



How did I GET that data?:

reproducing steps

Iain Emsley

University of Oxford

Oxford e-Research Centre

www.oerc.ox.ac.uk



Introduction

Irreproducibility

Digging into EEBO

Storing the Work

Citation

Collaborate with your future self

Principles of Irreproducibility

- Think “Big” Picture
- Be abstract
- Short and sweet
- The deficit model
- Don’t share

Above all:

Make sure that you **cannot** reproduce your own work

(Hong, N.P.C., et al, 2015)

Digging into EEBO metadata (Part 1)

- Started in Excel
 - Downloaded the CSV from TCP's Github repository

A1 ✕ ✓ *fx* TCP

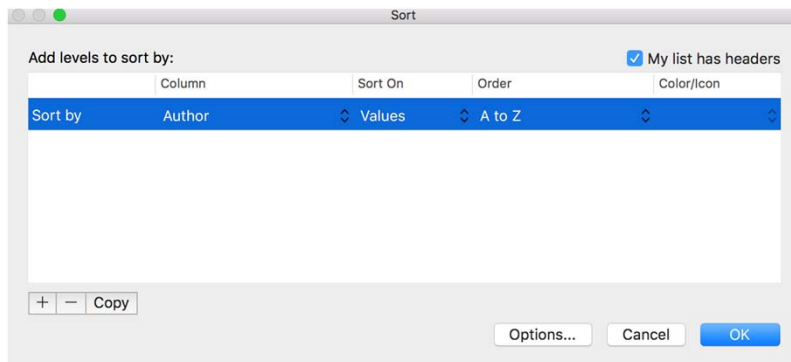
	A	B	C	D	E	F	G	H	I	J
1	TCP	EEBO	VID	STC	Status	Author	Date	Title	Terms	Pages
2	A00002	99850634	15849	STC 1000.5; I	Free	Aylett, Robert, 1583-1655?	1625	The brides ornaments viz. fiue meditations, morall and diuine. 1. Knowledge, 2. zeale, 3. temperance, 4. b		134
3	A00005	99842408	7058	STC 10000; E	Free	Higden, Ranulf, d. 1364. Polycroni	1515	Here begynneth a shorte and abreue table on th	Great Britain -- History -- To 1485 -- Early works to 1800.;	302
4	A00007	99844302	9101	STC 10002; E	Free	Higden, Ranulf, d. 1364. Polycroni	1528	The Cronycles of Englonde with the dedes of po	Great Britain -- History -- To 1485 -- Early works to 1800.;	386
5	A00008	99848896	14017	STC 10003; E	Free	Wood, William, fl. 1623, attribut	1623	Considerations vpon the treaty of marriage betw	Great Britain -- Foreign relations -- Spain.	14
6	A00011	99837000	1304	STC 10008; E	Free		1640	Englands complaint to Iesus Christ, against the	Church of England. -- Thirty-nine Articles -- Controversial	54
7	A00012	99853871	19269	STC 1001; ES	Free	Aylett, Robert, 1583-1655?	1623	Ioseph, or, Pharoah's fauourite; Joseph.	Joseph -- (Son of Jacob) -- Early works to 1800.	99
8	A00014	33143147	28259	STC 10011.6; F	Free		1624	Greate Brittaines noble and worthy counsell of	England and Wales. -- Privy Council -- Portraits.; Great Bri	1
9	A00015	99837006	1310	STC 10011; E	Free	Jones, William, of Usk.	1607	Gods vvarning to his people of England By the g	Floods -- Wales -- Early works to 1800.	16
10	A00018	99850740	15965	STC 10015; E	Free		1558	The lame[n]tacion of England; Lamentacion of E	Great Britain -- History -- Mary I, 1553-1558 -- Early work	26

Digging into EEBO metadata (Part 1)

- Started in Excel

A1

	A	B	C	D	E	F	G	H	I	J
1	TCP	EEBO	VID	STC	Status	Author	Date	Title	Terms	Pages
2	A00002	99850634	15849	STC 1000.5; I Free		Aylett, Robert, 1583-1655?	1625	The brides ornaments viz. flue meditations, morall and diuine. 1. Knowledge, 2. zeale, 3. temperance, 4. b		134
3	A00005	99842408	7058	STC 10000; E Free		Higden, Ranulf, d. 1364. Polycroni	1515	Here begynne th a shorte and abreue table on th Great Britain -- History -- To 1485 -- Early works to 1800.;		302
4	A00007	99844302	9101	STC 10002; E Free		Higden, Ranulf, d. 1364. Polycroni	1528	The Cronycles of Englonde with the dedes of po Great Britain -- History -- To 1485 -- Early works to 1800.;		386
5	A00008	99848896	14017	STC 10003; E Free		Wood, William, fl. 1623, attribut	1623	Considerations vpon the treaty of marriage betv Great Britain -- Foreign relations -- Spain.		14
6	A00011	99837000	1304	STC 10008; E Free			1640	Englands complaint to Iesus Christ, against the Church of England. -- Thirty-nine Articles -- Controversial		54
7	A00012	99853871	19269	STC 1001; ES Free		Aylett, Robert, 1583-1655?	1623	Ioseph, or, Pharaoh's faourite; Joseph. Joseph -- (Son of Jacob) -- Early works to 1800.		99
8	A00014	33143147	28259	STC 10011.6; Free			1624	Greate Brittaines noble and worthy councill of England and Wales. -- Privy Council -- Portraits.; Great Bri		1
9	A00015	99837006	1310	STC 10011; E Free		Jones, William, of Usk.	1607	Gods warning to his people of England By the gi Floods -- Wales -- Early works to 1800.		16
10	A00018	99850740	15965	STC 10015; E Free			1558	The lame[n]tacion of England; Lamentacion of E Great Britain -- History -- Mary I, 1553-1558 -- Early work		26



Digging into EEBO metadata (Part 2)

- Used Python to automate
 - What data did I use?
 - Retrieving data from URLs
 - <https://api.twitter.com/1.1/search/tweets.json>
 - <https://raw.githubusercontent.com/textcreationpartnership/A29026/master/A29026.xml>
 - Parsing the XML to find the publishing place
 - Find publicationStmt/pubPlace
 - Clean up some punctuation
 - CSV writing

	A	B	C	D	E	F	G	H	I	J	K
963	A00200	9985188	1833	STC 12197; E Free		vennecoo, jonanni van de, 1577-1645.	1602	A true discourse of all the sayes which were,	Francis	10	London
964	A00271	99851846	17138	STC 1038; ES Free		Verstegan, Richard, ca. 1550-1640.	1589	The copy of a letter lately wvritten by a S ^r	Great Britain	40	Antwerp
965	A01945	99841482	6068	STC 12091; E Free		Watson, Henry, fl. 1500-1518.; Fouquart, de Can	1510	The gosselles of dystaues; Evangelles des	Women -- H	63	London
966	A00275	99847131	12149	STC 10389; E Free		Whittingham, William, d. 1579, attrib. name.	1566	To my faythfull bretheren now affycted ..	Church vestn	58	Emden
967	A02379	99846673	11658	STC 12510.5; Free		William, of Ockham, ca. 1285-ca. 1349, attribute	1531	Disputatio inter clericum et miletem sup ^r	Church and s	29	Londini
968	A02381	99845542	10447	STC 12511A; Free		William, of Ockham, ca. 1285-ca. 1349, attribute	1533	A dialogue betwene a knyght and a clerki	Church and s	52	London
969	A00008	99848896	14017	STC 10003; E Free		Wood, William, fl. 1623, attributed name.	1623	Considerations vpon the treaty of marria	Great Britain	14	The Netherlands?
970	A00018	99850740	15965	STC 10015; E Free			1558	The lame[n]tacion of England; Lamentaci	Great Britain	26	Germany?
971	A02065	99853191	18560	STC 12207; E Free			1562	The true reporte of the forme and shape	Monsters -- f	2	London
972	A01878	99853699	19093	STC 12019; E Free			1569	A neue ballade intytuled, Good fellows	Ballads, Engl	2	London
973	A00489	99842419	7069	STC 10627; E Free			1580	A good exhortation to euery man what	he should doo	1	London
974	A02495	99842463	7116	STC 12626A; Free			1599	The principal nauigations, voyages,	traffi Voyages and	2072	London
975	A00026	29900271	28115	STC 10021.7; Free			1600	A Table of the cheiffest citties, and towni	Great Britain	1	London
976	A00923	99853346	18727	STC 11029; E Free			1600	Newes from Flaunders A new ballad of th	Nieuport, Ba	1	London
977	A02257	99853769	19164	STC 12392; E Free			1600	The groome-porters lawes at Mawe, to b	Card games	1	London
978	A01429	29900252	28108	STC 11548.5; Free			1604	Novum et solenne decretum ACEAS ^o Syno	Synod of Gar	1	S.I.
979	A02673	99839523	3952	STC 12805; E Free			1609	The liues, apprehensions, arraignments,	i Pirates -- Gre	62	London
980	A00476	99857028	22695	STC 10610; E Free			1610	A most notable and worthy example of	an vngratious s	1	London
981	A01558	99838709	3097	STC 11687; E Free			1612	The life and death of Lewvis Gaufredy a p	Gaufridi, Lou	38	London
982	A00025	29905398	28125	STC 10020.5; Free			1613	The Royall line of kings, queenes, and	pri Elizabeth, -- f	1	London?
983	A01001	99838148	2511	STC 11107; E Free			1619	The wonderful discourie of the vvitche	Flower, Marg	48	London
984	A00287	33143102	28231	STC 10406.5; Free			1621	The effect of a bill exhibited in Parliam	England and	1	London

Store it



Carpet Bag

By Sobebunny - Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=7274913>

Version Control

sourceforge

GitHub

 Bitbucket



Version Control



Version Data

Your own institution's data repository



TELLME!

- Erm, it's Humanities isn't it?
 - Digital is one part...
 - Are your reasons, inferences and assumptions visible?

- Write a narrative

- README
- Blog post, poster, paper.

- Literate programming

- Jupyter for Python
- Knitr for R

Event Attendance ¶

Searching for events in the calendar that the Institute has attended.

Meeting can be used for both project meetings and internal institute meetings, so this is an area that needs further work.

```
#find types of meetings attended (search for meeting, conference, meet)
confs = ["conference", "(?<!SSI) event", "Workshop(?:\s-)", "Symposium", \
        '(?<=ScalaLife)|(?<=(NGS|1|L)ab)|Pis|PDG|SI2|BSC)|(?<=CoSec)|(?<=ELIXIR)|(?<=ParaFEM)|(?<=INFRASUPP)|(?<=Infrasupp-4)|(?<=Data_Review)|(?<=IPCLC) meeting']

_create_graph(caldata_df, '|'.join(confs), "line", "Overall meetings attended by year")

#for conf in confs:
_getpeople(caldata_df, '|'.join(confs))
```



License

- Code and data should have a license file
 - LICENSE or LICENSE.txt
- DON'T try and write your own
- Use a service
- Does your institution or group recommend one?

Enable Citation

Allow others to cite your software!

Add a citation file to your code to show how you want to be cited.

Provide a Digital Object Identifier for your code and data.

Figshare and Zenodo offer ways of doing this.

If a paper uses your software, or part of it, without credit, contact the editor

Usher, 2017

Cite your software sources

Cite software or data in your research or paper

- Describe software that provides a critical part of, or unique contribution to, your research
- There are many ways of describing software: footnotes, acknowledgements, references, and method sections
- Be aware that the software licence may require you attribute the software
- Cite software papers as an aid not a replacement to software or data
- Recommended citations may need extra detail
- If the software has a DOI, use it.

(Jackson, 2016)

Two final questions

- Can someone in your group reproduce one of your results using available information on the web or in papers?
- Can you do this for their results?

Thank you for listening

Are there any questions?

@iainemsley

iain.emsley@oerc.ox.ac.uk

Acknowledgments

Gabriel Egan, Centre for Textual Studies, De Montfort University

Laura Fortunato, Kesson Magid, Philip Fowler, Pip Willcox, Martin Hadley, University of Oxford, Reproducible Research Oxford, <https://rroxford.github.io/>

David De Roure, Oxford e-Research Centre

Software Sustainability Institute

References

- Baker, Monya. “1,500 Scientists Lift the Lid on Reproducibility.” *Nature News* 533, no. 7604 (May 26, 2016): 452. doi:10.1038/533452a.
- Hong, Neil P. Chue, Tom Crick, Ian P. Gent, Lars Kotthoff, and Kenji Takeda. “Top Tips to Make Your Research Irreproducible.” *ArXiv:1504.00062 [Cs]*, March 31, 2015. <http://arxiv.org/abs/1504.00062>.
- Huang, Daisie and Gonzalez, Ivan (eds): "Software Carpentry: Version Control with Git." Version 2016.06, June 2016, <https://github.com/swcarpentry/git-novice>, 10.5281/zenodo.57467
- Jackson, Mike, How to cite and describe the software you used in your research - top ten tips, <https://www.software.ac.uk/blog/2016-10-07-how-cite-and-describe-software-you-used-your-research-top-ten-tips>
- Lowndes, J.S.S. *et al.* Our path to better science in less time using open data science tools. *Nat. Ecol. Evol.* 1, 0160 (2017), doi:10.1038/s41559-017-0160
- Morin Andrew, Urban Jennifer, Sliz Piotr , A Quick Guide to Software Licensing for the Scientist-Programmer, PLOS Computation Biology, <https://doi.org/10.1371/journal.pcbi.1002598>
- Piccolo, Stephen R., and Michael B. Frampton. “Tools and Techniques for Computational Reproducibility.” *GigaScience* 5, no. 1 (December 1, 2016): 1–13. doi:10.1186/s13742-016-0135-4.
- Usher, Will, How Not to Cite Software (and how to be cited), <https://www.software.ac.uk/blog/2017-03-30-how-not-cite-software-and-how-be-cited>

Useful Websites

- Software Carpentry, <https://software-carpentry.org/>
- Data Carpentry, <http://www.datacarpentry.org/>
- Software Sustainability Institute, <https://www.software.ac.uk/>