# OPEN DATA ISTITUTE OPEN DATA DIGITAL SERVICES: TH_NK RESPONSE

# TH_NK

AUTHOR
JONNY SYCAMORE

DATE
24TH MARCH 2014

VERSION
1.1

8-14 VINE HILL
LONDON
EC1R 5DX


**T** +44 (0) 207 504 6000
**F** +44 (0) 207 504 6002
**E** info@think.eu

55 DEGREES NORTH
PILGRIM STREET
NEWCASTLE UPON TYNE
NE1 6BF


**T** +44 (0) 191 241 7000
**F** +44 (0) 191 241 7002
**E** info@think.eu

## INTRODUCTION

We're very excited to be a part of the process to identify open data suppliers and we have really enjoyed pulling together our response.

Below you will find our responses to the different sections outlined in your brief. It's worth noting at this point that as an agency we follow the Agile SCRUM project methodology and will refer to its practices throughout our response.

## UNDERSTANDING DATA USER NEEDS

Understanding user needs requires well thought out research techniques across the different audiences to deliver comprehensive and well-understood information. For a brief such as this we would undertake research in the form of interviews and surveys with members of the public, as well as specialist data users to gather insights into their expectations and requirements. From a data perspective some of the example key questions we would want to understand could include:

- What is most valuable to them?
- What format do they want the data in?
- How will they filter it/search within it/interpret it?

Off the back of the research and requirements gathering process we would then develop user stories in line with Agile practice. The user stories would allow us to flesh out the requirements of our users and will start to inform the solution.

## DATA PROCESSING

It's important to point out that any backend system or CMS or data housing solution needs to be seen as a means to an end, it isn't the solution itself and should be treated as one element of an overall solution.

Any decision around technology or systems will of course need to be based on the requirements established in the early planning phases of the project. Those requirements will be linked to the user requirements as outlined in question 1 but also the specific data requirements. For example, it may be the case that a specific user type will deal with data in a certain way - perhaps a data re-user will analyse it using a specific software product. The data may also have certain characteristics that need to be factored in – format for example, word documents vs. data tables in this scenario. Those factors will have an important bearing on the backend solution that is used - any solution would need to be compatible with those user and data requirements.

On that basis it may also be the case that existing systems or platforms do not meet the requirements, in which case we would look to scope bespoke/customised backend systems.

Once the right system has been identified, the process of importing and cleaning up the data could begin. By analysing the different data sources and formats and establishing common reference points we would be able to understand how the data sets relate to each other and how they can be identified across the different sources. Depending on the systems in place we could then write algorithms to automate and aid the importing of data.

## DATA EXPLORATION

To establish meaningful and useful interfaces for users we would again need to refer back to the user stories developed in the first stage of the project to ensure any interface options cover the requirements described.

Posing and answering simple questions of the data is more likely to be a requirement of the citizen user type, it's likely that the more advanced data users would want to analyse the data at a deeper level. On that basis we would look to develop interfaces to align with the following data consumption principles:

- Passive / Reading and viewing data
  - o This could be in the form of headline figures and visual charts that tell a simple story from the data. E.g. Food type X is the most susceptible to contamination by chemical Y
- Interactive / Manipulating an interactive interface to affect the data
  - o This takes the relationship with the data to a more interactive level and allows the user to manipulate the data representation they are viewing on the screen. This might take the forma of an interactive graphic that moves as a user changes the data reference. In this scenario an example could be a sliding scale of contamination by chemical X, the user could change the food type and see how the contamination level changes.
- Creative / Delivery of bespoke data personalised to the user and their own data
  - o The final principle sees users adding in their own data to cross reference against the existing data set for the creation of personalised, bespoke or unique data. A good example in this scenario would be a user inputting their daily food consumption habits and viewing chemical contamination levels personalised to their own consumption.

The principles used and the interaction levels we would develop in an interface would depend on the user requirements we have already established.

At TH_NK we have experience of developing user interfaces across all the levels of interaction described above. Our award winning Utopia Inquiry campaign used a creative interface to allow users to receive personalised feedback and share it socially.

## LEGAL ISSUES

Data security and legality is of paramount importance for any digital project – whether it has a specific data focus or not. At TH_NK we take the security of data very seriously. As with any project, there will be specific requirements that will be established but as a general rule some of the security steps we would follow are below:

- All required data and quality certifications are put in place and any required qualifications are understood and undertaken by those responsible for them
- Due diligence and legal sign off. Ensure properly written privacy policies and licenses are all in place from the supplier side and from the client side
- Encryptions of proprietary data. In the case of opening data up to users to explore and manipulate for themselves we would ensure that all proprietary or source data is secured and encrypted in a way that does not compromise its availability of data but protects it from damage from external users and third parties

- Any personal data that is collected to secured and used only as the user permits and in line with the privacy policy
- Ensuring that all systems and platforms used are thoroughly security tested

## PRACTICAL ISSUES

In terms of ensuring the publication of the data is easily available we would follow SEO best practice, as we would do with the publication of any digital content on the web. All properties would be tagged and meta descriptions written to the proper standards to ensure all users looking for the data can find it easily.

Within the data itself we would always want to make sure that labelling, tagging and metadata descriptions have been written properly. The data should be searchable and available for all user types to filter and manipulate as required.

In answer to the second part of the question around documentation of data quality, there are two methods we would ensure are covered:

- PROACTIVE
    - o Ensure the scope and statistical significance of the data is documented and fully described.
    - o Highlight and describe and know limitations to the data
    - o Use FAQs to arm users with knowledge to common questions over he data
    - o Clearly describe the processes undertaken to establish the data and its purpose
- REACTIVE
    - o Create feedback areas for users to outline any issues or limitations they finding the data so these can be clearly communicated or fixed where possible
    - o Be open to conversation and questions surrounding the data

## TECHNICAL ISSUES

With regards to publishing data for the needs of different user types, once again we would refer to the extensive requirements gathering process outlined in the first question. Once these are understood we would publish the data in the corresponding formats in the right areas for each user type. For example, we might publish access to raw data for reuse by more advanced data users in scientific journals or blogs. For citizens, or everyday data consumers, that data would be more likely to be available within the standard press and newspaper websites. We would of course ensure that all published data is linked to each other and easily accessible across various sources.

In terms of formats and open standards, at this stage it's difficult to outline exactly what they would be as they will be dependent on requirements that are yet to be defined. However we can from an accessibility point of view confirm that we would use open standards in publication. For example it's highly unlikely that an interface or publication would be developed in flash. We would use HTML5 to develop and display data and any interfaces so they're as accessible as possible across platforms and devices.

We would carry out a thorough data planning and analysis process to understand the data sets to find persistent identifiers to link those different sources together. These identifiers could take on the

form of aggregators, common identifiers or keys that can be used to match data and provide a more in depth story across the data sets.

In answer to the final part of this question, again this would depend on requirements, but we would ensure there is a framework in place for additional metadata to be easily added to the DB. Examples could include attribution/authors/reference info/bibliography/related and external content links.

## SOCIAL ISSUES

The type of documentation or access would depend on the user type but there are various ways of providing this kind of advice. Examples could include:

- Open APIs for advanced users to access the data and manipulate it for themselves. APIs would be accompanied by:
    - How to guides
    - Example query strings
- Embed codes could be used for those users who wish to take a cut of the data and display it on their own web properties. This could work for journalists reusing the data in another publication
- Socialised content for citizen type users who have created their own personalised view of the data and want to share it on social networks

From a feedback perspective there are two different approaches and we would recommend both be covered to allow feedback for a broader audience across different user types:

- Open and persistent feedback in the form of:
    - Contact forms
    - Feedback tabs
    - Reviewing analytics
    - User rating opportunities
- Closed/finite feedback
    - User surveys

## TEAM

At this stage we could only give an outline of a team capable of planning this kind of project. Without fully understating the requirements it's impossible to say what resources would be required to deliver tasks that are yet to be defined. From a project planning/discovery/sprint 0 perspective we would recommend the following key team members:

- Business Analyst
- Database Analyst
- Data Strategist
- User Experience Designer
- Programme Manager

Once full detail and requirements are understood we could commit to a team for an engagement that could include other resources such as design, front and back end development, service desk for on-going support etc.