Web Technology Group

T +44 (0) 207 339 8600 • www.wtg.co.uk

ODI Open Digital Services

Scenario Response

20/03/2014
Version 1.1

5th Floor, 76 Hammersmith Road, London, W14 8UD

Web Technology Group

1.1 WTG Overview

Web Technology Group (WTG) are a UK based SME headquartered in Hammersmith. The business is built on four propositions:

- Application management (core systems not desktop, accredited secure up to and including IL3)
- Application migration (system replacement and 'channel shift')
- Data Migration (secure to IL3)
- Secure Working (build and integration of collaborative applications, accredited secure up to and including IL4)

The following submission outlines the approach proposed by WTG to address the needs of ODI as illustrated by the scenario provided.

1.2 Understanding data-user needs - *How would you go about understanding the need… users alike?*

WTG will include engagement tasks with DEFRA and representatives of the expected user groups as part of the service design to identify and confirm their individual needs

Based on the user groups identified in the scenario, the approaches we would adopt are identified in the table below.

DEFRA	Citizen	Scientific & other data re-users
Meetings, workshops or interviews with key stakeholders	User interviews with representative sample users	Meetings, workshops or interviews with sample users
Topics for exploration		
Target audiences for the data	Preferences for accessing data	Preferences for accessing data
Current expertise and existing knowledge of data users	Data presentation preferences	Target data
Degree of support available to the open data service	Data of special interest	Data quality requirements
Confirm legal and privacy requirements for the data		Desired levels of detail
		Systems used to process data (gain working context)
Identify obstacles to data access in order to address these issues when designing and building the service		

Where it is of benefit to consult a broader audience email and online questionnaires will be employed.

As development progresses we will make use of demonstrations to enable a more constructive visualisation for users and make interfaces available to early adopters as part of Alpha and Beta stages to obtain feedback to refine the solution. Online feedback forms will be made available to capture feedback through the year so that improvements can be made for the next year's publication.

The process followed by WTG is illustrated in figure 1.

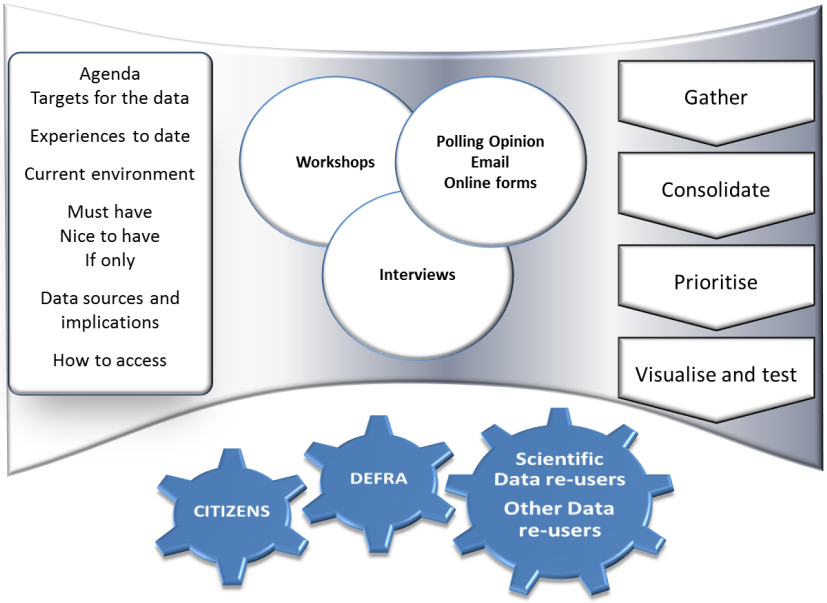

Figure 1: Data Gathering

1.3 Data processing - *How would you take messy data from a variety of sources and… drive a website?*

WTG will address of varying data quality from multiple sources through the following steps:

- Confirm data status and determine scale of quality and integration required. Aspects considered including data volumes, current quality status, identified issues and formats variance.
- Produce a target data model that maps existing data to the target backend incorporating validation rules, transformations, cleansing rules e.g. de-duplication and identify fields where quality issues are expected to persist despite cleansing aims.
- Propose backend technology. Consideration will be given to an appropriate relational database e.g. MySQL, Oracle, or a NoSQL database e.g. MongoDB. Depending on the data sources and the nature of the data they need to store and manage this may be an amalgam of both.

Data Integration and Quality tools will be employed (e.g. Talend), to configure and automate data transformation employing data quality rules for the delivery of Raw and Aggregate statistical data to the target data repository. WTG will use Talend tools to perform an annual Extract, Transform and Load (ETL) operations, which can be used rapidly and repeatedly, to handle data in Excel and existing data stores providing a clean format with retained audit trail to originating sources with versioning for reliable and efficient handling of each year's data. Where volumes of reference data in MS Word are not excessive, WTG will manually transform them into machine readable format (e.g.JSON) for integration. Any transformations will be agreed and tracked for provenance so that underlying data integrity remains intact to maintain its value & integrity.

1.4 Data exploration - *How would you create a useful and meaningful interface… from the data?*

Based on the scenario provided, WTG believe that two interfaces should be considered; one for the citizens and one for the Scientific data/other data re-users. Both will be accessed through a website for the service.

Citizen interface: A highly visual, intuitive and simple interface. We anticipate the data would be largely based on the aggregate and reference data sources and so will propose use of an open source search and visualisation framework such as elastic search and Kibana, or using the GDS visualisation frameworks to enable an efficient and configurable implementation. This approach enables predefined reports to be provided in addition to offering flexible search.

Web Technology Group

Scientific and Other data re-users: These users are likely to have more detailed requirements when compared with the Citizen, WTG propose to expose search and visualisation features to them through the methods already identified. By providing a browseable data hierarchy (e.g. using linked data) users will be able to explore and retrieve data for further analysis on their own systems.

1.5 Legal issues - *What activities would you undertake to tackle the legal issues around… and privacy?*

WTG will confirm with DEFRA (and their legal guidance) on how the data needs to be licensed at the earliest possible point and apply the following courses of action:

- Confirm that data is appropriate for Open license
- Ensure waivers of Copyright & Database rights statements are published with attribution stated.

WTG expect to deliver the data services through a website or microsite. Within the website we will ensure there is content to cover a rights statement including Licencing, Copyright, Privacy statement and Attribution requirements. During the process WTG will review and ensure that details that could be sued to identify or infer identification, are not published, with exclusion or anonymization rules to be agreed with the publisher and embedded in the process.

1.6 Practical issues - *How would you ensure that the data is easy to find… quality are documented?*

To ensure data is easy to find WTG will employ the following techniques:

- Ensure links from DEFRA website(s) & Gov.uk are in place to the original published data.
- Ensure the URLs for the website and data are dereferenceable URIs.
- Register data service with websites that provide online collections data services e.g. data.gov.uk
- We will incorporate linked data, to provide relationships between related data e.g. between aggregated summary and raw data entities.

To manage issues about quality we would:

- Include a section on the website where known data quality issues and limitations are clearly described. This will include information from DEFRA and also any quality issues identified during our transforming and processing of the data. The section will also provide details on how to report data issues.
- We will also provide online contact details so that feedback on data quality can be submitted and collated for review and, if appropriate, subsequent publication on the website. This would also explain our approach to managing data issues.

1.7 Technical issues - *How would you publish different types of data… provide?*

Approach to Publication: We will use the following approaches to publish different data.

- Reference data - published online as HTML and available as XML/JSON via web services API
- Raw data –XML/JSON via web service API and downloadable in CSV format.
- Aggregate statistical data –XML/JSON via web service API and downloadable in CSV format.

Formats and Standards: WTG will use recognised open machine readable formats and standards to expose data, avoiding proprietary formats such as MS Word and Excel, ensuring data can be exposed in JSON or XML, employing GDS, ODI and W3C supported standards to aid interoperability and integration.

Persistent Identifiers: If persistent identifiers (P-IDs) exist within data sets we will use those to uniquely identify and reference records. Examples of this may include ISBN, GUIDs or URNs. If no persistent ID exists WTG will apply a unique ID to all records so that they can be exposed via web services (e.g. a

unique URN and URL is used to return the specific record) and thus WTG will refer to and follow Government guidance on P-IDs e.g. http://standards.data.gov.uk/challenge/persistent-resolvable-identifiers

Additional Metadata: WTG will use metadata to provide accurate, comprehensive details to cover provenance and other information necessary to describe the data. The metadata will include:

- Descriptive data i.e. what the data content is, what language it is in, description
- Legal information i.e. copyright, how is the owner
- Scope i.e. what time and geography is covered
- Relevance i.e. when last updated, how current, update frequency and period of validity.

WTG will use existing metadata format standards & vocabularies including RDF, Dublin Core and DCAT.

1.8 Social issues - *How would you support users of open data… to the publisher?*

The website through which the data service is published will include:

- Data structure – Outline of data available & method of access – variance for version and release
- Provide User and technical Guides including examples for users in HTML and PDF format.
- Provide details of vocabularies or taxonomies present in the data, referring to resources already published and available for reference.
- Provide a FAQ to provide guidance and response to common questions with contact details on how to obtain support, including links to online form for queries.
- We will consider providing an online discussion forum – either on the website or as a community on Social Media e.g. Facebook to encourage and support user discussions on the data. This will also provide an essential source of feedback and commentary on the data.

Feedback to the publisher will be provided through regular service reviews where feedback from users is presented and priorities addressed with DEFRA Service Owner.

1.9 Team - *What team would you put together to provide this service?*

WTG will provide this service through two related teams, one responsible for implementation of the service and the second for the subsequent operation and support of the service.

We will expect DEFRA to provide a Service/Product Owner or manager for the service. This individual would be empowered to take design and implementation decisions on behalf of DEFRA.

Member	Role	Implementation	Operation
Delivery Manager	Manage implementation team	Y	
User Researcher	Lead user interviews and persona creation	Y	
Analyst/Architect	Manage requirements and specification	Y	Y
Developer	Build, integrate and develop services. Perform annual data updates	Y	Y
Tester	QA of service	Y	
Service Manager	Manage operational service		Y