

# 1 UNDERSTANDING DATA-USER NEEDS

Starting with the existing user groups and types of data identified in the brief, we'll create personas, scenarios and user journeys for each key audience, task matrix and measurable design criteria using the following approach:

1. Sampling a cross section of users who fit the groups identified we'll conduct interviews, preferably face-to-face, to discuss the type of information they're looking for, how they use it, and what they're next steps are.
2. Analysing the research gathered we'll validate existing user-groups, ascertain any new ones, and create personas for the most important. We'll identify key user journeys, scenarios and needs associated with each persona in order to bring them to life and guide design decisions throughout the project for both internal and external teams.
3. We'll identify key tasks conducted by each user-group plotting their importance and frequency to give us a priority level of tasks to focus on. By analysing user-journeys and tasks we can ensure efficiency in task completion by employing persuasion, emotion and trust techniques to fulfil both user and business requirements.
4. Within the research data we'll also identify usability design criteria. This criteria will enable us to measure the success of the end solution during usability testing, example criteria might be 95% of "citizens" will be able to share their findings with their social network.

## 2 DATA PROCESSING

To make sure we could take into account the different types of data we would audit the different data-sets and then create a matrix outlining the steps we would go through for each data-type. E.g. for word documents, excel files, SQL databases.

Our development approach would include interface specification and will be used to define source file formats, reference values and mandatory fields, and an ETL (Extract, Transform, Load) process used to map and transform source data to desired target structure. Within this specification, data dictionaries will define semantics of data fields, providing lookups for ETL process and a reference for data explorers.

Using a range of algorithms from direct lookups of reference data, to "fuzzy logic" offering best-option alternatives, we would map a variety of messy data sources to required functionality, providing opportunities to cleanse and improve data where appropriate during 'staging data' waypoints in the processing pipeline.

Unstructured data, documents and images would likely be stored in a high availability document store, with cataloguing and tagging through metadata which is stored in the central database alongside reference pointers to files. The technology platforms used would be selected as part of our initial consultancy.

Data processing issues we would anticipate in the context of an open data project include:

- Absent ("null") data in areas critical to visualisation functionality, such as a meta-data field which acts as a 'dimension' for refining datasets
- Data values which are outside of the defined, expected range, and system approach (cleans / badfile) for these entities
- Data quality variation, such as inappropriate capitalisation of text, which may not be appropriate for digital delivery and result in application design inconsistencies

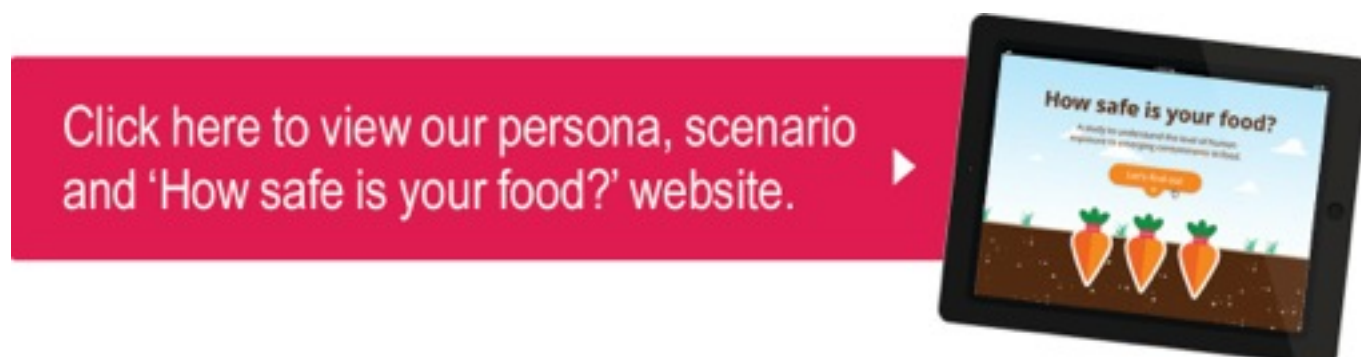
## 3 DATA EXPLORATION

From work understanding data-user needs we will create a representative selection of questions users might want to 'ask' the data. These and the personas would inform interface design and visualisations in two core ways.

For 'scientific data re-users' and some 'other data re-users', we would design an advanced interface that allowed flexible interrogation of the data online to create simple results (e.g. selecting areas of the data, filtering, and plotting data out visually in various table and chart forms), plus options to download the data-set relating to their area of interest. The design would therefore focus on how to engage users with chart tools and manage expectations of what can be done before offline data analysis becomes preferable. It would also provide guidance documentation for advanced users on accessing the raw data.

For 'citizens' and some 'other data re-users', we would look to define clear 'stories' the data communicates, especially where expert interpretation is helpful. These would be worked into pre-designed charts and infographics that help communicate these stories, which can

be delivered along with dynamically generated charts and tables from pre-filtered data. The focus of this interface would be around helping users gain clear insight from the data, and supporting this with education on the meaning of results and how to use the interface. We have provided an example scenario and visualisation for you to explore:



<http://opendata.concepts.mandogroup.com/>

## 4 LEGAL ISSUES

Legal advice can require specialist help and we have contacts that we can engage with if necessary, with regards to appropriate legal safeguards required for the provision of data. However we would recommend that there should be a page on the website which provides the terms and conditions (written in plain language so as to be accessible and simple to understand) for each of the three different types of user groups. Areas which should be covered include:

- Intellectual property: the terms should outline who owns the intellectual property for all elements of the data (including derived data) provided through the website.
- Licensing: the terms should make clear the basis by which the owner provides the data, such as through a non-exclusive royalty-free licence such as the OGL (Open Government Licence). To promote compliance we would ensure that attribution is easy for the user by surfacing and clearly signposting provenance to all data re-users, both in terms of raw data for scientific re-users and in relevant citation formats for others
- Privacy: the terms should outline any privacy issues, particularly if any data is requested from the users and stored for any purpose. An example would be a user being required to register on the website in order to download data.
- Warranty and liability: the terms should outline any warranties that may or may not be made in providing the data, such as its correctness or completeness. It would be recommended that the terms outline that the data is provided without liability and any subsequent use of the data is at the users own risk.

In order to gain acceptance of the terms and conditions from users, it would be recommended that any areas of the website where data can be downloaded contains a clearly signposted link to the terms and conditions page - along with a method of acceptance where required; for example an acceptance tick-box which the user is required to tick as part of a registration process; or an implied acceptance given by choosing to download a data-set.

## 5 PRACTICAL ISSUES

The ODI Open Data Certificate level we would target for each dataset would be "Standard", which would require at a minimum:

- That the data is **findable**. A link to the data would be provided from the main website. The Food and Environment Research Agency are also an approved publisher of data-sets within the data.gov.uk Open Data platform (<http://data.gov.uk/publisher/the-food-and-environment-research-agency>), so the data would be listed within this collection, and also should be referenced within other publications.
- That the data is **accurate**. The published data should be regularly updated, with explanatory accompanying release notes and clear communication channels to provide feedback and obtain support.

- An accompanying **quality** statement is made available on the website and could be linked to a data policy which is made open and available to data-users (<http://democracy.york.gov.uk/mgConvert2PDF.aspx?ID=19377> is an example of such a data policy). This statement would cover quality control processes, and any data that doesn't adhere would be flagged accordingly.
- That the data is openly classified with a level of **guarantee** of availability, by stating how long the data will be available, and how resilience is provided via backups.
- **Legal** and licencing information is clearly established and provided for the data-sets
- Data-users must receive support from the publisher, which would likely require the commitment of FERA resources to assist with user queries and ad-hoc guidance on usage. In addition, data-users should have provision for feedback to the publisher, through email, website forms and potentially extending to offline channels depending on availability of suitable resources.

## 6 TECHNICAL ISSUES

The **location** of the data-sets would be provided when the Open Data Certificate is created, and also clearly linked from the website, other open data catalogues (eg data.gov.uk) and blogs, publications etc. All of these contribute to the discoverability and trust in the data.

All data-sets will be made available in multiple **formats**, both human- and machine-readable depending on audience and likely use cases, using the simplest interfaces possible in each case. Examples of suitable data formats would be tabular data in CSV format or structured JSON or XML documents, all of which can be stored and parsed (and often used by developers) or Excel Spreadsheets, which have the advantage of allowing comments (often used by statisticians). Other emerging standards such as SDF (Simple Data Format) combine the benefits of both. There may be use cases around API's (Application Programming Interfaces), which would allow more dynamic data integrations and "mash-ups", and again we would look to use open standards such as ODATA (<http://www.odata.org/>) to implement these interfaces. For human readable documents we would look to open web standards such as HTML, or where necessary other common digital document formats such as PDF.

Where persistent identifiers are already available for the data domain, these would be utilised for the encoding of the data to facilitate combination with other data sources and to allow users to interrogate relationships between data. This could include provenance metadata and visualisation of such data where required.

## 7 SOCIAL ISSUES

Providing sufficient guidance on how to query and extract the data is a vital element of the project which will help to drive engagement with the data-sets amongst the data-user communities.

The complexity of documentation and advice depends on the complexity of the data-sets being made available. For a handful of methods and objects we may recommend self-producing documentation and performing some basic user-testing on its ease of use. Sample charts or infographics could be produced to show the potential application of the data. For more complex use-cases we may recommend the investigation of a platform like CKAN to power the API, leveraging their existing 'how-to guides' and 'extended social publishing' features.

Mechanisms such as the one used by data.gov.uk are good examples of encouraging collaboration and feedback on individual data-sets. Receipt of online feedback could automatically notify the publisher by email, giving the publisher the opportunity to respond.

To provide feedback routes to data publishers, we would incorporate tools to allow any data visualisation to be shared – for example via social channels – with comments, allowing tracking of what is being shared and how. We would also allow users to directly feedback to authors, by attributing contacts details to datasets used and by allowing simple response functionality on the site.

## 8 TEAM

We've got a talented group of people working at Mando with deep skillsets across disciplines and technologies. We employ more than 60 permanent staff working in Consultancy, Project Management, Graphic Design, Front End Development, Development & Test and Managed Services.

To help with your understanding of our resources and capacity to deliver we thought it might help to meet some of the team here at Mando. The people listed below are all employed by Mando and would be amongst those likely be on the team working with you (please note that we will not be able to confirm exact resources until formal project kick off).

**Jonathan Seal: Strategy Director / Head of Consultancy**

Jon has been involved with award-winning web and media companies throughout his 15 years in the digital sector, including running his own business specialising in user experience (UX) design and prototyping innovative user interfaces for a range of digital solutions.

**Joanne Finch: User-Experience Consultant**

Jo joined the company in 2005 with more than 10 years' experience in digital design. With a background in public sector design, Jo specialises in accessibility and user experience, working with organisations including the Equal Opportunities commission, Citizens Advice, Department for Transport, and the Office of the Deputy Prime Minister. While at Mando Jo has helped our clients understand their customers' needs and frustrations, conducting research and testing sessions to ensure the experience of the solutions we design meet the expectations of the people who use them.

**Kate Birch: Solutions Consultant**

Kate has been in the IT industry for 10 years, starting as an online applications tester but quickly moving into Project Quality Manager; focusing on gathering client requirements and managing the product release cycle. In 2007 Kate joined Mando as a Project Manager, qualifying as a Prince 2 Practitioner. Kate has managed a great many projects at Mando performing a joint-role of both Business Analyst and Project Manager gathering the requirements and producing information architecture and specification documentation. Her experience of day to day relationships with clients has given her the ability to explain all parts of the project process keeping the clients informed at every stage and this year Kate has moved to a role focused exclusively on user-experience

and solutions analysis.

**Matt Salmon: Head of Development**

Matt has worked in software implementation and development roles for over 10 years. He began his career at Atos Origin, a global systems integration provider, designing and developing Oracle data warehousing and management information systems which collated high volume overnight Receipt and Dispatch data into OLAP cubes suitable for self-service management reporting. Matt joined Mando in 2007 and leads a team of software developers delivering business-critical online solutions using Microsoft technologies.

**Mark Simpson: Account Director**

Mark is an Account Director with over 10 years' experience of delivering digital solutions to a range of private, public and not-for-profit organisations. Mark's experience includes working with the UK Research Councils on several data-driven initiatives; including the RCUK Research Outcomes System, the ESRC's Awards and Outputs database and the EPSRC's data visualisation proof-of-concept, 'epiphany'. A key driver for all of these projects being the collection, exposure and visualisation of data in several formats including oData, OAI-PMH, SWORD and CERIF-XML and the collaboration with the UKDA and the CIA (CERIF-In-Action) programme.