

CONSUMER DATA RESEARCH CENTRE

Safeguarded Data Project Proposal Form

INTRODUCTION

This form should be completed by those wishing to access the Consumer Data Research Centre's (CDRC) safeguarded data collections. You should consult the **CDRC Data Service User Guide** before completing the form. A CDRC data scientist will be assigned to you who can provide support in the application process.

Once submitted your proposal will be forwarded to the CDRC Research Approvals Group (RAG) for independent assessment. Projects will be assessed based on the criteria listed in Appendix 1.

Approval to access data will not be granted without evidence that the applicant has acquired ethical approval for the research through their institution, or supplied evidence that this is not applicable. For non-academic projects, where there is no approval process in place, the CDRC will assist the applicant in acquiring this.

PART A. PROJECT DETAILS

1. Contact Details

1.1. Lead applicant: [Thehuan Hoang](#)
Department and Institution: UCL
Institution Address: [Gower Street, London WC1E 6BT](#)
If a student, please provide your home address as well: [N7 8FB](#)
Email: the-huan.hoang.23@ucl.ac.uk
Telephone: [07587470653](tel:07587470653)

1.2. Initial Proposal Form Reference: [2097](#)

1.3. Title of Project

[\[Working title\] Predicting trip attraction for travel demand planning based on POI typology and network centrality](#)

1.4. Access Details

Which data sets? (more details in 3.1): [HUQ in-app location](#)

Date access requested from¹ (dd/mm/yyyy): [05/06/2024](#)

Date access requested to (dd/mm/yyyy): [30/09/2024](#)

2. Project Proposal Details

2.1. Abstract. Appropriate for a general audience. This may be used by the CDRC for reporting and publicity purposes.

[\[Work in progress\]](#)

¹ Please note the RAG review process takes approximately 4-5 weeks, so this is an estimated date.

This paper examines the relationship between the volume of non-work trips to urban subzones and various factors such as the typology and density of points of interest (POIs), centrality within the transport network, and socio-demographic characteristics of the area. By doing so, we can uncover patterns of human movement in cities beyond the traditional commute-focused models. We employ a deep learning method (ANN) to train the predictive model, using large-scale mobility dataset for training and validation, and SHAP to exact feature importance tpo the target variable. By identifying the key attractors for non-work trips, we aim to inform urban planning and transportation strategies that cater to the diverse and evolving needs of urban dwellers.

- 2.2. Project description.** A detailed description of the project, documenting the motivation, scope and aims of the intended research as well as the methods you will use in the proposed research.

[Work in progress]

This paper examines the relationship between the volume of non-work trips to urban subzones and various factors such as the typology and density of points of interest (POIs), centrality within the transport network, and socio-demographic characteristics of the area. By doing so, we can uncover patterns of human movement in cities beyond the traditional commute-focused models. We employ a deep learning method (ANN) to train the predictive model, using large-scale mobility dataset for training and validation, and SHAP to exact feature importance tpo the target variable. By identifying the key attractors for non-work trips, we aim to inform urban planning and transportation strategies that cater to the diverse and evolving needs of urban dwellers.

2.3. Research Category:

- 2.3.1. What sector do you work in?

Academic: Staff ☐, PhD Student ☐, Masters Student ☒, Undergraduate Student ☐,
Business ☐
Local Government ☐
Central Government ☐
Schools ☐
Third Sector ☐
Other ☐ (please specify):

- 2.3.2. Is your project funded, commissioned or sponsored by a funding body or any other organisation? Yes ☐ No ☒ Funding application in progress ☐

Please include the name, postal and web address of your current or prospective funder, and your grant/project reference number (if applicable).

The research question was derived from a related proposed project by Transport for London regarding identifying activity centres in London. TfL does not provide funding for the research but may be interested in the findings.

- 2.4. Why This Data?** Please explain why having access to this particular data set will help your project.

The dataset will serve as the ground-truth data to train a model that seeks to explain trip attraction into a specific urban area based on a variety of features to be explored in the research (POI typologies, centrality, employment density, etc.)

As the main purpose is to estimate total number of people attracted to a certain spatial unit from other units, the location data will be aggregated as such:

COUNT during most trafficked hour – COUNT during least trafficked hour of the day for each spatial unit across 7 days in one randomised week.

Moreover, there will be no need for unique identifiers or further socio-demographic variables to ensure maximum safety in dataset usage. Any count less than 10 to be suppressed

2.5. Project Impact. Please describe the anticipated scientific and societal benefits of the project and the ways in which you intend to maximise those benefits.

Uncover human mobility patterns in a city without explicit OD matrices, or using more resource-intensive methods such as travel diary surveys, or snapshots such as censuses, with possible broader generalisability for other cities by using only open geospatial data

2.6. End Users. Who are the main end users of this research (academic research, central government, consultancy, industry, local government, NHS, public sector, third sector? List all that are applicable.)

Academic research (full dissertation), and public sector/third sector (presentation of findings)

2.7. Outputs and Publications. What are the intended outputs or publications arising from the use of these data? (For example, journal articles, PhD thesis, report for government department, policy documents for a local authority, White Papers, new software or other tools, etc.)

Academic research (full dissertation), and public sector/third sector (presentation of findings)

2.8. Research Team

Please list the names, affiliation and email addresses of all known members of your research team who are requesting access to the CDRC safeguarded data. If you are a student please include your academic supervisor.

To access some of the CDRC safeguarded data products it is necessary for users to have successfully completed a safe researcher training course. Please see the terms of use information for the particular dataset.

Research Team *(including yourself as lead applicant, add more rows if required)*

Title, Name	Department / Organisation	Organisational email address	Have they completed a safe researcher training course? If yes, please specify course and date of completion.	Will they be accessing or viewing the safeguarded data? Y/N
Thehuan Hoang	UCL	the-huan.hoang.23@ucl.ac.uk	N/A	Yes
Elsa Arcaute	UCL	e.arcaute@ucl.ac.uk	N/A	Yes



2.9. Ethical Approval

2.9.1. Have you sought or are you seeking **ethical approval from an institutional ethical approvals panel** or any other appropriate body?

Yes ☒ No ☐

2.9.2. If *Yes*, please provide **evidence** of the status of the application or the outcome of the ruling issued. Please list what evidence you are enclosing below and return it as a separate attachment in PDF format when you return this application form. Feel free to add any comments below.

I will seek additional approval from UCL Ethical Research Committee to incorporate data from CRDC partners into my research

2.9.3. If *No*, please bring this to the attention of the CDRC as soon as possible so that routes for ethical approval may be discussed. We can provide advice if needed, particularly for non-academic applications.

[Click here to enter text.](#)

2.9.4. If you believe that ethical approval is not required for your research, please provide justification for this below.

[Click here to enter text.](#)

PART B. DATA REQUEST

3.1. Data Required. Please provide the following information for each dataset requested. Please add more lines if required.

Data Partner	Data Set Name & URL	Access to Full Data Set requested or specific variables	Geographic Extent (e.g. London)	Temporal Extent (e.g. 2015)
HUQ	HUQ aggregated in-app location dataset	Full Data Set	Greater London	Full (2016-2020)

4. Data Linkage

- 4.1. Data Linkage.** If your project will be linking more than one data source, describe which data sources will be linked and how the linkage will be done, including any specific variables that need to be linked (if known). If any of the data to be linked has identifying information as defined in the Data Protection Act 1998 or General Data Protection Regulations please provide details. Please note that no project that has the potential to re-identify individuals through data linkage will be approved.

n/a

PART C: DECLARATION

By completing this declaration I hereby declare that the information included in this application form is true and correct to the best of my knowledge. I understand that any false or misleading information given by me in connection with my application may result in termination of the application process and/or other sanctions.

I also agree that I will be the single point of contact for progress updates and communication regarding the progress of the application.

I agree for my personal information to be used for the purposes of processing this application in accordance with the relevant data laws of the UK.

☒ I consent to my contact details being added to the CDRC contacts database so that the CDRC can send me notifications of CDRC related activities.

☒ I understand that forwarding this form by email constitutes an electronic signature.

☒ I understand that final approval for this project may require the additional submission of project approval forms.

Name: [Thehuan Hoang](#)

Date: [06/06/2024](#)

APPENDIX 1: RAG CRITERIA FOR ASSESSMENT

The role of the RAG is to provide independent and transparent assessments of applications by researchers for access to data through both the CDRC Safeguarded and CDRC Secure services based on a set of standard evaluation criteria. RAG is independent to the CDRC and will include representation from the academic, big data, industrial sectors as well as the data partners concerned. For full Terms of Reference and membership see <https://data.cdrc.ac.uk/using-our-data-services>.

Criteria for Approval

- **Scientific advancement** – how the project has the potential to advance scientific knowledge, understanding and/or methods using consumer data;
- **Public good** – how the project has the potential to provide insight and/or solutions that could benefit society;
- **Privacy and ethics** – the potential privacy impacts or risks, and wider ethical considerations relating to the project
- **Project Design and Methods** – how the project will be conducted and who will be involved with a focus on demonstrating project feasibility.
- **Cost and resources issues** – what impact the project is likely to have on CDRC resources, including CDRC staff time and use of infrastructure, as well as any data acquisition costs. Resource requirements should be justified.