

## **PISA Digital Humanities Summer School, June 2022**

### **Morning Exercise, 11 am to 12:30, 16 June 2022**

#### **Meaning-making and storytelling: Investigating Open Data**

In this exercise in teams of 3-4 participants you will examine an open data set to develop a critical insight into the data set and an appreciation of its strengths and weaknesses as a resource. While goal of working with open data is to build stories that help us to understand or explain phenomena and to create meaning, this exercise focuses on the first step in this process—how should one look critically at an open data set. Think about the kinds of questions these data sets would allow you explore and the kinds of analyses you might do with them.

We will devote 70 minutes to this activity and each team will offer a 3-minute report back on their findings. Before you start the exercise, please read the whole exercise.

Each group will be allocated an open data portal (in this case these are national level resources, but they could equally well be regional or municipal resources). In the exercise you should spend about 15 minutes explore the different kinds of data that is available at the portal and the ways in which it made available. After about 10 minutes think of a question your group would like to explore using open data and find some data sets that might allow you to do that. Then you should focus on one data set about which you should consider the following key issues:

The issues you should consider relate to the provenance of the data set, the paradata and metadata that is provided to help you understand its origin, its potential uses, the level of documentation, accessibility, currency, context, reusability, interoperability, historicity.

Example questions might be:

- Is the description (title, statement about the data set) adequate? (Is there data dictionary support?)
- What is the data set about or intended to be used for?
- What is the currency of the data set?
- How comprehensive is the data set?
- Are metadata available and do they help you to understand the data set?
- Can you easily understand the structure of the data and its content?
- Are there sufficient paradata for you to assess how the data were collected, represented, and encoded?
- Are the data consistent?
- What is the overall quality of the data and the data set?
- What are limitations on use of the data set – legal, technical, quality ?
- Does the dataset have long term viability? (Is it updated regularly with earlier versions still available).
- What formats are the data provided in?
- What tools does the portal provide for exploring the data set in question?
- Do these tools enable you to use this data set in conjunction with others?
- Is the data clean? If not, how difficult would it be to clean it?

- If you wished to download the data set to use it in conjunction with data sets drawn from other portals how much effort would be involved in the process?
- Is there any information about the purposes for which the data set has been used in the past?

You should download the data set to explore it, if the tools that you need are not readily available on the portal.

Finally your team should leave enough time to explore what other data sets you might wish to use in conjunction with this data set.

*Report Back:*

Your report back should offer an overview of the story you wanted to use the data to construct. Why you chose the data set? And answers to the questions that you explored about the data set.

16 June 2022

Seamus Ross

**Open Data Sites to be Allocated:**

Group 1: Canada, <https://open.canada.ca/en/open-data>

Group 2: Ireland, <https://data.gov.ie/>

Group 3: New Zealand, <https://www.data.govt.nz/>

Group 4: Italy, <https://www.dati.gov.it/>

Group 5: United Kingdom, <https://data.gov.uk/>

Group 6: European Open Data Portal, <https://data.europa.eu/en>

(Aggregates data from across EU and also publishes European Commission data).

Group 7: Germany, <https://www.govdata.de/>