QUALITY METRICS NATIONAL TEST OPEN DATA GUIDE

Introduction

This document provides an overview of the open data set produced alongside the Quality Metrics National Test. It also provides some examples on how to begin looking at the data in its published format. The data set has been provided to enable interested parties to explore it beyond the analysis in the Quality Metrics National Test report where an extensive selection of tables has been published. The open data resource is owned by Arts Council England. Arts Council England is unable to provide further support for those wishing to interrogate the data.

This data set provides a summary of the aggregate data set and does not expose the raw data collected through the Quality Metrics National Test. For the purposes of anonymity, some of the detail in the raw data set has been excluded. The complete aggregate data set was analysed and results published in the Quality Metrics National Test Report.

Data Collection

This section outlines the data presented in the open data set. For more detail, please refer to the Quality Metrics National Test Report.

The data set is principally made up of evaluations between November 2015 and May 2016, of work produced by a broad range of organisations in the funded National Portfolio. Each evaluation is classed here as an event.

Event Choice

Choosing events for the trial was in the hands of the participating cultural organisations. It is reiterated here that whilst this sample is broadly representative of the whole funded portfolio it is inevitably incomplete.

Respondent Categories

A key element of this methodology was the triangulation of scores within respondent categories. Comparing and contrasting the respondent types provides a much more complete perspective on the work based upon the perspectives of the creators and production teams, other professionals, and the public audiences.

The respondent categories were as follows:

- Self (time category: prior)
- Self (time category: post)
- Peer (time category: post)
- Public (time category: null)

Prior responses were collected before the event was presented to the public, post responses were collected after the first presentation of the work.

Dimension Data

The core quality metrics chosen for the national test are outlined in Table 1 below. Nine dimensions were used in surveys designed for all respondent categories, with an additional three (risk, originality, and excellence) for self and peer respondents only. Data for the metrics are listed under their dimension name in the dimensions table in the open data set.

Table 1: Quality Metrics

| Dimension | Statement | Respondent Type | | |
		Self	Peer	Public
Concept	it was an interesting idea	✓	✓	✓
Presentation	it was well produced and presented	✓	✓	✓
Distinctiveness	it was different from things I've experienced before	✓	✓	✓
Captivation	it was absorbing and held my attention	✓	✓	✓
Challenge	it was thought-provoking	✓	✓	✓
Enthusiasm	I would come to something like this again	✓	✓	✓
Local Impact	it is important that it's happening here	✓	✓	✓
Relevance	it had something to say about the world in which we live	✓	✓	✓
Rigour	it was well thought-through and put together	✓	✓	✓
Risk	the artists/curators were not afraid to try new things	✓	✓	-
Originality	it was ground-breaking	✓	✓	-
Excellence	it is one of the best examples of its type that I have seen	✓	✓	-

Metadata

In addition to the core quality metrics data collected, metadata was also assigned to responses or events accordingly.

The metadata exposed in the open data set falls in to four categories:

Table 2: Open Metadata

Metadata	Field Examples	Process
Non-question Evaluation Data	Evaluation reference, respondent category, time category	Assigned by Culture Counts
Event Artform & Artform Attribute	Category types (e.g. artform, medium), Values (e.g. theatre, textile)	Supplied by organisations in the trial; structure developed by Culture Counts
Event Location	Event Region	Event address supplied by organisations in the trial, mapped against Arts Council England's regional categorisations
Geomapping Data	Rural Description	Office of National Statistics Open Data mapped against raw event postcodes[1]

Understanding Artforms & Artform Attributes

An evaluation can have any number of artform or artform attributes. The terminology was developed with data sourced from each of the organisations in the trial. Artforms and artform attributes are categorised. Categories for an artform or artform attribute can be found in the 'Field' column.

A process of inference has been applied to artforms. This means that when a specific artform has been described such as ballet, it also gets assigned with a broad artform (dance) and a sensory artform (movement). Inference has not been applied to artform attributes. For more information on this, refer to the Quality Metrics National Test report.

---

[1] For more information visit:
http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/geography/products/area-classifications/2011-rural-urban/index.html

Open Data Set

The open data set contains three data tables. The fields are as follows:

Table 3: The Open Data Tables

Table	Column	Data Type	Description
✔ all tables	Eval Reference	Integer	Numerical reference identifying individual evaluations in the dataset
dimensions	respondent-category	String	Filter this column for self, peer and public respondent categories
	time-category	String	Filter this column for prior and post event values (N.B. public rows are null in this column)
	key measure*	String	The values in this field represent the values for each measure in the interquartile range.
	Concept	Integer	Dimension value
	Presentation	Integer	Dimension value
	Distinctiveness	Integer	Dimension value
	Challenge	Integer	Dimension value
	Captivation	Integer	Dimension value
	Enthusiasm	Integer	Dimension value
	Local Impact	Integer	Dimension value
	Relevance	Integer	Dimension value
	Rigour	Integer	Dimension value
	Risk	Integer	Dimension value
	Originality	Integer	Dimension value
	Excellence	Integer	Dimension value
event location	Online Location Flag	Boolean	True = work is online, False = offline
	England Location Flag	Boolean	True = work is based in England, False = anywhere else
	Multiple Location Flag	Boolean	True = if the evaluation included responses from multiple locations, False = one location
	Event Region	String	These regions are based on the arts council regions and relate to the <u>event</u> location.
	Rural Description	String	These 6 values are matched from ONS open data rural categorisation
event artforms	Field	String	These are categories e.g. artform or types of artform attributes
	Value	String	These are the values of the artform fields such as 'theatre', or 'contemporary'

Using the Open Data Set

This section is aimed at those looking at the quality metrics open data set for the first time, and for those who do not regularly use data manipulation tools. The suggestions that follow provide a way in to enquiry and are by no means the only way in which the data can be interrogated.

All tables include an evaluation reference. This field functions as a unique ID for an evaluation and is attached to data in each of the three tables.

Dimensions Table

Each evaluation is summarised with 5 numbers by respondent/time category (self prior, self post, peer and public) per dimension. This provided an accurate summary of the raw data using medians and interquartile ranges. The interquartile range value can be selected by the Key Measure field: Q0, Q1, Q2, Q3, Q4. Q0 is the minimum value, Q2 is the median, and Q4 is the maximum. Q1 and Q3 are the 25th and 75th percentiles respectively.

For one evaluation, or event, there may be as many as twenty number per dimension as shown in Table 4. As can be seen, these rows all relate to evaluation reference 4, there are four respondent categories/ time categories, and five key measure for each respondent/time category combination.

If a respondent category is missing from the data, it is because that evaluation did not collect data from that group.

Table 4: example groups

Eval Reference	respondent-category	time-category	key measure
4	Self	prior	Min_Q0
4	Self	prior	25_Q1
4	Self	prior	Med_Q2
4	Self	prior	75_Q3
4	Self	prior	Max_Q4
4	Self	post	Min_Q0
4	Self	post	25_Q1
4	Self	post	Med_Q2
4	Self	post	75_Q3
4	Self	post	Max_Q4
4	Peer	post	Min_Q0
4	Peer	post	25_Q1
4	Peer	post	Med_Q2
4	Peer	post	75_Q3
4	Peer	post	Max_Q4
4	Public		Min_Q0
4	Public		25_Q1
4	Public		Med_Q2
4	Public		75_Q3
4	Public		Max_Q4

Null values in the dimension table are blank. This occurs when data for a particular dimension was not collected in that particular evaluation. It is important to remember to filter

the data based on the columns in Table 4. i.e. respondent category, time-category and key measure. A good place to start is to use the median value (Med_Q2) as the key measure. Each table enables selection of a given value by clicking on the little dropdown icon visible next to a column header, as circled across.

The next step is to filter again, this time on the respondent category. A good place to start is by selecting the respondent category: public. This enables a visualisation of the aggregate baseline; a crude dimension profile of the whole dataset.

Example 1: Creating the Aggregate Baseline

1. Download the open dataset to your local computer. These instructions will assume you are using Excel
2. Open the dimensions table by selecting the dimensions tab in the bottom left of the spreadsheet
3. Filter the data:
   a. key measure: Med_Q2
   b. respondent-category: public
4. Select the filtered table, and copy and paste it in to a new worksheet
5. Select a non-blank cell in the new worksheet where the filtered table has now been copied. Then click on the Insert ribbon[2] (top left) and select Table. This should now highlight all the newly copied cells and recognise the column headers. Click OK and Excel will now recognise the data as a new table.
6. Save the table as public_median by changing the table name in the top left corner of the Table Tools ribbon. It can also be useful to name the tab at the bottom of the sheet by the same name.
7. Insert a row above the table (not in the table).
8. In the cell above the Concept header type in the formula =AVERAGE(E3:E354) – *see image*

9. Copy this formula above all the dimension headers. It should adjust automatically so that in column F (Presentation), it averages data in column F i.e. the copied formula becomes =AVERAGE(F3:F354) and so on.
10. Insert two rows above and one below the formula row, and copy the dimension headers above the formula row. Add in labels for the two new rows: dimension and average. This keeps the most important information at the top of the sheet and makes the information easier to see. It should now look like this:

11. Select a non-blank cell in this new group e.g. D2 as shown above and return to the Insert ribbon near the top of the screen. Select charts > 2D column chart. This should

---

[2] https://msdn.microsoft.com/en-us/library/windows/desktop/dn742393(v=vs.85).aspx

recognise the new data and automatically generate a chart of the average values newly calculated.

12. The chart should look something like the chart below. You may wish to edit colours and chart elements in the design ribbon.
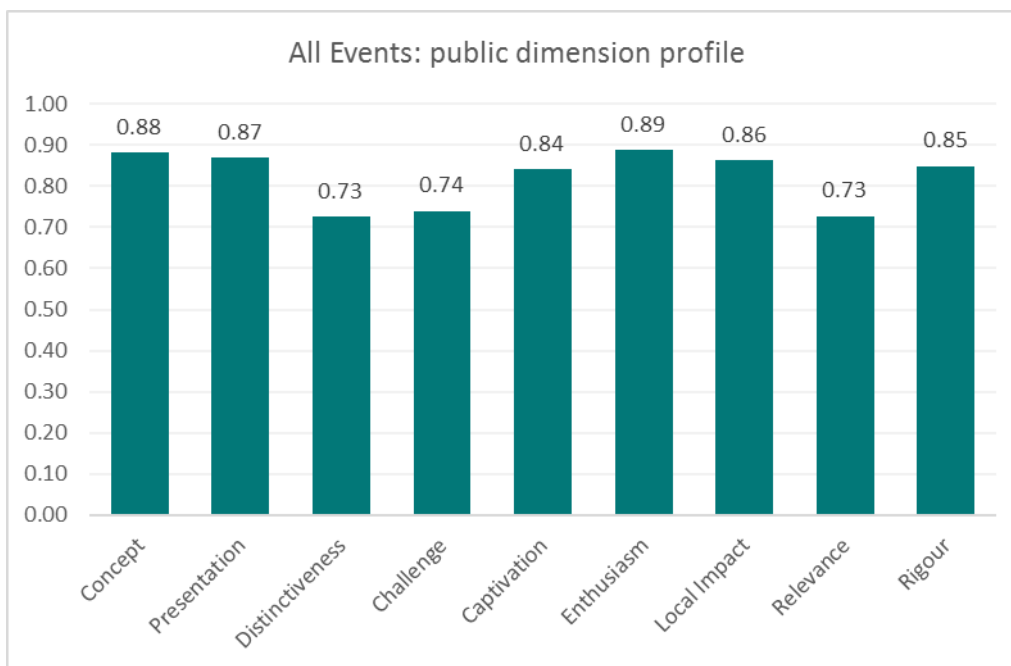
**Looking at dimension scores by location or artform**

There are a number of ways to cut and explore the data. Depending on how complex the exploration is somewhat dictates the method in which to organise the data. Below are some suggestions which can be used in Excel workbooks. Again, these are listed as starting points.

It is important when manipulating any dataset to keep track of the tables that are created. It is advisable when creating multiple tables and datacuts to do so systematically based upon the exploration planned.

Example 2: Filtering tables (skill level: easy)

Each table can be filtered using the dropdown icon on the table column headers. A simple way to get started is by filtering on a location (or artform) and noting which evaluation numbers remain. For example, to look at all work with the subject of Christmas:

1. In the event artforms tab dropdown the value column and select "Christmas". In this case, only the subject of Christmas exists. In some cases, a value can have multiple field types, so it is important to make sure the field column is exactly what is required as well as the value column.
2. Note each of the evaluation references. It can be useful to copy and paste them in to a notepad so they can be viewed at the same time as the Excel table being worked in.
3. Go to the dimensions table and use the dropdown for the Eval Reference column and select the first evaluation reference (26)
4. Then add further evaluations from the list and search and select both the evaluation reference being added to the filter and "Add current selection to filter" – See image:
5. Once all the evaluations are selected a summary can be created by following steps 4 onwards from Example 1.
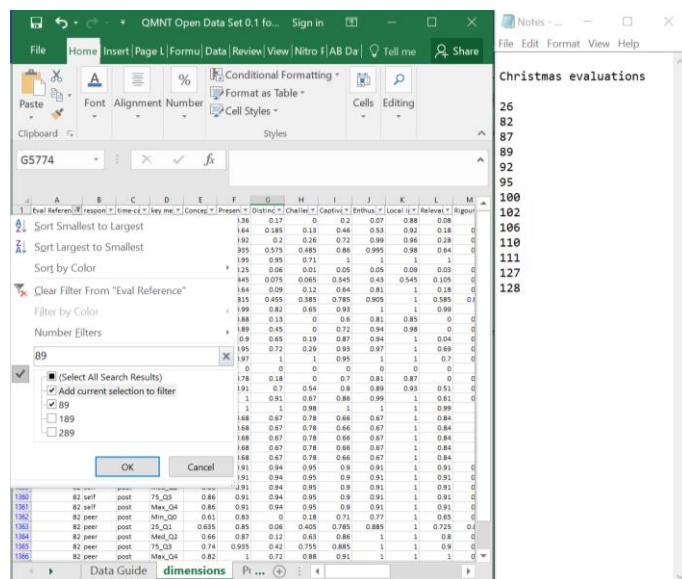
It can be useful to keep a main dimensions table for trying different data cuts, and copy/paste(values only) average summaries in to a new sheet with suitable labels so that they can be returned to and compared.

Example 3a: Using LOOKUP for location data (skill level: intermediate)

It is worth noting that multiple rows exist per evaluation in this dataset in the dimensions and artform tables. This means that simple stitching of tables together is not appropriate. In some cases it is more efficient to create tables based on a specific exploration. For example, if only the median values are of interest to start with, it is best to create a median value only table for each of the respondent/time categories (four tables in total).

Using the LOOKUP[3] formula can be a useful way of adding data from one table to another.

Attaching data from the event location table to the dimension table can be done as follows:

In the dimensions sheet, in cell Q1, add a column to the dimensions table by typing the name of a new column. In this example, "Region". In the cell below (Q2) type in the lookup formula:

=LOOKUP([@[Eval Reference]],event_location[Eval Reference],event_location[Event Region])

To add in a different column, repeat the step starting in R1, for example, by typing "Rural Description" and repeat the lookup formula, this time specifying the Rural Description column to be retrieved:

=LOOKUP([@[Eval Reference]],event_location[Eval Reference],event_location[Rural Description])

These columns should have by now automatically been populated and look like this:

The dimension data can now be filtered on Region and Rural Description.

To create representative charts based on a filter, follow the steps in Example 1, with the filters applied as required.

A separate table using the AVERAGEIF[4] function might be useful for some users.

---

[3] https://support.office.com/en-gb/article/LOOKUP-function-446d94af-663b-451d-8251-369d5e3864cb
[4] https://support.office.com/en-gb/article/AVERAGEIF-function-faec8e2e-0dec-4308-af69-f5576d8ac642

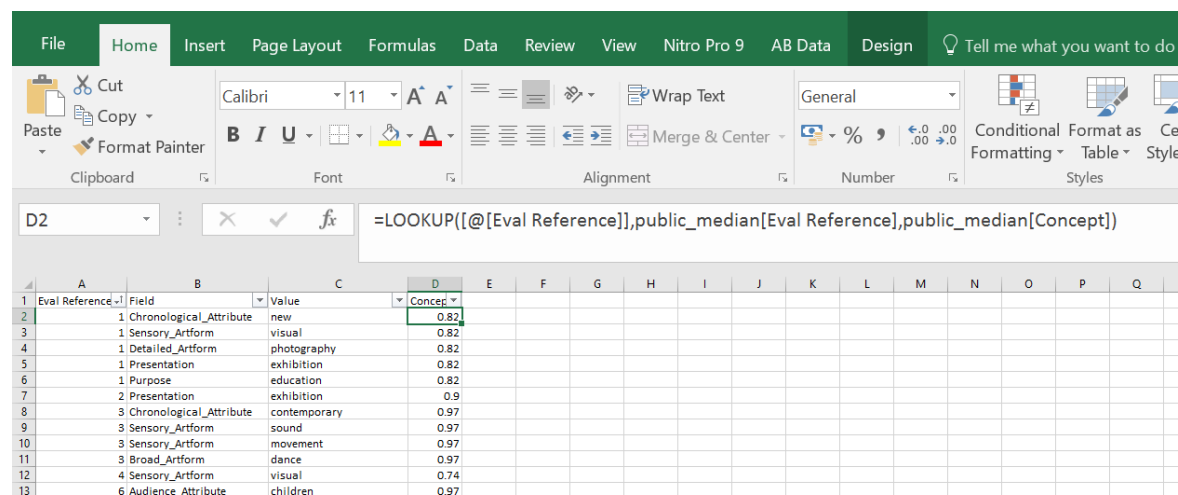**Example 3b: Using LOOKUP for artform data (skill level: intermediate)**

Combining data from the dimensions with the artform table is more complicated. It is advisable to join the dimensions data TO the artform table, and do this so that only one row exists per evaluation in the dimensions table. This is because multiple artforms and attributes can be assigned to an evaluation. In this instance, use the public median table created in Example 1. This table already only has one value per evaluation.

It might be useful to attach any event location data using LOOKUP, as explained above to the public median table before working with artforms.

Next, in the event_artforms table, starting with cell D1, type in the column header for attaching. e.g. Concept.

Type in the following LOOKUP formula in to cell D2:

=LOOKUP([@[Eval Reference]],public_median[Eval Reference],public_median[Concept])

Repeat this for each column, referencing the corresponding column header in the final part of the formula.

Now the table can be filtered based on artform and artform attributes. It is useful to remember here that the data transferred is only that for public responses. This process would need to be repeated for self prior, self post or peer responses.

It is also useful to remember that when working in the artform table, some evaluations feature more than others. This is due to the amount of metadata attached to a given evaluation. Only tags that were assigned to five or more evaluations are included in the open data set.

A Note on Analysis and Interpretation

The purpose of this document is to provide an overview to getting started with the open data set. For ideas on how to cut the data, many examples are shown in the Quality Metrics National Test Report.

For users who have collected their own quality metrics data, ideas on analysing the data can be found on the quality metrics YouTube channel, run by Culture Counts in response to tutorial requests from organisations in the national test cohort. Visit: https://www.youtube.com/channel/UCm6xwczzZuYL97REMqYCNPw

There are also a number of case studies on the national test website where interpretation for some specific evaluations is written in more detail. Visit: http://www.qualitymetricsnationaltest.co.uk/case-studies

This data guide was written by Alison Whitaker, Culture Counts. August 2016.