FWO DMP Template - Flemish Standard Data Management Plan

Version KU Leuven

Project supervisors (from application round 2018 onwards) and fellows (from application round 2020 onwards) will, upon being awarded their project or fellowship, be invited to develop their answers to the data management related questions into a DMP. The FWO expects a **completed DMP no later than 6 months after the official start date** of the project or fellowship. The DMP should not be submitted to FWO but to the research co-ordination office of the host institute; FWO may request the DMP in a random check.

At the end of the project, the **final version of the DMP** has to be added to the final report of the project; this should be submitted to FWO by the supervisor-spokesperson through FWO's e-portal. This DMP may of course have been updated since its first version. The DMP is an element in the final evaluation of the project by the relevant expert panel. Both the DMP submitted within the first 6 months after the start date and the final DMP may use this template.

The DMP template used by the Research Foundation Flanders (FWO) corresponds with the Flemish Standard Data Management Plan. This Flemish Standard DMP was developed by the Flemish Research Data Network (FRDN) Task Force DMP which comprises representatives of all Flemish funders and research institutions. This is a standardized DMP template based on the previous FWO template that contains the core requirements for data management planning. To increase understanding and facilitate completion of the DMP, a standardized **glossary** of definitions and abbreviations is available via the following link.

1. General Project Information	
Name Grant Holder & ORCID	Jo Seldeslachts 0000-0002-4326-4810
Contributor name(s) (+ ORCID) & roles	Albert Banal-Estanol: co-author Sven Maertens: co-author Wolfgang Grimme: co-author Christina Stadler: co-author Jonas Nieto: database RA
Project number [1] & title	G057524N - Government Ownership and Competition: Empirical evidence from the European Airline Industry following the COVID-19-Pandemic
Funder(s) GrantID [2]	
Affiliation(s)	X KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: ROR identifier KU Leuven: 05f950310
Please provide a short project description	This project aims to empirically assess the impact of government ownership on competition. The COVID-19 pandemic and subsequent governmental equity interventions in the European airline industry provide for a particularly ideal setting to investigate this topic, and this for several reasons. First, airline markets and competition therein are well-defined and well-understood. Second, European countries offered rescue packages differing in size and scope; several countries provided sizable equity interventions to their national airlines - in total, around 13.7 bn €. We aim to focus on governmental ownership's impact on entry and exit of airlines within routes, and ultimately on the impact on prices and welfare.

[1] "Project number" refers to the institutional project number. This question is optional. Applicants can only provide one project number.

[2] Funder(s) GrantID refers to the number of the DMP at the funder(s), here one can specify multiple GrantIDs if multiple funding sources were used.

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				ONLY FOR DIGITAL DATA	*ONLY FOR DIGITAL DATA*	*ONLY FOR DIGITAL DATA*	*ONLY FOR PHYSICAL DATA*
Dataset Name	Description	New or Reused	Digital or Physical	Digital Data Type	Digital Data Format	Digital Data Volume (MB, GB, TB)	Physical Volume
2018_EU_step1.dta, 2019_EU_step1.dta, 2020_EU_step1.dta, 2021_EU_step1.dta, 2022_EU_step1.dta, 2023_EU_step1.dta,	**O&D data from Sabre:** This dataset is a data extract from commercial dataset of **Sabre Airvision Market Intelligence**. The data is provided by the German Aerospace Center (DLR). The first dataset was collected from the **O&D page** and provides estimates of all tickets and enplaned passengers by fare class, marketing and operating airline that were sold for flights originating and ending in the European Union in a particular month of a given year. In detail, the dataset contains for each origin-destination itinerary demand-specific and ticket-specific information at monthly level, such as origin airport, origin city, origin country, destination airport, destination city, destination country, marketing airline, operating airline, connecting airports, flight type (non-stop, one-stop etc.), distance, number of passengers, fare class, average base fare per class, i.e. fare without taxes and surcharges, average total fare per fare class. The period of observations is from January 2018 to September 2023. Our data extracts comprise flights that have both origin and destination at an airport within the European Union.	Reuse existing data from commercial database, access via DLR	Digital	Numerical	raw data format: csv	< 1 TB	

2018_EU_schedule_mktoptcarrier_directionalM.dta, 2019_EU_schedule_mktoptcarrier_directionalM.dta, 2020_EU_schedule_mktoptcarrier_directionalM.dta, 2021_EU_schedule_mktoptcarrier_directionalM.dta, 2022_EU_schedule_mktoptcarrier_directionalM.dta, 2023_EU_schedule_mktoptcarrier_directionalM.dta,	**Schedules data from Sabre:** This dataset is an data extract from commercial dataset of Sabre Airvision Market Intelligence. The data is provided by the German Aerospace Center (DLR). The second dataset was collected from the **Schedules page** and contains all non-stop flights that took place in the European Union in a particular month of a given year. In detail, we see origin airport, origin country, destination airport, destination, country marketing airline, operating airline, indicator for code sharing, frequency, i.e. number of departures, total capacity, i.e. total available seats. With respect to codeshare flights, a marketing airline may not operate non-stop routes, but can still sell tickets for these flights.	Reuse existing data from commercial database, access via DLR	Digital	Numerical	raw data format: csv	< 1 TB	
DLR Airport Database.xlsx	**DLR Airport Database:** This dataset lists all for all airports, heliports, bus stations, military airfield with a iata code, the following information: iata code, name, city name, country countrycodeiso31661, type[airport, bus station etc.], geographical coordinates (latitude, longitude). We re-use this dataset provided by our co-authors form the DLR to identify tickets that contain ground traffic. We drop these tickets.	Reuse exiting data collected by co-authors from DLR	Digital	Numerical	xlsx	< 1 GB	
pso_OD.dta	**PSO Routes in 2018, 2019:** In this datasets we collect and digitized all routes under "Public service obligation (PSO) schemes, i.e., the non-profitable thin routes for which countries granted subsidies to airlines operating them in 2018 and 2019. The European Commission published the lists of all routes under public service obligations (LIST OF ROUTES (178) WITH PUBLIC SERVICE OBLIGATIONS (as of 09/2017), LIST OF PUBLIC SERVICE OBLIGATIONS - 176 ROUTES (as of 09/2018), LIST OF PUBLIC SERVICE OBLIGATIONS (176 routes as of	Generate new data	Digital	Textual	raw data format_ pdf	< 1 GB	

	09/2019)). We drop non-profitable ("thin") routes for which EU countries granted subsidies to an airline group flying them in 2018 or 2019.						
10052022_airlinename_airlinegroup.xlsx	**Airline name - Airline group - Business Model Dataset:** This dataset shows for each airline present in the Sabre datasets its ultimate parent, changes in ownership in our observation period and the business model. We collect information of the ultimate owner from (1.) CAPA (2023). Airline profiles. Available online: https://centreforaviation.com/data/profiles/airlines, (2.) Orbis Global (Bureau van Dijk). CAPA (2023) provides airline profiles for each current and historic airline showing the type of business model and the ultimate owner. Further, we collect the business model of the marketing airlines in our dataset from the CAPA dataset and European Commission (2021). European Commission (2021) provides a list of all low cost carriers in Europe.	Generate new data	Digital	Numerical & Textual	xlsx	< 1 GB	
catchment_areas_EuropeanCommissionSeoAmsterdam.xlsx	**Airport - Catchment Area - Dataset:** This dataset provides information of all airports that are part of a catchment area. We define the "market" as all the plane tickets for the same city-to-city route. More precisely, we use directional origin catchment area-to-destination catchment area pair. By using catchment areas, we make the implicit assumption that there is perfect demand and supply substitution between two routes with the same origin (destination) catchment area but different origin (destination) airports. We follow a recent study by the European Commission (2021) that defines a catchment area in the EU at a radius of two hours driving time of metropolitan areas and identifies the airports within these catchment areas.	Generate new data	Digital	Numerical & Textual	xlsx	< 1 GB	
17022022_StateAid_Export.xlsx	**Dataset of State Aid Characteristics:** We collected information on airline group's state aid implemented pursuant to section 3.11 of the TF during March 2020 until September 2023 from the decision texts published at the state aid register of the European	Generate new data	Digital	Numerical & Textual	xlsx	< 1 GB	

	Commission. Note that in general European countries must publish all relevant information on each individual government interventions above EUR 100,000 granted under the COVID-19 Temporary Framework (TF) under the European transparency requirements for State aid. The Transparency Award Module (TAM) dataset provides information on the state aid individual grants by European countries. However, until June 2024 the dataset has been very incomplete. To verify completeness, we cross-checked that information with the annual reports of all airline groups.						
ScoreboardData\new Scoreboard2022_update\aid_scb_inst_linear.csv	**State Aid Scoreboard Data:** This dataset shows for each EU member state the state aid expenditure by instrument type (e.g. equity intervention, grant etc.) as percentage of GDP for 2015 until 2019. Source: State Aid Scoreboard Data of European Commission.	Reuse existing data provided by European Commission	Digital	Numerical	csv	< 1 GB	
Crosswalk_airportNuts2016\airport_NUTS-2016.csv	**Crosswalk airport - NUTS3-region:** Crosswalk between airports and NUTS3 regions available at Tercet of the European Commission.	Reuse existing data provided by European Commission	Digital	Numerical	csv	< 1 GB	
metropolitan_regions\Metro-regions-NUTS-2016.xlsx	**NUTS 3 region - metropolitan regions link:** List of metropolitan regions as combination of NUTS 3 regions of Eurostat	Reuse existing data provided by European Commission	Digital	Numerical	xlsx	< 1 GB	
ARDECO_gdp-at-current-prices__gdp-at-current-prices-	**Population and GDP at NUTS3-region-level of EU27 countries:** Population and GDP data at NUTS3 regions from ARDECO, the Annual Regional Database of the European Commission	Reuse existing data	Digital	Numerical	csv	< 1 GB	

eur.csv, ARDECO_total-population-on-1-january-demographic-statistics__total-population.csv		provided by European Commission					
regionalgrossdomesticproductgdpallitlregions.xlsx	**Population and GDP at NUTS3-region-level of UK:** Population and GDP data at NUTS3 regions for the United Kingdom from Office for National Statistics.	Reuse existing data provided by Office for National Statistics.	Digital	Numerical	xlsx	< 1 GB	
AIR_DISTANCE_MATRIX_TOTAL.txt	**Airport distance matrix:** This dataset provides the geographical distances between all EU airports. We utilize the airport distance matrix from Tercet, provided by the European Commission, to create the control variable hubdistance.	Reuse existing data provided by Office for National Statistics.	Digital	Numerical	txt	< 1 GB	

GUIDANCE:
The data description forms the basis of your entire DMP, so make sure it is detailed and complete. It includes digital and physical data and encompasses the whole spectrum ranging from raw data to processed and analysed data including analysis scripts and code. Physical data are all materials that need proper management because they are valuable, difficult to replace and/or ethical issues are associated. Materials that are not considered data in an RDM context include your own manuscripts, theses and presentations; documentation is an integral part of your datasets and should described under documentation/metadata.
RDM Guidance on data

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type.	

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.	☐ Yes, human subject data; provide SMEC or EC approval number: ☐ Yes, animal data; provide ECD reference number: ☐ Yes, dual use; provide approval number: ☒ No Additional information:
Will you process personal data[3]? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).	☐ Yes (provide PRET G-number or EC S-number below) ☒ No Additional information:
Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.	☐ Yes ☒ No If yes, please comment:
Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements, research collaboration agreements)? If so, please explain to what data they relate and what restrictions are in place.	☒ Yes ☐ No The airline data comes from a commercial dataset, and cannot be made public.
Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain to what data they relate and which restrictions will be asserted.	☐ Yes ☒ No

[3] See Glossary Flemish Standard Data Management Plan

2. Documentation and Metadata	
Clearly describe what approach will be followed to capture the accompanying information necessary to keep **data understandable and usable**, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded). *RDM guidance on documentation and metadata*.	We will use the templates and procedures that are asked by the American Economics Association (AEA), as this the "gold standard" in economics: https://www.aeaweb.org/journals/data/data-code-policy
Will a metadata standard be used to make it easier to **find and reuse the data**? If so, please specify which metadata standard will be used. If not, please specify which metadata will be created to make the data easier to find and reuse. *REPOSITORIES COULD ASK TO DELIVER METADATA IN A CERTAIN FORMAT, WITH SPECIFIED ONTOLOGIES AND VOCABULARIES, I.E. STANDARD LISTS WITH UNIQUE IDENTIFIERS.*	☒ Yes ☐ No If yes, please specify (where appropriate per dataset or data type) which metadata standard will be used: We will most likely use the AEA depository and standards: https://www.openicpsr.org/openicpsr/aea

3. Data Storage & Back-up during the Research Project

Where will the data be stored? *Consult the interactive KU Leuven storage guide to find the most suitable storage solution for your data.*	☐ Shared network drive (J-drive) ☐ Personal network drive (I-drive) ☒ OneDrive (KU Leuven) ☐ Sharepoint online ☐ Sharepoint on-premis ☒ Large Volume Storage ☒ Digital Vault ☐ Other: We use Dropbox during the initial phases, copy the project in its entirety to OneDrive at fixed points of the project and finally migrate all materials to a digital vault at the time of publication in journals. The reason we also use DropBox is that this is the only way all members of the team have easy access to the data (whereas this is much more cumbersome with internal KU Leuven storage).
How will the data be backed up? *WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS?*	☐ Standard back-up provided by KU Leuven ICTS for my storage solution ☒ Personal back-ups I make (specify) ☐ Other (specify) The data will be stored in multiple places on the cloud: (i) Dropbox, (ii) OneDrive and (iii) a depository such as described by the AEA.
Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.	☒ Yes ☐ No The data in its entirety is about 1TB, which is within the limits of our storage spaces.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons? *CLEARLY DESCRIBE THE MEASURES (IN TERMS OF PHYSICAL SECURITY, NETWORK SECURITY, AND SECURITY OF COMPUTER SYSTEMS AND FILES) THAT WILL BE TAKEN TO ENSURE THAT STORED AND TRANSFERRED DATA ARE SAFE.* *Guidance on security for research data*	I do not consider this a threat, as this is (i) publicly available data or (ii) historical airline data that is not "sensitive" data. The main strategy to safeguard our data is to have it stored in multiple places, where reproduction of database and results is regularly checked.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	The expected costs are (i) Dropbox fees for the team and (ii) hardware (laptops). The costs will be covered through this research grant, and other funding sources.

5. Data Preservation after the end of the Research Project	
Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...). *Guidance on data preservation*	☒ All data will be preserved for 10 years according to KU Leuven RDM policy ☐ All data will be preserved for 25 years according to CTC recommendations for clinical trials with medicinal products for human use and for clinical experiments on humans ☐ Certain data cannot be kept for 10 years (explain)

Where will these data be archived (stored and curated for the long-term)? *Dedicated data repositories are often the best place to preserve your data. Data not suitable for preservation in a repository can be stored using a KU Leuven storage solution, consult the interactive KU Leuven storage guide.*	☒ KU Leuven RDR ☐ Large Volume Storage (longterm for large volumes) ☐ Shared network drive (J-drive) ☒ Other (specifiy): Journal repositories for published articles.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	None.

6. Data Sharing and Reuse	
Will the data (or part of the data) be made available for reuse after/during the project? Please explain per dataset or data type which data will be made available. *NOTE THAT 'AVAILABLE' DOES NOT NECESSARILY MEAN THAT THE DATA SET BECOMES OPENLY AVAILABLE, CONDITIONS FOR ACCESS AND USE MAY APPLY. AVAILABILITY IN THIS QUESTION THUS ENTAILS BOTH OPEN & RESTRICTED ACCESS. FOR MORE INFORMATION: HTTPS://WIKI.SURFNET.NL/DISPLAY/STANDARDS/INFO-EU-REPO/#INFOEUREPO-ACCESSRIGHTS*	☒ Yes, as open data ☐ Yes, as embargoed data (temporary restriction) ☒ Yes, as restricted data (upon approval, or institutional access only) ☐ No (closed access) ☐ Other, please specify: All public data will be made available after the project (i.e., after publication in journals). The commercial data (Sabre) will be explained in readme files; codes to handle this data will be made public, but the data itself not.
If access is restricted, please specify who will be able to access the data and under what conditions.	The conditions for using the Sabre database is to have a license to do so.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain per dataset or data type where appropriate.	☐ Yes, privacy aspects ☒ Yes, intellectual property rights ☐ Yes, ethical aspects ☐ Yes, aspects of dual use ☐ Yes, other ☐ No If yes, please specify: The Sabre database is a commercial database.
Where will the data be made available? If already known, please provide a repository per dataset or data type.	☒ KU Leuven RDR ☐ Other data repository (specify) ☐ Other (specify)
When will the data be made available?	☒ Upon publication of research results ☐ Specific date (specify) ☐ Other (specify)
Which data usage licenses are you going to provide? If none, please explain why. *A DATA USAGE LICENSE INDICATES WHETHER THE DATA CAN BE REUSED OR NOT AND UNDER WHAT CONDITIONS. IF NO LICENCE IS GRANTED, THE DATA ARE IN A GREY ZONE AND CANNOT BE LEGALLY REUSED. DO NOTE THAT YOU MAY ONLY RELEASE DATA UNDER A LICENCE CHOSEN BY YOURSELF IF IT DOES NOT ALREADY FALL UNDER ANOTHER LICENCE THAT MIGHT PROHIBIT THAT.* *Check the RDR guidance on licences for data and software sources code or consult the License selector tool to help you choose.*	☐ CC-BY 4.0 (data) ☐ Data Transfer Agreement (restricted data) ☐ MIT licence (code) ☐ GNU GPL-3.0 (code) ☒ Other (specify) The data usage licenses are determined by (i) the conditions of the commercial database we use (Sabre) and (ii) the conditions of the journals we publish in (see e.g. https://www.aeaweb.org/journals/data/data-code-policy)

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here. *INDICATE WHETHER YOU INTEND TO ADD A PERSISTENT AND UNIQUE IDENTIFIER IN ORDER TO IDENTIFY AND RETRIEVE THE DATA.*	☒ Yes, a PID will be added upon deposit in a data repository ☐ My dataset already has a PID ☐ No
What are the expected costs for data sharing? How will these costs be covered?	We do not expect that there will be costs for data sharing.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Jo Seldeslachts (PI) and Jonas Nieto (database RA).
Who will manage data storage and backup during the research project?	Jo Seldeslachts (PI) and Jonas Nieto (database RA).
Who will manage data preservation and sharing?	Jo Seldeslachts (PI) and Jonas Nieto (database RA).
Who will update and implement this DMP?	Jo Seldeslachts (PI) and Jonas Nieto (database RA).