Supplying Open Digital Services - Response

Working alone, as part of your team, or with my network of experienced associates, <u>Changeful</u>, my service design and innovation practice, offers:

- ◆ *people-centred service design;*
- ◆ *web and mobile service innovation; and*
- ◆ *capability development and transformation.*

UK public sector organisations may be able to access these services through the G-Cloud <u>Cloudstore</u> or the <u>Digital Services Framework</u>. I am also proud to work in partnership with <u>Actuated Futures</u>, a future city technology company which is the ODI-Node for the Leeds City Region. This means I have additional domain expertise to draw upon relating to open data, environment and technology.

The Open Data Institute has put together a <u>scenario</u> that involves providing an open digital service. This document explains how I would expect to work as part of a team to provide it.

Matt Edgar, 24 March 2014

07964 763252

<u>matt@mattedgar.com</u>

Making things open makes them better. Specifically this scenario for the Food and Environment Research Agency (FERA) holds out the prospect of more informed citizens and businesses, new scientific insights, and other benefits that will only be uncovered by the process of making the emerging contaminants data freely available through an open digital service.

But this is not a case where "if we build it, they will come." Poorly handled, the first-time release of such a high-profile dataset could set back public understanding and weaken industry relationships, both things on which the agency relies to perform its duties. Conversely, the data could easily pass its potential users by. To date, half the Defra datasets listed on data.gov.uk have been viewed fewer than 40 times each. Our ambition for the emerging contaminants service must be much greater than has been the case for those.

People-centred service design provides a powerful toolkit for understanding user needs and meeting them effectively with open data. All the other headings requested (data processing and exploration, legal, practical, technical, social issues and team) are derivative of user needs and as such are touched on in the iterative service design process described below.

As a supplier of user research on the Digital Services Framework, I have focused this proposal on the people-centred aspects of the service. On the data processing side, I expect to work closely with technical architects and developers as part of an agile, multi-disciplinary team. Given the importance of making the data understandable to a non-specialist audience, a content designer should be included from the outset. As well as product and delivery managers, the team should have at least one scientific subject matter expert and a department or agency specialist in policy and communications.

Understanding data-user needs

In line with the Government Service Design Manual, the service will run through discovery, alpha and beta phases before going live. I recommend these phases proceed as follows.

During the discovery phase we will:

Publish the scope of the data in question. Essentially this is the reference data specified in the brief – as a declaration of intent and to give interested parties a clear view of the data that is to be released. This may also be accompanied by either sample or dummy data, along with a feedback mechanism for interested parties to contact the service team and volunteer as potential alpha users.

Engage with users in each of the three groups identified – citizens, scientific users and other data re-users. The agency will already have good connections with some of these groups, but should also think laterally about groups that have not previously been on its horizon. I have experience of a wide range of insight techniques, including running contextual interviews and co-creation workshops to uncover needs and spot relevant new service opportunities. Where possible I encourage clients and colleagues to take part in user engagement to build their own first hand empathy with users and a nuanced appreciation of their needs. Out of this research a series of user personas will be created to inform the design and development process.

Co-create possible use scenarios and opportunities relating to the data sets in question and establish a framework to test these scenarios and ways to curate the data-sets that encourage and accelerate the use of the data-sets.

Collaboratively draft the Open Data Certificate to which we aspire. This drafting process will help stakeholders develop their shared understanding of the status of the data, including the license under which the data will be issued (assumed to be the Open Government License). It will also flush out any potential intellectual property or privacy factors to be addressed during the phases that follow.

Map the service ecosystem. More than typical "user experience" methodologies, service design emphasises the diverse nature and roles of the external and internal actors, and how they interact with each other. For example, we might consider the roles of scientists, the food industry, the media and citizens, and how they might influence each other around the first release of the data.

Write and prioritise user stories. At this stage we will consider which user personas should be supported in the alpha. User stories will be backed by insight and prioritised according to their potential for learning more about user needs. These in turn may be combined into "epics" which illuminate users' end-to-end service journeys.

During the alpha we will:

Make a first release of some or all of the data in a number of formats. Given the high-profile nature of the data, this may be in the form of a private alpha to selected users. Data released in a private alpha will not be true open data, so this should be seen as a prototype on the path to openness, not a formal release of data in its own right.

From the outset the data released should be:

Trustworthy: while the means of accessing the data may be novel, the data itself must be accurate and of the high standard expected of FERA as the national reference laboratory for chemical safety in food.

Large and varied *enough*: if only a subset of data is released in alpha, it should be at an order of magnitude and representative variety that allows for specialist users to put it through its paces in the statistical models and tools of their choice.

Human-readable and machine-readable: As a minimum, data should be released as human-readable web pages and through a RESTful API. A useful pattern is for the human-readable pages to call on the API. Links between the two will help people understand the data; use software to deepen their understanding of it; answer specific questions; and automate repeatable processes. Besides predictable programmatic access to every data point, the API must allow rapid bulk download of the datasets as a whole. Depending on the user needs identified, the service may also offer CSV files that can be imported for manipulation into a wide variety of software packages such as SPSS, Microsoft Access and Excel.

Native to the web: For both humans and machines, the data should follow web standards for persistence of URLs, discoverability by traversing hyperlinks and use of web standards such as microformats. While these can appear to be technical considerations they can have a big impact on the user experience. Done right, they allow users to make the most of modern browsers and search engines. Patterns to be encouraged include using understandable, nested URL structures and the ability to drill down from summary to detailed data so the content becomes self-describing.

Contextualised: the data and its provenance must be well documented. Specialist users may require detailed descriptions of the methods used in the study, and the statistical factors to be considered in interpreting the data. Non-specialist users have greater difficulty understanding statistical concepts, especially those around risk. They may need more simplified (but not dumbed down) interpretations. This context must be preserved as metadata when making the data machine-readable so that data re-users never present data shorn of context.

Social: Users must be able to share their questions and findings among themselves. This may mean the agency hosting and moderating discussion on its own service, or simply making it easy for users to share, link to and reference the data elsewhere on the web. If the alpha is run in private, users may require a temporary space to discuss the data among themselves even if this will be less of a requirement once the data is made open.

Engage with alpha users. As well as offering feedback links throughout the service, we will invite alpha users for one-to-one and group sessions to understand how they are using the alpha data and what barriers they face to using it to meet their needs. Hack days and challenges *may* be part of the mix to build social capital with developers. If they are, extreme care must be taken to avoid a zoo atmosphere in which dead-end services are created just because the data is there. Since the alpha is for learning about genuine user needs, it is better to invest the available time and money supporting real users developing in the wild.

Re-write and re-prioritise user stories. Based on the insights from users in the alpha, we may adjust the personas and user stories, and change their priority for the development of the beta.

During the beta, we will:

Prepare to publish. Significant public interest in the data may be triggered here, so the agency must be prepared to handle inquiries from the media and others outside the scientific and industry communities. By the end of the alpha stage, the data will be well presented and contextualised so it can be released on equal terms to everyone at the same time, without media pre-release or embargo.

Publish the whole dataset under the appropriate open data license and certificate. The disruptively innovative nature of open data means you cannot truly learn how people will use it until you make the same unrestricted data publicly available to everyone. As with the alpha, the beta label signifies the status of the *means* of releasing the data, not the quality of the data itself.

Engage with beta users. There will be more users than in the alpha, including some unexpected ones who could not have been identified during the discovery. User feedback will be closely monitored and contact permissions gathered so that user researchers can follow up with beta users as individuals and in groups. In addition, data analytics will be deployed to monitor use of the datasets. This must extend to monitoring of API calls from machines as well as requests for the human-readable versions of the data.

Iterate in the open. The team must have capacity to update the service very frequently during its early weeks and months. In so doing, we will consider the impact on users of changes in the way the data is presented. Machines in particular can be baffled by even the smallest change in data format. Known users and re-users must be alerted in good time to any changes; version control may be used to preserve deprecated functionality; and feedback channels and analytics monitored for evidence of any unintended impacts after changes are made.

When the service moves from beta to live, we will:

Publish a service dashboard for the data. This "data about the data" helps the service manager and others to understand how all the formats are being used and make improvement plans accordingly.

Encourage people to tell stories with the data, often by linking it to other datasets that provide fresh insights. This can apply at the specialist scientific and industry levels and with the general public and media. Visualisations should be a doorway inviting users deeper into the data, not a shop window that only shows a few top items.

Nurture the community around the data, enabling users to support each other and to interact with other stakeholders who may have different perspectives on the data.

Adopt the release of open data as an on-going commitment, not a one-off publication. This could involve working with FERA's researchers who run the study to automate and improve the flow of data from lab to API, so that future updates can be published efficiently. The service must be architected so that new contaminants and samples can be added to the dataset as soon as test results become available.

Allow open data to transform the whole organisation. If the release of data proves valuable to users and re-users there may be a business case for moving away from a single annual study to more frequent monitoring, or a rolling programme able to check a wider range of products. This could entail working very differently on the inside and interacting very differently with people on the outside. Over time, as a result of open data and people-centred service design, the agency will become more responsive and more valuable to citizens, the food industry and other stakeholders alike.