Open data needs for researchers and academics

data.europa.eu
The official portal for European data

This study has been prepared as part of data.europa.eu. Data.europa.eu is an initiative of the European Commission. The Publications Office of the European Union is responsible for contract management of data.europa.eu.

For more information about this paper, please contact:

European Commission
Directorate-General for Communications Networks, Content and Technology
Unit G.1 Data Policy and Innovation
Email: CNECT-G1@ec.europa.eu

data.europa.eu
Email: info@data.europa.eu

Written by:
Oscar Corcho (https://orcid.org/0000-0002-9260-0753)

Last update: 12 July 2024
Internet: https://data.europa.eu/

Luxembourg: Publications Office of the European Union, 2024

ISBN: 978-92-78-44244-6 doi: 10.2830/603563 Catalogue number: OA-02-24-794-EN-N

Contents

Executive summary

This report is the first in a series, the overall objective of which is to provide a better understanding of the links between the worlds of open government data, open science and education, which are tightly interrelated but seem to be often disconnected from each other in their current activities and recommendations. As a result of the research carried out in this series of reports, we will identify current limitations and challenges and will provide recommendations on how government officials may increase their awareness of developments and trends in the research world (e.g. the European Open Science Cloud), and how researchers and academics may increase the use of open datasets in their scientific or academic work. Our work will also focus on the main limitations that researchers and academics have when trying to make use of open datasets, so as to provide government data providers with better insight into how they could make their data more easily available to researchers and academia.

In this first report we characterise three types of (potential) users and providers of open data in the research and academic contexts. This characterisation is done by developing three 'personas' (a climate change researcher, a lecturer on data journalism and an open data officer).

We also provide four initial recommendations to make the data from data.europa.eu more readily available and easier to use by researchers and academics: (a) create a specific searchable tag or subsite with data stories where examples of the usage of open government data for research or academic activities are shown; (b) allow academics to share their learning materials (and teaching experiences) on the use of open datasets on open data portal sites; (c) provide funding opportunities for academics to create academic content associated with open datasets; and (d) provide some form of federation of some of the content from data.europa.eu into services related to the European Open Science Cloud for data archival (e.g. Zenodo).

1. Introduction and motivation

Government data plays a crucial role in academic research across various disciplines, from ICT, engineering and the natural sciences to the social sciences and the humanities. Government data, be it combined or not with data sources from other types of providers, may provide sufficient evidence to support or refute hypotheses, theories or research questions. Moreover, transparent and well-documented data allows other researchers to validate and reproduce experiments or studies. Data providers and data reusers must have a clear understanding of the other's wants and needs.

However, on the one hand, providers of government data are not always aware of the developments and trends in data infrastructures or in common practices used in the context of research. The European Open Science Cloud (EOSC) ([1]) is an example of such an infrastructure designed to create a virtual environment for sharing and accessing research data across borders and scientific disciplines (ICT, natural sciences, social sciences, humanities). The EOSC can potentially contribute to the increased reuse of data, stimulating collaborations between government agencies and researchers.

On the other hand, researchers and academics from many different disciplines may not always be aware of the existence of open government data portals such as data.europa.eu or of the opportunities that those datasets may provide in bringing value to research activities and educational contexts, such as lectures, hands-on projects and other academic activities at all levels of education (from primary and secondary school to university). Increasing this awareness may result in a higher degree of data reuse in research and academia.

This report is the first in a series, the overall objective of which is to provide a better understanding of the links between the worlds of open government data, open science and education, which are tightly interrelated but seem to be often disconnected from each other in their current activities and recommendations. As a result of the research carried out in this series of reports, we will identify current limitations and challenges and will provide recommendations on how government officials may increase their awareness of developments and trends in the research world (e.g. the European Open Science Cloud (EOSC)) and on how researchers and academics may increase the use of open datasets in their scientific or academic work. Our work will also focus on the main limitations that researchers and academics have when trying to make use of open datasets, so as to provide government data providers with better insight into how they could make their data more easily available to researchers and academia.

In this first report we characterise several types of (potential) users and providers of open data in the research and academic contexts. This characterisation of types of users and providers is done by developing 'personas' for a variety of researchers and academics, along with government officials working with open data initiatives. A 'persona' is a personalised fictional character created to represent a user type that might use a service. It encapsulates the common characteristics found within consumer groups or segments, derived from the collection of demographic and behavioural data sourced through user interactions, qualitative interviews and participant observation. In the case

([1]) https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en.

of researchers and academics, they span several scientific and academic fields, including the natural and social sciences, engineering, etc.

In this report we also provide some initial recommendations (in the short, medium and long term) to make the data from data.europa.eu more readily available and easier to use by researchers and academics.

The report is structured as follows. Section 2 provides a glossary and some relevant definitions used throughout the report. Section 3 describes the methodology used for the research that is covered by this report, including the selection of interviewees (academics, researchers and government officials) and the main parts covered in the structured interviews. Section 4 provides the main highlights from the interviews and includes the 'personas' that have been developed. Finally, Section 5 provides an initial set of actionable recommendations to make the data from data.europa.eu more readily available to the community of potential users from academia and research.

2. Glossary and relevant definitions

In this section we compile some relevant terms and definitions that are used throughout the report, with references to relevant literature sources where such definitions have been coined or are further explained.

Open science 'is a framework for how scientists interact with one another and how the public engages with, and is engaged in, science' (European Commission, 2018). Data is a key pillar of any open-science strategy, alongside methods and research management tools. This includes open access, which is driven by the understanding that publicly funded research should be accessible to all members of society. The open-science imperative of sharing information and results from publicly funded research has led to the promotion of the open-access publication model (where scientific publications are freely available rather than subject to expensive subscription rates) and open scientific data repositories (where datasets are made freely available to other potential users).

Personas are fictional characters that are created in the context of service and product design, based upon research, to represent the different user types that might use such service, product, site or brand in a similar way. Creating personas helps data providers understand users' needs, experiences, behaviours and goals. It can also help them recognise that different people have different needs and expectations and identify the types of users that the service or product is being designed for.

3. Methodology

The research that is reported in this document has been carried out following a methodology based on four main steps.

Step 1. User research. We have gathered data using structured interviews, so as to identify the specific challenges and opportunities that academics, researchers and government officials see when they need to use or to provide open data. The selection of the interviewees was carried out as follows.

— We distributed an open call for volunteers, to identify potential interviewees, in three different forums: on the network of the members of the EOSC Association, so as to identify researchers; on the networks of the members of the association of universities Conference of European

Schools for Advanced Engineering Education and Research and of the research infrastructures Common Language Resources and Technology Infrastructure and Digital Research Infrastructure for the Arts and Humanities, so as to identify academics at the university level; and on our own network of contacts related to open government data, so as to identify government officials related to open data initiatives (asking them as well to identify teachers in primary and/or secondary schools that may be using open data datasets, according to their knowledge). In all of our communications, we asked for expressions of interest from people with some experience in using open data for their research or academic activities, and government officials with some experience in offering data for these types of activities.

- We received a total of 32 expressions of interest (15 from researchers, eight from university academics, two from secondary schoolteachers and seven from open government data providers), from 12 different EU Member States, at different levels of seniority and engagement with open data initiatives.
- We carefully analysed the expressions of interest and selected six researchers from different disciplines, five academics working in education at different levels (one in secondary school and four in universities) and three government officials at different levels (local, regional and national).
- We ran the interviews in the period between 20 February and 16 April 2024, using videoconference or phone facilities, depending on the availability of the interviewees, and in two specific cases in the context of in-person events (the 2024 Transport Research Arena conference, in Dublin, Ireland, and the Spanish Federation of Municipalities and Provinces meeting for the presentation of the data governance ordinance template, in Madrid, Spain). The script used for the interviews (with the required variations depending on the profile group) is available in the Annex.

Step 2. Identification of common patterns and themes in the interviews. The transcripts of the interviews have been analysed to identify common patterns and themes, recurring traits, behaviours, demographics and characteristics among the target users. The objective of this analysis is to inform the development of the persona profiles and to derive an initial set of recommendations for each user group.

Step 3. Creation of persona profiles. We decided to generate one persona profile for each of the selected groups. Such descriptions are delivered in a visual manner using a common template, following common practices in persona development. We have also used storytelling techniques (e.g. ‘a day in the life of person X’). The template covers aspects like education, professional activities, academic and research activities (when applicable) and general interests.

Step 4. Actionable recommendations. As a result of the analysis of the interviews conducted in step 2 and the work done towards the development of the persona profiles, an initial set of recommendations for data.europa.eu was generated, with the objective of improving the provision of data so that it can be more easily reused by researchers and academics.

It is important to note that the research methodology followed and the number of interviews carried out were not aimed at providing statistically significant quantitative results, for which we would have needed a larger set of interviewees or for which we should have used other techniques, such as an online questionnaire or survey, but rather at identifying some commonalities and differences across

some carefully selected individuals, and revealing challenges and opportunities at the intersection of open government data, open science and education.

4. Results and discussion

This section provides the results of the analysis of the interviews carried out and the resulting personas that were generated after the analysis. The analysis of the interviews is organised according to the three main groups that were discussed above. This analysis informed the creation of the personas as representative (but non-exhaustive) profiles that can guide further system/functionality development or decision-making in this area.

4.1. Summary of the interviews with researchers

As discussed in Section 3, we carried out six interviews with researchers from a variety of disciplines. We considered different levels of seniority in their academic careers, ranging from PhD students to senior researchers, along with their various nationalities and research areas (two in engineering – computer science and transport; two in the natural sciences – biodiversity and climate; and two in the social sciences and humanities – cultural heritage and political sciences). Three of the researchers were employed as researchers full-time, and three others were performing research as part of their activities as academics (i.e. they also had teaching duties as part of their primary activities).

They all considered themselves users of open data for part of their research (two of them also mentioned that they were using open data for personal reasons, but this was considered outside of the scope of our analysis). This is not a surprise since, in our open call for volunteers, this was one of the aspects that we had pointed out. In most cases, they pointed to data portals that are specific to their field, such as HuggingFace ([2]), Kaggle ([3]), GBIF ([4]) or Europeana ([5]), but also to general-purpose portals, such as Zenodo ([6]) or Wikidata ([7]), or to general services for obtaining earth observation data that they would process afterwards. Only in three cases was the use of open government data portals mentioned, one for each of the main blocks of disciplines that we had selected (for demography and statistics, for environmental data and for traffic).

Data.europa.eu was well known by all of the interviewees. However, only one of them had commonly used datasets retrieved from data.europa.eu for her research, with datasets from Eurostat. **The other five interviewees claimed that they had never used this portal to find datasets for their research, or in those cases when they had used it, they had been unsuccessful in their search for datasets since they did not exist or were not at the appropriate level of detail required for their research.** They instead used other open government data sources such as local open government data portals (and application programming interfaces (APIs)) and data portals from meteorological agencies to find environmental and traffic data. The main conclusion from this analysis is that researchers commonly look for open data sources on the research data portals from their communities, in ad hoc data services that serve their data-related needs or on generic research data portals. They only seek out open

([2]) https://huggingface.co/datasets.
([3]) https://www.kaggle.com/.
([4]) https://www.gbif.org/.
([5]) https://www.europeana.eu/.
([6]) https://zenodo.org/.
([7]) https://www.wikidata.org/.

government data when the characteristics of the datasets that they are looking for make them think that they would be covered by public governmental institutions (including statistical offices or environmental agencies).

In terms of the citation of datasets, the six respondents were aware of good practices in data citation, including data.europa.eu recommendations (Publications Office of the European Union and Jessop, 2022).

Finally, in terms of finding open government data (and more specifically the datasets described and published on data.europa.eu) more useful for their use and reuse, **two of the interviewees proposed the creation of a specific searchable tag or subsite with data stories where examples of the usage of open government data for research are shown, since this may provide examples for them and their colleagues of the usefulness of open government data for research purposes**.

4.2. Summary of the interviews with academics

As discussed in Section 3, we carried out five interviews with academics working in education at different levels (one in a secondary school and four at universities). The one working in a secondary school taught subjects relating to computer science, and more specifically a subject relating to programming languages for big data (e.g. Python), and those at universities worked in various disciplines (computer science, health informatics, environmental sciences and information sciences – journalism).

They all considered themselves users of open data for part of their academic activities (as in the case of researchers, three of them also mentioned that they used open data for personal reasons, but this was considered outside of the scope of our analysis). Again, as in the previous case, this is not a surprise, since in our open call for volunteers this was one of the aspects that we had pointed out. Those working in computer science pointed to two of the data portals that were also identified by the researchers (HuggingFace and Kaggle). This is completely understandable, given that many computer-science educational materials in this area make use of resources on these two sites. In four of the cases, they also pointed to open government data portals at different administrative levels (local, regional, national and data.europa.eu), from which they were taking datasets to be used in their lectures or in hands-on activities. Some examples are:

- local datasets on the business census and on traffic accidents, used for lectures and hands-on activities on improving data quality with tools like OpenRefine ([8]), some of which were even used in exams;
- local datasets on tree inventories, used for demonstration purposes in lectures related to data visualisation (e.g. showing trees on a map);
- national datasets on weather, used for lectures on data access for students with no computer science background;
- national and European datasets on budgets and public contracting, used for lectures (at universities and as parts of workshops and tutorials in special interest groups) in sessions on data journalism;

([8]) https://openrefine.org/.

— national and European datasets relating to COVID-19, used for lectures on data analysis.

As in the case of the researchers, data.europa.eu was well known by all of the interviewees. However, only two of them commonly used datasets retrieved from data.europa.eu for their academic/lecturing activities. The other three interviewees claimed that they had never used this portal to find datasets since they preferred using datasets from field-specific portals that included pre-prepared educational materials, or they found it sufficient to use datasets at the local level, from cities close to the places where they lectured. Similar to the case of the researchers, the main conclusion from this analysis is that academics are more prone to use open government datasets for their lectures and other academic activities, and that they usually prefer those portals where some educational materials are already attached to the data sources, so that they can reuse them more easily. **It is interesting to note that one of the academics sought out datasets with errors, which, as he said, are common at the local level, to show how to improve their quality with existing tools and guides.**

In terms of the citation of datasets, none of the academics used any particular way to cite the datasets, and they generally simply included links to the dataset's URL as part of the learning materials offered to students. This was surprising, given that some of these academics (those at universities) also perform research activities, although in all cases they are mostly focused on the academic aspects of their curriculum.

Finally, in terms of finding open government data (and more specifically the datasets described and published on data.europa.eu) more useful for their use and reuse, **all the interviewees proposed the possibility of being allowed to share (on data.europa.eu or on the local/regional/national data portals where they obtain datasets) their experiences using datasets and the associated learning materials**. They also would appreciate, similarly to the researchers' case, the creation of a special site with learning materials, in their own languages, where open government datasets are being used.

4.3. Summary of the interviews with government officials

As discussed in Section 3, we carried out three interviews with government officials at different levels (local, regional and national). All of them were part of the team responsible for an open data portal at their administrative level and had a good knowledge of all aspects related to open data, including the EU directives and their corresponding national transpositions and regional/local laws relating to open data, the ongoing work on high-value datasets, etc. They all had more than 5 years of experience in open data initiatives. All of them also had the capacity within their organisations to make decisions on the implementation of open data initiatives and had already had some interactions with academics and researchers to understand their potential needs and propose a set of measures towards increasing the (re)use of open government data as part of their activities.

In all cases, the interviewees also had a good knowledge of the activities being performed at the European level by data.europa.eu, including not only the usual processes for metadata federation, but also the activities of the data.europa.eu academy, which they followed regularly.

When consulted about their relationships with academics and researchers, and how they knew about their activities involving the data published on the open data portals for which they were responsible, there were different types of answers, which showed some heterogeneity in how these groups of reusers can be considered.

- In two of the cases, academics were one of the priority reuser groups in the context of their open data initiatives, and therefore the government officials regularly performed activities in order to reach academics, including talks as part of university courses or hackathons organised by universities, along with the international open data day in March every year.
- Most of the activities geared towards academics, as pointed out above, were focused on universities, although the government officials expressed interest in reaching the secondary school level as well, with less success. In fact, one of the interviewees had had some experience bringing open government data into secondary schools as a side activity, but not as part of their main activities in the open data initiative. They claimed it was difficult to reach the secondary school level, given the fact that curricula at this educational stage are stricter than those at the university level; therefore, there is a need to contact the institutions and departments that manage the curricula, instead of individual academics.
- **The activities with researchers normally stem from the researchers themselves, from many different disciplines, who approach the open data portal team** through the contact forms available for reusers, directly by email or at in-person events, especially before the COVID-19 pandemic, when government officials were more active at in-person events.

Because of the relationship with researchers attested to by the three interviewees, all of them reported having knowledge about data citation practices, including the recent data.europa.eu report on that topic (Publications Office of the European Union and Jessop, 2022), which has been mentioned previously.

In terms of continuing outreach to researchers and academics as part of their open data initiatives, the three interviewees reported that they had identified measures in this respect, including, in some cases, the issuing of public contracts to help them better structure their offer so that it becomes more attractive to academics and has a stronger commitment that extends over a number of years.

Finally, two of the interviewees shared the feeling that more must be done to make open government data more appealing for use in research and academia. In the research context, they agreed that the level of granularity of open government data must be improved, but they all claimed that they needed help from the research community on how to anonymise the data and how to make it available in forms that are more suitable for researchers (APIs, bulk downloads, archived datasets in Zenodo, etc.). **In the academic context, the interviewees had a clear view of the fact that educational content should be made available for academics to replicate it, especially at lower levels of education.** They would be happy to include such materials on their open data portals, but they do not feel as comfortable when creating those materials and think that if there are no clear measures or funding for this purpose, the academics will not create them easily either.

As a final note, the three interviewees did not have a strong position as to whether the data that is published on their open data portals should be replicated on research data portals or not. On the one hand, they see the value of reaching more researchers, but on the other hand they are conscious that this may mean a decrease in the visitors to their open data portals.

4.4. Personas

Taking the people who have been interviewed and the answers that they have provided during the interview as inspiration, together with the commonalities and differences across the various

interviewees, we worked on the creation of three personas, one for each category (researcher, academic, government official), to potentially help us in the development of services or systems around the use of open government data in research and academia and to better understand some of the needs that have been identified.

It is important to note that, as in any persona-development exercise, the selection of characteristics for each persona profile does not aim at being exhaustive, and that it is always fictitious, although inspired by reality. That is, none of the personas that are described in this section is real, nor is their combination of characteristics based on a single real person.

4.4.1. Persona 1.

Ingrid
The climate change researcher

Name: **Ingrid Johansen**

About: **As a researcher in climate change, I am passionate about my research area and at the same t me concerned about the effects that climate change will have on our lives in the coming decades.**

Age: **46**
Gender: **Female**
Organisation: **Norway's Climate Change Research Institute**

Bio: Ingrid has worked as a researcher for 20 years in three different European countries since finishing her PhD in environmental sciences at Graz University. Her background is in civil engineering, although she specialised in environmental sciences during her master's and PhD. Now she is back home in Norway, her country of origin, as a senior researcher, and her research is focused on the effects of climate change on biodiversity. She leads a group of five researchers and teaches at university occasionally.

Motivation: After having witnessed a clear decline in biodiversity in many regions in Europe, Ingrid is exploring with her team the consequences of climate change on such biodiversity decline. While scientific publishing is relevant for her, as for any other researcher, she is also interested in changing practices in science, especially in her community, towards a more open-science and reproducible-science approach.

Goals: To advance research on climate change and provide scientific support to measures towards mitigating climate change.

To demonstrate to her peers that reproducibility in science is possible by following an open-science approach: publishing openly all the data that is produced as a result of her research and uploading the software produced (mostly Python scripts) in open-source code repositories, etc.

To convince public administrations and private companies to make their data more open and easier to use.

Frustrations: Many of the datasets that Ingrid needs to use (e.g. on biodiversity) are not at the appropriate level of granularity, and are spread over many different open (research) data portals. Some of them are available but difficult to understand and process, with no further documentation attached to them and no clear methodology on how they have been acquired. Nonetheless, she would love to cooperate with open data producers in improving the quality of datasets, using techniques that she knows well.

4.4.2. Persona 2.

Sergio
The lecturer on data journalism

Name: **Sergio Mancini**

About: **As a data journalist (and lecturer on that topic at university), I love organising ad hoc groups in the lab of my city to create nice data stories that can be published on the websites of local and national newspapers.**

Age: 3**2**
Gender: **Male**
Organisation: **University of Siena**

Bio: Sergio obtained his degree in information sciences, and soon decided to dedicate his life to journalism, with a special focus on data journalism, which he learned by working in one of the city labs in Milan. He has collaborated with most of the national newspapers in Italy and has contributed more than 50 data stories (on corruption, on social good, on education, etc.), some of which have had a large impact. Recently, he was offered a temporary position at the university, as a lecturer, which he accepted and through which he has started a course on data journalism.

Motivation: After having finished his degree, Sergio realised the potential of data when communicating news to citizens, so he took a couple of courses on data wrangling and did a bit of programming. He wants to help change society for the better, and therefore he is very motivated to train more journalists to perform this type of job.

Goals: To show how the availability of data and the possibility to tell stories based on data can provide more impactful results in society.

To improve the capabilities of journalists when dealing with heterogeneous data sources from many different data providers.

To convince public administrations to get engaged in the publication of data stories based on the data sources that they manage and that they can publish on their open data portals.

Frustrations: Sergio still finds some difficulties and barriers when talking to public administrations and requesting that new datasets be opened, since some of them are reluctant to do so because they fear that they may be criticised due to the datasets' low quality or because a case of corrupted data is identified. One of the excuses that he receives for not opening data is related to the anonymisation of the data. Moreover, he would love for all of his materials (and the ones from other colleagues doing a similar job) to be more easily accessible to other people in his country and worldwide, so that more people can be trained on these topics. And obviously, he would like to receive proper recognition for this contribution to academia and society.

4.4.3. Persona 3.

Philippe
The open data officer

Name: **Philippe Dubois**

About: **As a civil servant working in the open data and transparency department of a city, Philippe is actively talking to all city departments to convince them of the usefulness of publishing open data and trying to create an ecosystem around the already mature open data portal to increase knowledge about it and its use.**

Age: **52**
Gender: **Male**
Organisation: **Open data and transparency department, city of Lyon**

Bio: Philippe studied law and soon decided to become a civil servant. He worked for regional governments in France, in the areas of health and social services, and then decided to move into the local administration, where he started a bit more than 10 years ago working as part of the open data team. He also regularly advises the national government and data.europa.eu on steps to be taken to move forward in open data adoption across public administrations.

Motivation: Philippe is an enthusiast of open data and has been influential in changing many policies in his city to have an even stronger approach to open data. He also has a strong interest in education, and whenever possible he goes to universities and secondary schools to talk about open data in general, and the open data initiative of his city in particular. He has been influential in setting up a working group in the city with secondary schoolteachers, so as to create materials that can be used at that level of education and would like to do the same at the university level and for researchers.

Goals: Facilitate the generation of learning materials that can be used by schoolteachers and university lecturers, based on the open data that is being published by the city.

Facilitate the work of researchers that need to use open data, and understand better their needs in terms of granularity, frequency updates, and quality in general, so as to provide the means via public contracts to provide adequate data sources.

Make sure that all the departments in the city are conscious of the opportunities provided by open data and make them publish even more datasets

Frustrations: In his endeavour to facilitate training on open data in schools, Philippe always finds it difficult to reach a larger group of teachers, beyond the enthusiastic ones who are already collaborating in the working group that he contributed to creating. They always claim that they are too busy and that the curricula that are proposed by the national and regional administrations do not allow them to be very flexible on what to teach. He would also like to have more training material delivered by other open data initiatives, including other local open data portals, but also national and European ones, so that the reuse of those materials can facilitate this work. He is also frustrated to see that many researchers are using open data, which is nice, but without letting the department know, which would be very helpful in order to show the benefits of open data not only for economic reasons but also for society at large and for research.

5. Recommendations

This section provides a list of initial recommendations that have been extracted from the interviews and the development of persona profiles that were done during this work. This list aims at providing an initial set of specific measures that can be undertaken to improve the availability of open government data as part of research and academic activities.

R1. Create a specific searchable tag or subsite with data stories where examples of the usage of open government data for research or for academic activities are shown.

R2. Allow academics to share their learning materials (and teaching experiences) on the use of open datasets on the open data portal sites. An academic course on this topic may be also relevant.

R3. Provide funding opportunities for academics to create academic content associated with open datasets. This may include challenges or annual prizes.

R4. Provide some form of federation of some of the content from data.europa.eu into EOSC-related services for data archival (e.g. Zenodo). For instance, this could be only the content that is identified by a reuser as potentially useful for research purposes. This way, the dataset can get a digital object identifier, and the specific version used for a specific research project can be archived.

References

European Commission: Directorate-General for Communications Networks, Content and Technology (2021), *Towards a European strategy on business-to-government data sharing for the public interest – Final report prepared by the High-Level Expert Group on Business-to-Government Data Sharing*, Publications Office of the European Union, Luxembourg, https://doi.org/10.2759/406717.

Publications Office of the European Union and Jessop, P. (2022), *Data Citation: A guide to best practice*, Publications Office of the European Union, Luxembourg, https://data.europa.eu/doi/10.2830/59387.

Annex. Interview scripts

The following script has been used for the interviews with the selected participants:

'As you know, we are conducting this interview in the context of a research activity done for the organisation in charge of the data.europa.eu portal and services (the Publications Office of the European Union), where we are making research on the intersection of open government data, open science and education, trying to understand the opportunities and challenges for academics and lecturers to make use of open datasets available in data.europa.eu and for government officials in charge of open data initiatives to make these datasets more reused.

The interview will be recorded and transcribed, for the purpose of conducting this research. Once the research is completed, the recordings and transcriptions will be destroyed. And the data that you will provide in this interview will not be presented as such in the generated report, but aggregated so as to preserve anonymity.

Next, we start with the interview, which will be divided into several blocks.

General information. Can you give me your name, age, nationality, organisation(s) where you work, position in those organisations and primary activity? Please note that none of this data will be disclosed neither internally nor externally, as discussed above.

Relationship with open data. Are you an open data provider, an open data user or both? How often do you access open data portals for any purpose (as a user, provider or both)? Which are the data portals that you normally have access to? What for? For how long have you been involved in activities that have to do with open data?

Relationship with data.europa.eu. How often do you use data.europa.eu to look for datasets that you may reuse or to look for datasets that may have been published by your own organisation or by other organisations? If you are only using it scarcely or not using it at all, is there any reason for this?

(For researchers) Use of open data in research. Do you use open government data as part of your research activities? If yes, which ones? Can you provide me, in an approximate order, the places where you normally look for the data to use in your research (e.g. domain-specific portals, generic research data portals, open government data portals, no data portals at all)? If you are using (or not using) open government data portals, is there a specific reason for it? How do you cite the datasets that you use?

(For academics) Use of open data in education. Do you use open government data as part of your educational activities? If yes, which ones? can you provide me, in an approximate order, the places where you normally look for the data to use in your educational activities (e.g., domain-specific portals, educational portals, open government data portals, no data portals at all)? If you are using (or not using) open government data portals, is there a specific reason for it? How do you cite the datasets that you use?

(For government officials) Understanding the use of open data in research or education. Do you know of any researcher or academic who is using open data from the portal that you manage for their research or educational activities? If yes, how did you come to know that? Is this a type of reuse that

you or your organisation would be interested in? Are you planning or have you planned any actions to increase the use of open datasets in these types of activities? If yes, which ones?

(For academics/researchers) Improving data (re)use. How would you find open government data more useful to be (re)used? Are there any specific datasets that you would like to see in open government data portals for your common activities? According to open science / education principles, do you think that the data that you produce in the context of your activities should be republished in open government data portals or elsewhere? Are your colleagues also familiar with these practices?

(For government officials) Improving data (re-)use. How would you like to see open government data used in research or academic activities? Do you think that you have to improve the way in which data is being published in open government data portals for this purpose (e.g., improving metadata, providing documentation, examples of usage, etc.)? do you think that your data should be replicated in the resources that academics and researchers are commonly using, similarly to how data is federated by national and European data portals?

Finally, do you have **any other statement on this topic that you would like to share**?'

Publications Office
of the European Union