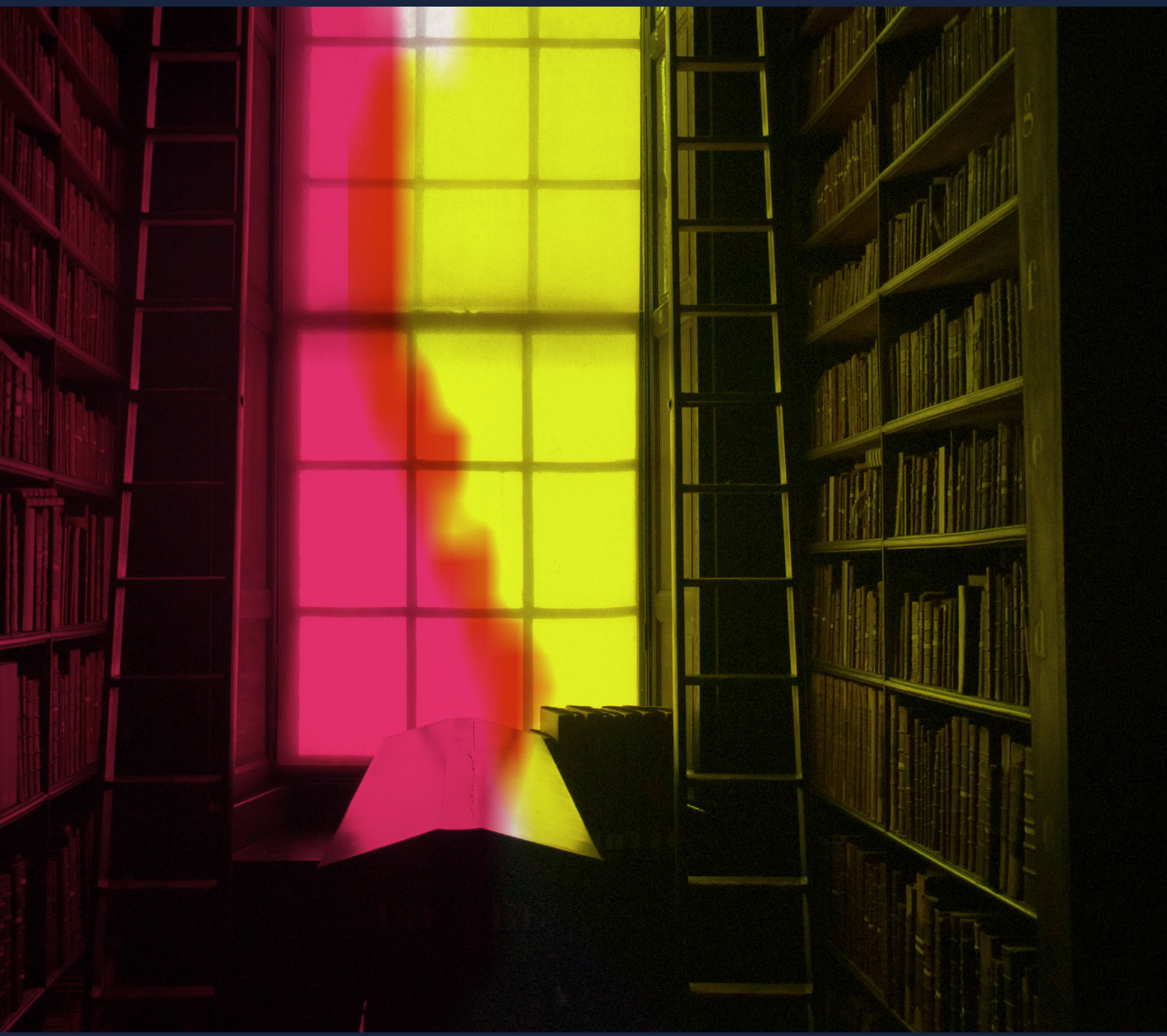
THE UNIVERSITY of EDINBURGH
Centre for Data, Culture & Society



ARCHIVES, COLLECTIONS & ACCESS

Guidance for Researchers

DCS DATA CULTURE SOCIETY

INTRODUCTION

For many researchers working with archives and collections the coronavirus pandemic has accelerated the use of online research practices. Some have been unable to physically access collections they planned to use. Many have had to adapt their methods, learning new tools and processes, and even rethink research questions and whole research projects.

Researchers have always had to work within time, resource and access limitations. However being prompted to rethink methods, practices and processes offers a chance for reflection. Pivoting to working remotely and digitally with archives and data collections will involve continuities and ruptures as changing methods of accessing primary materials come with both challenges and opportunities.

This document provides high-level guidance on remote working with archives and collections, and contains case studies focused on the recent experience of researchers working on archival projects. It is written for researchers, primarily those working at the University of Edinburgh, and identifies the methodological, practical, technological, legal, ethical and accessibility considerations that are likely to arise, and provides an overview of some of the tools, resources and support available.

PLANNING

ACCESSING MATERIALS

CREATING DATASETS

WORKING WITH DATA

ACCESSIBILITY ISSUES

CASE STUDIES

FURTHER READING AND RESOURCES

ACKNOWLEDGEMENTS

PLANNING

Planning archival research

Library, archives and collections staff will all stress the importance of getting in touch with them in the planning stages of a project, and they can offer a wealth of support. They can provide information about which materials are catalogued and which are not, conditions of access (such as whether photography is allowed), whether materials have already been digitised or if they can be, and suggestions of comparable materials in their collections or held in other archives. Crucially, they will be able to advise on the timescales associated with accessing and potentially cataloguing or digitising the required materials. The benefits of contacting library, archives and collections staff as early in the planning process cannot be overstated, their collaboration will contribute to a successful project.

Planning digital research

When working with data collections or creating one's own dataset, a good starting point is to formulate a Data Management Plan (DMP): indeed, this is something that many funding bodies now require. An effective DMP provides a framework for thinking through all of the processes involved in the research project. The University of Edinburgh offers **advice on writing a DMP** as well as **MANTRA, an online course on good practices in research data management**, and has developed a **course on research data management and sharing** in collaboration with University of North Carolina. US Historians Alan MacEarchern and William J. Turkel have written a useful article called **'A Time for Research Distancing'** which advises on practices to make a digital research project go more smoothly, including tips on how to work with and manage data effectively. Related resources are also gathered in the University of Edinburgh's **SERCH webpages** (Support for Research during Covid Hub).

PLANNING

ACCESSING MATERIALS

CREATING DATASETS

WORKING WITH DATA

ACCESSIBILITY ISSUES

CASE STUDIES

FURTHER READING AND RESOURCES

ACKNOWLEDGEMENTS

ADVICE ON WRITING A DMP

MANTRA: RESEARCH DATA MANAGEMENT TRAINING

RESEARCH DATA MANAGEMENT AND SHARING COURSE

A TIME FOR RESEARCH DISTANCING

SERCH WEBPAGES

PLANNING

Ethics of Data Use

When it comes to using digital datasets (whether already published or curating one's own), researchers are responsible for taking ethical and legal issues into account. These will vary according to each project, but are likely to include copyright, access and privacy questions. Researchers must also take into consideration their University's **rules around research ethics and integrity.**

While library, archives and collections staff can assist in making the researcher aware of the legal and ethical considerations of using specific materials, for example the Data Protection Act 2018, creating one's own digital datasets will require consideration tailored to that project. Subjects to consider include whether you have adequate permission to gather and use data, whether you need to anonymise it or otherwise adapt it, and how you will store it and manage access.

Data taken from online sources can require particular care, since those who published it are unlikely to have intended its use for research. The blog 'Empirical Data' has published a useful post with advice on **ethical web scraping**. The **Association of Internet Researchers** offers **comprehensive advice on research using data collected over the internet in a set of regularly reviewed guidelines.**

For a shorter document that attempts to synthesise some of the issues around ethics and using social media data, known colloquially as the 'Aberdeen Document', see Townsend & Wallace in the **Further Reading and Resources** section of this document.

The University of Edinburgh has recently published **ethical guidance for social media research.**

PLANNING

ACCESSING MATERIALS

CREATING DATASETS

WORKING WITH DATA

ACCESSIBILITY ISSUES

CASE STUDIES

FURTHER READING AND RESOURCES

ACKNOWLEDGEMENTS

RESEARCH ETHICS AND INTEGRITY

ETHICAL WEB SCRAPING

ASSOCIATION OF INTERNET RESEARCHERS

INTERNET RESEARCH: ETHICAL GUIDELINES 3.0

SOCIAL MEDIA RESEARCH: ETHICS GUIDANCE FOR RESEARCHERS (EDINBURGH)

ACCESSING MATERIALS

National and institutional archives and collections have been increasingly accessible to online researchers in recent years, and many organisations have fast-tracked their plans in response to the coronavirus pandemic. Most now offer some form of remote access, as well as some provision for digitisation, and some digital collections. Alternatively, online datasets can often offer immediate remote access to research materials.

Access to Physical Collections

The conditions under which access to physical collections is available is determined by national guidelines in response to the changing coronavirus pandemic situation, and it is recommended that you check each archive's website for up-to-date information on how to access their collections.

Most archives have some provision for digitising their physical materials on a case-by-case basis, and the process for this is usually detailed on the archive's website. Getting in touch with the individuals at the archive who are responsible for digitising / scanning is recommended, as they can offer practical and project-appropriate advice. They may also be able to provide information on comparable resources that are already accessible in a digital format.

Be advised that scanning / digitising materials may incur a significant cost and will take time. Some materials may not be able to be digitised because they are too fragile, or the process may need to involve initial conservation work. The overarching advice is to contact library, archive and collections staff as early as possible in the planning of a research project to negotiate the best strategy.

Depending on social distancing restrictions, the **University of Edinburgh Library Services**, may offer services such as Scan & Deliver, temporary access to additional e-resources and / or physical access to their print collections. Please check for **up-to-date advice on using the library**. The Centre for Research Collections may be able to offer face-to-face services and / or **remote access**, which can include:

- searching catalogues, indexes and collections handlists
- advice on potential items of interest in the collections or elsewhere
- photographing materials
- quotations for digital imaging services
- online appointments to view relevant collection items on a visualiser

Please see their **visitor information page for up-to-date advice.**

It is important to note that taking photographs of archival material or requesting digitised copies will result in a collection of material that should then be considered as a created dataset, and managed as such: see **Creating Datasets**.

PLANNING

ACCESSING MATERIALS

CREATING DATASETS

WORKING WITH DATA

ACCESSIBILITY ISSUES

CASE STUDIES

FURTHER READING AND RESOURCES

ACKNOWLEDGEMENTS

Physical Collections

UOE LIBRARY SERVICES

USING THE LIBRARY

REMOTE ACCESS TO CRC

CRC VISITOR INFORMATION

ACCESSING MATERIALS

Access to Digital Collections

Many archives also have collections of material that they have already digitised as well as born-digital data collections. Currently, archives are devoting more resources to digitisation and so the availability of digital collections will increase quickly.

Each library, archive, data repository or dataset may require different methods to access, interact with and collect data. 'Access' can mean:

- Viewing files online
- Downloading selected files and metadata to one's own computer
- Downloading large collections of data (for example in a zip file)
- Using an API (Application Programming Interface)

The volume of data can be very large, often too large for the storage and processing capacities of a personal laptop. In cases such as these, University of Edinburgh researchers can access extra processing power through **the University's high performance computer cluster, Eddie**, available through the **Edinburgh Compute and Data Facility,** which is part of the resources offered by Information Services.

Accessing Online Data Sources

There are many online data sources pertinent to researchers in the arts, humanities and social sciences, which include published datasets as well as repositories of digital and digitised material from which to create a new dataset.

Local resources include:

- **Edinburgh DataShare**, a digital repository of research data produced at the University of Edinburgh
- **The Angus McIntosh Centre for Historical Linguistics**, which offers historical and modern Scots and Scottish English language corpora
- **Collections as Data**, the University of Edinburgh's downloadable datasets
- The National Library of Scotland's **Data Foundry**

Further resources include:

- **Europeana** aggregates materials in European archives, libraries and museums
- The **Digital Public Library of America** which aggregates digital materials in repositories across the United Stated of America
- The **African Online Digital Library**, currently under development, will provide cultural heritage materials from and about African countries
- The **HathiTrust Digital Library** contains an enormous volume of digitised items from academic and research libraries

More online data sources are listed on the webpages of the Centre for Data, Culture & Society.

Although issues of rights and reuse for these materials need to be considered, online data sources potentially enable access to a huge variety of materials by bypassing geographical limitations. As a response to the coronavirus pandemic, some repositories have made their materials more freely available, though this may not be permanent.

PLANNING
ACCESSING MATERIALS
CREATING DATASETS
WORKING WITH DATA
ACCESSIBILITY ISSUES
CASE STUDIES
FURTHER READING AND RESOURCES
ACKNOWLEDGEMENTS

EDDIE, HPC
EDINBURGH COMPUTE AND DATA FACILITY

Online Data Sources

EDINBURGH DATASHARE
ANGUS MCINTOSH CENTRE
COLLECTIONS AS DATA
NLS DATA FOUNDRY
EUROPEANA
DIGITAL PUBLIC LIBRARY OF AMERICA
AFRICAN ONLINE DIGITAL LIBRARY
HATHI TRUST DIGITAL LIBRARY

CREATING DATASETS

If the required data is not already available, then creating a dataset from scratch is an option and the processes involved will depend on the type of data being collected, the methods of analysis being used, and the guiding research questions.

Creating a bespoke dataset may involve:

- Using an API to query an existing data repository
- Web scraping to collect data from the internet
- Collecting data from a social media site
- Combining and curating existing datasets
- Developing a dataset of new digitised materials

There are step-by-step **guides to creating a dataset from scratch**, and advice on **creating a machine learning dataset**. For more detailed guidance on working with social media data, the University of Edinburgh has recently published **ethical guidance for social media research.**

PLANNING

ACCESSING MATERIALS

CREATING DATASETS

WORKING WITH DATA

ACCESSIBILITY ISSUES

CASE STUDIES

FURTHER READING AND RESOURCES

ACKNOWLEDGEMENTS

CREATING A DATASET

CREATING A MACHINE LEARNING DATASET

SOCIAL MEDIA RESEARCH: ETHICS GUIDANCE FOR RESEARCHERS (EDINBURGH)

WORKING WITH DATA

Working with digital and digitised materials can require a whole set of tools and competencies. This can seem daunting but support and training are available.

Required skills and competencies can include:

- Accessing materials and data online
- Using search queries and image viewers
- Using APIs and web-scrapers to create a dataset
- Familiarity with multiple file formats and programs for working with them
- Software and tools for cleaning data
- Software skills for visualising data
- Analysing data
- Learning a programming language
- Using online collaboration tools for keeping track of multiple researchers' work and versions of the data and outputs

There are many training resources designed to help humanities researchers develop these skills and competencies:

- The Centre for Data, Culture & Society provides an extensive **training** programme, and **bursaries** for external training
- The **Programming Historian** offers tutorials on a range of digital methods in English, Spanish, French and Portuguese
- The **data, library and software training** available through The Carpentries is another widely-used resource
- **LinkedIn Learning,** to which the University of Edinburgh holds a subscription, as well as **Coursera,** both offer a wide variety of online courses on digital skills and methods aimed at all ability levels
- The University of Edinburgh's **Digital Skills** programme offers courses

For learning specific programming languages, code camps like Codeacademy, freeCodeCamp and Codewars can be a good way to quickly progress.

Other good starting points for advice on working with digital data include:

- The University of Edinburgh's **Research Data Service**
- **Digital Research Services**, which acts as a single point of access to all of the data and computing services available to researchers at the University of Edinburgh

PLANNING
ACCESSING MATERIALS
CREATING DATASETS
WORKING WITH DATA
ACCESSIBILITY ISSUES
CASE STUDIES
FURTHER READING AND RESOURCES
ACKNOWLEDGEMENTS

CDCS TRAINING
CDCS BURSARIES
PROGRAMMING HISTORIAN
DATA, LIBRARY AND SOFTWARE TRAINING
LINKED IN LEARNING
COURSERA
UOE DIGITAL SKILLS
RESEARCH DATA SERVICES
DIGITAL RESEARCH SERVICES

ACCESSIBILITY ISSUES

Finding and retrieving published research materials remotely and digitally may require specific accessibility considerations and researchers who are creating data collections should consider the wider accessibility of their collection as part of their broader data management plan.

There are a number of assistive technologies that can enable better access to online resources. **AbilityNet** provides extensive advice on customising a computer's accessibility options for various vision, hearing, motor and cognitive reasons. Examples of the **assistive technology available through Information Services** are: hardware and software for viewing texts, enlarging or digitising printed materials; speech-to-text software; text-to-speech software; and dyslexia-friendly fonts. Staff and students at the University of Edinburgh also have access to **SensusAccess** assistive technology, which enables conversion of digital material into a variety of accessible formats.

Public sector organisations, such as the University of Edinburgh, have a legal duty to make websites accessible and to publish an **accessibility statement** describing how to access websites and report accessibility problems, such as the accessibility statement for the University of Edinburgh's **DiscoverEd search portal**.

However, although many archives and libraries in the UK, have an accessibility statement, the format of each digital resource can vary. The Open University maintains a very useful list of **accessibility tips for using online databases**. For example, their information on the online repository JSTOR describes the order in which a screen reader encounters search results and how to access the full text of articles.

Researchers with a disability, learning difference or health condition are advised to communicate with archives and collections staff as early in the process as possible so that reasonable adjustments and a suitable access plan can be put in place. For PhD researchers at the University of Edinburgh, the **Student Disability Service** can assess requirements and provide support.

The University of Edinburgh **Centre for Research Collections** has a range of access options. Depending on pandemic restrictions, they may be able to offer face-to-face services and / or remote enquiries, which can include searching catalogues, indexes and collections handlists, advice on potential items of interest in the collections or elsewhere, photographing materials, quotations for digital imaging services, or online appointments to view relevant collection items on a visualiser. Their visitor information page contains up-to-date advice.

PLANNING

ACCESSING MATERIALS

CREATING DATASETS

WORKING WITH DATA

ACCESSIBILITY ISSUES

CASE STUDIES

FURTHER READING AND RESOURCES

ACKNOWLEDGEMENTS

ABILITYNET

INFORMATION SERVICES

SENSUS ACCESS

UOE ACCESSIBILITY STATEMENT

OU ACCESSIBILITY FOR DATABASES

STUDENT DISABILITY SERVICE

CRC ACCESS OPTIONS

CASE STUDIES

The following case studies are from University of Edinburgh researchers whose work has been affected by the coronavirus pandemic. They each describe their projects, some of the challenges and opportunities they have faced, and how they have successfully pivoted to remote and digital ways of working.

Sarah Van Eyndhoven
Alternative Digital Materials & Opportunities

Sarah Van Eyndhoven, a PhD candidate in Linguistics & English Language who researches 18th-century Scots and identity, discusses the challenge of her planned research materials in the National Library of Scotland, National Records of Scotland and University of Edinburgh Special Collections being made inaccessible.

She describes how locating alternative digital materials through the National Library of Scotland's Data Foundry and Scottish History Society Publications enabled her to continue her research. She also discusses how the inaccessibility of her planned research materials led to delays and changes in her research timetable, which she used as an opportunity to learn new technological skills she hadn't found time for before (for example, Transkribus).

WORKING WITH ARCHIVES DURING COVID-19

Dr James Cook
Alternative Digital Materials & New research outputs

Dr James Cook, Lecturer in Early Music, describes facing similar problems accessing historical documents, and the ways in which digital catalogues and digitised materials offered a suitable alternative for his purposes. He also discusses creating an online database as a research output, an activity which has the benefit of increasing the quantity and quality of information that is accessible to future researchers.

TWO CASES STUDIES ON RESEARCH ADAPTATION

PLANNING
ACCESSING MATERIALS
CREATING DATASETS
WORKING WITH DATA
ACCESSIBILITY ISSUES
CASE STUDIES
FURTHER READING AND RESOURCES
ACKNOWLEDGEMENTS

CASE STUDIES

Barbara Dzieciatko
Locating a suitable data collection tool

Barbara Dzieciatko, a PhD candidate in Education who researches teachers' communities of practice, gives an in-depth account of trying to find a suitable tool to collect data from online forums and evaluates rvest using RStudio, Textblob using Notable, and Beautiful Soup using Anaconda. She outlines some of the challenges she encountered, which included a steep learning curve, a large time cost associated with learning and trialling new tools, issues with tools and packages being outdated, and online tutorials not always providing adequate detail and support.

DATA AND TEXT MINING TO MAP TEACHERS COMMUNITIES OF PRACTICE

Dr Kate Miltner
Challenges of searching through data sources

Dr Kate Miltner, Train@Ed Postdoctoral Fellow at the Centre for Research in Digital Education, discusses running into difficulties finding relevant research materials about mid-century computing schools and describes her journey through the data sources ProQuest Historical Newspapers digital archive and the Internet Archive. Kate summarises some of the challenges she encountered: an overwhelming amount of material and yet archives not being exhaustive, a bias towards English-language and Western sources, the requirement of digital fluency and keyword search skills, and the politics of which archives are digitised.

DIGITAL ARCHIVAL RESEARCH ON MID-CENTURY COMPUTING SCHOOLS

PLANNING

ACCESSING MATERIALS

CREATING DATASETS

WORKING WITH DATA

ACCESSIBILITY ISSUES

CASE STUDIES

FURTHER READING AND RESOURCES

ACKNOWLEDGEMENTS

FURTHER READING AND RESOURCES

Bond, Paul. 'Anger Mounts at UK Government Proposal to Halve Arts Education Subsidy'. World Socialist Web Site, 19 May 2021
www.wsws.org/en/articles/2021/05/20/aruk-m20.html

Cox, Christopher. 'Changed, Changed Utterly'. Inside Higher Ed, 5 June 2020
www.insidehighered.com/views/2020/06/05/academic-libraries-will-change-significant-ways-result-pandemic-opinion

Eckstrand, Nathan. 'Is It Time for the Humanities to Strike?' LSE Higher Education, 6 June 2019
blogs.lse.ac.uk/highereducation/2019/06/06/is-it-time-for-the-humanities-to-strike

Fiesler, Casey, and Nicholas Proferes. '"Participant" Perceptions of Twitter Research Ethics'. Social Media + Society, vol. 4, no. 1, SAGE Publications, Jan. 2018, SAGE Journals
doi:10.1177/2056305118763366

Milligan, Ian. 'How Archival Research Morphs in the Digital Age'. Social Science Space, 21 Aug. 2019
https://www.socialsciencespace.com/2019/08/how-archival-research-morphs-in-the-digital-age/

Moore, Niamh. 'The Politics and Ethics of Maming: Questioning Anonymisation in (Archival) Research'. International Journal of Social Research Methodology, vol. 15, no. 4, Routledge, July 2012, pp. 331–40. Taylor and Francis+NEJM
https://www.tandfonline.com/doi/full/10.1080/13645579.2012.688330

Skov, Malte, and Andreina Carrillo Espinoza. 'The Challenges and Opportunities of Researching Migration in a Pandemic'. Global Development Institute Blog, 21 Jan. 2021
blog.gdi.manchester.ac.uk/the-challenges-and-opportunites-of-researching-migration-in-a-pandemic

Tebeau, Mark. 'A Journal of the Plague Year: Rapid-Response Archiving Meets the Pandemic': Collections: A Journal for Museum and Archives Professionals, SAGE Publications, Jan. 2021, SAGE Journals
https://journals.sagepub.com/doi/full/10.1177/1550190620986550

Tesar, Marek. 'Ethics and Truth in Archival Research'. History of Education, vol. 44, no. 1, Routledge, Jan. 2015, pp. 101–14. Taylor and Francis+NEJM
https://www.tandfonline.com/doi/abs/10.1080/0046760X.2014.918185

Townsend, Leanne, and Claire Wallace. Social Media Research: A Guide to Ethics. 2016
https://www.gla.ac.uk/media/Media_487729_smxx.pdf

PLANNING

ACCESSING MATERIALS

CREATING DATASETS

WORKING WITH DATA

ACCESSIBILITY ISSUES

CASE STUDIES

FURTHER READING AND RESOURCES

ACKNOWLEDGEMENTS

ACKNOWLEDGEMENTS

With thanks to the participants at the Research Adaptation Workshop: Archives & Access on 22 June 2021 and Francesca Baseby and Rachel Hosker who provided valuable insight and informed the content of this document.

This was part of a broader series of guidance documents created by the Centre for Data, Culture & Society and the College of Arts, Humanities and Social Sciences Research Office for researchers in CAHSS who are pivoting to remote and digital research methods. This document was written by Suzanne Black, with support from Lisa Otty, Geoffrey Edgson, and Valentina Andries.

PLANNING

ACCESSING MATERIALS

CREATING DATASETS

WORKING WITH DATA

ACCESSIBILITY ISSUES

CASE STUDIES

FURTHER READING AND RESOURCES

ACKNOWLEDGEMENTS

Other guidance documents

PARTICIPATORY, OBSERVATION & FACE TO FACE RESEARCH

PUBLIC ENGAGEMENT, EVENTS & EXHIBITIONS

SOCIAL MEDIA RESEARCH: ETHICS GUIDANCE FOR RESEARCHERS

THE UNIVERSITY *of* EDINBURGH
Centre for Data, Culture & Society

CONTACT

E: CDCS@ED.AC.UK

WWW.CDCS.ED.AC.UK

@EDCDCS

CITE THIS PAPER

BLACK, ET AL. ARCHIVES, COLLECTIONS & ACCESS GUIDANCE FOR RESEARCHERS AT THE UNIVERSITY OF EDINBURGH (SEPT 21)

Report, Centre for Data, Culture & Society, University of Edinburgh:
https://doi.org/10.5281/zenodo.5575417

If you require this document in an alternative format, such as large print or a coloured background, please contact Lisa Otty, cdcs@ed.ac.uk