

A
M
A
Z
I
N
G



spreadsheet

100%

madness for tei/cbml

airplane

train

MILLIONS OF USERS, TWO LIBRARIANS AND NO TIME!

communication-tech

a structured note field for comm telephone

THE INCREDIBLE STORY OF HOW A JMU COURSE OVERCAME THE IMPOSSIBLE AND CHANGED THE WORLD...

gender-markers

a structured note field for (binary) #cleavage

#promotion
#barefoot
#barbershop
#pride
#bride
#engagement
#necklace

EASY
no CODING!
no <>> ⚡

A four-part description of how to “compile” visual and textual comic book data into spreadsheet templates that can be turned into a TEI/CBML XML document.

Goal of this guide: Enable the student to “encode” the data necessary to create a TEI/CBML XML file using Google Sheet spreadsheet templates.

Required tools: web browser, Google Sheets.

The Text Encoding Initiative, and its Comic Book Markup Language extension, are natively written in XML (eXtensible Markup Language) format, a W3C standard from those wonderful folks who gave us the Internet. XML is a powerful and efficient language for database construction, and the TEI standard is incredibly flexible, but that power and flexibility comes with a steep learning curve. Even with a “schema-aware” XML editor, creating a TEI file from scratch requires a substantial familiarity with the TEI standard for ready mastery.

There is another way....

Did you know that almost any metadata standard can be encoded using spreadsheet technology? Nothing but tables made up of columns and rows! In order to maximize the time you spend on comic book analysis in this course, we have developed four spreadsheets that, with some agile post-processing, can be turned into TEI/CBML XML documents. Neat, eh?

BUT: while it is not strictly necessary that you know anything about XML to populate these spreadsheets, you MUST have a sound working knowledge of the building blocks of conventional comic books: stand-alone stories, title pages, pages of panels, individual panels, panel transition types (think Scott McCloud) speech balloons, captions, character names. In order to successfully fill out these spreadsheets, you will have to “nest” information correctly and use controlled vocabularies for comic books characters, as well as other elements that you will subject to analysis using other tools.

SPREADSHEET MADNESS PART 1

Open the TEI_Data spreadsheet. Navigate to the PL17_1_Panel_Coding worksheet. Your initial data entry will use 9 column headings:

panelGrp_Number	[Column A]
panel_Number	[Column B]
panel_Characters	[Column D]
panelNote_Desc	[Column E]
panelNote_Resp	[Column F]
balloon_Type	[Column I]
balloon_Who	[Column J]
balloon_Text	[Column K]
caption_Text	[Column L]

There are other column headers in the worksheet -- ignore them for now!

NICE OF YOU TO LOAN BETTY MONEY TO GET STARTED IN BUSINESS, SANDRA. HOPE SHE CAN COOK...												
A	B	C	D	E	F	G	H	I	J	K	L	
panelGrp_Num	panel_Number	panel_Transitic	panel_Charact	panelNote_Des	PanelNote_Res	PanelNote_Ger	Gender_marke	balloon_Type	balloon_Who	balloon_Text	caption_	
1.1												
1.1.1	#phantom_lady	# Phantom Lady,	#SWH	#cleavage	#pro	phantom_lady					WHO WO	
				#fedora	#suit	anonymous_crimal						
1.1.2	#subject-to-sub	#sandra_knight	Sandra Knight	a #SWH	#dress	#long-ha	sandra_knight	#dialogue	#don_borden	NICE OF YOU T	SANDRA	
				#suit	#short-ha	don_borden	#dialogue	#sandra_knight	SHE'S AN OLD FRIEND, DO			
1.1.3	#subject-to-sub	#betty	#anonymo	In a circular par	#SWH	#dress	#apron	# betty	#dialogue	#anonymous_c	WE WARNED Y BUT INSIDE	
						#necktie	#fedor	anonymous_cri	#dialogue	#betty	I CAN'T PAY FOR PROTECT	
									#dialogue	#anonymous_c	OH, YOU CAN'T EH?	
2.1												
2.1.1	#phantom_lady	Phantom Lady,	#SWH	#cleavage	#pro	phantom_lady					PHANTOM	
				#uniform-jocke	dishonest_jockey							
2.1.2	#subject-to-sub	#don_borden	Headshot of Do	#SWH	#suit	#necktie	# don_borden	#dialogue	#don_borden	CAN'T TELL YOU SANDRA		
2.1.3	#subject-to-sub	#sandra_knight	Sandra Knight,	#SWH	#long-hair	#lipstick	#blouse	#sk	#dialogue	#sandra_knight	A SURPRISE, EH? BETTER	
								#thought	#sandra_knight	AND SO WILL THE PHANT		
2.1.4	#subject-to-sub	#sandra_knight	Full length shot	#SWH	#long-hair	#dre	sandra_knight	#monologue	#sandra_knight	WELL, HERE I A	THAT AFT	

SPREADSHEETMADNESS PART 1

The panel numbering system we have adopted for this course records the story number, panel group number, and the individual panel so, for instance, 1.1.1, translates to story number one, panel Group Number 1, and panel number 1. In most cases, each panel group corresponds to a single page of comic book panels. Each Phantom Lady #17 story (there are three) consists of ten pages. The number of panels per page varies, but is usually 7-9. So, PL#17/1 runs from panel group 1.1 to 1.10. Think of the individual panels as “children” of their “parent” panel group, and each panel group as a “child” of its story “parent”. You will need to “nest” the panel numbers correctly under their respective panel group numbers for the eventual XML conversion to function correctly. “panel_Characters” lists all of the characters who figure in the panel, whether or not they have individual speech balloons -- these names should be used consistently, another controlled vocabulary. Enter the names in a single spreadsheet cell, separated by a whitespace, NOT a comma.

“Panel Note: Description” (panelNote_Desc) is a brief description, in your own words, of the key events in each panel: actions, characters, objects, significant design elements. Depending on your research focus, writing the panel description can help you grapple with -- character development, classes of objects, scene transitions -- you name it. It can also help other researchers to better understand both the comic book and your approach to unpacking it. “panelNote_Resp” should be the initials of the person responsible for writing the panel description.

SPREADSHEETMADNESS PART 1

Speech balloon conventions are observed by most, but not all, comic books. Dialogue and auditory speech is usually set in rounded balloons, while inner dialogue is conventionally lodged in “pillowy” balloons. “balloon_Type” values should be a limited controlled vocabulary that is recognizable by other researchers. “balloon_Who” is the balloon character “owner” -- these names should precisely correspond in format to the “panel_Characters”. “balloon_Text” is the raw text in the balloon, transcribed exactly as found. Most comic book balloons are written in all caps -- that is how they should be transcribed.

“caption_Text” is where you will transcribe the panel captions, those extra bits of narrative that the comic book writers add to the other textual and visual elements. They too should be transcribed exactly as found.

Your assignment: PL17/1, panel groups three and four (1.3, 1.4). The tiny page numbers, bottom right, are 2 and 3, respectively.

Teamwork! We recommend you split up the chore of data-entry, panel description and so forth, and then proof each other’s work.

SPREADSHEETMADNESS PART 2

A successful digital humanities project, using a standard as malleable as TEI, rises or falls on the quality of the tools you engineer for analysis. Complex artifacts like comic books can be explored in countless ways, BUT: your wheels will spin in the sand unless you are lucidly clear upfront about your research objectives. Are you interested in visual diversity modeling? Then you will need to devise a controlled vocabulary of diversity markers and use them consistently throughout the document markup.

For this exercise, we ask that you concentrate on gender marker conventions in the character development of PL#17. Open the TEI_Data Google sheet and navigate to the “PL17_InterpGrp” worksheet. Column C, “groupElements,” nested under the Column A entry “gender-markers,” has a few entries, but you will need others to satisfactorily capture the in-your-face binary gender markers that Matt Baker used for this comic. Please use only lower-case letters, begin each term with a hash-mark, connect separate words with underscores, and be careful of the nesting for each list.

SPREADSHEETMADNESS PART 2

a structured note field for (binary) gender marking

A	B	C	D
groupName	groupDesc	groupElements	
transportation-tech	a structured note field for transportation	automobile bus hero_vehicle airplane train	
communication-tech	a structured note field for communication	telephone radio microphone	
gender-markers	a structured note field for (binary) gender markers	#cleavage #prominent-breasts #bare-legs #bare-arms #prominent-eyelashes #bracelet #earring #necklace #dress #blouse #skirt #dress-gloves #mink-coat #lipstick #dress-hat #nail-stockings	

SPREADSHEETMADNESS PART 2

Near the bottom of Column A, “panel-transitions” uses the panel transition typology developed by Scott McCloud in *Understanding Comics*, pp. 70-74. The section above it, “speech” lists six terms for tagging the “types” of speech balloons. “Transportation-tech” “communication-tech” “weapon-tech” and “uniforms” identify other classes of visual objects that may or may not be of interest to you, research-wise. The flexibility of TEI allows you to create as many such note-classes as you need. Using a schema-aware XML editor and laboring directly at the coal-face, as it were, the editor will conveniently autofill the XML elements and attributes with the controlled vocabulary items. Using a spreadsheet, you will have to copy and paste the values where they belong.

Be careful to scrupulously follow the patterns you find in the “PL17_InterpGrp” worksheet -- it will be converted into an important section of the TEI/CBML document.

Before you can enter these values into the PL17_1_Panel_Coding workbook, however, you will need to establish fixed values for the names of the characters. Open the “Dramatis_Personae” worksheet. The TEI standard permits you to enter as much information about each character as you require. We have added values for most of the major characters in PL#17 -- Phantom Lady, Sandra Knight, Betty, Don Borden, with key information about their relationships and occupations. It is necessary to create a standardized ID (Column A) for each character of interest -- that should take the same format as the gender markers: lower-case letters, given name and family name connected with underscores, with an initial hash-mark. Here is where clarity about your research project is paramount!

SPREADSHEETMADNESS PART 2

For example: the comic book illustrator is pretty consistent in delineating the various criminal gang members, even if the writer does not give them individual names. Yes, you can tag them individually -- but does that matter for your analysis? Is there enough speech content for the minor criminals to warrant giving them individual identities? For my analysis, no, so I have labelled them "#anonymous_criminal" whenever they appear in a given panel. It is entirely up to you -- the TEI standard will support all the character atomization that heart could desire. We have left off the name of the criminal gang leader -- add it to the list, with similar contextual description. Pay attention to the nesting in the "Occupation" heading for #phantom_lady, #sandra_knight, #dishonest_jockey (PL#17/2) and #frankie (PL#17/3). In order to run one of the analysis tools we have developed, you will have to enter either 1 or 2 under the "Sex_Value" heading for the binary sex identifications.

crime-fighting hero										
A	B	C	D	E	F	G	H	I	J	
ID	Name	Additional_Narr Note		Affiliation	Occupation	Role	Sex	Sex_Value	Nationality	
#phantom_lady	Phantom Lady	Sandra Knight			crime-fighting	hero	female	2	American	
					socialite	human				
#sandra_knight	Sandra Knight	Phantom Lady			crime-fighting	hero	female	2	American	
					socialite	human				
#don_borden	Don Borden				boy-friend of Sandra Knight		male	1	American	
#anonymous_criminal					criminal		male			

Now go back to the "PL17_1_Panel_Coding" worksheet. Column G, "PanelNote_Gender_markers" is where you will use the controlled vocabulary you developed in Column C of the "PL17_InterGrp" worksheet. Use as many of the gender marker descriptors for each character as you need, separating each by a whitespace. Column H, "Gender_markers_Who" is where you will enter the ID of the characters whose gender marker you are recording. Unfortunately, here only when you enter the IDs into column H, do NOT precede them by a hash-mark. When you use these IDs for the "panel_Characters" and "balloon_Who" columns, do use the initial hash-mark. This is a XML convention for which there is no workaround, one of the genuine drawbacks of using spreadsheets to record XML values instead of an XML editor.

SPREADSHEET MADNESS PART 2

panelGrp_Num	panel_Number	panel_Transitio	panel_Characte	panelNote_Des	PanelNote_Res	PanelNote_Ger	Gender_marke	balloon_Type	balloon_W
1.1	1.1.1	#phantom_lady	# Phantom Lady, #SWH	#cleavage #pro	phantom_lady	#fedora #suit	anonymous_criminal		
	1.1.2	#subject-to-sub	#sandra_knight Sandra Knight a #SWH	#dress #long-ha	sandra_knight	#dialogue	#don_borde		
	1.1.3	#subject-to-sub	#betty #anonymmc In a circular par #SWH	#suit #short-ha	don_borden	#dialogue	#sandra_kn		
				#dress #apron	# betty	#dialogue	#anonymou		
				#necktie #fedor	anonymous_cri	#dialogue	#betty		
						#dialogue	#anonymou		
2.1	2.1.1	#phantom_lady	Phantom Lady, #SWH	#cleavage #pro	phantom_lady	#uniform-jocke	dishonest_jockey		
	2.1.2	#subject-to-sub	#don_borden Headshot of Do #SWH	#suit #necktie	# don_borden	#dialogue	#don_borde		
	2.1.3	#subject-to-sub	#sandra_knight Sandra Knight, : #SWH	#long-hair #lipstick	#blouse #sk	#dialogue	#sandra_kn		
	2.1.4	#subject-to-sub	#sandra_knight Full length shot #SWH	#long-hair #dre	sandra_knight	#monologue	#sandra_kn		

This is also an opportune moment to make the values you inserted for the character names under the “panel_Characters” and “balloon_Who” headings match the “Dramatis_Personae”, hash-marks and all.

SPREADSHEETMADNESS PART 3

Using Google Cloud Vision API demo, try OCRing panel groups 1.5 and 1.6 -- follow the directions in the course guide to OCR using Google Cloud Vision API demo. Any luck? You may have to "double-key" (manually transcribe) most of the text, a real drag. I used a two-column Word table for each panel group, with the panel number and speaker in the first, and the transcribed text in the right. Why Word and not Google Docs? Because Word preferences, unlike Google Docs, can turn off smart (curly) double and single quotes, and standardized ellipses (...) all of which will complicate the extraction and analysis of the text further on. Example:

2.1.1·CAP: [»]	PHANTOM·LADY·HAD·LITTLE· EXPERIENCE·WTH·BANGTAILS· BUT·PLENTY·WITH·MURDER· YET·IT·TOOK·FAST·THINKING· AND·ACTING·TO·KEEP·UP· WITH·THE·ADVENTURE·OF· "THE·STINGING·WHIP!" [»]
2.1.2·CAP [»]	SANDRA·KNIGHT'S· BOY·FRIEND,·DON· BORDEN,·IS·BUBBLING· OVER·WITH· SURPRISE... [»]
2.1.2·#don_borden [»]	CAN'T·TELL·YOU·NOW,· SANDRA,·BUT·MEET· ME·AT·THE·GRAND· FAIRGROUNDS·TRACK· BEFORE·THE·FIRST· RACE·THIS· AFTERNOON! [»]
2.1.3·#sandra_knight [»]	A·SURPRISE,·EH?· BETTER·BE·GOOD!·I'LL· BE·THERE... [»]

SPREADSHEETMADNESS PART 4

Before you complete this assignment, please add all of the missing column heading values to PL#17/1, panel groups 1.3-1.6. That includes Column C, "panel_Transition" following Scott McCloud's typology (do the best you can!), and Columns M-P, "uniform_Type" "weapon_Type" "transportation_Type" and "communication_Type". Be sure that the values for Columns C-D and G-J correspond exactly with the values in the "PL17_InterGrp" and "Dramatis_Personae" worksheets, remembering to leave off the initial hash-mark on the character IDs for Column H, "Gender_markers_Who".

TEIHeader: All TEI documents must consist of two parts, the TEIHeader and Text. The former records the creator of the TEI document, bibliographic details about the original documents, and editorial decisions about how the TEI markup is rendered. "Text" includes lists of characters ("Dramatis_Personae" or listPerson in TEI-speak), controlled vocabularies (interGrp blocks), and the text of the document itself, which in our case is the panel groups, organized by TEI "div" elements.

The "TEIHeader" worksheet is an oversimplified smattering of key elements required for the TEIHeader section. The headings should be self-explanatory -- change the values to fit your team.

Phantom Lady, Vol. 17, April 1948							
A	B	C	D	E	F	G	H
comic_Title	responsName	responsID	publisher1	date1	publisher2	date2	comicURL
Phantom Lady,	Steven W. Hollo	SWH	James Madison	2017	Digital Comic N	2010	http://digitalco

Before submitting the "PL17_InterGrp" and "Dramatis_Personae" for processing into an XML document, there is a final simple step. On their respective worksheet tabs on the TEI_Data Google Sheet, right-click "duplicate" and rename the duplicated tabs "PL17_InterGrp_Export" and "Dramatis_Personae_Export". In each of the worksheets, using the Edit=>find and replace function, delete all of the hash-marks.

