

Strengthening the National Information Infrastructure (NII)

Open Data User Group Review

SCOPING / TERMS OF REFERENCE

1. Purpose

In October 2013 the Government published the first iteration of the National Information Infrastructure (NII) intended to 'contain the data held by the government which is likely to have the broadest and most significant economic impact if made available and accessible outside of government'.

ODUG's view, shared by many in the data community, is that the Government should clearly articulate what constitutes underlying core reference data within the NII, together with a structured release schedule, to maximise the medium term impact of new open data. To assist with an improved definition of the NII and help identify suitable priorities for data release ODUG is developing a position paper on how the NII should be structured.

2. Context

The Government Response to the June 2013 Shakespeare Review set out the aim to create a National Information Infrastructure (NII). In October 2013 the Government published the first iteration of the NII intended to 'contain the data held by the government which is likely to have the broadest and most significant economic impact if made available and accessible outside of government'.

The current NII on data.gov.uk lists 233 datasets, of which 66 are unpublished. The criteria for selecting these 233 datasets are not clear and many datasets which are fundamental to a NII are not listed. For those datasets which are identified there is no indication as to the priority for their release.

The NII as currently defined appears to assume that all datasets are equally important and will form the 'core reference data' the Shakespeare Review referred to when it recommended a twin-track approach to the release of Public Sector Information.

The NII document includes a number of NII actions and deadlines for the government, including Departments and their Arms Length Bodies (ALBs). ODUG has recently questioned the level of progress against these actions since nothing has yet been reported against them. Given the progress to-date we do not expect to see a full definition of the data available inside government, a comprehensive release of new open data, or the NII as currently defined coming to fruition by June 2014.

Open Data User Group (ODUG)

Over the last two years ODUG, via the demand led Data Request mechanism¹, has been able to identify and prioritise key datasets for release, based on user demand. Our work has prompted good progress in some individual cases, but has not addressed the wider issue of public data release by default or what the NII should look like. Our view, shared by many in the data community, is that the government should articulate what constitutes underlying core reference data within the NII, together with a structured release schedule, to maximise the medium term impact of new open data. Therefore, to assist with an improved definition of the NII and help identify suitable priorities for data release, ODUG is developing a position paper on how the NII should be structured.

Whilst a detailed analysis of the framework of public sector responsibilities and tasks is beyond the remit and resources of the group, ODUG will seek to identify important datasets currently omitted from the NII, and their data holders. We will also aim to define core reference data for the NII more tightly – as the set of data and identifiers required to provide the connective tissue across all open datasets. We will also group some of the datasets into collections by sector to show how a combination of datasets, with their underlying metadata, can make newly released open data more immediately beneficial to the user community. The overall scope of ODUG's review will be to:

- Outline a principled foundation of what information we *should* know to support transparency, social & economic growth, efficient & effective public services, and the underlying 'connector' data needed to link different datasets together.
- Identify gaps in the NII as currently defined, and any datasets in the NII that are less likely to provide significant impact.
- Explain the underlying connectivity and interaction between the datasets proposed for inclusion in the NII.
- Set out a tighter definition of core reference data with an explanation of how it delivers a platform to enable the maximum benefit to be derived from open data.
- Determine what metadata is essential to open data users.
- Set out some proposed collections of datasets, by sector, which will deliver high value benefits.

3. Overview of Approach

- 1 We will start by outlining a basic structure of what we should know (5.1), not a reflection of the data currently available. This is a practical exercise to produce a principled foundation for the NII. This structure will cover the requirements for: transparency, social & economic growth, efficient & effective public services and the underlying connector information. We will take a thematic approach, starting with an identified theme (Health & Social care was identified at the 13/5 ODUG meeting)
- 2 The next step is to identify what data is actually available as either open or closed data (5.2) – particularly the connector data.
- 3 We will then identify what is missing from the NII as currently defined as a result of a gap analysis (5.3).

¹ <http://data.gov.uk/data-request>

Open Data User Group (ODUG)

- 4 We will set out proposed priorities for future open data releases (5.4) based on our view of what the NII should contain and proposed sector priorities.
- 5 Metadata is important, and will be identified (5.5) along with any other data publication requirements (use of core identifiers, minimum publication standards etc.).
- 6 Finally we will make suggestions for future improvements to the NII (5.6).

4. Our view of the purpose of the NII

The NII primarily fulfils a strategic purpose; setting out what information we need to know about the UK, to maintain and have insight into the society we live in. The datasets listed in the NII are funded through taxation and are primarily developed to deliver and commission public services. However, these datasets also have huge, and as yet largely untapped, potential as a resource to deliver greater transparency, improved efficiency & effectiveness in the delivery of public services, innovation and social and economic growth.

Specifically we see that the NII should set out:

1. *Structure*: What we should know about the UK, to support transparency, social & economic growth and efficient and effective public services.
2. *Datasets*: Datasets (open and closed), metadata, and the underpinning 'connector' data and identifiers used to link together different datasets.
3. *Priorities for data development*: Based on the gaps in the available data against the structure and the priorities for publishing & improving open data, and providing secure access to closed data.
4. *Data publication*: Recommendations for data publication, for example minimum standards for metadata, use of common identifiers, vocabularies, public registers of core reference components etc.
5. *Future improvements*: Suggested future improvements, for example exploring whether non-government data sources should be included.

5. Activity Plan

5.1 What we should know - NII Structure

Tasks: (1) Review CO themes; (2) Produce draft set of themes; (3) Select a number of themes to focus on (13/5 meeting proposed Health & Social Care) and develop set of 'personas' identifying the data needs under each theme – including core reference data needs; (4) Send out to ODUG for comments; (5) Final version of themes, and data needs / personas under each selected theme.

Precursors: None.

Support needed from CO: None.

5.2 Available data – current NII content

Tasks: (1) Organise the existing 233 NII datasets under the themes produced above.

Precursors: Final version of themes (4.1).

Support needed from CO: CO to provide complete metadata for all existing NII datasets.

5.3 Gap Analysis – the difference between what we should know and the NII as defined

Tasks: (1) Review areas of government by department/ALB and by sector to identify significant datasets which are missing; (2) Review and cross-check the themes/ data needs identified in 4.1, against the NII datasets in 4.2; (3) Based on these reviews, identify gaps in the NII including any gaps in core reference datasets; (4) Based on the review, identify existing NII datasets that are less likely to provide significant impact (and could be dropped).

Precursors: Final versions of themes and data needs (4.1). Current NII content under the themes (4.2).

Support needed from CO: None.

5.4 Priorities for open data release

Tasks: (1) Review the existing NII datasets, and gap analysis in 4.3; (2) Draft priorities for publishing & improving open data, and/ or providing secure access to closed data. To include updates (or 'real-time' access) to existing datasets, as well as new data releases; (3) Send out to ODUG for comments; (4) Final version of priorities, by theme.

Precursors: Gap analysis (4.3). Existing NII indicators by theme (4.2).

Support needed from CO: None.

5.5 Metadata and data publication requirements

Tasks: (1) Review existing NII metadata and datasets, and identify data fields that (should) use core reference identifiers; (2) Review existing metadata and produce minimum specification of metadata that is essential to open data users (including use of core reference identifiers).

Precursors: None.

Open Data User Group (ODUG)

Support needed from CO: CO to provide complete metadata, and actual dataset examples for all existing NII datasets.

5.6 Future improvements to the NII

Tasks: (1) Produce recommendations for future developments to strengthen the NII.

Precursors: None.

Support needed from CO: None.