Predicting cause and effect of data strategies – internal question scoping

This document summarises work done to internally scope the policy questions we want to answer during the predicting cause and effect of data strategies project.

Approach to forming questions

Based on a workshop within the ODI, we uncovered a lot of potential questions that we would like to answer in the form: **'What happens to *[feature of the data economy]* if *[a particular data policy is implemented]*?'**.

From this exercise and follow-up conversations, we identified that we should;

1. **Limit the scope of the *'features of the data economy'* we want to understand**, and focus on the different policy questions we might want to ask
2. **Limit the number of *'data policy areas'* we focus on, at least initially,** in order to build simpler models which could then be iterated or expanded.
3. **Identify generic questions and later apply them to specific sectors,** this will allow us to narrow the scope, if required, during the model building process and also allow us to build on existing knowledge.

Scoping potential *'features of the data economy'*

As we identified, we need to scope down the number of features of the data economy that might be affected by different data policies. From the workshop session, some clear features of interest emerged:

1. **Economic / social value**
2. **Data availability / sharing / reuse**
3. **Trust**
4. **Competition**
5. **Innovation**

By choosing to monitor one, or a number of these features, in response to policy changes or changes in initial conditions, we should be able to understand potential impact.

Scoping potential *'data policy areas'*

There were many different types of data policies and strategies that were identified, in order to understand the potential generic questions we could ask, we have summarised the broad areas of questions. **What happens if:**

1. **more/particular types of data is shared / opened?** (by government or business)? (e.g. open activity data, OS Mastermap, high value data)
2. **data sharing / opening is imposed by regulation or incentivised by persuasion?** (e.g. Buses Bill, OpenActive)
3. **different methods of opening / sharing data are used?** (e.g. data trusts, logins required, APIs vs downloads)
4. **data literacy and skills are improved?** (e.g. introduce programmes)
5. **innovation programmes are launched?** (e.g. startup accelerators)

There is also another area of questions that primarily relate to the **bestowing or exercising of individual rights over data** (e.g. portability). These are being treated somewhat separately as the implication of asking these questions is that they would have a significant impact on the shape of the model (they are based primarily on individual behaviour and less on company behaviour). We need to make a decision about whether we should pursue these questions or those focused on business, or business and government, focused models (see the 'scoping our approach' document).

Scoping potential sectors

Based on initial discussions and the workshop, we have identified some of the sectors we might wish to use to scope down questions during the model building phase, subject to requirements for data/models/theories of change and whether they are useful for examining the questions we want to answer. Key sectors identified include;

1. **Physical activity** (OpenActive)
2. **Geospatial**
3. **Agriculture** (GODAN)

Also suggested, though less well defined are:

1. **Artificial Intelligence**
2. **Transport**
3. Antimicrobial Resistance (AmR)