

Preservability Self-Assessment Tool

Jonathan Greenberg, NYU Libraries; Thib Guicherd-Callin, LOCKSS Program, Stanford University; Karen Hanson, Portico, ITHAKA; Angela T. Spinazzè, ATSPIN consulting; Scott Witmer, University of Michigan Library

March 2025

DOI: <https://doi.org/10.33682/r23i-9zg2>

This work is licensed under CC BY 4.0. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

[Introduction](#)

[Part 1: Creating Context](#)

[Part 2: Defining Core Intellectual Components](#)

[Part 2 - For Publishing Platforms: Defining Core Intellectual Components](#)

[Part 3: Identifying Relevant Guidelines](#)

[Part 4: Developing an Action Plan](#)

[Closing](#)

[Appendix A: Preservation Method](#)

Introduction

This self-assessment tool aims to help authors, publishers, and platform developers examine the state of their publications or platforms in terms of how “preservable” they are. In this context, preservability is a measure of how likely it is that the essential features of a digital publication can be preserved. The assessment is designed for enhanced non-traditional digital publications that have features beyond text and images such as embedded audio or video, interactive elements, data supplements, annotations and more. Though individuals can complete this assessment, it is designed as a series of conversations for two or more people to work through together and is based on the process used by the embedding team during the Embedding Preservability for New Forms of Scholarship project. It is to be used with the [Guidelines for Preservability in New Forms of Scholarship](#), a set of 72 guidelines written by the creators of this tool. The tool includes question sets with links to relevant guidelines that, when implemented, can lead to enhanced preservability.

This tool is available online as a Google Doc and accompanying Google Sheets as well as a set of printable PDFs. In developing the tool, the authors found that it was often effective to print out the tool and tables to complete offline.

Note: This tool is designed for use with web-based platforms and publications. While it may also be helpful to those delivering content in other forms such as via mobile apps, it has not been tested for that purpose.

The Self-Assessment tool has 4 parts.

- **Part 1** is focused on **creating context**. It aims to collect some foundational information about the publication(s).
- **Part 2** is focused on **defining the core intellectual components** that make up the publication(s) that should be preserved, **identifying the Risk Category** associated with them, and identifying your likely **preservation approach**.
- **Part 3** is focused on exploring the Risk Categories and **identifying the recommendations** related to them from the Guidelines for Preserving New Forms of Scholarship.
- **Part 4** is focused on using the relevant guidelines to **develop an action plan** for improving preservability.

Part 1: Creating Context

This section is focused on creating context. It collects some basic information to encourage you to consider what is important about your publication(s) and what your objectives are for preservation.

Please complete the following questionnaire.

- a) **Which of the following most closely describes you or the organization/group you're working with? Some content in this assessment will be marked as specific to one or more of these categories.**
- i) **PUBLISHING PLATFORM:** We are involved in developing a publishing platform to support multiple publications, potentially across multiple organizations e.g. a journal or ebook publishing web platform.
 - ii) **PUBLISHER:** We work at a publisher that uses a web-based platform for multiple publications and want to consider our general approach. We cannot change the functionality of the platform.
 - iii) **PUBLICATION (ON PLATFORM):** We are working on a single publication (as an author, publisher, or other supporting role) that will be published on a shared web-based platform.
 - iv) **PUBLICATION (CUSTOM):** We are working on a single publication (as an author, publisher, or other supporting role) for which an entirely customized website will be developed. This will require a new website, or use of a more general purpose CMS such as Wordpress.

The category that best describes you for this assessment is: _____

- b) **Describe the publication(s) (theme and functionality):**
- PUBLISHING PLATFORM:** At a high level, describe the options for structuring individual publications on your platform. How does it handle supplements? What features can an author use to enhance publications?
- PUBLISHER:** Describe a typical publication, what are some of the characteristics your platform/approach offers authors to enhance their publications.
- PUBLICATION (ON PLATFORM OR CUSTOM):** Describe the publication: What is the working title? What is the publication about? What enhancements do you anticipate, or have already been identified, for inclusion in this publication? For each enhancement: why this enhancement? How was it decided to include this type of enhancement?

- c) **What is your current preservation approach, if you have one? If you work with a platform, you may need to ask about the preservation service options for the platform level. If you work with a preservation service or team, what is their name/contact information?**

(note: consider discussing this assessment with your preservation service/team)

Part 2. Defining Core Intellectual Components

If you're a PUBLISHING PLATFORM, please use [this modified version of Part 2](#).

This step pertains to [Guideline 10](#), which recommends defining the [core intellectual components](#) of your work(s) - the aspects of the publication(s) that are considered integral to the understanding of it and therefore important to preserve.

For this section, you will complete **TABLE 1**, using **TABLE 2** as a reference.

To do this, follow these steps:

- a) Using your own words, create a list of all of the components/features that you may include in your publication(s), regardless of whether you think they would be important to preserve. Enter these into the first column of **TABLE 1**. Try to describe each feature in abstract functional terms so that it might describe similar features elsewhere in the publication(s), rather than referring to a specific instance e.g. "YouTube video embedded in the flow of the main text" rather than "Video of an interview with x." These features will be mapped to more general features using **TABLE 2** shortly, so you may find it helpful to review that table for ideas for what to include in the list.
- b) Evaluate each feature you listed in **TABLE 1** to determine if it needs to be preserved and enter a value in the "REQUIRED FOR PRESERVATION" column.
- c) For those features that are *Preferred* or *Required* for preservation, use **TABLE 2** as a reference to find their closest *Risk Category*. As you go through this process, you may find you need to split some of the rows in **TABLE 1** into multiple rows. Note, for example, that metadata for the text and additional resources are separate categories from the resources themselves. Enter the matching Risk Category letter in the third column of **TABLE 1**.
- d) Consider your options for a **preservation approach**. We cover 3 options in this tool: *export package*, *web archiving*, and *website-as-software*.
Export packages are generated by a platform and contain the files that represent the important features of a publication. For example, this could consist of an EPUB, some bibliographic metadata, and supplemental files.
Web archiving is a process in which a tool automatically "visits" the web pages that make up a publication and converts them to a WARC (web archive) file. This method retains the look, feel, and some of the interactivity of the publication.
Website-as-software is our shorthand term for preserving an entire web application as a piece of software so that it can be recreated on a new web server—that means code, databases, and other supporting resources that are required to run the website. This is an advanced method of preservation and is rarely used in publishing.

We recommend you **consult with your platform provider, technical team and/or preservation service** to understand the preservation options for your content and what features each approach is likely to cover. With their input, complete the **{METHOD} SUPPORTS** columns in **TABLE 1** for the preservation approach(es) you are considering as best you can, so that you can understand what might be at risk of loss when using different methods. To assist with your understanding of what is more likely to be at risk using each method, **TABLE 2** has columns that highlight the risk categories that *tend* to be challenging to preserve when using a *web archiving* or *website-as-software* method.

If you would like to consider your preservation method options in more depth, [Appendix A](#) offers guidance for this. Alternatively you could consider all guidelines for each Risk Category to highlight potential risks across all preservation methods.

If you have identified a preferred preservation approach, based on its compatibility with your required features, please mark it here. You may wish to choose more than one:

EXPORT: _____
WEB ARCHIVING _____
WEBSITE-AS-SOFTWARE: _____

Consider **PART 3** of this assessment for each of the preservation methods selected.

Part 2, for Publishing Platforms: Defining Core Intellectual Components

This step pertains to [Guideline 10](#), which recommends defining the *core intellectual components* of your work(s)—the aspects of the publication(s) that are considered integral to the understanding of it and therefore likely important to preserve.

This section is tailored to Publishing Platforms. For Platforms, the core intellectual components of the publications will be defined by the publishers who use it, so any feature offered by the platform could be of potential interest for preservation.

For this section, you will complete **TABLE 1**, using **TABLE 2** as a reference.

To do this, follow these steps:

- a) Using your own words, create a list of all of the components/features that may be included in a publication on your platform, regardless of whether you think they would be important to preserve. Enter these into the first column of **TABLE 1**. Try to describe each feature in abstract functional terms so that it might describe similar features in other publications, rather than referring to a specific instance e.g. “YouTube video embedded in the flow of the main text” rather than “Video of an interview with x.” These features will be mapped to more general features using **TABLE 2** shortly, so you may find it helpful to review that table for ideas for what to include in the list.
- b) Evaluate each feature you listed in **TABLE 1** to determine if it needs to be preserved and enter a value in the “REQUIRED FOR PRESERVATION” column. As a platform, you might consider evaluating all features as required, so that you can accurately communicate preservation capabilities to users.
- c) For those features that are *Preferred* or *Required* for preservation, use **TABLE 2** as a reference to find their closest *Risk Category*. As you go through this process, you may find you need to split some of the rows in **TABLE 1** into multiple rows. Note, for example, that metadata for the text and additional resources is a separate category from the resources themselves. Enter the matching Risk Category letter in the third column of **TABLE 1**.
- d) Consider your options for a **preservation approach**. We cover 3 options in this tool: *export package*, *web archiving*, and *website-as-software*.
Export packages are generated by a platform and contain the files that represent the important features of a publication. For example, this could consist of an EPUB, some bibliographic metadata, and supplemental files.
Web archiving is a process in which a tool automatically “visits” the web pages that make up a publication and converts them to a WARC (web archive) file. This method retains the look, feel, and some of the interactivity of the publication.

Website-as-software is our shorthand term for preserving an entire web application as a piece of software so that it can be recreated on a new web server—that means code, databases, and other supporting resources that are required to run the website. This is an advanced method of preservation and is rarely used in publishing.

Export packages or *web archiving* are most commonly used with publishing platforms. If your platform supports an export feature, you should evaluate this in PART 3. If not, *web archiving* is the likely approach used for preservation. Your platform might have an existing relationship with a preservation service that favors one method over the other.

As a platform, even if you have an export option, it may be beneficial to evaluate both the export and web archiving options so that you can communicate with your users about the compatibility of the platform with different preservation methods, and possibly identify areas to improve compatibility. We recommend you **consult with your technical team and/or preservation service** to understand the potential limitations of each approach. With their input, complete the **{METHOD} SUPPORTS** columns in **TABLE 1** for each approach, so that you can understand what might be at risk of loss when using different methods. To assist with your understanding of what is more likely to be at risk using each method, **TABLE 2** has a column that highlights the risk categories that *tend* to be challenging to preserve when using *web archiving*.

If you would like to consider your preservation method options in more depth, [Appendix A](#) offers some guidance for this. Alternatively, you could consider all guidelines for each Risk Category to highlight potential risks across all preservation methods.

If you have identified a preferred preservation approach, based on its compatibility with your required features, please mark it here. You may wish to choose more than one:

EXPORT: _____
WEB ARCHIVING _____
WEBSITE-AS-SOFTWARE: _____

Consider **PART 3** of this assessment for each of the preservation methods selected.

Part 3: Identifying Relevant Guidelines

This step will look at the Risk Category Keys you identified in column 3 of **TABLE 1**. For each Risk Category Key, find the matching section of **TABLE 3** and consider the questions within that section. You will see that some of the questions in **TABLE 3** are scoped to a particular method or format, so take note of that to determine their relevance to your situation as you move through the list.

If you selected one or more preservation methods in **PART 2**, please also review the corresponding Risk Category section in **TABLE 3** for general guidance related to your selected method(s). Review Risk Category X for advice about Exports, Y for Web Archiving, and Z for Website-as-Software.

Respond “Yes,” “No,” “Partial,” or “Don’t Know” to each question. Where the answer is “No” or “Partial” review the corresponding guideline. If helpful, use the Notes field to record your initial reaction to it. We will think about the guidelines in more depth in **PART 4**, where we will consider their priority and formulate an action plan.

You will note that some questions are technical - if you are not in a technical role, we encourage you to talk with a relevant expert for clarification. Some guidelines may also suggest changes that are not in your control. Part of the purpose of this process is to first develop some understanding of what aspects of the publications might be at risk of loss when preserved and then, if you can, take steps towards improved results through workflow changes and conversations with those in charge of your platform.

Part 4: Developing an Action Plan

In **PART 3** you will have generated a list of relevant guidelines. Copy those guidelines into **TABLE 4** so that you can develop an Action Plan for them.

- a. Read each guideline and consider the following prompts. If useful, use the Notes column in the table to record your responses.
 - i. What is the guideline asking us to do?
 - ii. What would it take to shift our process to fully embrace all that this Guideline requires?
 - How long would it take to implement this/these new action(s)?
 - Where might we get assistance?
 - Who needs to be involved?
 - Who will be responsible for this/these action(s)?
 - When would we expect to complete this work?
- b. Once you have thought about how the guideline fits into your context, complete the PRIORITY and EFFORT columns. You can use this information to help organize which might be most impactful and easy to address.
- c. Finally, use the NEXT STEPS column to write notes on specific steps you can take to fulfill the guideline, and designate responsibility in the final column.
- d. Come back and update the Action Plan as you move forward.

Closing

You have now identified areas to work on to improve preservability and have developed an action plan. Consider coming back and updating your assessment over time to see how things are evolving.

Appendix A: Preservation Method

Here we will attempt to determine which preservation approaches are likely to be an option for your publication(s). If you work with a platform, start by checking the documentation to determine what preservation options are supported or reach out to the platform developers and ask. If you know the name of a preservation service they work with, you can also ask that service about the method they use and how they work with the platform. Regardless of whether you can talk to a publishing platform or preservation service, it is helpful to understand the options as it can impact the nature of the work required to support preservation.

There are 3 main preservation approaches for enhanced publications. You may choose to use one or more of them depending on the circumstances:

1. **Export.** The platform you work with may have a feature that allows you to package some parts of the publication to be sent to another location for preservation. The export may, for example, generate an EPUB or PDF of the text, a set of supplements, and a metadata file that describes publication. For enhanced publications some components may be lost if they are not included in the export.
2. **Web archiving.** A harvesting tool “visits” the URLs that make up your publication and creates a copy of those web pages. The web pages are saved into a file called a WARC file (“Web Archive”) that can be viewed and navigated in the absence of the original web pages using a WARC viewer. You can navigate the pages of the publication similar to how they appeared on the Web. If you are familiar with the Internet Archive or Webrecorder, this is the process they use to preserve websites. When successful, this can support preservation of the appearance and experience of the website including any embedded components such as videos. The process can work well in many cases where an export is unavailable or does not cover all features, but also has some limitations that are dependent on the design of the website and your rights to copy and preserve the material.
3. **Website-as-software archiving.** This is where the code and data underlying the website are preserved in such a way that they can be recreated on a new web server and potentially run on an emulator in the future. This is typically reserved for the most complex publications where other methods are not appropriate, and will likely require time and attention from a digital archivist and website administrator. This approach might be used for sites that are very interactive e.g. with a search feature or other dynamic interactive feature that reacts to user input and is core to the publication. It is not effective if key components of the website depend on third party services (e.g. YouTube, Google maps, 3rd party data APIs) to function.

Quick Evaluation

You will find a longer evaluation of your options in the next section but in some cases this quick evaluation may suffice and allow you to continue with the Self-Assessment process.

If you manage or use a publishing platform:

Publishing platforms, such as PubPub and Manifold, are designed to hold many publications, often across multiple publishers. *Export packages* or *web archiving* are the approaches most commonly used for publishing platforms. Is there an export feature built into the platform? If not, *web archiving* is the likely approach. Does the platform have existing relationships with preservation services that favor one method? This can inform the available preservation method(s).

If you have a website that was custom-built for a publication:

Web archiving and *website-as-software* methods may be considered. Web archiving is more widely supported in the preservation community but if the website has highly interactive data-driven features, it may warrant preservation of the entire web application and any underlying data as a piece of software.

Detailed Evaluation

Please answer the following questions to help identify the approach that fits best.

Does the publication platform support an export option?

If yes: Populate the "Export Supports" column of **TABLE 1** for each feature. If using a publishing platform, you may need to consult with your platform provider or documentation to get answers to these.

If you would like to consider another method besides Export, continue to the next question.

If no: *Continue to next question*

Is the publication in the form of a web page or pages on the Internet or will it be?

If yes: Web archiving can copy individual web pages, or crawl your website's HTML hyperlinks with the constraints you set such as specific domain names. It can generally collect fonts, stylesheets, javascript, formatted text, and audio or video that are embedded in a standard way with HTML tags. Its ability to collect dynamic content such as interactive maps or streamed video is dependent on the design of the features. You may need to consult with a web archiving professional and/or your platform developers for a more precise evaluation. If you are willing to run tests on specific publications, [Webrecorder](#)'s tools include a browser plugin called [ArchiveWeb](#) that enables you to record a page/pages, then open the file in [ReplayWeb](#) to see if it works. To explore whether this method might work, consider the following questions and populate the "Web Archiving Supports" column of TABLE 1.

- a. *Are all features that you marked as “Required” in Table 1 available in some form via a URL that is public or otherwise accessible if you allow authenticated access?*
Enter “NO” in the “Web Crawl Supports” column for any features that cannot be accessed via a URL. An example would be if you wanted to preserve an EPUB copy of the publication, but there is no URL on your website that would allow a crawler to download the EPUB directly. The EPUB may only be available if you run a special process to generate one.
- b. *For features that are available via a URL, use the “Web Archiving Typically Supports” column in TABLE 2 to consider whether it could be compatible with this method.* This is a general guide; the only way to know for certain is to test it or work with a web archive expert, but this column can help identify potential challenges to using web archiving. In general, anything that could use an open ended combination of user inputs to display the data are not well preserved with web crawling. Examples of this are IIF viewers, interactive maps, 3D visualizations, search engines, and other dynamic content that changes within the page as you interact with it. These are features that will need to be tested with web archiving before deciding on a way forward.
- c. You can also try this tool to see if it raises any issues with your web pages: <https://archiveready.com/>

If no: Web archiving will not be an option. You may need to explore website-as-software preservation options. Continue to the next question.

ADVANCED: Do you (a) have access to the code for the website and (b) have the ability to seek support from a website administrator to retrieve the necessary resources for rebuilding the webpages on a different server?

If yes: Consider each of the features you require to be preserved. As you populate the “Website-As-Software Supports” column of [TABLE 1](#), you will need to find out if any of the features depend on anything that is not on your main web server and within your control? You may need to work with your website administrator to determine this. Examples of potentially problematic content for this method are features that depend on fonts (e.g. Google Fonts), JavaScript or CSS that are not on the same server as the application. Also a problem is where third party video streaming (YouTube, Vimeo, SoundCloud) or visualization services (ArcGIS) form major components of your site. If there are significant resources outside of the main web server, this approach may not be the best option or some combination of methods may be required. If, however, the main web server contains all content required for the web application to function, this method may be appropriate.

If no: This method is unlikely to be an option - you may need to use a combination of the other methods to fulfill your preservation requirements.

If it is clear from this analysis that one method is preferred, select that method in Part 2 and continue the assessment. Once you analyze your options, you may discover that most of the features you want to preserve are not covered by any one method, and a combination of methods may be required. For example, you may export most of the publication, but create a web archive file (WARC) of a particular webpage to be added to the export file.