



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

The Lothian Diary Project

Citation for published version:

Hall-Lew, L, Cowie, C, McNulty, SJ, Markl, N, Liu, S, Lai, C, Llewellyn, C, Alex, B, Fang, N, Elliott Slosarova, Z & Klingler, A 2021, 'The Lothian Diary Project: Investigating the impact of the COVID-19 pandemic on Edinburgh and Lothian residents', *Journal of Open Humanities Data*, vol. 7, 4, pp. 1-5. <https://doi.org/10.5334/johd.25>

Digital Object Identifier (DOI):

[10.5334/johd.25](https://doi.org/10.5334/johd.25)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Journal of Open Humanities Data

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.





The Lothian Diary Project: Investigating the Impact of the COVID-19 Pandemic on Edinburgh and Lothian Residents

DATA PAPER

LAUREN HALL-LEW 

CLAIRE COWIE 

STEPHEN JOSEPH MCNULTY 

NINA MARKL 

SHAN-JAN SARAH LIU 

CATHERINE LAI 

CLARE LLEWELLYN 

BEATRICE ALEX 

NINI FANG 

ZUZANA ELLIOTT 

ANITA KLINGLER

]u[ubiquity press

**Author affiliations can be found in the back matter of this article*

ABSTRACT

The ongoing Lothian Diary Project consists of 125+ audio/video recordings collected since May 2020 from residents of Edinburgh and the Lothian counties in Scotland. The diaries comprise self-recorded monologues or semi-structured interviews in which participants discuss their experiences during different stages of the COVID-19 pandemic. Recordings were uploaded to an online survey that also collected consent, demographic information, and opinion regarding Covid-related policies. All data marked for reuse are and will be housed in the University of Edinburgh's DataShare and DataVault repositories. A partial deposit is available now and another will be made available upon completion of data collection. Data from consenting participants will form an oral history archive with Museums and Galleries, Edinburgh.

CORRESPONDING AUTHOR:

Lauren Hall-Lew

University of Edinburgh,
Edinburgh, Scotland, United
Kingdom

Lauren.Hall-Lew@ed.ac.uk

KEYWORDS:

COVID-19; Edinburgh; oral
history; Scotland; self-
recording; video diary

TO CITE THIS ARTICLE:

Hall-Lew, L., Cowie, C.,
McNulty, S. J., Markl, N., Liu,
S.-J. S., Lai, C., Llewellyn, C.,
Alex, B., Fang, N., Elliott, Z., &
Klingler, A. (2021). The Lothian
Diary Project: Investigating
the Impact of the COVID-19
Pandemic on Edinburgh and
Lothian Residents. *Journal
of Open Humanities Data*, 7:
4, pp. 1–5. DOI: [https://doi.
org/10.5334/johd.25](https://doi.org/10.5334/johd.25)

(1) OVERVIEW

REPOSITORY LOCATION

The University of Edinburgh, DataShare, DOI: [10.7488/ds/3009](https://doi.org/10.7488/ds/3009).

CONTEXT

The data for the Lothian Diary Project are part of *Edinburgh Speaks*, a project of the Language Variation and Change Research Group in Linguistics and English Language at the University of Edinburgh. The dataset comprises 125+ individual contributions with varying levels of consent for sharing. Each contribution consists of a video or audio recording of 1–22 minutes and most also include answers to a 20-minute survey on participant demographics and COVID-19 experience. The video/audio recordings are mostly vlogs of individual adults or children speaking, responding to project website prompts (e.g. “How has your life changed during lockdown?”), but some are interviews with a parent, volunteer, or social worker behind the camera asking similar questions. These data are useful for both qualitative and quantitative social science research interested in any area of COVID-19 experience at the individual level. The focus is on Edinburgh and the Lothian countries because of the project’s original motivation to conduct dialectological and sociolinguistic research on the local community. A subset of the data has been used for an MSc dissertation on how different notions of audience influence messaging and presentation style (Lee, 2020).

(2) METHOD

STEPS

(1) Participants self-record a digital audio/video diary. (2) Participants give consent for data usage via the opening page to the survey on Qualtrics (level 1, other researchers; level 2, oral history archive; level 3, radio/television/online) and are given the option to waive anonymity. (3) Participants upload the diary file to a secure, temporary repository on [Box.com](https://www.box.com) only accessible by the research team. The file is manually transferred by a Research Assistant to a protected server space for processing (‘DataStore’). (4) Participants answer questions about their COVID-19 experience through a Qualtrics survey, adapted from questions created by the University of Edinburgh’s CovidLife project.¹ (5) Participants chose an option for a £15 payment: bank transfer, local business voucher, or charity donation. All diaries are backed up in original formats and as WAV files (mono, 16 KHz and mono/stereo 44 kHz).

Data processing steps: Transcripts of English language recordings are generated using custom automatic speech recognition models implemented in Kaldi (Povey et al, 2011) and then hand-corrected in ELAN (ELAN, 2020). Transcripts of non-English speech will be hand-transcribed at a future date. For contributions marked for sharing with other researchers, mono 16 KHz WAV files, MP4 files, and accompanying transcripts and surveys are manually deposited into the Lothian Diary Project repository on DataVault; one deposit is available now and another will be made available upon completion of data collection. Data collection began on the 25th of May 2020 and will continue until the 31st of July 2021.

SAMPLING STRATEGY

Stage 1 of data collection followed convenience sampling. Any resident of Edinburgh and the Lothian counties was welcome to participate, regardless of vulnerability (e.g., children, vulnerable adults). Participants were recruited via social media platforms, adverts in local newspapers and radio, press releases, and word-of-mouth. Stage 2 of data collection was introduced to recruit underrepresented participants. The sampling strategy at this stage was to contact and partner with charities representing homeless, disabled, or other vulnerable individuals, as well as caregivers of any group. We established charity partnerships by running a targeted social media advertising campaign and an online interactive workshop for the Economic and Social Research Council’s ‘Festival of Social Science’. We also rented a local community space for three days to allow digitally excluded members of the public (e.g., those without reliable Wi-Fi access) to participate in person. At the time of writing,

¹ <https://www.ed.ac.uk/generation-scotland/covidlife-volunteers/what-is-covidlife>.

most of the participants are from Edinburgh City, and twice as many are female as are male. Most have an undergraduate degree or technical qualification, and most are between the ages of 16–64. Relative to the Edinburgh population, the sample has a greater representation of participants who are of colour (20%) and born outside the UK (27%). 15% are disabled and over 20% are LGBTQA+.

QUALITY CONTROL

The survey questions, many of which had already been piloted by the CovidLife research team, went through two rounds of piloting with this project, including input from archivists at Museums and Galleries, Edinburgh, prior to public release. Based on the piloting, we created a website with a FAQ page addressing practical advice and ethical advice (e.g., descriptions of data use, risk, an example consent form).² The survey was also revised several times to ensure clarity and ease of use with regard to the consent process, the *Box.com* file upload, and the survey questions.

(3) DATASET DESCRIPTION

OBJECT NAME

The Lothian Lockdown Diary Corpus

FORMAT NAMES AND VERSIONS

MP4, MP3, M4A, MOV, WAV, CSV, TXT

CREATION DATES

2020-05-25 to 2021-07-31 to match prior text (and the facts)

DATASET CREATORS

Lauren Hall-Lew: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – original draft

Claire Cowie: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing

Stephen McNulty: Data curation, Investigation, Methodology, Writing – review & editing

Nina Markl: Data curation, Investigation, Software, Visualization, Writing – review & editing

Shan-Jan Sarah Liu: Conceptualization, Funding acquisition, Supervision, Writing – review & editing

Catherine Lai: Data curation, Resources, Software, Supervision, Writing – review & editing

Clare Llewellyn: Investigation, Writing – review & editing

Beatrice Alex: Writing – review & editing

Nini Fang: Writing – review & editing

Zuzana Elliott: Methodology, Data curation

Anita Klingler: Data curation

LANGUAGE

English, Scots, Scottish Gaelic, Mandarin Chinese, Cantonese, and British Sign Language. Scottish Gaelic contributions have been translated into English. All non-English languages will be transcribed faithfully prior to translation into English. Many recordings include local place names and local terminology, often in Scots.

LICENSE

The sharable portions of the Lothian Diary Project are and will be deposited under the Creative Commons open license (CC0).

² <https://lothianlockdown.org/faqs/>.

REPOSITORY NAME

All data marked for reuse by other researchers (as per participant consent) are placed in the Lothian Diary Project collection either housed in the DataShare repository or the DataVault repository at the University of Edinburgh. The DataShare repository is Open Access and includes the audio or video recordings and their corresponding transcripts. The DataVault repository contains the same as well as participants' survey answers, including demographic information. These more sensitive data are accessible by contacting the data manager, Catherine Lai, at C.Lai@ed.ac.uk. For data collected at the time of writing, the DOI for DataShare is [10.7488/ds/3009](https://doi.org/10.7488/ds/3009) and the DOI for DataVault is [10.7488/7a22cc4b-87ec-4df3-a549-3b347fd4bca5](https://doi.org/10.7488/7a22cc4b-87ec-4df3-a549-3b347fd4bca5).

PUBLICATION DATE

Data collection for the Lothian Diary Project will continue until approximately 31 July 2021, with publication to follow soon thereafter.

(4) REUSE POTENTIAL

Data that have been marked for reuse, as per the participant's consent, will be housed as an oral history archive with Museums and Galleries, Edinburgh, allowing researchers, policymakers, civil organisations, and community members to access. Those marked for wider reuse are available for other purposes (e.g., aggregation, reference, validation, teaching) to any interested individual or organisation, with the permission of the Lothian Diary Project research team. Our data can be reused beyond the scope of our project for other researchers, as well as non-researchers to identify common themes, topics and issues discussed in the pandemic. For example, the diaries capture facial, gestural, and speech components of affectively charged topics, enabling individuals who are interested in multimodal expression to further explore this area. Policymakers and civil organisations may also access our data to conduct further research to advance future responses to health emergencies, as well as to provide support for vulnerable groups. We also welcome collaborative opportunities, especially for creative ways in which we could share and reuse our data beyond the current scope of our project.

ACKNOWLEDGEMENTS

We would like to thank Clare de Mowbray, School of Political and Social Science (SPS), University of Edinburgh, for web support, and Eleanor MacCallum (SPS) for some initial assistance with manually correcting automated transcripts. The Centre for Speech Technology Research (CSTR) at the University of Edinburgh have also helped with the automatic speech recognition software used to generate the video diary transcripts.

FUNDING INFORMATION

This research has been supported financially by the University of Edinburgh's College of Arts, Humanities, and Social Sciences (CAHSS) College Research Office; an ESRC Impact Acceleration Grant awarded to the University of Edinburgh (grant reference: ES/T50189X/1); the School of Philosophy, Psychology, and Language Sciences' Knowledge Exchange and Impact Office; and the ESRC Festival of Social Sciences. This work was supported in part by the UKRI Centre for Doctoral Training in Natural Language Processing, funded by the UKRI (grant reference: EP/S022481/1) and the University of Edinburgh, School of Informatics and School of Philosophy, Psychology & Language Sciences.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR INFORMATION

Lauren Hall-Lew: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – original draft

Claire Cowie: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing

Stephen McNulty: Data curation, Investigation, Methodology, Writing – review & editing
Nina Markl: Data curation, Investigation, Software, Visualization, Writing – review & editing
Shan-Jan Sarah Liu: Conceptualization, Funding acquisition, Supervision, Writing – review & editing
Catherine Lai: Data curation, Resources, Software, Supervision, Writing – review & editing
Clare Llewellyn: Investigation, Writing – review & editing
Beatrice Alex: Writing – review & editing
Nini Fang: Writing – review & editing
Zuzana Elliott: Methodology, Data curation
Anita Klingler: Data curation

AUTHOR AFFILIATIONS

Lauren Hall-Lew  orcid.org/0000-0003-3111-0352

University of Edinburgh, Edinburgh, Scotland, United Kingdom

Claire Cowie  orcid.org/0000-0003-0591-340X

University of Edinburgh, Edinburgh, Scotland, United Kingdom

Stephen Joseph McNulty  orcid.org/0000-0001-6607-5265

Linguistics and English Language, School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, Edinburgh, Scotland, United Kingdom

Nina Markl  orcid.org/0000-0001-9906-9961

Linguistics and English Language, School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, Edinburgh, Scotland, United Kingdom;
Institute for Language, Cognition and Computation, School of Informatics, University of Edinburgh, Edinburgh, Scotland, United Kingdom

Shan-Jan Sarah Liu  orcid.org/0000-0001-7997-1827

Politics and International Relations, School of Social and Political Science, University of Edinburgh, Edinburgh, Scotland, United Kingdom

Catherine Lai  orcid.org/0000-0003-2411-8954

Linguistics and English Language, School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, Edinburgh, Scotland, United Kingdom

Clare Llewellyn  orcid.org/0000-0002-3589-9517

Neuropolitics Research Labs, School of Social and Political Science, University of Edinburgh, Edinburgh, Scotland, United Kingdom

Beatrice Alex  orcid.org/0000-0002-7279-1476

Edinburgh Futures Institute, School of Literatures, Languages and Cultures University of Edinburgh, Edinburgh, Scotland, United Kingdom

Nini Fang  orcid.org/0000-0001-8730-8582

Counselling, Psychotherapy and Applied Social Sciences, School of Health in Social Science, University of Edinburgh, Edinburgh, Scotland, United Kingdom

Zuzana Elliott  orcid.org/0000-0001-6024-5489

Linguistics and English Language, School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, Edinburgh, Scotland, United Kingdom

Anita Klingler

History, School of History, Classics and Archaeology, University of Edinburgh, Edinburgh, Scotland, United Kingdom

REFERENCES

ELAN (Version 5.9) [Computer software]. (2020). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from <https://archive.mpi.nl/tla/elan>

Lee, Y. S. (2020). *Audience design in self-recorded monologues: How video and audio diary makers assign participant roles in the Lothian Diary Project*. [Unpublished MSc dissertation]. University of Edinburgh.

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., & Vesely, K. (2011). The Kaldi Speech Recognition Toolkit. In *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society.

TO CITE THIS ARTICLE:

Hall-Lew, L., Cowie, C., McNulty, S. J., Markl, N., Liu, S.-J. S., Lai, C., Llewellyn, C., Alex, B., Fang, N., Elliott, Z., & Klingler, A. (2021). The Lothian Diary Project: Investigating the Impact of the COVID-19 Pandemic on Edinburgh and Lothian Residents. *Journal of Open Humanities Data*, 7: 4, pp. 1–5. DOI: <https://doi.org/10.5334/johd.25>

Published: 16 April 2021

COPYRIGHT:

© 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Journal of Open Humanities Data is a peer-reviewed open access journal published by Ubiquity Press.