



# Documenting Long-Form Data

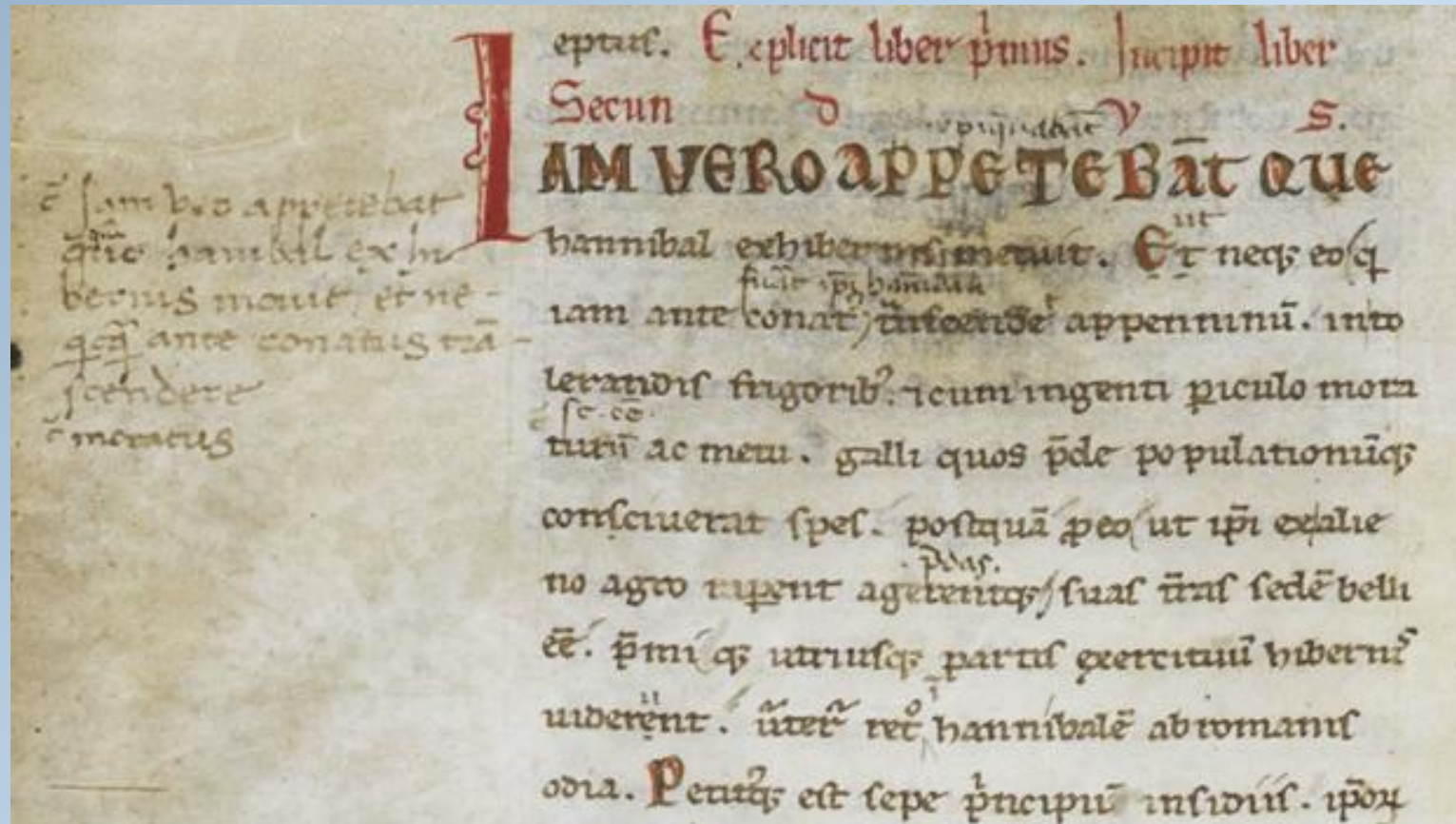
M. H. Beals

*Loughborough University & The Software Sustainability Institute*

# Understanding the Long Form

- What does it include?
  - Newspapers, Letters, Diaries, Memorials, Broadsides, and so on
- What should you record?
  - Textual Content
  - Paratext, Marginalia
  - Layout, Typography, Colour
  - Contextual Notes
- How much should you record?
  - Transcription
  - Recreation
  - Summary

# What Do I Record?





# Brain Storm Session

What sort of materials do you use ?  
What do you need to document ?



# Getting Meta About Data

- Bibliographic Details
  - Author, Date(s), Title, Publisher, Printer, Language, Subject
- Physical Details
  - Dimensions, Weight, Media Type
- Holder Details
  - Location, Shelf Number, Access Restrictions
- Rights Details
  - Copyright, Photographic Reproduction, Reuse in Publications
- Provenance
  - Version of Consultation, Method of Digitisation

# Brain Storm Session

How Does Your Workflow  
Incorporate Metadata ?



# Format Wars

- Word (.docx)  
For When You Like Giving Bill Gates Money
  - ßÔ <À#å\*– ÎKÛîÁ[•¿|Ã2)¶ 5-\\ãéþæ1ß...r6€Å-|©ã &£
- Plain Text (.txt)  
For When Text is Enough
  - Suddenly, a shot rang out.
- Markdown (.md)  
For When Basic Formatting is Enough
  - *\*Suddenly\**, a shot rang out.

# Format Wars

- XML (.xml)

For When You Need to Tag Things

`<italics>`Suddenly`</italics>`, a `<noun>`shot`</noun>` `<verb>`rang out`</verb>`.

- Text-Encoding Initiative (.xml)

For When You Need to Tag Specific Things for Wider Consumption

`<p>`

`<hi ref="italics">`Suddenly`</hi>`, a shot rang out. `<lb/>`

`</p>`



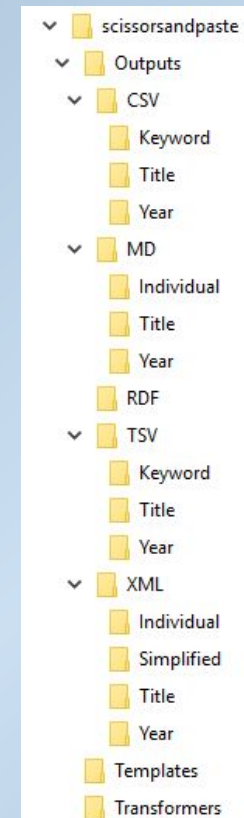
# Practical Session

Finding a Balance with  
Historical Newspapers



# What's in a Name ?

- Directory Organisation
  - Type
  - Purpose
  - Provenance
  - Category
  - Date
- Filenames
  - Short or Full Title of Document
  - Spaces or Punctuation
  - Date of Composition / Revision
  - Author(s) and Revisers
  - Category or Purpose



Name

- 1800.01.02\_CaledonianMercury\_0001
- 1800.01.02\_CaledonianMercury\_0002
- 1800.01.02\_CaledonianMercury\_0003
- 1800.01.02\_CaledonianMercury\_0004
- 1800.01.02\_TheDerbyMercury\_0001
- 1800.01.02\_TheDerbyMercury\_0002
- 1800.01.02\_TheDerbyMercury\_0003
- 1800.01.02\_TheDerbyMercury\_0004
- 1800.01.02\_Trewman'sExeterFlyingPost\_0001
- 1800.01.02\_Trewman'sExeterFlyingPost\_0002
- 1800.01.02\_Trewman'sExeterFlyingPost\_0003
- 1800.01.02\_Trewman'sExeterFlyingPost\_0004
- 1800.01.04\_CaledonianMercury\_0001
- 1800.01.04\_CaledonianMercury\_0002
- 1800.01.04\_CaledonianMercury\_0003
- 1800.01.04\_CaledonianMercury\_0004
- 1800.01.04\_Jackson'sOxfordJournal\_0001
- 1800.01.04\_Jackson'sOxfordJournal\_0002
- 1800.01.04\_Jackson'sOxfordJournal\_0003
- 1800.01.04\_Jackson'sOxfordJournal\_0004
- 1800.01.04\_ThelpswichJournal\_0001
- 1800.01.04\_ThelpswichJournal\_0002

# Brain Storm Session

How Do you Store, Organise  
& Name Your Files?





# Versioning

- Documenting:
  - Date of Creation
  - Date of Revision
  - Changes Made
  - Rationale for Changes
- Naming Conventions
- Software Tracking

2916	2916	1801.01.06_MorningChronicle_0004,1800.12.27_JacksonsOxfordJournal_0002,Advert,Medical,DrJamesAnaleptic
2917	2917	1801.01.06_TheHullPacket_0001,1800.12.25_TheDerbyMercury_0001,Advert,Publication,MoralManagement
2918	-	1801.01.08_TheDerbyMercury_0003,1800.12.20_TheIpswichJournal_0004,,,
	2918	+ 1801.01.08_TheDerbyMercury_0003,1800.12.20_TheIpswichJournal_0004,Domestic,Charity,MongewellPoor
2919	2919	1801.01.08_TrewmansExeterFlyingPost_0002,1800.12.27_JacksonsOxfordJournal_0002,,,
2920	2920	1801.01.08_TrewmansExeterFlyingPost_0003,1800.11.10_TheAberdeenJournal_0003,,,
2899	2899	1801.01.03_JacksonsOxfordJournal_0001,1800.12.20_TheIpswichJournal_0001,Adverts,Insurance,SunFireOffice
2900	-	1801.01.03_JacksonsOxfordJournal_0004,1800.12.29_PortsmouthTelegraphorMottleysNaval andMilitaryJournal_0003,,,
	2900	+ 1801.01.03_JacksonsOxfordJournal_0004,1800.12.29_PortsmouthTelegraphorMottleysNaval andMilitaryJournal_0003,International,Naval,ShipSirius
2901	2901	1801.01.05_TheAberdeenJournal_0001,1800.12.13_TheIpswichJournal_0001,,,
2902	2902	1801.01.05_TheAberdeenJournal_0001,1800.12.16_TheHullPacket_0001,,,
		ups,NewnwickSociety
2913	2913	1801.01.05_TheAberdeenJournal_0003,1800.12.29_CaledonianMercury_0002,Domestic,Charity,StockMarshallDonations
2914	-	1801.01.06_TheHullPacket_0001,1800.12.25_TheDerbyMercury_0001,,,
2915	2914	1801.01.06_MorningChronicle_0004,1800.11.29_JacksonsOxfordJournal_0003,Advert,Medical,DrJamesAnaleptic
2916	2915	1801.01.06_MorningChronicle_0004,1800.12.13_JacksonsOxfordJournal_0001,Advert,Medical,DrJamesAnaleptic
2917	2916	1801.01.06_MorningChronicle_0004,1800.12.27_JacksonsOxfordJournal_0002,Advert,Medical,DrJamesAnaleptic
	2917	+ 1801.01.06_TheHullPacket_0001,1800.12.25_TheDerbyMercury_0001,Advert,Publication,MoralManagement
2918	2918	1801.01.08_TheDerbyMercury_0003,1800.12.20_TheIpswichJournal_0004,,,



# Practical Session

Documenting Change with  
Github Desktop



# Online Documentation

- Creating a Private Repository on Github (or other sites)
  - Permanent (\*caveats\*) location for data
  - Easy backup and retrieval
  - Full versioning record
  - Ability to release datasets publically
- Creating a Project Workspace on Open Science Framework
  - Store all your files in one place
  - Collaborative workspace for different elements of project
  - Share what you want, when you want

# Practical Session

Maintaining your Data with Github

