key. Athough it comes with challenges, this focus on knowledge co-creation and networking is facilitated by personalisation, participation and productivity (McLoughlin and Lee 2008).

There is personalisation in the form of learner choice, learner agency and customization. The open-space aspect of Datalabs provided these features, particularly during the day two sessions where participants could work autonomously, or in self-selected groups, on their own datasets.

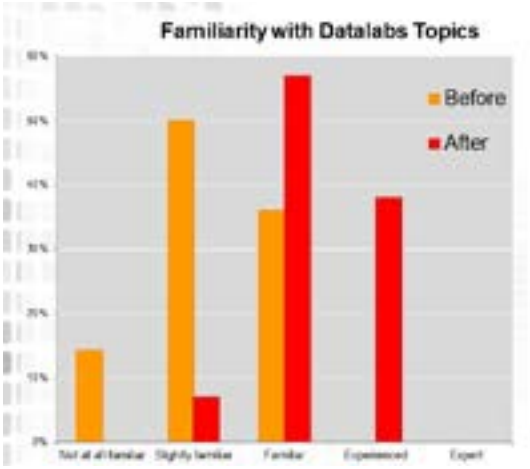
Furthermore, autonomy did not hinder participation. As there was no instruction on the second day, Datalabs' participants encouraged communication and collaboration. This autonomous, hackerlab style time and space, encouraged partners and stakeholders to test and apply their knowledge and skills gained during the scaffolding learning sessions. Here they moved beyond using the provided sample datasets to generate and experiment with their own datasets (or indeed generating new ones) both individually and in collaboration with other Datalabs participants.

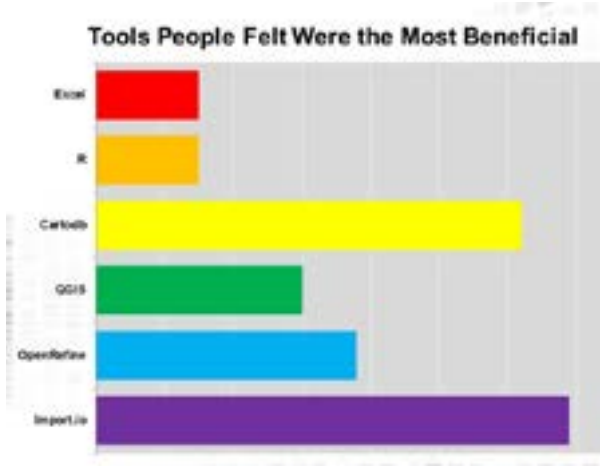
McLoughlin and Lee (2008) also emphasise that the purpose of this approach to pedagogy "is to enable self-direction, knowledge building, and learner control by providing options and choice while still supplying the necessary structure and scaffolding". As the community "spend[s] time together, they typically share information, insight, and advice. They help each other solve problems. They discuss their situation, their aspirations, and their needs. They ponder common issues, explore ideas, and act as sounding boards." (Wenger et al., 2002 p4).

This description by Wenger (2002) was reflected in our experience during the self-directed, open-space learning during day two of Datalabs. Overall, and as evidenced in participant feedback, we found that although the mixed format of scaffolded learning on day one plus open-space learning on day two seemed quite challenging at the beginning, once participants became comfortable with the format, the blend of taught and autonomous lab sessions worked well.

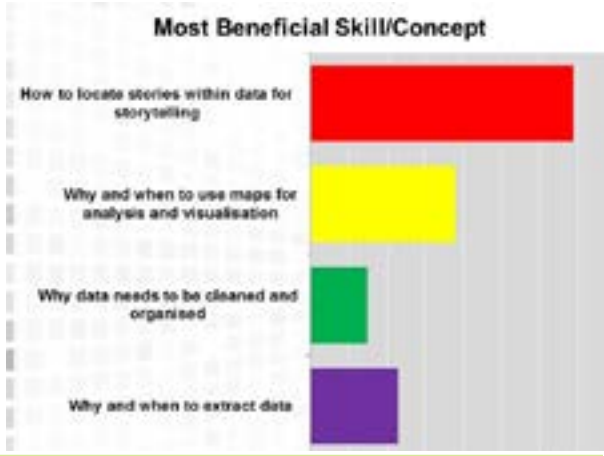
Feedback from our Datalabs participants showed that most people's familiarity with the tools and concepts taught during the workshops increased substantially, with most participants becoming familiar or experienced by the end of their sessions (below left).

In terms of the tools participants found CartoDB and Import.io most beneficial (below right).





Whereas, in terms of concepts and skills participants found it most beneficial how to locate stories within data for storytelling (right).



Stakeholder Output

Our stakeholders' output varied as different organisations have different needs and resources in terms of adopting tools and techniques covered in the Datalabs. Here we want to highlight two organisations that stand out in terms of building on the knowledge gained in the Datalabs workshops and applying these skills to better illustrate the issues they are concerned with.

Omega Research Foundation

Omega's focus is on the manufacture, trade and use of military, security and police technology. They work with news feeds, arms expo data and trade data. This made tools such as Import.io for data extraction and Open Refine for cleaning datasets, most benefical for them. By adopting these tools, they are now pulling data from exhibition lists and RSS newsfeeds, and comparing datasets from different sources. This has allowed them to create more responsive blog posts and press commentaries that can follow the news in real-time. For example, using StorymapJS, Omega aggregated stories on use of force in the Occupied Territories during an escalation of conflicts in September-October 2015. Using geo-location and photography to construct a narrative of force escalation, the storymap brought the feel of force escalation to life.

IRIN News

A humanitarian news organisation, IRIN focuses mainly on crises and humanitarian catastrophes. Although having made use of tools for extracting and cleaning datasets which were part of the early Datalabs session, IRIN primarily uses existing statistics from humanitarian organisations. Their primary aim is however, to narrate and contextualise data, embedding visual and interactive elements into their stories. This helps build audience engagement with what they are reporting. Therefore, tools such as CartoDB and Tableau were most suitable to capitalise on visualisation and narrative aspect of data storytelling.

Since our Datalabs workshops, IRIN has increased its use of data visualisation and data-driven storytelling. Using data provided by the World Health Organisation, IRIN News was able to create a map indicating regions and countries where the Zika virus is being transmitted and those where the outbreak is over. In addition, using data provided by Nature, IRIN News was also able to create a map that shows the spread of the Aedes aegypti mosquito, the mosquito responsible for transmitting the Zika virus. The use of these mapping tools allowed IRIN to link two different data sets, that of the Zika virus outbreak and that of the spread of the Aedes aegypti mosquito.

Key Insights

Creating visualisations and stories with data comes with a number of challenges. Throughout our experience with the Datalabs project, we aimed to confront, explore and address these challenges, leading to crucial insights into both how we think about data storytelling, as well as how we put it into practice together. Below we highlight four areas that emerged from our work with Datalabs that warrant further consideration in relation to data storytelling - either within the context of journalism education or practical application within industry.

Different Tools for Different Stages

Teaching tools and techniques does not only mean training people in the use of each individual tool, but also, helping people understand how these tools are used in combination with each other along the process of data-driven storytelling. Different stages of the data storytelling process require different tools. Sometimes there is a linear flow—from scraping and cleaning to visualising—but other times you must return to your dataset and dig for new stories or angles. This requires a flexible approach to the use of different tools at different stages, as well as the ability to identity which tool or technique is most appropriate for the data available.

Choosing the Right Type of Visualisation

Different data stories require different visual languages and storytelling techniques. Not every dataset can be or should be visualised into a map just because it has geo-locations in it. Nor does a good visualisation necessarily have to be digital. Analogue visualisations can be just as powerful as digital ones, if not more so (Cohen 2015, Rivas 2015). The challenge is not only to be able to use the tools proficiently, but also to know when to use which tools and what its limitations are for the story you are trying to tell.

One of the most significant challenges in data-driven storytelling is sequential and cumulative distortion. In every stage of the data storytelling process, from gathering information to circulating a visualisation on social media, distortion almost inevitably comes into play. This might arise from missing data, mis-recorded data, or information displayed out of context. It can happen when designs exaggerate representative symbols, for example the sizes of bubbles or the scale of bars. Distortion might also occur in the analysis, where subject positions shape the interpretation of findings and this becomes codified in the data story.

Distorted data

In the process of our project we began to imagine this problem as a layer cake. Each stage of the data storytelling process can add these new layers of distortion. While from a storytelling and design perspective, the motivation can be to frost the cake beautifully, covering over crumbs and smoothing out surface. But

ethics and responsible practice demand that while we want to make the cake as presentable as possible, we should not be afraid to reveal the layers inside. It is only by slicing the cake open for our audiences, and providing them with the recipe, that we can foster a transparent and accountable culture of data storytelling.

Visualising Sensitive Subjects

In addition to these data challenges, working with data that deals with human rights abuses, corruption, torture and crisis, provides important insights into the particular challenges that telling data stories about sensitive subjects brings. For these sensitive issues, in which the subjects represented already face difficulty in having their voices heard and experiences shared, digital storytelling comes with additional risks. When working with charts, maps, iconography and pictograms, it is particularly challenging to represent and visualise complex issues like death, torture or suffering in a way that is both accurate and ethical. When giving visual narrative to such sensitive topics, it is crucial to approach visualisation design with respect for those affected, to protect privacy when necessary, and to avoid turning victims into numbers in ways that lose their humanity.

Conclusion

While steps have been taken in recent years to build data journalism curriculum, there remains a need for more reflection on how best to teach data storytelling skills and technologies diverse journalism education environments (Hewett 2016, Splendore et al 2015). Within this article we have sought to outline our experiences of developing an interdisciplinary model, alongside external stakeholders and journalism students, to co-create data storytelling curriculum for journalism education. Our aim has been to illustrate both the pedagogical goals and practical considerations to their implementation, so as to illustrate how such a model could be adapted by other universities and organisations in order to encourage the growth of data journalism within journalism education and training.

At the heart of developing data journalism education is the task of balancing truth, accuracy and transparency with engaging storytelling for news audiences. It is crucial that we educate next generation journalists not just to tell data stories, but to tell them responsibly. In other words, to encourage journalists and journalism educators not to just serve the cake, but to serve the cake sliced open. To mitigate distortion, we advocate transparency around data storytelling processes and data sources. Where it does not unduly jeopardise privacy, publishing a guide to your process of data storytelling, along with the raw datasets used, is good practice and helps the data storytelling community reflect and grow. Readers should be able to see what goes into the data storytelling process. This includes notes on your sources (collection), transparency about limitations on what the data covers and a reflection on any assumptions that you made about data validity, calculations and correlations (interpretation), as well as notes on how design and usability choices have influenced the presentation of data (visualisation).

This ‘slicing open’ is not meant to predetermine audience interpretations, but rather should be seen as a virtue of data-driven storytelling. To further capitalise on the potential for different audience interpretations, source data can be made available in ways that encourage people to create their own visualisations and analyses. Taking personalisation and drilling-down further, readers are able to mash-up and manipulate original data. These user practices can create further insights and initiatives. The question then, is how to best celebrate these of forms of remediation, while mitigating the pitfalls of distortion? One potential solution for this would be to encourage users to display the same transparency and caveats in their re-mediations as are included in the source material. Such a system would be akin to, or potentially even utilise, creative commons attribution licences.

References

Amin, A., 2002. “Spatialities of globalisation.” *Environment and Planning,* 34(3), pp.385 – 399.

Anderson, C., 2008. “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete.” *Wired Magazine* 16(7).

Borland, J., 2007. “Hacker Space” Movement Sought For U.S. *Wired*, [online] 8 November. Available at: <http://www.wired.com/2007/08/us-hackers-moun/> [Accessed 27 July 2015].

Boyd, Danah, and Kate Crawford. 2012. ‘Critical Questions For Big Data’. Information, Communication

& Society 15 (5): 662-679.

Bowles, N., Hamilton, J. T., and Levy, D. A. L. (eds) (2014) *Transparency in Politics and the Media: Accountability and Open Government*, London and New York: I. B. Tauris & Co

Bradshaw, P. (2015) 'Data journalism' in Zion, L. and Craig, D. (eds) *Ethics for Digital Journalists: Emerging Best Practices*, New York and Oxon: Routledge, pp. 202-219

Cavalcanti, G., 2013. Is it a Hackerspace, Makerspace, TechShop, or FabLab? *Maker*, [online] 22 May. Available at: <http://makezine.com/2013/05/22/the-difference-between-hackerspaces-makerspaces-techshops-and-fablabs/> [Accessed 27 July 2015].

Churchill, R.R. and Stege, E.H., 2006. From Afghanistan to Iraq in media maps: Journalistic construction of geographic knowledge. Cartographic Perspectives, (54), pp.55-68.

Coddington, M., 2015. Clarifying Journalism's Quantitative Turn. Digital Journalism. 3(3), pp.331-348.

Cohen, B., 2015. Art Exhibit Perfectly Captures the Insanity of America's Criminal Justice System. *The Daily Banter*, [online] 13 May. Available at <http://thedailybanter.com/2015/05/art-exhibit-perfectly-captures-the-insanity-of-americas-criminal-justice-system/> [Accessed 27 July 2015].

Crampton, J. W., 2009. 'Cartography: Maps 2.0'. *Progress In Human Geography* 33 (1), pp.91-100.

Demirci, A. and Suslu, Z., 2015. The Use of Maps in Turkish Newspapers. Journal of Geography, pp.1-10

Dodge, M. and Kitchin, R., 2013. *Crowdsourced cartography: mapping experience and knowledge.* Environment and Planning A, 45 (1). pp. 19-36.

Elwood, S. 2009. 'Geographic Information Science: Emerging Research On The Societal Implications Of The Geospatial Web'. Progress In Human Geography 34 (3), pp.349-357.

Gershon, N. and Ward P., 2001. 'What Storytelling Can Do For Information Visualization'. *Commun. ACM,* 44 (8),pp.31-37.

Gray, J., Lämmerhirt, D. and Bounegru, L., 2016. Changing What Counts: How Can Citizen-Generated and Civil Society Data Be Used as an Advocacy Tool to Change Official Data Collection?. Available at SSRN 2742871.

Green, M. and Myers, K., 2010. Graphic medicine: use of comics in medical education and patient care. BMJ, 340 (mar03 2), c863-c863.

Greenwood, D.J. and Levin, M., 2007. Introduction to Action Research: Social Research for Social Change: Social Research for Social Change. SAGE publications.

Hewett, J. (2016). Learning to teach data journalism: Innovation, influence and constraints. *Journalism*, *17*(1), 119–137.

Hmelo-Silver, C.E., Duncan, R.G. and Chinn, C.A., 2007. Scaffolding and Achievement in Problem-Based and Inquiry Learning: A Response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42(2), pp.99-107.

Hullman, J., and N. Diakopoulos. 2011. 'Visualization Rhetoric: Framing Effects In Narrative Visualization'. IEEE Transactions On Visualization And Computer Graphics 17 (12): 2231-2240.

Kennedy, H. 2015. Part 1: What we found and what this means for how we think about 'effective' visualisations? Visualisingdata.com. Available at: http://www.visualisingdata.com/2015/10/views-from-seeing-data-research-part-1/

Kirk, A. 2012. Data Visualization: a successful design process. Birmingham: Packt Publishing.

Kirk, A. 2016. Part 3: What are the implications of this study for visualisation practitioners? Visualisingdata.com. Available at: http://www.visualisingdata.com/2015/10/views-from-seeing-data-research-part-3/

Kosara, R. and Jock M., 2013. 'Storytelling: The Next Step For Visualization'. *Computer* 46 (5).

Lancy, D., 2001. 3D Data Management: Controlling Data Volume, Variety and Velocity. *META Group Inc*, 949, pp.1-4.

Larsen, M. and Walby, K., eds., 2012. Brokering Access: Power, Politics, and Freedom of Information Process in Canada. Vancouver & Toronto: University of British Columbia Press.

Lewis, S.C., 2015. Journalism In An Era Of Big Data. Digital Journalism. 3(3), pp.321-330.

McConway, K., 2016. Statistics and the media: A statisticians view. *Journalism*, *17*(1), 49–65.

McGrath, L. and Guglielmo, L., 2015. Communities of Practice and Makerspace: DMAC's Influence on Technological Professional Development and Teaching Multimodal Composing. *Computers and Composition*, 25, pp.44-53.

McLoughlin, C. and Lee, M.J.W., 2008. The Three P's of Pedagogy for the Networked Society: Personalization, Participation, and Productivity. *International Journal of Teaching and Learning in Higher Education*, 22(1), pp.10-27.

Nguyen, A., & Lugo-Ocando, J. (2016). The state of data and statistics in journalism and journalism education: Issues and debates. *Journalism*, *17*(1), 3–17.

Rivas, J., 2015. L.A. art show wants to change hearts, minds – and criminal records. *Fusion*, [online] 4 May. Available at <http://fusion.net/story/129884/at-this-l-a-art-show-felons-were-the-special-guests/> [Accessed 27 July 2015].

Pandey, Anshul Vikram, Anjali Manivannan, Oded Nov, Margaret Satterthwaite, and Enrico Bertini. 2014. 'The Persuasive Power Of Data Visualization'. IEEE Transactions On Visualization And Computer Graphics 20 (12): 2211-2220

Rogers, S., 2013. Facts Are Scared. London: Faber and Faber.

Segel, E. and J Heer., 2010. Narrative Visualization: Telling Stories With Data. IEEE Transactions On Visualization And Computer Graphics 16 (6): 1139-1148. doi:10.1109/tvcg.2010.179

Splendore, S., Di Salvo, P., Eberwein, T., Groenhart, H., Kus, M. and Porlezza, C., 2016. Educational strategies in data journalism: A comparative study of six European countries. Journalism, 17(1), pp.138-152.

Weber, Wibke, and Hannes Rall. "Between data visualization and visual storytelling: The interactive information graphic as a hybrid form." Conference of the International Communication Association (ICA). 2012.

Wenger, E., McDermott, R. and Snyder, W.M., 2002. *Cultivating Communities of Practice: A Guide to Managing Knowledge*. Boston, Massachusetts: Harvard Business School Press.

Wilson, M. W., 2011. Training The Eye: Formation Of The Geocoding Subject. Social & Cultural Geography 12 (4): 357-376. doi:10.1080/14649365.2010.521856.

Wilson, M., 2015. What Killed The Infographic? *Fast Company*, [blog] 6 May, Available at: <http://www.fastcodesign.com/3045291/what-killed-the-infographic> [Accessed 27 July 2015].