

CDCS RAPID PROTOTYPING SANDPIT

Dr Lisa Otty | Dr Jessica Witte | Dr Lucia Michielin

25th June 2023

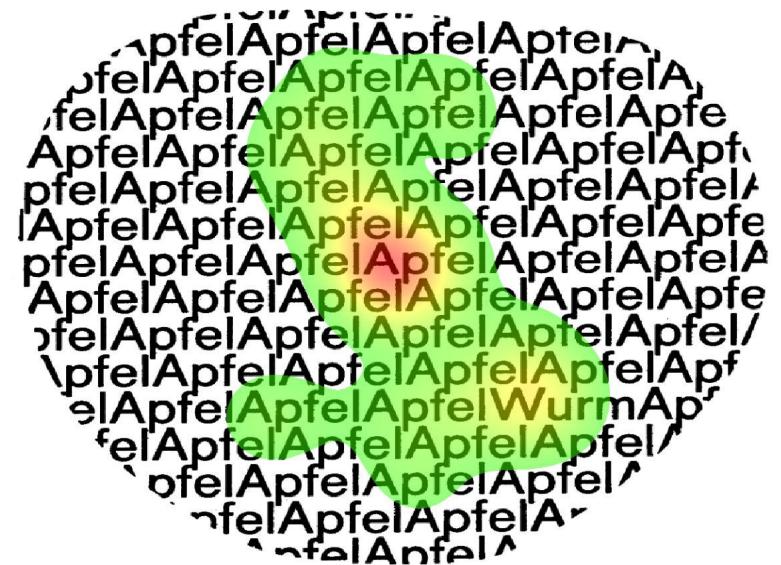


www.ccds.ed.ac.uk



LISA OTTY

- Book Historian – experimental publishing
- Postdocs using novel digital methods and visualisation / Digital Humanities
- Bibliographic data projects and research data services at EDINA (Jisc national data centre)



A graphic composed of the word "Apfel" repeated in a grid pattern, with a central cluster highlighted in green and yellow.



CENTRE FOR DATA, CULTURE & SOCIETY

- Founded 2019, by Prof Melissa Terras
- Steered by multi-disciplinary Advisory Board
- Based within the Edinburgh Futures Institute
- Building capacity for data-led and applied computational research across the arts, humanities and social sciences



www.ccds.ed.ac.uk



THE UNIVERSITY of EDINBURGH
Centre for Data, Culture & Society

"SHAPE research and skills are particularly valuable to the 21st-century workplace – they are vital to the health, wellbeing and prosperity of the nation and to tackling grand challenges. They teach us to analyse, interpret, create, communicate and collaborate with rigour, clarity and energy – crucial skills for today. And together with STEM subjects, they help us make innovation work harder for the benefit of everyone."

- The British Academy www.thebritishacademy.ac.uk/this-is-shape



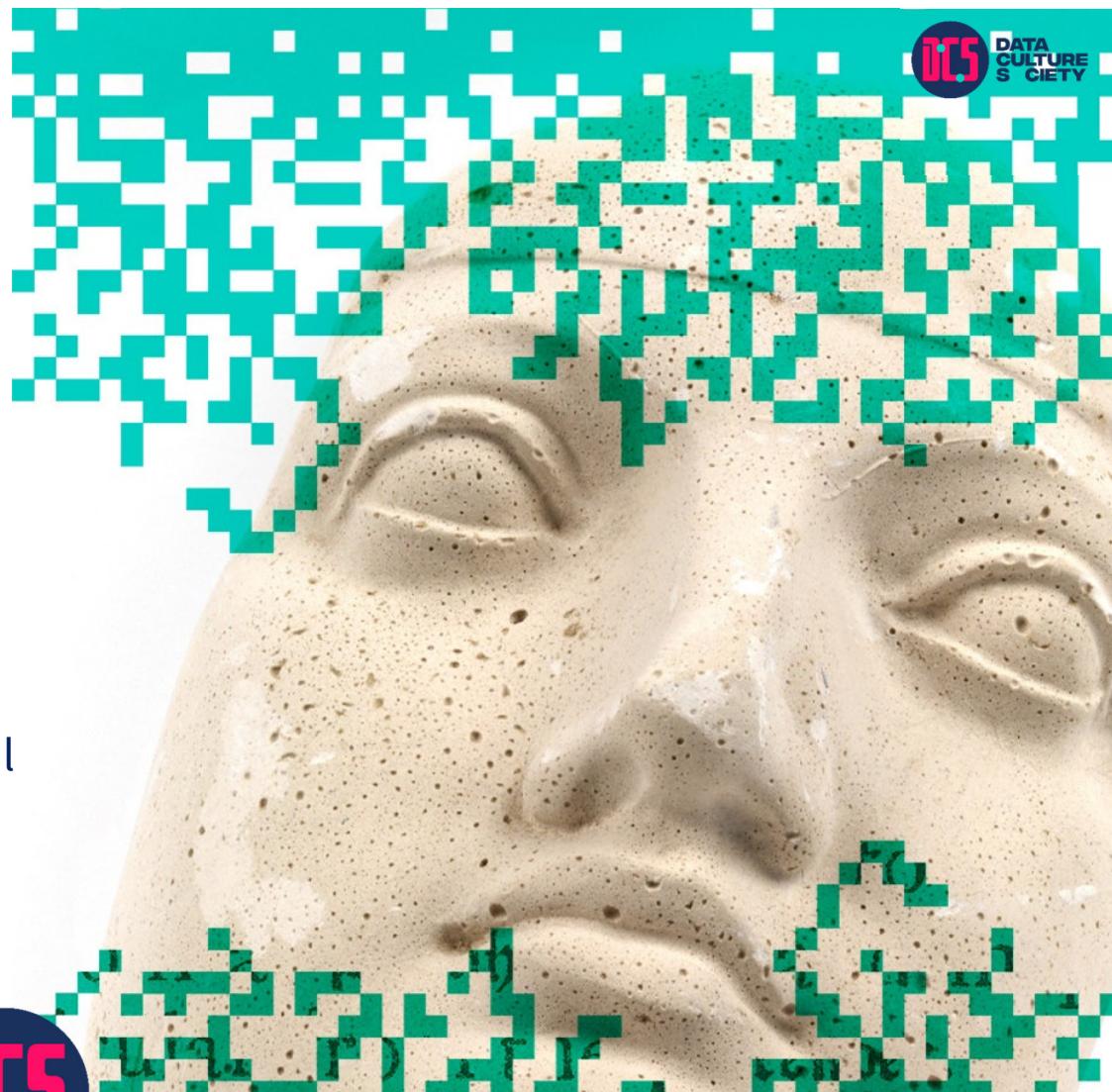
www.ccds.ed.ac.uk





Challenges

- Highly multi-disciplinary community: economics to classics, law to literature, design to international relations.
- Lack of a shared language / language barriers
- Many disciplines with no tradition in teaching digital skills (but some pockets of excellence)
- Participants at different stages of research career



www.ccds.ed.ac.uk



ACTIVITIES

- Upskilling
- Network building
- Showcasing
- Developing infrastructure

IMPACT

Inspire and embed
computational research
across the arts, humanities
and social sciences



OUTPUTS

- Skilled staff and students
- Well developed and funded projects
- Innovative collaborations
- Responsive infrastructure

OUTCOMES

- Communities of practice
- Effective institutional support
- Leading change



www.cdcsl.ed.ac.uk



- Seminars & Events
- Digital Research Prize
- Helping shape the ecosystems
- social events
- Cross department introductions
- Hosting communities of practice

COMMUNITY BUILDING

RESEARCH SUPPORT

network building

upskilling

developing infrastructure

showcasing

TRAINING

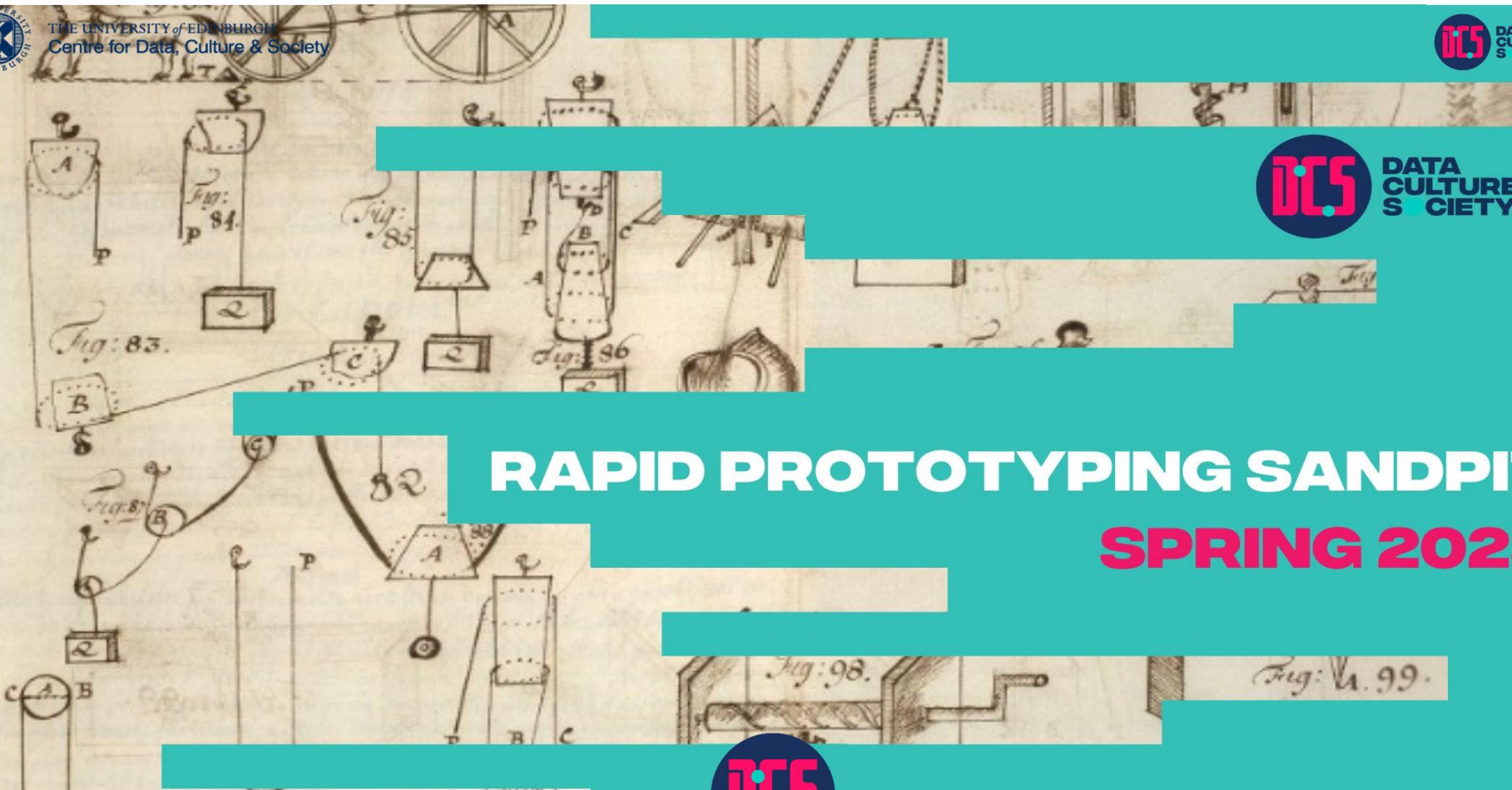
- Training Programme
- Online materials
- Skill sharing
- Challenge-led Skills Development
- Cross sectoral training

- Advice and Guidance
- Seed funding
- Data clinics
- Identifying support gaps and demand
- Developing prototype services
- **New RSE service**





THE UNIVERSITY of EDINBURGH
Centre for Data, Culture & Society



RAPID PROTOTYPING SANDPIT

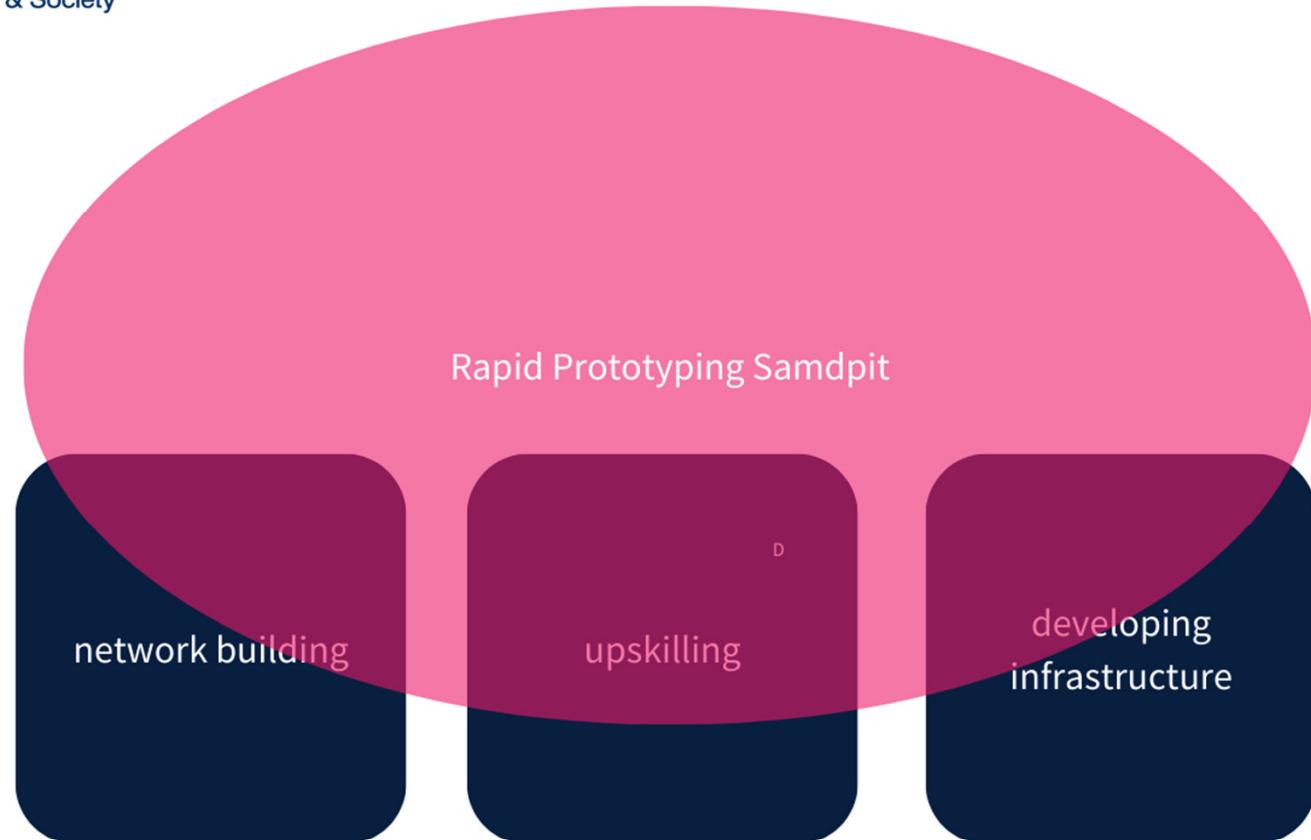
SPRING 2023



www.ccds.ed.ac.uk

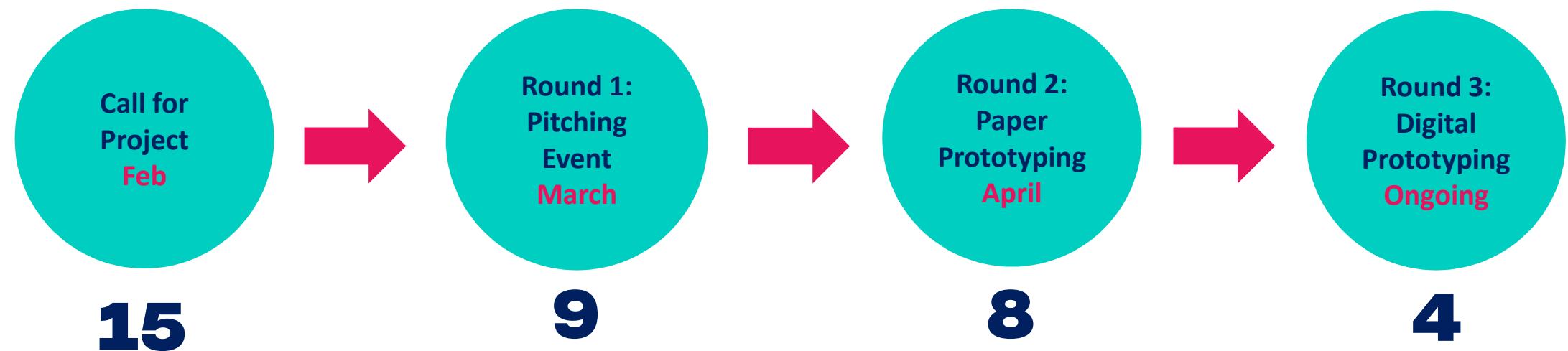


THE UNIVERSITY of EDINBURGH
Centre for Data, Culture & Society



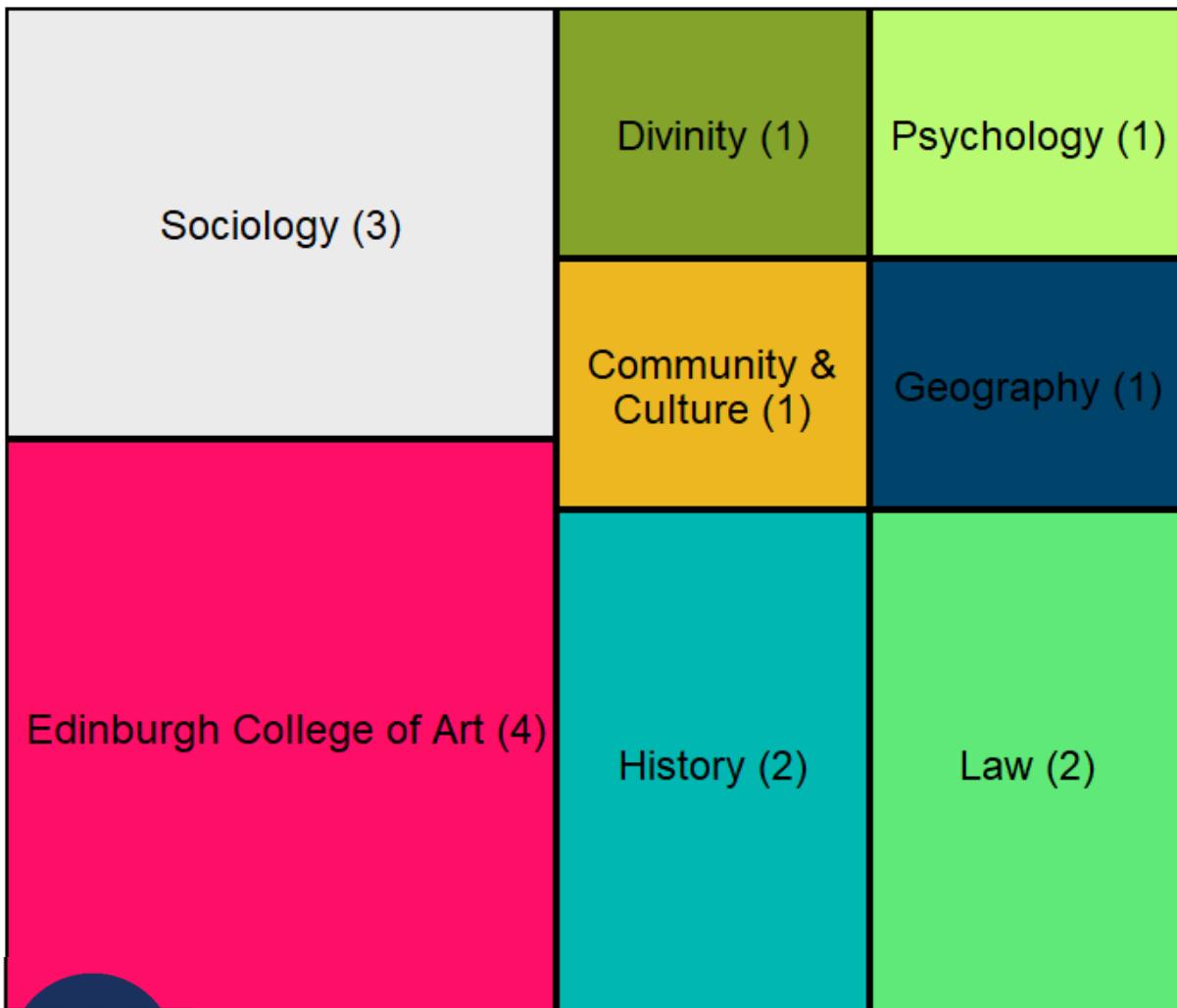
www.ccds.ed.ac.uk

2022/2023 SANDPIT STRUCTURE



www.ccds.ed.ac.uk

- **15** projects
- **8** different School/Disciplines
- **6** support staff involved
- **75%** staff
- **25%** PhD students
- **40%** part of a founded project
- **60%** unfunded projects



METHODS & TOPICS

- Databases
- GIS
- Web Developing
- Data Cleaning
- Web Scraping
- OCR
- Digital Archiving
- Linked Open Data
- NFTs

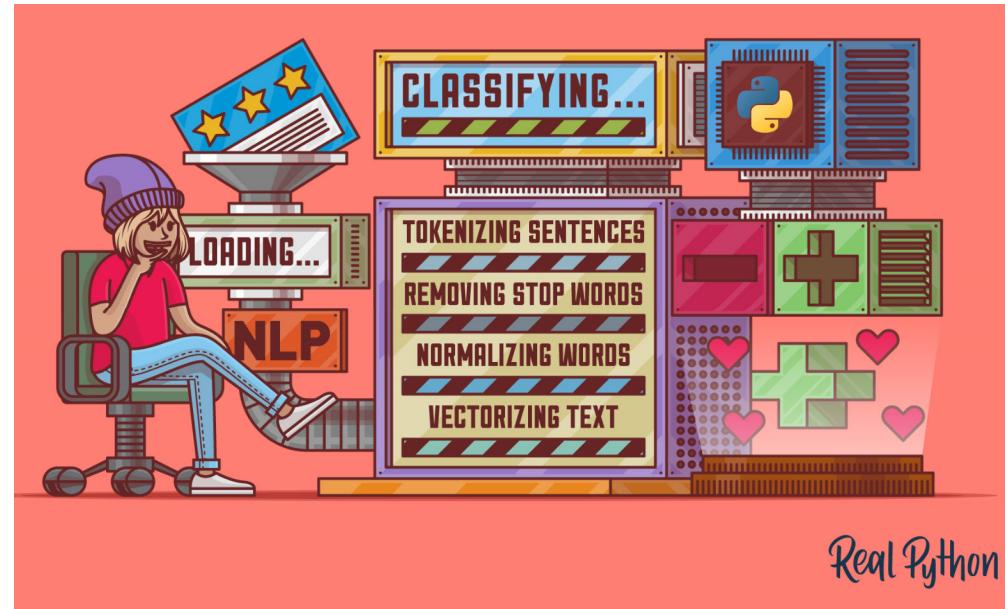


www.ccds.ed.ac.uk



JESSICA WITTE

- Postdoctoral fellow in text & data mining (CDCS and the School of Literatures, Languages & Cultures)
- PhD in nineteenth-century literature
- Research interests include computational literary studies, NLP for humanistic inquiry, digital archives, and improving text mining tools for specific domains



PROJECTS: EXAMPLE 1

AUTOMATING THE CONVERSION OF OLD
PEACE AGREEMENTS TO MACHINE-
READABLE TEXT



www.ccds.ed.ac.uk

Project Overview

- Lead researcher based in the School of Law
- Final product will be implemented in the context of an ongoing collaborative, funded project
- Aimed to expedite an existing workflow with automatic text recognition and machine translation
- Users have varying levels of digital skills

Example 1: Automating the Conversion of Old Peace Agreements to Machine-Readable Text



www.ccds.ed.ac.uk

Benefits & challenges

- Potential for cost saving, as project currently relies on human translation
- Potential for time saving
- Risk of errors: OCR, translation
- Some documents in the dataset are classified or sensitive
- Project team's digital skills literacy

Example 1: Automating the Conversion of Old Peace
Agreements to Machine-Readable Text



www.ccds.ed.ac.uk

Minimum viable product

1. Take as input an image of Arabic text
2. OCR the image
3. Translate the Arabic into English
4. Compare the translated text to a lexicon of keywords/phrases associated with peace agreements
5. If keywords/phrases match, identify the image as a potential peace agreement
 - Jupyter/Colab notebook

Potential next steps

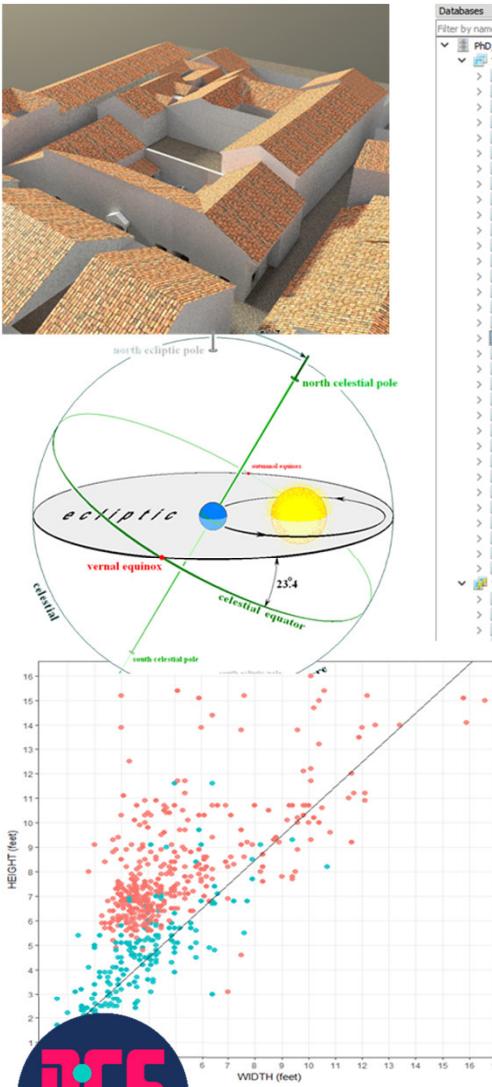
- Widget-style GUI with "drag & drop" functionality
- Exploration (and comparison) of other OCR and translation engines
- Including HTR-specific OCR engine (Transkribus)
- Embedding statistical measures (e.g. confidence interval, number of lexicon matches)



THE UNIVERSITY of EDINBURGH
Centre for Data, Culture & Society

LUCIA MICIELIN

- CDCS Training Manager
- PhD in Classical Archaeology on computational analyses of doors and windows in Roman Private houses
- Specialised in text & data analysis, GIS, photogrammetry, data visualisation, 3D reconstructions



A screenshot of a database management interface, likely SQLite, showing the structure of a table named 'OPENINGS'. The table has 18 columns with the following details:

Name	Data type	Primary Key	Foreign Key	Unique	Check	Not NULL
OPENING_ID	INTEGER (4)					
TYPE	TEXT (8)					
PECULIAR_TYPE	TEXT (20)					
HOUSE_ID	TEXT (30)					
BIGGER_AREA_ROOM	TEXT (25)					
PASSING_STANDING_BIGGER	TEXT (20)					
BIGGER_AREA_ROOM_TYPE	TEXT (30)					
BIGGER_AREA_ROOM_SIZE	INTEGER (4, 2)					
SMALLER_AREA_ROOM	TEXT (25)					
PASSING_STANDING_SMALLER	TEXT (20)					
SMALLER_AREA_ROOM_TYPE	TEXT (20)					
SMALLER_AREA_ROOM_SIZE	INTEGER (4, 2)					
WIDTH	INTEGER (4)					
HEIGHT	INTEGER (4)					
LENGTH	INTEGER (4)					
LEVEL_OF_PRESERVATION	INTEGER (1)					
ID_THRESHOLD_SILL	TEXT (5)					
LD_LINTEL	TEXT (20)					



PROJECTS: EXAMPLE 2

BLACKNESS AND PERFORMANCE IN SCOTLAND 1850–1950



www.ccds.ed.ac.uk

Project Overview

- Will explore representations of blackness in Scottish culture in the XIX and early XX centuries.
- It will examine the ways in which blackness is imagined and performed in theatre, music hall etc.
- Project in development phase
- Lead researcher based in the School of History, Classics, and Archaeology



Benefits & challenges

- Aim to produce new resources for the study of Black History in Scotland (collaborative perspective + public facing)
- Easy implementable options only
- Still on planning phase lot of different possible road to take
- Difficult to see the final outcome at this stage



Minimum viable product

1. Identify the different focuses of the data collection
(sources, venue, actors, plays, etc.)
2. Work on the data relationship
3. Data input from a selection of case studies
4. Geographical visualisation example

Database in SQLite with geographical coordinates that could be used in a GIS environment

Example 2: Blackness and Performance in Scotland
1850-1950



www.ccds.ed.ac.uk



Mr T. Rice, The Original Jim Crow as drawn by Edward Williams Clay' c.1830s.

Possible Next steps

- Use the minimal viable product result to help securing funding for a larger project and to dialogue with potential partners
- Generate a website with embedded Web-GIS that could be used both by the public and by project partners to further develop the project



Example 2: Blackness and Performance in Scotland
1850-1950



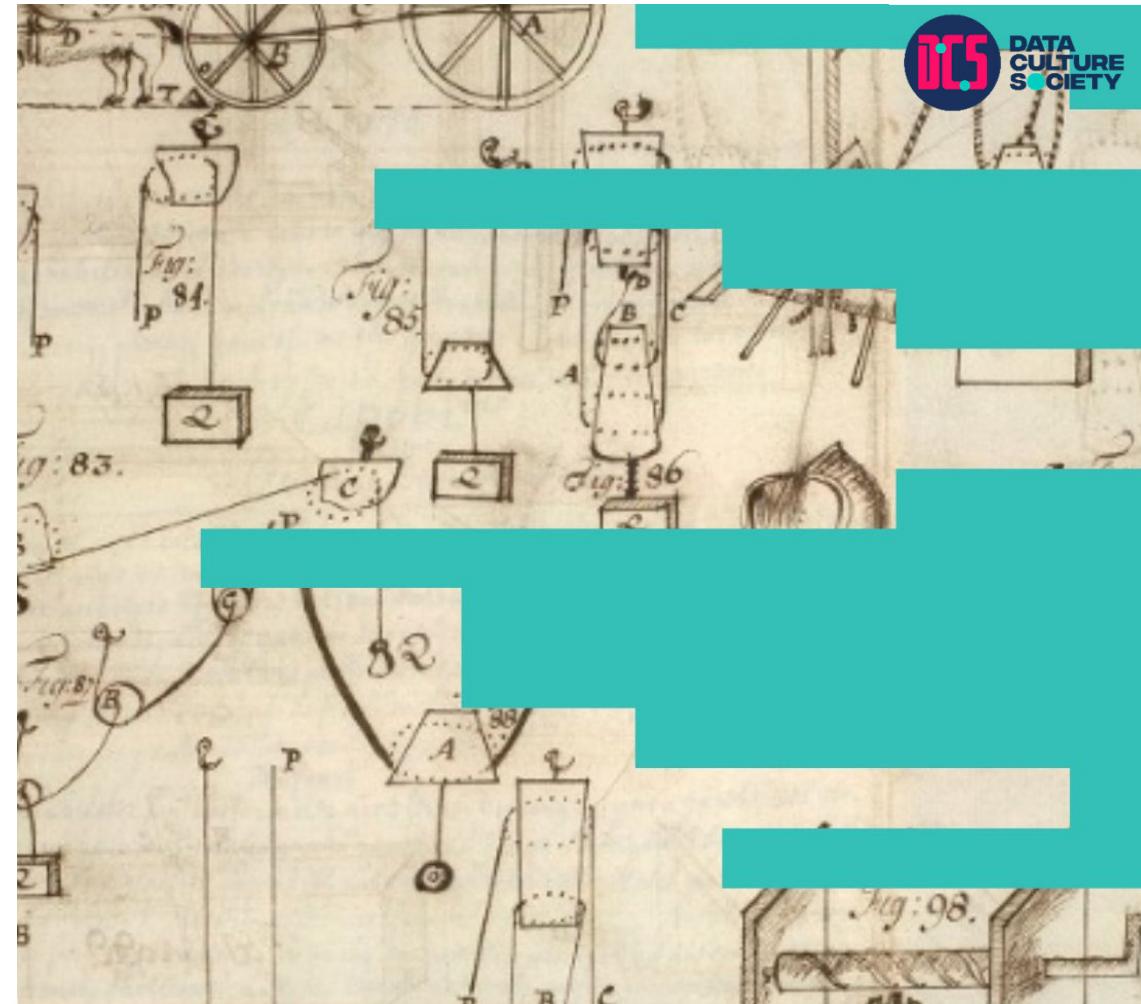
www.ccds.ed.ac.uk

Mr T. Rice, The Original Jim Crow as
drawn by Edward Williams Clay'
c.1830s.



FUTURE OF THE RAPID PROTOTYPING SANDPIT

- Is prototyping the right name?
- More time for early-stage communication
- Rolling rather than time limited?
- Cap on number of projects?
- Separate service for staff?





THE UNIVERSITY of EDINBURGH
Centre for Data, Culture & Society



THANK YOU & QUESTIONS



www.ccds.ed.ac.uk