Data Management Plan

StatusCities:
Migrant legal STATUS diversity and diversity dynamics in European CITIES

Dr Fran Meissner

Contents

Acknowledgements:
This project has received funding from the European Union's Marie Skłodowska-Curie Actions Grant Agreement No. 707124.

I would like to thank the Research Data Services Team at the Delft University of Technology (TU Delft) for providing exceptional support and feedback on earlier drafts of this deliverable.

0. Overview of Project

Name of group/project	**StatusCities** Migrant legal STATUS diversity and diversity dynamics in European CITIES
Name of researcher(s)/student(s)	Dr Fran Meissner

Description of Research

Main research question:

How are legal status differentiations relevant for urban migration-driven diversity?

Focused sub-questions:

1. How is regulating migration at the national level relevant for understanding urban superdiversity?
2. What dynamic spatial patterns of legal status diversity can be locally identified and mapped both at the city level and at the neighbourhood level?
3. Do legal status tracks impact on residential decisions, mobilities and personal networks in systematic ways?
4. Do dynamics at the city and neighbourhood level differ if compared across neighbourhoods and between cities?

Abstract:
StatusCities will provide a comprehensive investigation of the city level implications of migrants being differentiated by a myriad of legal status tracks. Conceptually and empirically focussing on superdiversity and diversity dynamics – rather than cross-sectional configurations - the specific aim is to link debates about the national level management of migration with debates on urban migration-related diversities. This is extremely timely given unprecedented migration to Europe and especially to Europe's cities. Asking where migrants subject to different legal status conditions live and what their residential biographies look like, StatusCities will **combine an analysis** of both unique **longitudinal register data** and **qualitative data** derived through mobile-phone supported long term elicitation techniques. The former will be used to map and visualise migrant residential patterns in light of status differentiations, the latter to analyse city and neighbourhood level patterns by paying attention to individual mobility and sociality decisions at the time of legal status transitions.
The project will be focused on four urban centres located in one of Europe's major conurbations the Dutch Randstad. The research will consider multiple scales of analysis through a focus on different types and differently sized cities and within them neighbourhoods that show a relatively high turnover of different status migrants. StatusCities will thus contribute knowledge and methodological innovation by drawing on debates from various disciplinary fields and using a sophisticated and innovative mixed methods approach. Research has not yet investigated how the multiplicity of status differences - with their associated eligibility criteria and the parameters of presence that they set out - are relevant to dynamic changes in migration-related diversity. StatusCities is devoted to filling this highly policy relevant knowledge void.

Funding body(ies)	European Commission Call: H2020-MSCA-IF-2015
Grant number	707124
Partner organisations	-
Project duration	Start: **2017-09-01** End: **2019-08-31**
Date DMP created	**2018-04-10**
Date last update	2018-04-30
Version	1.1
Name of researcher(s) with responsibilities for data management	Dr Fran Meissner (PI)

1. Data Collection

Secondary Data (Data Type A)

Source of Data: Dutch Statistics office **(CBS)**.
Accessing Data: Only within a secure research and analysis environment *set up and monitored* by CBS[1]
Further Details:
Focus will be on the following catalogue items:

VRLMIGMOTBUS2015 variables about immigration statuses of foreign born
GBAMIGRATIEGEBEURTENISBUS variables about im- and emigration
Woonruimteregister verrijkt 2012V1 variables about dwelling e.g. size; owner occupied
Adresgwbtab geospatial variables needed to map outputs

Exporting Output: An export folder is used to take information outside the secure research and analysis environment (including substantive questions, syntaxes as well as results).

For ANY "(draft) results [exported] from the secured environment, CBS will verify that the *results do not contain any identifiable data*"[2]

Particular character of Data:
Within the secure environment administered by the CBS, I will be working with micro-level data (individual level data). I will be abiding by the strict data-protection rules and regulations of the CBS.

CBS approved output, exported to TU-Delft servers (tables, figures etc.) will be for the purpose of inclusion in publications and for saving draft results rather than for conducting further analysis.

Primary Data (Data Type B)

Source: Individually collected data
Sub-Types and sources of Data:

B1 – Personal Identifying Data (longitudinal design: need to save contacts of respondents)
B2 – Interview data from repeat interviews with key informants
B3 – Diary Entries made by key informants
B4 – Visual Data (Photo and Film) recorded by key informants
B5 – Geospatial Data collected by key informants

Data Collection Process

B1 – Personal data will be elicited as part of the sampling procedures, personal data of potential respondents who opted out of being interviewed will be deleted immediately, for study participants it will be kept until it is no longer necessary to contact the respondent to achieve the project objectives.
B2 – Audio recordings of face to face interviews – subsequently transcribed by professional transcription service or a trained research assistant

[1]See: https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research/applying-for-access-to-microdata [accessed 31-03-2018] for more information about process of accessing data.
[2]See https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research/export-of-information [accessed: 31.03.2018] for more information about exporting research results.

B3 – B5 Respondents will independently collect this data with an app on their mobile phone – two apps have been identified that would offer both the needed functionality to record all the data types and that were programmed for academic purposes: epicollect5[3] or GeoODK[4]

In an upcoming meeting with ICT support it will be determined which is better suited to ensure privacy by design standards.

Particular character of Data
Data and partially outputs from Data Source B will be personal and/or sensitive (see Ethics application). All raw Data of Type B will need to be encrypted and password protected. TU Delft ICT will support the project by setting up a project drive on the TU-Delft Server. Access to the project drive will be restricted to the PI. Only the PI will be able to request access to the drive or sub-folders of the drive should there be a need for this (e.g. hiring the research assistant to do transcription of interviews).

Please note:
B1 (personal data) will at all times be kept separately from the B2 – B5 (research data) and it will be subject to additional security measures to ensure that personal data such as names and email addresses cannot be linked to research data.

B1 will be deleted from the data corpus upon completion of the research.
Selected, sharable (de-identified) Research Data will be deposited at DANS EASY repository (see section 5 “Data Sharing and Reuse”).

Handling Version Control:
Git Environment will be set up for research drive covering both files of Data Type A or B and outputs derived from them. To set up Git Environment an appointment with TU Delft ICT will be set up.

Reproducibility
The project will aim for the highest level of transparency and reproducibility possible. While the sharing of code used to analyse Data Type A is part of those efforts, it will not be possible to share micro-data held by the CBS. By requesting access to the CBS data through their own institution other researchers should be able to reproduce the results of the study. The code will further be available to researchers who may wish to adapt it for other data and research questions.

For Data Type B achieving reproducibility will be less straight forward due to the qualitative nature of the data and its analysis. On the spectrum of reproducibility it will for instance not be possible to share data due to privacy concerns (see archiving). What will be possible is to clearly explicate the research methods (e.g. share the XML File that will be used to create the GeoODK questionnaire, the question guide used during semi-structured interviews and to document the coding process of the data. This information will be shared in DANS EASY repository and on the project website so that it is accessible to a wide audience.

[3] https://epicollect5.gitbooks.io/epicollect5-user-guide/content/developers/intro.html [accessed: 30-03-2018)
[4] http://geoodk.com/

Further Details:

Type of data	Format	Software	Estimated data size	Specific character
Secondary Data (Type A)	Output Tables, Maps, Figures R Scripts *.rproj *.R *.csv *.svg *.pdf *.png *.tiff	R-Studio Inkscape Adobe Illustrator	10 GB	No personal data included in output files, Any output exported from the CBS environment is checked by CBS
Primary Data (Type B)				
B1	*.csv	Excel	30 kb	Personal Data as per the GDPR.[5] Encrypted and password protected not linkable to research data (B2-B5)
B2	*.mp3 *.rtf	Dedoose / Atlas Ti (QCA analysis) Transcription software	3 GB	May contain personal data (e.g. addresses, names of friends, etc.) relevant for analysis
B3 and B5	*.csv *.kmz	Epicollect5 OR GeoODK Dedoose /Atlas Ti; R-Studio (mapping with sp package)	10 GB	Will contain personal data (e.g. addresses, names of friends, etc.) relevant for analysis
B4	*.mp3 *.jpg *.png *.mp4	Epicollect5 OR GeoODK Dedoose/Atlas Ti	7 GB	May contain personal data (e.g. addresses, names of friends, etc.) relevant for analysis
Estimated Total Project-Server space required:			~ 30 GB	will depend on which QCA Software is used more with AtlasTi

[5] https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en [accessed: 04-04-2018].

2. Data Storage and Back-up

Storage of Data Type A:
During the research micro-data always remains within the CBS environment (where they are automatically backed up) only output files and analysis code will be stored on TU-Delft Servers.

Storage of Data Type B:

Data	Storage medium and location	Backup location and backup frequency
Raw data (only applicable for Data Type B)	Project network drive with restricted access on TU Delft server in separately encrypted folder (with Vera Crypt)	Backed up regularly and automatically via TU-Delft ICT server
Processed data (de-identified)	Project network drive with restricted access on TU Delft server	Backed up regularly and automatically via TU-Delft ICT server
Models/code	Project network drive with restricted access on TU Delft server	Backed up regularly and automatically via TU-Delft ICT server
Other? (e.g. informed consents)	Project network drive with restricted access on TU Delft server in separately encrypted folder (with Vera Crypt)	Backed up regularly and automatically via TU-Delft ICT server

3. Data Documentation

Documentation during research:
The data folders of the project will all contain a ReadMe File outlining the project set up and crucial documentation information (following DDI 3).

For Data Type B there will be a folder for each key-respondent collating all the data files obtained from the respondent. Each of these folders will also contain a text file outlining containing relevant meta data for the interview (Interviewer, Interview Time and Place, Respondent ID; respondent characteristics; field notes contextualising interview and other micro-level data B3-B5).

Documentation Standards:
DDI 3.1 for Project documentation.

For the documentation of qualitative interviews (B2) and respondent collected data (B3-B5) no standard will be used but key elements highlighted by DDI 3.1 will be included in the relevant field notes. In addition a data listing template[6] will be use to maintain an overview of different data files and meta-data describing them.

[6] https://www.ukdataservice.ac.uk/manage-data/document/data-level/qualitative [accessed: 06-04-2018]

Directory and file naming convention:
A consistent system for saving the different data objects will be implemented following this convention:

RespondentID_DateCreated_TypeOfData_Language_City_Version

Example:
fm01_180404_InterviewRecording_en_Delft_v1-01

The respondent ID is created by using Interviewer Initials & Two Digits consecutively numbering the respondents

A readme file in the main data folder will explain abbreviations used in file names

There will be four main data folders on the encrypted and restricted access drive:

(1) Outputs and R Files based on **Data type A**
(2) **Data Type B2-5 Raw data**-folder (password protected folder)
(3) **Data Type B2-5 Anonymised data** and analysis folder
(4) **Data Type B1 and Consent forms** (password protected folder – different password than Folder 2 – Passwords will be managed with KeePass)

Subfolders:

(1.1) Outputs
(1.2) R Files

(2.1) – 2.n Subfolders for each Respondent named by respondent ID (s. a.)
(3.1) Analysis Folder (Atlas TI Directory)
(3.2) - 3.n Subfolders 2.n Subfolders for each Respondent named by respondent ID (s. a.)

Project and/or Data Identifiers:
The file naming convention (described above), will facilitate that all data files will have unique and descriptive names. These file names will also serve as unique file identifiers.

4. Data Access

Copyright and Intellectual Property Rights
I will abide by the guidelines set out by the TU Delft on Intellectual and Property rights.[7]

The TU Delft and specifically Maarten van Ham are the beneficiaries of the Marie Curie Grant funding this project. The rights and responsibilities of data access are set out in the grant agreement.

Restrictions on the Data access during the project?
By virtue of being stored on TU Delft Servers in a dedicated project drive it will only be possible for the PI to access the data and to give temporary access to the should a research assistant be employed to translate and/or transcribe Data of Type B2-B5.

Control of data access
Data Type A: CBS
Data Type B: PI

[7] https://intranet.tudelft.nl/en/targeted-info/valorisation-centre/intellectual-property/?login=1 [accessed: 06-04-2018].

Data Accessibility in case of staff changes, illness etc.
It will be agreed with IT that if for unforeseen circumstance the PI is no longer able to manage the data, that the Project Drive will be accessible to the Grant Holder and Project Mentor Prof Maarten van Ham. Prof van Ham will then be able to decide how the Data and output saved in the Folder are to be treated, if outputs in the folders at such a time are used for publication or similar, the PI needs to be credited. This has been agreed with Prof van Ham on 11-04-2018.

Data Accessibility after end of contract
As it is likely that the PI will need to continue working with the data after her contract ends. Given that the nature of her employment as a post-doc on an external grant means that a continued employment at the TU Delft cannot be guaranteed, the PI shall be able to maintain an affiliation with her research group in order to finalise the analysis without compromising the data security. The details of this will be reviewed as soon as it is clear that the PI is moving on to another job, but no later than 2 months before the end of her contract.

5. Data Sharing and Reuse

Audience for reuse of data

Now: Primarily the data will be used by the *PI*.

Data Type A: Only R-Code is sharable and will be available as soon as it is committed to GitHub and can be used by *researchers interested in further developing the code or applying it to their own data sets*.

For Data Type B: *Individual Respondents* will be able to access their own data (but not that of other respondent's) to review and if desired partially revoke their data for the use in the research.
Potential *co-authors* may be granted access to a de-identified version of the data if this is necessary to collaborate for publication.

Once the project is concluded
Same audience as during the research project. In addition researchers interested in accessing the de-identified data of Type B2 or B3 (Interview Transcripts or Diary Entries) will be able to request such access with the PI (s. Section 6 Data Preservation and Archiving):

Reasons for not sharing all data:

It will not be possible to share raw data for both data sources.

Reason Data Type A: Data is managed and owned by the CBS. Other researchers have to request access with CBS.

Reason Data Type B: Confidentiality and Privacy concerns.

For geospatial (B5) and audio- visual data (B4) the procedures for de-identification are by no means straight forward and the implication of sharing identifiable data cannot be estimated and would violate confidentiality and rules of protecting personal data.

Sharable Data:
Only textual Data (B2 and B3) are amenable to de-identification and sharing after the completion of the research project. De-Identified data is sharable and will be deposited in DANS EASY, but may also only have limited usability.

Together with the de-identified data the following analysis documents will be shared: 1) Questionnaires and other research tools (*.xml for GeoODK); 3) A fieldwork report and 3) a description of research (information about population, sample procedure, response/non-response, data collection method).

Funders requirements for data sharing:
Within the Horizon 2020 framework open access to publications is to be ensured, whereas data should be made openly available, if possible. The sharing of data should not be in conflict with confidentiality obligations or with rules and obligations of protecting personal data. The approach to sharing data is summarised by the commission: "As open as possible, as closed as necessary "[8]

Licensing for reuse
Sharable data (see above) will be deposited in DANS EASY under a CC BY licence.

When will you publish your data and where? Will they be linked to one or more scientific publications?

Output will be published in peer-reviewed articles (min. 3 Articles), in reputable working paper series (further 2 Articles) and in an interactive project report accessible to the public via a project website. As per funder requirements if possible all publication will be open access. Publications will have a clear statement on data availability and once available a link to the sharable data (de-identified data, as described above) which can be deposited in DANS EASY.

6. Data Preservation and Archiving

Criteria for what data to archive:

(1) Compliant with confidentiality commitments (see Research Ethics application and Informed Consent Template).
(2) Data must be de-identified

File formats for long term preservation:

Depending on which repository the data will be kept in, the data will be saved in the data formats deemed most suitable for long term preservation. The formats chosen will be accessible and those for which software is widely available on a variety of platforms.

Appropriate data repository
It is planned to deposit the Data with DANS-Easy[9]. To access the data it will be necessary to first inform the PI via the DANS-Easy Platform how the data will be used. The PI will then estimate if this use is in-line with the intended use agreed with respondents before granting access to the data.

Estimated total costs for archiving the data:
DANS has a flexible costing model, the actual cots can only be determined on request, once the exact data files to be deposited are known. The PI will receive support in liaising with DANS for cost negotiation from the Research Data Services team at TU Delft Library.

[8] http://ec.europa.eu/assets/eac/msca/funded-projects/how-to-manage/funded-projects/how-to-manage/itn/05-open-access_en.pdf [accessed 06.04.2018].
[9] https://dans.knaw.nl/en/deposit/information-about-depositing-data [accessed: 06.04.2018].