University of Southampton's response to the Cabinet Office's consultation on open data

***Do the definitions of the key terms go far enough or too far?***

***Where a decision is being taken about whether to make a dataset open, what tests should be applied?***

***How do we get the right balance in relation to the range of organisations (providers of public services) our policy proposals apply to? What threshold would be appropriate to determine the range of public services in scope and what key criteria should inform this?***

The University of Southampton is committed to openness, accountability and transparency, and we understand the importance of these values in maintaining our credibility with students and the broader public. The consultation covers how public bodies and providers of public services might be held to account for delivering open data. It is unclear whether there is an intention for universities to be included in this definition. Universities have a range of different governance models, but most are exempt charities or registered charities, and not classified as public sector entities in national statistics. However, a significant proportion of their activity is publicly funded, and some (but not all) of their activities might be deemed a public service.

Universities are organisations built on the generation and sharing of knowledge, and we believe that they should be at the forefront of developing and shaping the open data revolution. The University of Southampton has already been at the forefront of developing the technologies to enable the open data revolution to be possible, and academic staff from our University have been helping the Government develop this world-leading initiative.

Whether or not they are officially classified within the definition of "public bodies or providers of public services", we believe that universities should therefore seek to place as much of their own data as possible in the public domain. The University of Southampton is already doing this with its open data site, data.soton.ac.uk. We believe that this will promote efficiency and drive improvement, and make universities more competitive, as well as allowing others to add more value to the data. For example, we see students writing applications that use the data and provide useful services for everyone.

There are clearly some data held by universities that cannot be shared. These include all research data generated from projects funded by the private sector, which clearly belong to those funding that research.

***What would be appropriate mechanisms to encourage or ensure publication of data by public service providers?***

The FOI Act requires all public authorities to have an approved publication scheme, which is a means of providing access to information proactively. The Information Commissioner's Office (ICO) has developed a model publication scheme that all public authorities must adopt. The scheme defines the types of information that must be published routinely, how it must be published, and what charges can be levied for it.

Further definition of this information is provided by a definition document. The ICO has been working with the HE sector to update the definition document for universities, and the document requires the following information to be published:

- Who we are and what we do (organisational information, structures, location and contacts)

- What we spend and how we spend it (e.g., accounts, procurement, financial audit)
- What are our priorities and how we are doing (e.g., strategies, performance indicators, audits)
- How we make decisions (e.g., minutes of governing bodies)
- Our policies and procedures (including research policy and strategy)
- Lists and registers (e.g., asset registers, registers of interests, etc.)
- The services we offer (e.g., prospectuses, fee-based services, etc.)

We believe that this should be extended to include more. For example,

- Performance data (UCAS codes, course codes, HESA data, student satisfaction data, HEFCE KIS data etc)

We consider the ICO's model publication scheme to be an appropriate mechanism by which to ensure the publication of data by public service providers, and would urge building on this existing system rather than starting from scratch as open data for universities is taken forward.

***Is there a role for government to stimulate innovation in the use of open data? If so, what is the best way to achieve this?***

We note that the consultation mentions that open data has the potential to drive economic growth and quotes the view of the Royal Society's Working Group on open data, that the meta-analysis of the raw data from clinical trials is a fine example of the benefits of data-sharing (paragraphs A1.46 and A1.47).

Without pre-judging the final report of the Royal Society Working Group, we believe that in some cases, open data sharing of publicly-funded research at too early a stage could actually harm, and be detrimental to, economic growth. We would expect that universities would have a period of time from which to extract value from their data, including (if appropriate) protecting any IP by securing a patent.

From our own experience with data.southampton.ac.uk we would like to detail the following as good practice with respect to the issue of ***setting transparency and open data standards***

1. Processes be established from the outset to ensure Open Data is maintained and that this occurs close to where the data is collected and generated.
2. Develop mechanisms to enhance and improve data – some of it will be wrong.
3. Use an open license (for example the Open Government Licence).
4. Aspire to 5* data (linked, with URIs) but at least ensure it is machined readable— i.e. not a summary document claiming to be a spreadsheet.
5. Link data to documents describing the policies on update frequency, and any qualifications about the data.
6. Design good and persistent URIs for all entities, ideally linking to other datasets, and try and use the best practice and URI designs of others.
7. Provide access to data about each open data item (e.g. a course) in RDF and JSON. For important resources – courses, facilities, transport access points etc., make these available in structured formats such as CSV, no matter what the underlying data is.
8. Provide a human readable HTML view of items of data suitable for use by normal citizens.
9. Provide at least one service or utility that returns value to the originating data provider – for example by enhancing their website or otherwise helping them meet an obligation of their office.
10. Always look to enhance an existing current process in the course of publishing data and do not create new work for the data provider.