

*flyswot: garden-variety machine
learning applications*

Catherine Cronin – Digitisation Workflow Officer

Daniel van Strien – Digital Curator

Andrew Longworth – Digitisation Workflow Manager

What do all these images
have in common?

BRITISH LIBRARY

Or 8816

18th century, An Ethiopian manuscript containing: ff. 3r-155v. Psalms, መዝሙር ዘዳዊት. ff. 155v-171r . ff. 171r-178r. The Canticles of the Prophets of the Old and New Testaments ff. 179r-178r. Song of Songs of

Content

show

View

Open book

Direction

Left to Right

f.ii-v - iii-r

f.181v - 182r

f.182v - 183r

f.183v - 184r

f.184v - 185r

f.185v - 186r

f.186v - 187r

f.187v - 188r

f.188v - 189r

f.189v - 190r

f.190v - 191r

f.191v - 192r

f.192v - 193r

f.193v - 194r

f.194v - 195r

f.195v - 196r

f.196v - 197r

f.197v - 198r

f.198v - 199r

f.199v - i-r

f.i-v - ii-r

f.ii-v - iii-r

f.iii-v - n/a

[Digital object for Or 8816](#)

- f.198v - 199r
- f.199v - i-r
- f.i-v - ii-r
- f.ii-v - iii-r**
- f.iii-v - n/a
- front-i - back-i
- n/a - front
- back - spine

BRITISH LIBRARY

Or 12467

1660, Psalter, ጳጳሮች. ff. 1r-2v. Prayer for the dying. ff. 3r-7v. A fragment of the Miracles of the Virgin Mary containing five miracles, the end of the last missing. ff. 8r-129r . Psalms, Incipit: ፍካፊ ዘጸድቃን ወዘጋጥላን መዝሙር ዘላዊት ፩ ብፁዕ ብሉክ ዘኢሓረ. ff. 129v-142v. The Canticles of the Prophets of the Old and

Content

show link

View

Open book

Direction

Left to Right

f.i-v - ii-r

- f.151v - 152r
- f.152v - 153r
- f.153v - 154r
- f.154v - 155r
- f.155v - 156r
- f.156v - 157r
- f.157v - 158r
- f.158v - 159r
- f.159v - 160r
- f.160v - 161r
- f.161v - 162r
- f.162v - 163r
- f.163v - 164r
- f.164v - 165r
- f.165v - 166r
- f.166v - 167r
- f.167v - 168r
- f.168v - 169r
- f.169v - i-r
- f.i-v - ii-r**
- f.ii-v - iii-r
- f.iii-v - n/a

[Digital object for Or 12647](#)

f.168v - 169r

f.169v - i-r

f.i-v - ii-r

f.ii-v - iii-r

f.iii-v - n/a

front-i - back-i

BRITISH LIBRARY

Add MS 38132 c 1536-1545, ff. 1r-39v: Fragments from John Leland's Collectanea. The manuscript seems to have become separated from the rest of the Collectanea at some point before the middle of the 17th

Content

show link

View Direction

Navigation icons: back, forward, search, zoom in, zoom out, full screen.

Table of contents on the right:

f.i-v - ii-r
f.ii-v - iii-r
f.iii-v - iv-r
f.iv-v - v-r
f.v-v - vi-r
f.vi-v - 1r
f.1v - 2r
f.2v - 3r
f.3v - 4r
f.4v - 5r
f.5v - 6r
f.6v - 7r
f.7v - 8r
f.8v - 9r
f.9v - 10r

LIBRARY
HS11B

[Digital object for Add MS 38132](#)

Add MS 48002 1475-1581, Collected copies of treaties and documents concerning England and France between 1478 and 1581. ff. 1r-24v: Blank.f. 25r: Title page: 'Tractatus francici antiquiores'.f. 26r-26v: Paper of

Content

show link

View

Single

Direction

Left to Right



[Digital object for Add MS 48002](#)

Add MS 15736 1612-1623, Friendship album or 'Stammbuch' of George Andrew, Freiherr von Herberstein. Containing autographs and coloured arms, collected principally at Strasbourg and Steir, in the years 1612-1623. With an alphabetical index of writers (ff. 96v-104r).f. 1r Parchment fragment with ownershi

Content

show link

View Direction

f.vii-r

f.106v
f.106ar
f.106av
f.106br
f.106bv
f.107r
f.107v
f.i-r
f.i-v
f.ii-r
f.ii-v
f.iii-r
f.iii-v
f.iv-r
f.iv-v
f.v-r
f.v-v
f.vi-r
f.vii-r
f.viii-r
front-i
back-i
front
back
spine


[Digital object for Add MS 15736](#)





BRITISH LIBRARY

Add MS 11829 1500-1599, Torah scroll.Title: Sefer Torah.Title: ספר תורה.Note: Consists of a number of fragments: Genesis 1:1-49:7, Exodus 5:14-17:6, Deuteronomy 22:3-25:9, Leviticus 4:20-7:27, 13:37-17:12, 9:22-
Content
show link
View Single Direction Left to Right
f.i-r
f.i-r
f.ii-r
f.iii-r
f.1r
f.2r
front
f.i-r
f.ii-r
f.iii-r
f.1r
f.2r
front



↶ 🔍 🔍 ↶ ↷

f.i-r

f.i-r
f.ii-r
f.iii-r
f.1r
f.2r
front

[Digital object Add MS 11829](#)

BRITISH LIBRARY

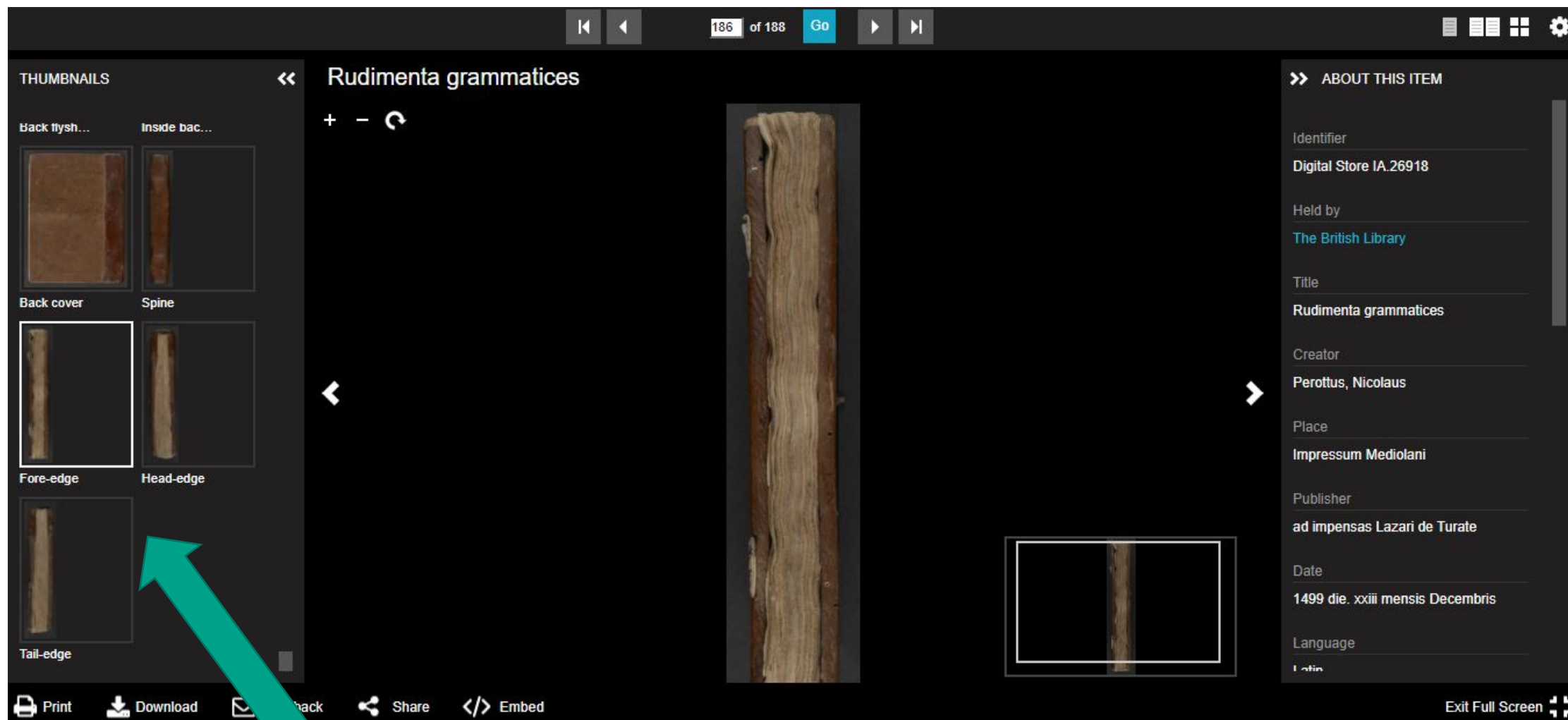
They are all FAKE FLYSHEETS

- Limited set of visible image labels
- Inflexible order sequence of images
- Tied to an old digital object platform
- Labels derived from image file names
- Many images incorrectly labelled as a flysheet



New IIF Universal Viewer

- Digital objects migrated to new viewer
- Standardised vocabulary for image labels
- Flexibility in order sequence of images
- Ability to apply appropriate labels to 'fake flysheet' images



[Digital object for IA.26918](#)

Fore-edge

Head-edge

Tail-edge

THUMBNAILS

xvi verso Back flysh... Back flysh... Inside bac... Inside bac... Inside bac... Inside bac...

Inside back case unfolded 2

British Library Ashley MS 5023 Letters (6) from Mary Shelley to Claire Clairmont; 27...

88 of 91 Go

UNBOUND MSS.

Letters (6) from Mary Shelley to Claire Clairmont; 27 Oct. [1844], 7 [July 1845], [12 Oct. 1845], [1

Identifier Ashley MS 5023

Held by The British Library

Title Letters (6) from Mary Shelley to Claire Clairmont; 27 Oct. [1844], 7 [July 1845], [12 Oct. 1845], [1

Creator Shelley, Mary Wollstonecraft, 1797-1851, writer

Date 1844-1848

Language English

Catalogue record

[Digital object for Ashley MS 5023](#)

Inside back case unfolded 2

THUMBNAILS

f. 198v

f. 198r

Back flysh...

Back flysh...

Back cover

Inside bac...

MSS Aves...

Spine

British Library MSS Avestan 17 Yasna sādah

+ - ↺

Navigation arrows: < >

>> ABOUT THIS ITEM

Identifier

MSS Avestan 17

Held by

[The British Library](#)

Title

Yasna sādah

Date

Language

Avestan; Gujarati

Catalogue record

[View the catalogue record](#)

Citation

British Library MSS Avestan 17,
<http://access.bl.uk/item/viewer/ark:/81055/vdc_100113993998.0x000001>

Digitised by

[Digital object for MSS Avestan 17](#)

MSS Avestan 17



We have a mountain of digitised content:

- Tens of thousands of images
- About 1 petabyte (1,000 terabytes)

How do we identify which images are ‘fake flysheets’?



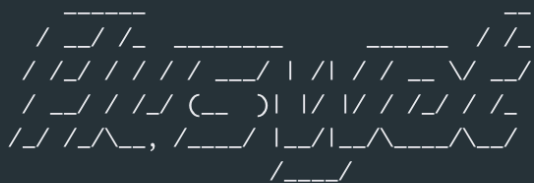
Using computer vision to detect fake flysheets

- Created training data by putting images into folders (integrating into existing workflows)
- Trained 'boring' CV models (ResNets)
- DVC and other tools used to make pipeline (semi) reproducible when new training data is available

flyswot



- Command Line Application used to 'deploy' our models
- Searches for images with a flysheet filename
- Predicts whether they are an actual flysheet or 'other' via a CSV report



📄 CSV report 📄

CSV report file at:

file:///Users/dvanstrien/Documents/DS/2021_11_17_10_15.csv

🕒 Time taken to run 🕒

0:00:19.819429

📁 files checked 📁

```
- flyswot searched for image files by matching the patern *fs*
- flyswot search inside: `/Users/dvanstrien/Documents/DS/flysheet_detect
- In total the directory contained **50** images
- There were **46** files matching the fs* pattern which flyswot checked
```

Prediction Summary

Prediction summary 1

Class	Count	Percentage
other	46	100.0
Total	46	

Prediction summary 2

Class	Count	Percentage
EDGE + SPINE	45	97.83
CONTAINER	1	2.17
Total	46	

Incorporating flyswot into the workflow

- Silver bullet!
- Unfortunately we have a large and varied collection
- flysheet vs other
- Interpret results using Power BI
- Allows us to assess the state of the image set and the amount of resource required

TIFF prediction 1		
1 flysheet vs other		Total
flysheet		434
other		58
Total		492

TIFF prediction 2		
2 label category		Total
flysheet		446
page + folio		28
cover		10
other		4
container		2
edge + spine		2
Total		492

Number of items analysed which contain flysheets
48
Total with "other" or flysheet < 80%
37

Interpreting the results

- Allows us to focus on the items with fake flysheets

Items and files = "other" and/or flysheet with less than 80% confidence

directory	1 flysheet vs other	confidence	2 label category	confidence
<input type="checkbox"/> \\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_10302\Tiff				
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_10302\Tiff\add_ms_10302_fs001r.tif	other	91.08%	page + folio	71.10%
<input type="checkbox"/> \\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_11414\tiffs				
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_11414\tiffs\add_ms_11414_fs001r.tif	flysheet	68.07%	flysheet	90.01%
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_11414\tiffs\add_ms_11414_fse001r.tif	flysheet	50.88%	flysheet	83.79%
<input type="checkbox"/> \\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_15705\tiffs				
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_15705\tiffs\add_ms_15705_fse007r.tif	other	95.38%	page + folio	94.66%
<input type="checkbox"/> \\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_18144\Tiffs 300 ppi				
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_18144\Tiffs 300 ppi\add_ms_18144_fs001r.tif	flysheet	76.97%	flysheet	85.57%
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_18144\Tiffs 300 ppi\add_ms_18144_fs001v.tif	other	93.15%	other	48.20%
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_18144\Tiffs 300 ppi\add_ms_18144_fse005r.tif	other	99.93%	edge + spine	95.56%
<input type="checkbox"/> \\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_18719\Tiffs 300 ppi				
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_18719\Tiffs 300 ppi\add_ms_18719_fs001r.tif	other	98.79%	cover	97.64%
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_18719\Tiffs 300 ppi\add_ms_18719_fse003r.tif	other	90.11%	page + folio	79.97%
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_18719\Tiffs 300 ppi\add_ms_18719_fse012v.tif	other	99.53%	cover	92.90%
<input type="checkbox"/> \\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_19553\Tiff				
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_19553\Tiff\add_ms_19553_fs001v.tif	other	99.85%	page + folio	99.02%
<input type="checkbox"/> \\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_21247\Tiff				
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_21247\Tiff\add_ms_21247_fs004v.tif	other	75.18%	page + folio	62.36%
<input type="checkbox"/> \\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_25435\tiffs				
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_25435\tiffs\add_ms_25435_fse002v.tif	flysheet	71.88%	flysheet	92.67%
<input type="checkbox"/> \\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_27695\tiffs				
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_27695\tiffs\add_ms_27695_fs002r.tif	other	97.07%	page + folio	96.34%

Continuous improvement

- Highlights any unexpected predictions

file	1 flysheet vs other	confidence	2 label category	confidence
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_27695\tiffs\add_ms_27695_fse007v.tif	other	63.67%	flysheet	49.16%
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_65194\tiff\add_ms_65194_fse001v.tif	other	50.83%	flysheet	88.31%
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_65194\tiff\add_ms_65194_fse002r.tif	other	86.22%	flysheet	57.86%
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\add_ms_89309\300ppi tiffs\add_ms_89309_fs001r.tif	other	74.94%	flysheet	61.65%
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\cotton_ms_tiberius_a_viii1\tiffs\cotton_ms_tiberius_a_viii1_fs001r.tif	other	50.04%	flysheet	76.67%
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\egerton_ms_2709\tiffs\egerton_ms_2709_fs003r.tif	other	51.47%	flysheet	59.65%
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\egerton_ms_3018\tiffs\egerton_ms_3018_fs001r.tif	other	87.07%	flysheet	54.23%
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\harley_ms_6205\tiffs\harley_ms_6205_fse004v.tif	other	57.01%	flysheet	50.62%
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\royal_ms_17_a_xliv\tiffs\royal_ms_17_a_xliv_fs001v.tif	other	71.32%	flysheet	77.77%
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\royal_ms_17_a_xliv\tiffs\royal_ms_17_a_xliv_fs002v.tif	other	75.53%	flysheet	81.42%
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\royal_ms_17_a_xliv\tiffs\royal_ms_17_a_xliv_fse003v.tif	other	73.33%	flysheet	81.51%
\\ad\collections\Preservation Digitisation Bids 2014-15\Finished Manuscripts 2\royal_ms_17_a_xliv\tiffs\royal_ms_17_a_xliv_fse004v.tif	other	67.30%	flysheet	82.32%

- Refine the thresholds, give it examples to learn from and improve accuracy

1 flysheet vs other

☐ other

2 label category

☐ Select all

☐ flysheet

1 flysheet vs other ...

50.04%

99.99%

2 label category con...

45.32%

99.98%

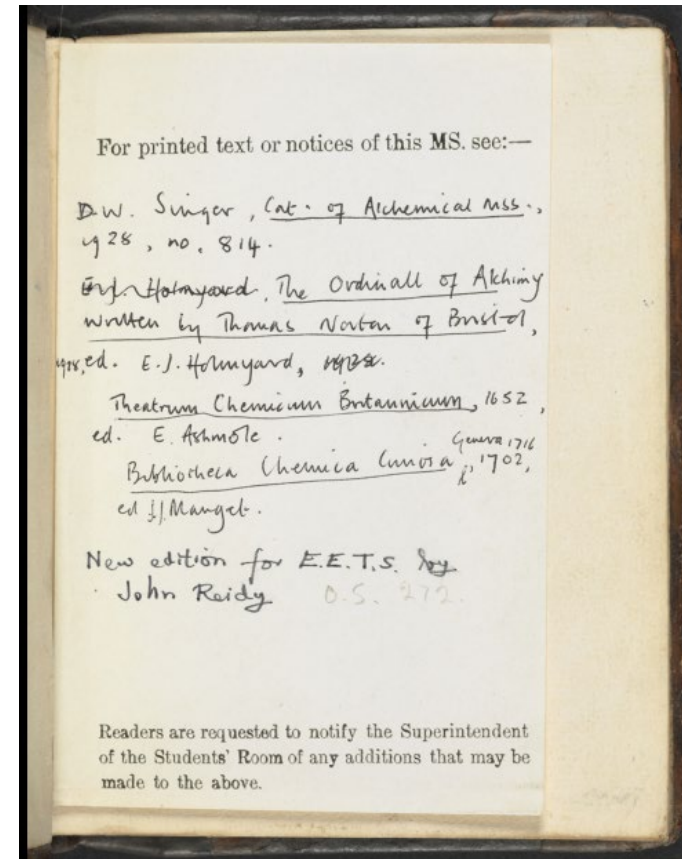
Comparing predicti...

☒ Mismatch

☐ Same

Flexible and understanding

- One of many tools we use as part of an adaptable workflow
- Manual intervention is still required but this is desirable
- Building trust and confidence in machine learning



add_ms_10302_fs001r.tif

3 | Flysheet | f. [i-r]

Thank you

Catherine.Cronin@bl.uk

Daniel.Van-Strien@bl.uk

Andrew.Longworth@bl.uk

