# Resources and Technical Guidance for CESTA Projects

Winter Quarter URI Edition; prepared by Alix Keener, Digital Scholarship Coordinator (alixkee@stanford.edu)

## Introduction

What does "sustainability" mean for digital humanities projects? It depends on your project and how you want it to function, and for how long. All of the sections here will contribute toward creating a sustainable DH project, from storage to documentation to web hosting. So, before diving in to more specific recommendations and resources, it's important to ask yourself several orienting questions at the outset of the project--the answers will influence the choices you make to best set yourself up for success:

- What phase of development is your project in? (see more at https://sites.haa.pitt.edu/sustainabilityroadmap/a2-longevity/)
  - Active Creation (including planning/renewal/recreation),
  - Ongoing Maintenance, or
  - Retirement
- Do you want to take the project with you if you're no longer affiliated with Stanford? If not, what do you want to happen to the project, especially if it involves a hosted website?
- If you're working with external collaborators, who will take care of and host the project data? Which institution's repository, hosting, resources, etc. will you make use of?

Version 20211214

- How long do you want the project to last? (the current academic quarter, three years, maintained in some active form for perpetuity)?
- Does the project have dynamic, interactive components that require ongoing maintenance? Who will maintain those and for how long? Do you want to plan for creating a "flattened" static version of your project for preservation? If so, at what point (or points, or versions) will you archive a copy, and how will you decide?

## Project Team Check-in

The start of the quarter is a great time to check in with yourself and your team about where the project is at, and where you hope to get by the end of the quarter. Broadly speaking, most DH projects follow a series of stages, and may iterate within or between stages at different points in their development:

1. **Scholarly question:** what question(s) in your field is this project trying to answer? And/or, what is the contribution your project is making to your field? Has that changed since last quarter? Is there more data or different kinds of analysis you need to pursue in order to answer the questions you currently have?
2. **Research materials:** what are the materials that you are working with in your project? Are they in the format(s) you need for your analysis, or do you need to change that through scanning, data cleaning, restructuring spreadsheets, etc? How are your research materials organized and stored for immediate use with the project, and (if appropriate) with an eye towards longer-term preservation? The beginning of the quarter can be a good time to evaluate the state of your research materials and invest some time in getting them organized.
3. **Analysis**: how far have you gotten on the actual analysis of your research materials? What is the state of the documentation for the steps you've taken for your analysis? If you received an influx of new materials, would you be able to repeat the process you previously used for your analysis? If you're using code to analyze your materials, is that code backed up anywhere, or versioned anywhere (e.g. on GitHub), or does it only exist on someone's laptop? Has anyone other than the code's original author looked at it to make sure it's well-commented and understandable by other people on the team? Where are you storing any outputs of your analysis (result spreadsheets, PDFs, images, maps, etc.)? Sometimes the results of an early analysis can lead to new questions or a need for more data; it can be worth taking a moment to evaluate where you are in your analysis, and whether you want to move forward with your current path or try another direction.
4. **Visualization**: many projects involve some form of visualization as a way to convey the results of their analysis, or provide an interface for navigating research materials. Similar to the "analysis" step, make sure you've documented the process you've used to visualize your materials or results, and that any code involved is documented and stored in an accessible place (see more in the Documentation and Data Storage sections below).
5. **Presentation:** how are you sharing the results of your research? This may look like a website, a single webpage, a data set, an exhibition, or any number of other things.

Have you chosen the way(s) you will share your work? If you've selected a medium, have you explored the technical or logistical options for implementing it? Is your project team ready to take on the work of creating and/or maintaining it, or will you bring in external partners? If there is code or design work involved, is it documented and put in an accessible place? What are your plans for sustaining this mode of presentation and/or the underlying data over the short, medium, and long-term?

6. **Dissemination**: how are you going to get the word out about your project's completion or current findings? What conference(s) have CFPs coming up where you could present something about the project? Are there online or in-person communities where you can share your work less formally? Have you explored publication venues for different aspects of your project (and are there parts of that work that other members of your project team could take the lead on writing up)?

## Data Storage

A good rule of thumb is **3-2-1:**
- ☐ At least 3 copies of the data
- ☐ On at least 2 different types of storage
- ☐ And at least 1 copy offsite (including cloud storage)

Storage options at Stanford:
- Local computer, external hard drive, USB flash drive
- Stanford Google Drive, Google Shared Drive (L,M,H)
  - https://uit.stanford.edu/service/gsuite/drive
  - https://uit.stanford.edu/service/gsuite/teamdrive
- Stanford Shared Computing Environment: Farmshare
- Available software: https://ssds.stanford.edu/software-resources/where-access-software-stanford
- Stanford UIT File and Data Storage
- Stanford Servers and Data
- School/Departmental Servers
- Cloud: AWS, Microsoft Azure, Google Cloud

## Documentation

It's important to create and update documentation about your project from the outset and along the way. This can be in whatever format works best for you and your team and as simple as a shared Google doc. Some basic information that you'll want to capture and display alongside any project outputs (thanks to Dr. Jessica Otis):

1. Project contact(s) and contributors and their roles
2. Project purpose

Version 20211214

3. Essential project functionality (technical functionality necessary to support the intellectual goals of the project which need to be sustained)
4. Access and permissions information (which team members have access to which technologies, and for how long) (consider if relevant)
5. Content Management System (CMS)/technical stack description
6. Workflows used for for data cleaning, analysis, and visualization
7. External dependencies (include external links)
8. Funders/grants list
9. Sustainability commitment given to funders/other legal entities

## Web Hosting

If you choose to create a website for your project--whether that's to publish or disseminate your findings or share or host your research materials for a broader audience; see point 5 in the "Check-In" section above--it's important to keep in mind that static websites are generally preferred. This is because static websites present minimal to no security vulnerabilities because of their "flat" nature and lack of a database for an attacker to exploit by injecting malicious code. Static sites (think early Internet) are hard-coded HTML files, while static generated sites are full static HTML websites based on raw data and a set of templates (see below for a link to a list of static site generators). Beyond security, static sites are also much more sustainable--there are no plug-ins or modules to worry about constantly patching and updating. And the format is more easily exportable to other platforms or hosting services if you need to make a change.

However, static sites are a bit more complicated than creating a "dynamic," database-driven website that has the benefit of a CMS interface like WordPress or Drupal. Creating a static website usually requires some knowledge of HTML, CSS, Markdown, git and GitHub, and more. Workshops and consultations are available through CIDR (contact Alix) and numerous tutorials exist online to help you get started.

### Stanford-supported

(Be sure to check for whether a SUNet ID is required, if your account owner may not be affiliated with Stanford for the desired life of the project.)
- [Stanford Web Services](#)
  - [Stanford Sites](#): Free, self-service, Drupal out-of-a-box. No app integration (e.g. Omeka)
- [Stanford Domains](#)
  - Free (fee for custom domain)
  - Reclaim Hosting platform: can install CMSes like WordPress or Omeka, 50+ other apps to choose from
  - 1:1 account owner to site
- More information:
  - [Website hosting comparison table](#)

Version 20211214

- - Stanford UIT recommendations for storage service
    - Custom stanford.edu URLs
    - Stanford Infrastructure Tools
- Spotlight at Stanford is an application for showcasing digital content in easy-to-produce exhibits. It is available for use by the Stanford community and its partner organizations. Users can quickly build their own online exhibits featuring digital content stored within and outside of the Stanford Digital Repository without having to develop their own project interface or use a proprietary system.

External Providers and Platforms

- GitHub Pages: good for hosting static sites, for free; you can still use a Stanford domain name. Lots of different Jekyll themes available.
- Netlify: good for hosting web apps, has a free tier
- And more static site generators (a list)
- Academic + open source applications (Wordpress, Scalar, and Omeka can be deployed out of hosted Stanford Domains service) (this is not an exhaustive list!)
    - WordPress: basic website platform that easily supports pages and blog posts with simple multimedia
    - Scalar: designed to create branching, non-linear paths through multimedia digital research materials
    - Omeka: designed for presenting different kinds of digital objects (especially images) and exhibits that combine them
    - Mukurtu: similar to Omeka, but with an eye towards more sophisticated access controls to support the needs of communities whose cultural heritage is on display
    - CollectionBuilder-GH: another tool for creating digital exhibits, CollectionBuilder is a template for creating small digital collection exhibits on GitHub Pages designed for teaching digital library skills.
    - Juncture: free and open source web tool, from JSTOR labs; billed as a "visual essay" that can integrate multimodal software like Markdown, IIIF, GeoJSON, D3, Timeline.js, and more. Uses GitHub to host files.
    - StoryMap JS: a visualization that combines maps and timelines; can be embedded as part of a bigger website. Can also be used with a super-high-resolution image, as a way to create a tour through different parts of the image.
    - Jupyter Book: combines code (in Jupyter notebook) and narrative (written in Markdown) into a structured book-like format with multiple levels of navigation.
- ArcGIS StoryMaps: Digital storytelling with custom maps.

## Web Archiving

If you have a URL for your project, you can send it to Alix to submit it to the quarterly crawl for the [Stanford University Website Collection](#). You can read more about the library's web archiving program on their [website](#).

For websites with interactive components that can't be successfully crawled, [Webrecorder](#) provides a suite of open source tools and packages to capture interactive websites and replay them at a later time as accurately as possible.

## Stanford Digital Repository

[Stanford Digital Repository](#) provides long-term storage for data and research materials. SDR has a DIY self-deposit form where you can simply upload a file, choose a license, and provide some metadata, and the file will be stored in SDR. This is similar to many other discipline-oriented digital repositories (e.g. [Humanities Commons CORE repository](#), or [Zenodo](#), which is more science-oriented but many DH projects deposit things there.) The SDR offers services that are complementary to, not competitive with, other data repositories. Benefits of depositing in the SDR include the association with Stanford University and trust in the institution's commitment to long-term stewardship of Stanford-produced research and other scholarly materials of value to Stanford scholars.

The SDR intake form requires that you assert that you have permission from the copyright holder to accession the files to SDR, so it is not a good option for long-term storage for images you have gathered from the internet (or other libraries and archives), text you may have scraped from websites, or other materials that exist in a copyright gray zone.

If you want to submit more than a few files to the repository (e.g. as part of the wrap-up phase of your project), don't try to do it piecemeal through the self-deposit interface. You can work directly with the SDR team to upload your materials in bulk, with proper metadata, as a "collection" for your project that can be findable in SearchWorks. Contact Alix to set up a meeting with Stanford Libraries repository and metadata specialists.

Here are two examples of the kinds of things humanities researchers might deposit (thanks to Hannah Frost):

1. You have created a visualization of movements of enslaved people that are recorded in documents held in a national archive.
- Definitely deposit:
    - compiled digital database used to create the visualization
    - information describing the contents of the database
- Maybe deposit:
    - code or web content used to create the visualization
- Do not deposit:

- original documents held in the national archive

2. Your work involves an analysis of tax documentation. You are not allowed to share the tax documents.
- Definitely deposit:
    - information about the research process
    - tools you created to perform the analysis
    - scripts you created to analyze the documents
- Do not deposit:
    - original tax documentation

## Hiring External Vendors

If you have funding available, sometimes it may be best to contract a designer for UI/UX design if your team is short on time and/or web design proficiency. Beyond your budget considerations, it's vital to agree up front on technical specs with the vendor, especially if they will be doing any backend work beyond UI/UX design. Please contact Alix if you have questions about particular vendors or for support for looking over a contractor statement of work. CESTA also has a small informal list of recommended vendors.

## Resources

- Stanford Libraries' Center for Interdisciplinary Digital Research (CIDR) offers [consulting](#) on a variety of topics.
- The [Stanford Office of Digital Accessibility](#) serves the Stanford community by providing technical guidance, techniques for achieving accessibility, and best practices for accessible electronic content.
- University of Pittsburgh's Visual Media Workshop [Socio-Technical Sustainability Roadmap](#), "a module-based workshop intended to help you and your team approach the seemingly daunting task of sustaining your web-based, user-facing, digital humanities project over time."

## Acknowledgements

Thank you to Quinn Dombrowski and other colleagues in CIDR for their input and feedback in drafting this guide.