Real-Time Thoughts on the DH Process

The Corona Papers Post #6.

This post is a collection of notes taken during the exploration of various online DH materials, including a scholarly article, video games, and software tools.

notes on article

* before even reading the article I realize I’ll finally get an answer to a question I’ve had since the beginning
  + what to do, or what is being done, with digital projects that are outdated/old/in need of repair/look like they were made in 2005/etc.
  + as the put it nicely in 2nd paragraph: “problems of maintainability and sustainability”
* early theme of “as funding dies, so does the project” is rather depressing
  + and also surprising: once online, how much maintenance does each project need? I suppose it varies, especially important for projects focused on current issues
  + or maybe for projects that are very old and were built during a very different era of the internet
* interesting points in paragraph 4: sustainability work never fundamentally ends, and some projects have to be let “die”—what does that mean in this case?
* one thing I’ve found interesting across the semester is the sheer amount of growth and change of the nomenclature around DH
  + King’s Digital Lab itself reflects these changes: first “Centre for Computing and the Humanities,” then “Centre for eResearch in the Humanities”
  + part of the broader struggle in DH to find terms that truly capture the work being done
* reminder to self to check out *Aphrodisias in Late Antiquity* based on title alone
* interesting point in par. 11—older projects come from a time when “technological optimism was somewhat higher”
  + is there pessimism now? whence came the optimism? why the shift?
  + maybe in 2012 (five years before article) and pre-2012, the limitations (security, capability) on computing weren’t fully realized
* what’s WannaCry—google later
* a bit alarming that they “promptly removed” projects built using WordPress…
  + update: in par. 25 they again dismiss WordPress, calling it “simple”
  + what’s their beef with WordPress is what I want to know
* really agree with the point in par. 17 that the tech world is obsessed with newness
  + even at the detriment of sustainability; it’s all about having the “next big thing”
  + seems to fit their general point that many projects, old and new, are created with little to no regard for their use in the future or sustainability five, ten, twenty years later
* good point in par. 23 that the digital environment is no less fraught with danger than the physical one for projects
  + like the use of methods utilized in physical world for preservation of projects applied to digital
* finally get to concrete advice
  + consider end-of-funding options before the project has even begun, and build those options into the project
  + DH projects benefit from having long-term considerations built into them from the start
  + utilize operational milestones long after publication date
  + and not all projects need be permanent—as they say learning that you don’t have to worry about a project forever can be something of a relief
* they call digital curation a “new discipline”—citing an article now eleven years old
  + not exactly new anymore
* that their measures were all admittedly finite (and only two to five years, which means many of them have now, as of 2020, already passed!) seemed strange to me
  + I assumed that the projects they deemed necessary for preservation would want to be preserved, well, indefinitely
* this article could be half as long if they didn’t repeat themselves so much—I think I’ve read the word “funding” ~25,000 times
  + update: I did a quick command-f and found 43 mentions of the word “funding”
  + what this means I’m not sure
* love the typical understated British humor in par. 35— “if a tone of frustration is detected in this article”
* for sites that are “archived”—what exactly does that mean?
  + where do they go? moved to different physical hardware? hidden in the bowels of the internet? taken offline entirely? placed in remote places inaccessible to the public? it seems to be a very vague and misleading term
  + their clarification paragraph 37 does little to clarify
  + par. 38 is better—vast “range of solutions”
* final thoughts: I’m glad that KDL experienced “uniformly positive” results from their process and progress toward sustainability—but they’re from one of the most well-known and best-funded universities in the world. could their success be replicated elsewhere? on what scale? at UNL? how much funding, work, transparency, computing power, etc., would it take to ensure successful digital preservation for years to come?
  + also I appreciated that they finally gave WordPress some credit by acknowledging they moved some projects to be based on it

notes on internet archive games

* tried “Prince of Persia” for twenty minutes and then gave up because I kept either getting impaled on spikes (pretty graphically, I might add) or stabbed through the heart by a turbaned guard
  + I have little talent for video games but also there was zero instruction on how to play the game or what the controls were—which I think could greatly aided me in avoiding the impaling and stabbing thing
* pac-man also didn’t work but I scrolled through the rest of the games for a while and liked looking at the old graphics. Oregon Trail game is a classic

notes on Ms. Dussault’s lectures

different threats faced by any digital projects

* physical (hardware) limitations
  + dropping phone in lake—I actually watched someone do that this summer and it was hilarious. he spent thirty minutes trying to dive down (into a completely opaque lake) and blindly retrieve his phone using a fishing net. result: unsuccessful.
  + losing stuff
  + component failure, files themselves degrade
    - the dreaded “corrupted file”
  + things rot—this is why I don’t use CDs
  + the point about tornados killing my files is a pleasant thought. as though we don’t have enough to worry about in 2020
  + solutions: back up stuff! which I try to do in the cloud for everything I do
  + migrate files every few years—okay this I’ve never done, and I’ve never used controlled storage
* software limitations
  + my iPhone 6 that lasted for four years and barely worked during the last one can confirm this. the battery life was approximately six minutes on a full charge during the last few months before I upgraded
  + can’t open a file format
  + unable to run a script
  + website broken—I find these way more often than I used to
  + options: completely redo project in new tech or upgrade
    - informed by how something broke or whether you can still access some or none of the project (can they be salvaged?)
* provider limitations (a third party providing some service to project)
  + a website hosting your project (like my great friend WordPress, slander of which will not be tolerated)
  + unexpected expenses
  + down for maintenance (I run into this all the time)
  + company goes out of business
  + changes to laws/regulations/policies
  + malicious actors (sounds ominous)
  + retiring products or services
  + solutions: make a plan for the worst case, be aware of correspondence from provider, and keep a copy of your data
* human limitations
  + original developer is gone
  + no instructions
  + not enough knowledge to fix or redo
  + solutions: write everything down and put it in a place to be found, choose commonly used tech (WordPress!), make a plan for the future
    - I suppose this all relies on the idea that I’ll eventually abandon my own work to someone else

planning your digital project’s life

* at the beginning/early in the project:
  + what’s the minimum viable product?
  + choose or create standards (metadata, file naming conventions, where you put stuff, scanning resolution, photo quality)
  + definitely should have standardized my own conventions before I began the DH Publications for this class, kept finding myself going back to edit things that I should have standardized initially
* collaborate
  + identify collaborators (organizations, people)
  + who is responsible for what
  + how will you communicate
  + how do you credit everyone
  + boy do I love group projects
  + positions of power: careful about how everyone gets credit
* set up documentation and issue tracking
  + find location for documentation (where things are)
  + think about tech for documentation
  + who will be able to access that documentation
* choose tech carefully
  + this seems very important—in this class we’ve been told what to use but moving forward which tech/software do I want to use for different things?
  + what do you need it to do?
  + is it widely used and supported?
  + how much maintenance will it need?
  + can you easily train someone to use it?
  + will your data be stored in the system or outside of it?
  + can you get that info with plain text and non-proprietary formats?
    - the concept of proprietary formats is new to me and very interesting
    - I miss the days of mp3 files on the iPod being the only way to listen to music
* money
  + figure expenses now AND in the future
  + staff/software/website and file hosting
  + how much will that cost and how much will it cost down the road
* make plans
  + will the project ever “end”
  + what does ending mean? how will you add/edit things or fix mistakes?
  + what if you’re asked to take things down?
  + what if part of the site stops working?
  + how might you gracefully retire it?
    - all things turn to dust is very true and also very depressing

notes on Voyant

* voyant: basic text analysis
* okay let’s learn what this does
* I’m going to use my blog posts from the Corona Papers and see what voyant comes up with
* I like how they call the data my “corpus”
  + I don’t know what corpus means
  + okay I looked it up: “a collection of written texts”
* between my four blog posts, my most used word is “post”— seems anticlimactic
  + followed by “corona,” “papers,” “posted,” “log,” and “metadata”
  + probably skewed by headers/footers
  + most common words not appearing in headers/footers: “account,” “change,” “commenting” (at 12, 13, and 14 respectively)
* other stats from “summary”
  + 2,957 total words and 871 unique word forms
  + very interesting data: average words per sentence
    - highest (24.8) in “Visual Narratives” post and lowest (14.1) in “The Andrews Protocol” post
    - apparently I was feeling syntactically long-winded writing the visual narratives post
* had an insanely surreal “conversation” with Veliza, the AI software on Voyant that you can “chat” with about your documents
  + none of the conversation made sense—Veliza would certainly not pass the turing test—but it was very entertaining nonetheless
* watched in fascination as the TextArc feature created a very visually striking and ever-changing representation of the most common words
* thinking about the questions Ms. Dussault posed:
  + how a project might incorporate voyant into its workflow: I’m not sure how I would use it personally, but it seems like it would be very useful to identify trends and patterns in your own data that you yourself might not have noticed. For example, I wouldn’t have guessed that “commenting” is my 14th most-used word—that kind of data isn’t something that the naked eye can pick out. Having Valorant identity those stats for you could be relevant in drawing the correct conclusions from a large pile of data that can’t be easily read by the human eye. I also think the graphs and visualizations are compelling and interesting enough to be worthy of merit simply for their striking nature alone. Plus, they could be useful in making pictures of the data that would reveal hidden connections otherwise undetectable.
  + how to describe the steps of using voyant: It was pretty self-explanatory—I just copy and pasted the links to my post into the box and it generated everything for. The really crucial thing to remember will be what each tool actually does, which I can use this link for:
    - <https://voyant-tools.org/docs/#!/guide/tools>

notes on GitHub and GitHub Desktop

* Spent a significant amount of time working through GitHub and GitHub desktop
* after creating a new repository I made my first .md file, using markdown code to try to have some fun. Here’s a screenshot of the final result:

Graphical user interface, text, application, email

Description automatically generated

* GitHub desktop was simple enough to download, although I don’t have or understand any of the software necessary to read documents in markdown format, so instead I just uploaded a few of my Dev. Report spreadsheets containing image metadata into my repository on my desktop. I pushed this to my main site and sure enough it appeared in my browser as part of the repository online
* Created a few issues in the repository, mostly questions
  + resolved one (how to insert links to external sites in markdown) and have a few unanswered (how to create a new README .md file without it intruding on the old one because GitHub didn’t let me do that)
* overall I enjoyed the markdown stuff and I can really see how GitHub desktop would be especially useful for collaboration but currently in my own individual project I don’t see much need for it

*Copyright Ben Rhodes,*[*CC BY 4.0*](https://creativecommons.org/licenses/by/4.0/)