**Lessons from Social Anthropology**

Tacey Laurie,

Data Science and Technology, HM Revenue & Customs

Notes to accompany PowerPoint slides presented at

22nd GSS Methodology Symposium, 17 July 2017

Background

HM Revenue and Customs took the decision to move customer interaction online in 2014. Since then, the department’s web presence has expanded at a rate unparalleled in even the private sector, earning it an industry award for best new online initiative in 2016. Analytical interest in the web service has ranged from A/B testing of “nudge” emails and customer segmentation for marketing purposes to analysis of mouse movement and page viewing time.

However, statisticians within the department struggled to obtain raw output from the website let alone understand it. The problem seemed to be part cultural, part physical: the majority of programmers sat in other buildings and were not included in the departmental address book and when they were found, they seemed to speak another language (chiefly Scala!) and thought the interest in statistics was limited to summary tables.

It seemed the only way to crack this conundrum was to plant someone on the inside, so in 2016, I was seconded to work in the Digital Centre in London for one year.

Requirement

My aims were to identify and document digital datasets to help statisticians scope projects, enable analyst access to digital data and influence outputs to enable easier analysis. It sounded easy enough: find out where the data dictionaries and variable lookup tables are, negotiate a clutch of analyst accounts and ask developers to please use similar data structures to make analysis easier.

Conclusions

I soon found out that there were no data dictionaries or if there were, they were written during the initial phases of development and had since been allowed to become obsolete. Some of the documentation was in flowchart form with arrows pointing to other outputs to various degrees of precision – one was a photo of a flowchart drawn on the back of a Taz takeaway bag!

Reports back to my team drew gasps of horror. I couldn’t find any specifications for what we could see online. Did this mean the business wasn’t giving any steer to developers or that developers weren’t paying attention to their input? When I produced extracts for statisticians, I found gaps in the data for certain months. Did this mean the service was interrupted or the feed was faulty? Many of the variables had names that were either really similar or completely incomprehensible and there were no data dictionaries. How would anyone know which variable to use when conducting analysis? Would I have to start from scratch?

These questions can be summed up as basic concerns about the quality of documentation and the impact of agile work practices. In a world where statisticians are routinely asked to scope projects before writing project plans and are evaluated on their ability to deliver outputs according to details in the project plans, agile started taking on a sheen of “making it up as you go along.” The only course of action for me seemed to be to write as many data dictionaries as possible – and as quickly as I could.

I set up a meeting with the developers working on corporation tax for small companies, as this was a relatively limited “product” and explained what I was trying to do and was met with polite interest mostly because of a link between my team and a former employee. They answered my questions and soon I had finished my first data dictionary.

I started on my second data dictionary, confident I could solve this problem by being a bit more efficient. Problems began when I started using my first data dictionary as a template and found that the second website area had a different structure and my sample queries didn’t work. When I checked the queries by running them for the original website area, I was shocked to find different variables for different time periods. I couldn’t keep revising my data dictionaries but worried that after my secondment was done, future statisticians would be back to square one.

I decided it was time to go back to the developers and have a serious conversation about how things statisticians deemed as essential were not the same things that developers deemed as even necessary. The many conversations that followed turned out to be the start of my learning about what kinds of people build websites, how they make decisions and how good websites get that way. I’d say it was less a lightbulb moment than dawn breaking on a cloudy day.

In communicating my learning, I am borrowing methodology from the world of advertising which has increasingly used social anthropology to understand the consumer in order to increase sales. If you can tap into what’s important to a consumer, you can market your product to address that need. If statisticians can understand how developers think, they might have a better chance at communicating with them and ultimately getting the kind of help that moves analysis projects forward instead of stopping them dead the moment the scoping is over.

Artifacts

If everything in your office was swept away tomorrow, what would archeologists have to guide them in their understanding of your world? For government statisticians, we would hope our publications would remain as well as our responses to Parliamentary Questions and Freedom of Information requests and all of the analysis work which underpins government policy. For software developers working in digital industries, their legacy is the online service: the user interface, the speed of the website and the efficiency of the product in delivering the requested outcome whether it’s a list of hotels or access to the saved version of a tax return.

The main difference is that the statisticians work with outputs mainly with administrative outputs created by other people and which are not subject to change. The fact that the data can be counted on to stay the same is why it is possible to scope projects and submit project briefs for senior sign off with confidence.

In digital, the work is all about creating something which has never been built before. This means that people who succeed in digital work are by nature pragmatic, output-oriented (who wants a website that works well in theory?) and website developers are solely focussed on the product they are tasked with building.

Not so long ago, customers used to approach computer programmers with ideas for a new database, booking system or website and many months later, the finished product was presented for testing which invariably worked in a completely different way to that which the customers intended. This is because customers aren’t computer programmers and don’t think about design the way a programmer does and computer programmers can only guess at the aspects the customer takes for granted and forgets to include in the specification.

Agile working practice came about to avoid this problem. It’s about building and testing in chunks so that each chunk is useful and complete in itself. This is the point at which all of those sketchy flow charts I found were used. The problem for people who like end-to-end specifications is that the moment they are written, they are obsolete as obstacles are encountered and new ideas are explored to move forward. As the HMRC website is committed to continuous improvement, this would make maintaining any end-to-end specification into a full-time job. Small wonder I found exactly zero for the digital products but in the pre-agile age, HMRC was awash with giant dataset specifications.

Identity

What makes us, “us” and what makes them, “them?” What kind of hierarchy operates in your world and how do members of the community move within it? Government statisticians have clearly-defined departmental hierarchies and promotion processes. Membership is by interview and subject to prerequisite training or work experience. Most members have a strong sense of belonging to a team/directorate when socialising within the department and outside of it, define their identity by their profession.

By contrast, members of digital teams identify strongly with the product they are building and the team they work with or the software they use or possibly the agency which employs them. They have little connection with the organisation they work for. Team have a flat structure, with team managers doing less formal team managing and more facilitating so that fellow team members can help each other out. Digital teams are called scrums because they bind together to move work forward.

This difference in perspective is why statisticians find it difficult to get buy in for their work when presented as part of the directorate’s strategy or supporting departmental objectives. The teams they are speaking to generally don’t have access to the intranet and so have no exposure to either of these let alone all of the acronyms which pepper the language of civil servants. The best way to gain traction with a digital team is to show that your analytical work will produce insight that will help them to improve their product. The best way to pitch your proposed analysis is to make a face to face appointment.

Communication

How do we communicate and what are different medium used for? Government statisticians use email for most communication. This has the advantage of attachments where complex work is being discussed and producing an audit trail should it be needed. Other forms of communication are the pieces of documentation that arise from traditional project management, meetings and all the written documentation associated with meetings. Generally speaking, email is used for day-to-day work either contributing to project management material or doing the actual work while meetings are reserved for confirmation of project methodology prior to sign off or knowledge sharing following completion of projects.

Digital teams routinely start the day with a short 9:30 meeting called, “Stand Up.” The aim is to report work completed the previous day, work being scheduled for today and any anticipated blockages so that team members can contribute ideas. These can range from volunteering to sit and review code to suggesting contacts for staff to call as staff lists seem to go out of date as soon as they are circulated. The meeting is held standing up only because there simply aren’t enough meeting rooms to accommodate everyone having a meeting first thing in the morning.

Many developers use an app called, “Slack” on their phones to communicate and either don’t have gsi email accounts (and so do not appear on the Global Address List) or they use their personal email accounts and rarely check the gsi email address. Just because you send an email doesn’t mean it gets read and just because it gets read doesn’t mean recipients feel it needs a response. Many times, I’ve spoken to digital teams after emailing them only to find out they read the email, got confused and moved on to do the work they are paid to do.

Add to this situation the high rate of turnover in digital teams and it can seem almost impossible for statisticians to contact digital teams. Short of showing up at the offices and having a wander, the next best thing is finding someone who has some digital staff contacts or someone with access to Confluence where you may find non-gsi email addresses and mobile phone numbers.

Lessons learned

Understanding how developers think helps in many practical ways (not depending on emails, not using acronyms, showing your work will result in actionable results) and can guard against projects of no return (sometimes literally!) such as trying to produce definitive documentation for a product which is still developing. There are also lessons statisticians can learn from digital when not working with digital products.

When starting a new project, don’t expect it to be the same as the previous one. Why should all data have the same documentation? If data is missing or has data quality issues, ask yourself whether this actually impacts on your ability to make a decision or produce recommendations. Your time is probably better spent finding a way to benchmark or impute than looking for the perfect dataset.

Meetings are boring if you aren’t finding out or contributing useful information but just because they are a fact of life for government statisticians doesn’t mean we can’t make them more efficient. Stakeholder engagement is necessary but documents can be circulated beforehand, contentious areas agreed and the meeting reserved for selected discussions. Presentations of completed work don’t need to examine code if the work has already been completed. If in doubt what to present, ask yourself, “and I care because?”

One reason developers are able to spend such a small amount of time in formal meetings is because they have lots of unscheduled discussions. If you have a question, it’s faster to ask it in person. If you go out regularly for lunch, you get to know what team members are doing, both during and outside of work.

Many developers studied computer science at university but most didn’t acquire the coding skills they use for work during formal study, partly because the moment a course is finished a new piece of software is launched or app circulated. The vast majority of developers learn a programming language by self study and choose which language based on how viable they believe the language will be at the point at which they gain accreditation, usually in the form of online tests resulting in certification.

Government statisticians can easily become institutionalised, looking only to courses offered by their department and often these are only approved if the skills are directly related to the work in hand. However, in a Civil Service whose size is continually shrinking, given the choice between candidates with requisite skills and paying for training, having the skills needed to do work when applying is an asset. Learning a new skill doesn’t have to be expensive. There are many free online courses and for those that aren’t free, you can think of the money as an investment in your future.