

### Repurpose, remodel and recast: the National Library of Scotland's Data Foundry

Dr Sarah Ames Digital Scholarship Librarian National Library of Scotland @semames1 | #NLSdata | sarah.ames@nls.uk

## Today's talk

- Data Foundry and Digital Scholarship at NLS
- Engagement and collaborative activity and outcomes
- The way forward
- Challenges





### The Scottish context

#### 2015 Open Data Strategy, Scottish Government

### 2018 City Region Deal

- £1.3bn/15 years (government, universities, private & third sector)
- £60 million for data driven innovation
- Skills/Entrepreneurship/Growth

### University of Edinburgh

- Bayes Institute
- Edinburgh Futures Institute
- Creative Informatics
- Centre for Data, Culture and Society

#### University of Glasgow

• Information Studies/Digital Humanities

#### Universities of Aberdeen, St Andrews, Stirling, Dundee

• Big data/data science

### CodeBase/CodeClan







## The National Library of Scotland

Over 31 million items in collection (excluding web archive, and owned-in-perpetuity digital collections)

'One Third Digital' by 2025 In-house mass digitisation programme

Strategy 2020-2025, 'Reaching People': strong focus on engaging with new audiences through the collections



## **Digital Scholarship Service**



ENCOURAGE, ENABLE & SUPPORT USE OF COMPUTATIONAL RESEARCH METHODS WITH THE COLLECTIONS ENSURE THAT THE COLLECTIONS ARE USED TO THEIR FULL POTENTIAL ESTABLISH A LIBRARY CULTURE WHICH UNDERSTANDS DIGITAL SCHOLARSHIP

PRACTISE AND PROMOTE TRANSPARENCY IN OUR DATA CREATION PROCESSES ANTICIPATE THE FUTURE OF RESEARCH



## Data Foundry



111111 11 2021111

01111111. 11 73

## No-nonsense data

オオイジナリアト ビング じんてんしんせん

uu (Minik 🤨 Minih



## Identifying user needs

Pro	Intermediate	Beginner
All the tech skills!	Limited tech skills	No tech skills
Will find a way to get the data no matter how presented	Understands value of different formats and approaches for research	Wants to use online tools to explore datasets
	questions: theoretical	Just wants the text
But – has expectations of standards (where they exist) & consistency	rather than practical understanding	
	Wants to get hold of the data easily to check what's there	
	Likely to employ an RA to do the work	

Plus broader audience includes other libraries (standards, presentation of data etc)



## Identifying user needs

Pro		Intermediate	Beginner	
All the tech sl	ills!	Limited tech skills	No tech skills	
Will find a wa data no matte presented But – has expe	r how	Understands value of different formats and approaches for research questions: theoretical rather than practical	Wants to use o to explore data Just wants the	sets
of standards exist) & consi	where they	Wants to get hold of the data easily to check what's there		
		Likely to employ an RA to do the work		

Plus broader audience includes other libraries (standards, presentation of data etc)



## Changing processes: digitisation to data







## Making decisions

- Standards:
  - METS/ALTO and Plain text
  - MARC/Dublin Core
  - Tiered downloads
- Image sizes/quality to include
- Storage (local/cloud)
- What metadata to include and what is available
- How to be transparent: gathering and presenting dataset context
- Now moving on to more metadata collections, web archive data, spatial data

National Library of Scotland

CONTACT

## Data Foundry

Data collections from the National Library of Scotland

https://data.nls.uk/



## Our principles

### National Library of Scotland data

## C



The National Library of Scotland publishes data openly and in re-useable formats.

Open

We take the provenance of our data seriously, and are open about how and why it has been produced.

Transparent



Practical

We present datasets in a variety of file formats to ensure that they are as accessible as possible.

### A whole-Library effort

MATIONAL LIERARY OF SCOTLAND

Developers Curators Metadata Rights Access Digitisation

### ...and the National Librarian! ('Data Foundry')

tional Library of Scotland

Leabharlann Nàiseanta na h-Alba



# Engagement activity and outcomes



## Teaching and learning









## **TDM Carpentry**

- A Medical History of British India dataset
- Dr Bea Alex (University of Edinburgh)

#### 2.2 Downloading and Processing Data

#### For a Single Document

Instead of providing a string of text we can download one from an corpus. Here we will, load a file from your local machine, tokenise and lowercase it.

Firstly we download a data set (and make a note of where it is saved). We will use the Medical History of British India collection provided by the National Libarry of Scotland as an example:

#### https://data.nls.uk/data/digitised-collections/a-medical-history-of-british-india/

We can use the open() method to open a file in this collection. You need to specify the path to a file in the downloaded collection (this will be different to the one below depending on where you saved it) and the mode of opening it ('r' for read).

The read() method is used to read the file. It is then stored as a string variable called india\_raw.

We then tokenize it as above and normalize it into lowercase. We can check it has worked by printing out a slice of the list lower\_india\_tokens .

[19]: 1 file = open('../../../Downloads/nls-text-indiaPapers/74457530.txt','r')

2 india raw = file.read()

- 3 india\_tokens = word\_tokenize(india\_raw)
- 4 lower\_india\_tokens = [word.lower() for word in india\_tokens]
- 5 lower\_india\_tokens[0:10]

:[19]: ['no', '.', '1111', '(', 'sanitary', ')', ',', 'dated', 'ootacamund', ',']



itial, exploratory analysis of some of the Library's datasets. No prior prog access and use these Notebooks.



## Jupyter Notebooks

- Series of Jupyter notebooks exploring some of the text and metadata datasets
- For students, learners, non-coders
- By @lucy\_havens





## Creative partnerships





### the dataset is not the map is not the territory

- Martin Disley @martin\_disley
- Funded by Creative Informatics (University of Edinburgh)
- Using GAN techniques with the collections
- <u>https://data.nls.uk/projects/artist-in-residence/</u>



Images: Martin Disley, used with permission





National Library of Scotland

How do we construct an archive? How do we construct truth? Using Artificial Intelligence to create a fake archive, our Artist in Residence @Ma rionC asks these questions using #NLSdata in her new work, Selective Memories.



2:55 PM · Nov 11, 2021 · Twitter Web App

### Is it true? The post-truth archive factory

Marion Carré @Ma\_rionC •

...

- Collaboration with Goethe-Institut Glasgow, Alliance Française Glasgow and Institut Français d'Ecosse
- Challenge: how does AI open up new ways of interacting with library and archival collections and what are the challenges and dangers of using this technology in archival research?
- Will you archive 'fact' or fake news? •
- Take part: <u>https://data.nls.uk/projects/artist-in-residence-</u> ۲ marion-carre/



'Is it true? The Post-truth Archiva Factory'

Marion Cané

7th December 2021 — 6th January 2022

11111624

ational Librar

Alter Alter

Images: Marion Carré, used with permission

sps-.udios.org.uk/shop

97

National Library of Scotland

## Research collaborations



Farmer's Son

THE CAPTAIN OF LOVE. MYMOTHER DIDSOBFFORE ME. JUST THE THING. THE WILD ROVER.





## Annual fellowship 20-21

- Dr Giles Bergel (University of Oxford)
- VGG computer vision tools
- Chapbooks dataset
- Exploring what can be learnt from their illustrations about chapbooks' origins; about relationships between chapbook printers, publishers and distributors; and about the type and range of imagery available to their readers.

https://data.nls.uk/projects/the-national-librarians-researchfellowship-in-digital-scholarship

	- ALCHATE_NEEDED / INNER SHE HESHELSY CK_SCH / ATTENDEDE
	urticles:
	ACQUEST: or Acquist, in law, signifies goods got by purchase or donation. See
	Collt JUEST
6 v	ACOUI: "a town of Italy, in the Dutchy of Montferrat, with a bishop\u2019s see,\
	\ and commodious baths. It was taken by the Spaniards in 1745, and retaken by\
	A the Piedmontese in 1746; but after this, it was taken again and difrcantled.
	\ by the French, who afterwards forsook it. It is seated on the river Bormio, \
10 11 - V	\ 25 milesN.W. Gence, and 385. Cafal, 8, 30.E. 44. 40. lat."
12	ACQUITSCINCE: in commerce, is the consent that a person gives to the determination given either by arbitration, orbyaconful
ű.	ACQUIETANDIS plegiis: in the English law, is a writ that lies for a surety, againd
14	a creditor, who refules to acquit the complainant after the debt is paid.
15 🗸	ACQUIETANTIA de Uniris et hundredhs in England, signifies the privilege of being
16	free from suit and service in filires and hundreds.
	ACOUISITION: in general, denotes the obtaining or procuring something. Among lawyers,
18	it is used for the right or title to an estate got by purchase or dona
19 🗸	ACQUITABE: in ancient law-books, signifies to difcliarge or pay off the debts
18	of a person deceased.
21 0	ACQUITTALE a discharge, deliverance, or setting of a person free from the guilt
12	or suspicion of an offence.
28 24 - v :	ACQUITTANCE: a release or discharge in writing for a sum of money. ACRA: a town of Africa, on the coast of Guinea, where the English, Dutch, and
15	Danes, have strong sorts, and each sort its particular village, o. 2.W. 5. o.
26	
27	ACRASIA: among physicians, signifies the predominancy of one quality over another.
18 👳	ACRE: "or Acra, a sea-port town in Syria. It was formerly called Ptolemais, and
29	\ is a bishop\u2019s see. It was very famous in the time of the crusadoes, and\
30	\ underwent several sieges both by the Chrif Hans and Saracens. It is now an\
31	\ inconsiderable town, being entirely supported by its harbour, which, is frequented
32	A by (hips of several nations, It is 20 miles5. Tyre, and 37N. Jerusalem, 39.)
13	1 25.E. 32. 40. lat. Acre, in the Moguilu2819s dominions, the same with lack,
34 35	\ and signifies the sum of 100,000 rupees ; the rupee is -of the value of the\ \ French crown of 3 livres, or 30 fols of Holland; an 100 lacks of rupees make\
36	\ a couron in Indoftan, or 10,000,000 rupees; the pound Sterling is about 8\
37	\ rupees; according to which proportion, a lack of rupees amounts to 12,500\
18	> pounds Sterling. Acre, a measure of land used in several provinces of France.
39	\ particularly in Normandy. It is larger or less according to the different\
4Ø.	A places ; but commonly contains 160 perches. The Acre of woods in France, consists
	\ of four roods, called verg es; the rood is 40 perches, the perch 24 feet, \
62	\ the foot 12 inches, the inch 12 lines. Acre, the universal measure of land\
8	\ in Britain. An acre in England contains 4 square roods, a rood 48ACR or poles\
14	\ of 16\xA3 feet each byftatute. Yet this measure does not prevail in all parts\
45 46	\ of England, as the length of the pole varies in different counties, and is\ \ called cujiosnary measure, the difference running from the 16feet to 28. The\
*** 67	\ acre is also divided into an square chains, of 22 yards each, that is 4848\
18	\ square yards. An acre in Scotland contains 4 square roods : 1 square rood\
19	\ is 40 square falls; 1 square fall, 36 square ells; 1 square ell, 9 square\
	A feet, and 73 square inches: 1 square foot, 144 square inches. The Scots acrel
	\ is also divided into 10 square chains ; the measuring chr. in Ihould be 24\.
	A ells in length, divided into 100 links, each link 8 r \u2019 c 7I and sol\
53	\ chain will contain 10,000 square links. The Englith (latute acre is about)
54	X 3 roods and 6 falls standard measure of Scotland."
55 56	ACREME: in old law-books, signifies ten acres of land.
50 57	ACRIBELA: signifies great accuracy. ACRID: a name for any thing that is of a (harp or pungent taste.
58. v	ACRIDOPHAGI: "signifies locujl-eaters. It has been much disputed whether the inhabitant
59	\ of Arabia, Ethiopia, <bc eat="" ever="" give="" ihall="" locusts.="" of<="" substance="" th="" the="" we=""></bc>
58	\ what Haffelquill says on this fubject, who travelled in Syria and Egypt so\
	\ late as the year 1752. This ingenious gentleman, who travelled with a view\
	\ to improve natural history, informs us, that he asked Franks, and many other\
53 54	x people who had lived long in these countries, whether they had ever heard
64	\ that the inhabitants of Arabia and Ethiopia, 6c. used locusts as food. They\
55	A answered that they had. He likewise alked the same question of Armenians.



## Annual fellowship 21-22

- Dr Rosa Filgueira (St Andrews University)
  - AI toolkit for the collections: bringing AI tools to those who can't code
  - Initially with the Encyclopaedia Britannica dataset

https://data.nls.uk/projects/the-national-librarians-research-fellowship-in-digital-scholarship-2021-22/





## Annual fellowship 22-23

- Dr Gustavo Candela (University of Alicante)
  - Exploring the Semantic Web and Wikidata with Library metadata collections

https://data.nls.uk/projects/the-national-librarians-researchfellowship-in-digital-scholarship-2022-23/



## PhD studentships



#### Joe Nockels

'Adopting Transkribus in the National Library of Scotland: Understanding how handwritten text recognition will change management and use of digitised manuscripts'.

Supervisors: Professor Melissa Terras (University of Edinburgh), Dr Paul Gooding (University of Glasgow), Dr Sarah Ames and Stephen Rigden (National Library of Scotland).

Funder: Scottish Graduate School for Arts and Humanities, AHRC Collaborative Doctoral Award



### Ash Charlton

'Slavery and Race in the Encyclopaedia Britannica (1768-1860): A Text Mining Approach'.

Supervisors: Professor Melissa Terras and Professor Diana Paton (University of Edinburgh), Dr Sarah Ames and Robert Betteridge (National Library of Scotland) Funder: Scottish Graduate School for Arts and Humanities, AHRC Collaborative Doctoral Award

Copyright Asad Khan, used with permission

National Librar

cotlar



National Library of Scotland Leabharlann Nàiseanta na h-Alba

Copyright Asad Khan, used with permission



Copyright Asad Khan, used with permission





**High Performance Computing meets** Encyclopaedia Britannica



HOME

ABOUT -

DATA -

Data visualisation student projects

3,000 Scottish chapbooks...as music!

TOOLS CONTACT



### Projects



Artist in Residence



The National Librarian's Research Fellowship in Digital Scholarship

### https://data.nls.uk/projects/

Examples of collaborative and individual projects over



Finding lost footpaths using GB1900



Teaching digital humanities with A Medical History of British India



AI in Residence



Mapping quarries and collieries



Geoparsing the Gazetteers of Scotland



the past 6 months



## The way forward





## Beyond the 'collection'

- Mix 'n' match datasets
  - New DAMS will enable staff to create non-collection-driven datasets
  - Eventually extend this to users?
  - User-driven





## 'Dark' Data Foundry

- Non-open data
- Gradually increase organisational appetite for risk
- Data Safe Haven?





## Improved engineering

- Ingest of modified datasets
- More sophisticated search
- In-house tech support
- Library as data: Library itself as a 'data space'

A dictionary of arts and sciences': ets/eb\_test/144850366

, signifies goods got by purchase or donation. See

the Dutchy of Montferrat, with a biihop\u2019s see, t was taken by the Spaniards in 1745, and retaken by ; but after this, it was taken again and difrcantled \mards forsook it. It is seated on the river Bormio, 305. Cafal, 8. 30.E. 44. 40. lat."

is the consent that a person gives to the determination on, orbyaconful

e English law, is a writ that lies for a surety, againd o acquit the complainant after the debt is paid. undredh: in England, signifies the privilege of being e in fluires and hundreds.

notes the obtaining or procuring something. Among lawyers, or title to an estate got by purchase or dona ooks, signifies to difcliarge or pay off the debts

iverance, or setting of a person free from the guilt e.

ischarge in writing for a sum of money.

the coast of Guinea, where the English, Dutch, and and each sort its particular village, o. 2.W. 5. o.

signifies the predominancy of one quality over another. town in Syria. It was formerly called Ptolemais, and It was very famous in the time of the crusadoes, and s both by the Chrif Hans and Saracens. It is now an\ in the Mogul\u2019s dominions, the same with lack,\ f 100,000 rupees ; the rupee is -of the value of the\ s, or 30 fols of Holland; an 100 lacks of rupees make\ r 10,000,000 rupees; the pound Sterling is about 8\ ich proportion, a lack of rupees amounts to 12,500\ a measure of land used in several provinces of France, y. It is larger or less according to the different\ ontains 160 perches. The Acre of woods in France, consists\ erg es; the rood is 40 perches, the perch 24 feet, inch 12 lines. Acre, the universal measure of land England contains 4 square roods, a rood 40ACR or poles\ atute. Yet this measure does not prevail in all parts th of the pole varies in different counties, and is\ re, the difference running from the 16feet to 28. They to xo square chains, of 22 yards each, that is 4840\ n Scotland contains 4 square roods ; 1 square rood\ quare fall, 36 square ells; 1 square ell, 9 square\ hes; 1 square foot, 144 square inches. The Scots acre square chains ; the measuring chr.in Ihould be 24\ into 100 links, each link 8 r \u2019 c ?I and sol\ 00 square links. The Englilh (latute acre is about)

ignifies ten acres of land.

ccuracy.

that is of a (harp or pungent taste. ujl-eaters. It has been much disputed whether the inhabitants\ c ever eat locusts. We Ihall give the substance of n this fubjedt, who travelled in Syria and Egypt so\ This ingenious gentleman, who travelled with a view ory, informs us, that he asked Franks, and many other ng in these countries, whether they had ever heard Arabia and Ethiopia, &c. used locusts as food. They He likewise alked the same question of Armenians. o lived in Arabia, and had travelled in Syria and f whom said they heard of such a pradtice, and others\ n the people eat these infedts. He at last obtained n this head from a learned Iheck at Cairo, who had a. This gentleman told him, in presence ofM. Grand,\ terpreter at Cairo, and others, that a famine frequently re is a scarcity of corn in Egypt, which obliges the\ coarser food than ordinary : That when corn is scarce, loculls in hand-mills, or Hone mortars, and bake them se cakes in place of bread: That he has frequently e Arabians, even when there was no scarcity of corn;\ itew them with butter, and make them into a kind\ ays is not disagreeably tailed; for he had sometimes caffees out of curiosity. From this account, we may ispute among divines about the nature of St John\u2019\ Some of them say that loculls were the fruits of certain\

#### A C R

ACQUEST, or Acquist, in law, figsifies by purchafe or donation. See Conquist, ACQUI, a town of Italy, in the Dutchy of rat, with a bishop's fee, and commodious was taken by the Spaniards in 1745, and r the Piedmontefe in 1746; but after this, it again and difinantled by the French, who forfook it. It is feated on the river Bormic N. W. of Genoa, and 30 S. of Cafal, 8. 30 44. 40. lat.

ACQUIESCENCE, in commerce, is the co a perfon gives to the determination given sit bitration, or by a conful

ACQUIETANDIS plegin, in the English writ that lies for a furety, against a creditor fuses to acquit the complainant after the de ACQUIETANTIA de firir et hundredir, in fignities the privilege of being free from fui vice in fhires and hundreds.

ACQUISITION, in general, denotes the ob procuring fomething. Among lawyers, it is the right or title to an effate got by purchaltion.

ACQUITARE, in ancient law-books, fignificharge or pay off the debts of a perfon decea ACQUITTAL, a difcharge, deliverance, or

a perfon free from the guilt or fulpicion of a ACQUITTANCE, a release or difeharge i for a fum of money.

ACRA, a town of Africa, on the coaft of Guin the English, Dutch, and Danes, have firand each fort its particular village, o. 2. 5. o. lat.

ACRASIA, among phyficians, fignifics the pro cy of one quality over another.

ACRE, or ACRA, a fea-port town in Syria. formerly called *Ptolemain*, and is a bilhop' was very famous in the time of the crafadoer derwent feveral fieges both by the Chriftians cens. It is now an inconfiderable town, bei ly fupported by its harbour, which is freq fhips of feveral nations. It is 20 miles S. and 27 N. of Jerufalem, 39. 25. E. I 40. lat.

Acκε, in the Mogul's dominions, the fame ' and fignifies the fum of 100,000 rupees; th of the value of the French crown of 3 livre fols of Holland; an 100 lacks of rupees ma ron in Indoltan, or 10,000,000 rupees; t Sterling is about 8 rupees; according to w portion, a lack of rupees amounts to 12,50 Sterling.

 ACRE, a meafure of land ufed in feveral pre France, particularly in Normandy. It is lefs according to the different places; but c contains 160 perches.

The ACRE of words in France, confifts of fe called verg es; the rood is 40 perches, the feet, the foot 12 inches, the inch 12 lines. ACRE, the univerfal measure of land in Brit acre in England contains 4 fquare roods, a



### Analysis layers

- Dr Rosa Filgueira Fellowship project
- Potential for platforms to be built on top of Data Foundry, given the standardised nature of the data

|--|

### National Library of Scotland Leabharlann Nàiseanta na h-Alba

## (Inter)national infrastructure?

- The success of Data Foundry depends on community
- Project led by Universities of Glasgow/Sheffield



## Challenges

### Change! Workflows, processes, culture...

### Resourcing and sustainability

### Identifying user needs

### (Inter)national funding/partnerships





## Thank you!

sarah.ames@nls.uk @semames1 | #NLSdata