

[Curaturae] FAQ Page

Written by [Jono Brandel](#), Project Lead. Last updated on April 22, 2022.

How does “Writing with Open Access” work?

Connecting your writing to images from the Smithsonian’s large collection of artifacts requires complex technical orchestration. This page aims to clarify and illustrate how this orchestration operates. The purpose of walking through this is twofold. First, it is meant as a companion document to the project. It contextualizes the project and the project’s processes. This project raises questions about: technology, algorithms, and the images on display to name a few. By contextualizing the project, the team aims to better facilitate your questions. Second, this case study provides concrete references for developers and researchers as they explore their own projects using Smithsonian’s Open Access collection. References include third party technologies worth using in conjunction with the Open Access collection. In both cases, this information is meant to equip you. To allow you to form your own opinion about both the collection and the technologies at play. In this regard, the project is a provocative design artifact to empower the public.

Preamble: why even write with Open Access?

Visiting the museum is a unique experience unlike other leisure activities today. Through the careful selection and placement of objects in space, the foundation is set for visitors to think about their world in new ways. This is first framed for visitors by an introductory statement when entering a gallery. The text offers visitors a sturdy point of departure for new thoughts to take flight. This experience represents the spirit of the museum and can be transformative for many. It does, however, require visitors to physically go to the museum. This inherently leaves out people who do not visit museums.

Open Access collections, like the one maintained by the Smithsonian for this project, constitute a key step to reach people unable to visit the museum in person. The primary function on Smithsonian’s Open Access homepage is a blank input field to query the collection. This function is useful like a search engine, but the spirit of the museum is left behind.

Writing with Open Access takes the next step to reach more people by bringing the spirit of the museum to you through writing. Like visiting the museum, this project offers new ways of thinking about ideas. Unlike the museum, however, the ideas in question are your own. This makes the collection immediately personal. It creates a link between your perspective and the museum’s. It brings the possibility to familiarize new audiences to the museum.

The life of a keyword

When the application connects your writing to images, it does so in 5 discrete steps that happen in a matter of milliseconds. At the core of these steps is a keyword. So, this case study is going to show how a keyword is selected and transforms into an image. The life of a keyword consists of these steps: translate text, analyze syntax, search collection, filter results, and place image.

Each step consists of a brief explanation and a list of associated questions. Afterwards there is a section with questions that do not fall in a particular step. If you do not see your question, please contact inquiries@curaturae.com so the team can address it.

1. Translate text

After writing some text down in the editor area on the left side of the screen (or top part of the screen in portrait mode on your phone), the application takes every paragraph and translates each sentence into the English dictionary form of the word. In linguistic studies, this form is known as a word's [lemma](#). For instance, the lemma for "I am", "he is", "you are" are all "to be". The translation of each word into their English lemma works across Chinese, French, German, Italian, Japanese, Korean, Portuguese, Russian, and Spanish; meaning "je suis" and "それは" also are represented by the lemma "to be". The result of this step gives the application a long list of keywords to query the Open Access collection.

Translation Q&A

- What happens if I write in an unsupported language?
 - The application will attempt to translate your text, but will return a red dot indicating that it failed.
- Why are only 10 languages available?
 - The application relies on none other than [Google Translate](#) in order to do translations. While Google Translate offers many more languages than 10, it can only do syntax and grammar analysis on the aforementioned languages. For more information on how to use Google Translate in your own projects see: <https://cloud.google.com/translate>.
- Why is the translation done by paragraph?
 - By splitting translations on carriage return characters, it allows you to write in different languages in the same session. Conversely, the more text Google Translate has to work with, the more accurate it can be with translations. From the team's perspective, splitting up by paragraph was the logical compromise of the two opposing ways of translating.

2. Analyze syntax

With your writing now in the form of a list of keywords, the application deconstructs the syntactical significance of each word by grammar. Think nouns, adjectives, verbs, prepositions, and articles among other parts of speech. The application selects all the nouns and verbs and highlights them green or blue, respectively, for you to see in the text editor. Every highlighted word's associated keyword will become a query to find images. In aggregate these images are meant to reflect your writing. The result of this step refines the list of keywords.

Syntax Q&A

- Why does the application only select nouns and verbs?
 - In early tests the application used all words. Pronouns and articles ("he", "the", "whose", etc.) became the majority of the keywords, but did not yield images that

resonated with the writing. So, the team needed to apply some level of algorithmic curation or bias in order to raise the relevance of the images displayed. Inspired by [Oulipo writers](#), who applied simple writing constraints in order to discover new perspectives on writing, the team opted for a simple and obtuse constraint: select nouns and verbs. Nouns represent the subjects and objects of your ideas, while verbs represent the actions. Subjects, objects, and actions cover the basics of most ideas yielding more emotionally relevant images. Yet, nouns and verbs are still common enough to result in numerous connections. Here, artistic liberties are taken in an attempt to create a balance. On one side images that represent the meaning of your writing and on the other a collage filled with images.

- Are there any words that are blacklisted?
 - There is no blacklist. However, if a word is not marked as either a noun or a verb, then it will not be used to query the Open Access collection.

3. Search collection

With keywords selected, the application sends each keyword into Smithsonian's site, known as the Open Access Application Programming Interface (API). It looks very much like if you were to input the keywords yourself manually on the Smithsonian's website. Here is a set of [results](#) for searching the collection for "watermelon" on their site. Instead of manually typing them in and searching, the application automates this step and can do thousands of searches in a second. For each search, Smithsonian's Open Access collection can return up to 50 different artifacts. The result of this step, however, does not return the image, but rather information about the image: when it was made, the creator, the material, the dimensions, description, etc..

Search Q&A

- Do all results have the same information?
 - No. While many results have a lot of information, the amount of data associated with each artifact is not uniform. In fact, since this data is input manually by museum employees and affiliates, certain information is input differently. For instance some artifacts' dimensions are in centimeters, while others are in inches.
- Why do results only go up to 50?
 - This is a creative liberty taken on by the team. Because there are only so many searches that can be asked at a time by our application, there is an intermediary step to cache all results on private servers. Due to cost implications, all data could not be reduplicated and 50 was chosen. In the filtering process, there is more information about how those 50 results are used.
- How do you configure the queries?
 - [The Smithsonian's Open Access portal](#) is the easiest and most comprehensive way to see ways you can configure your own query.
- Are there restrictions on the queries?

- Yes, Open Access collection is a database run by the Smithsonian. Like all databases that open their access to third party products and applications, they need to limit how many times organizations request information. Open Access follows the US government's guidelines to enforce limits. See their [agency manual](#) for more information.
- In addition to rate limiting, Writing with Open Access also only displays artifacts that are marked as Creative Commons Zero (CC0). This is a license called “No Rights Reserved” for free use in personal and commercial settings. For more information on CC0, see the [Creative Commons site](#).

4. Filter results

At this point in the life of a keyword, the application has a list of results from the Smithsonian which is associated with a keyword. These results, up to 50, are filtered for the application to select an image to be displayed on the right side of the collage interface. First, the application removes results that do not have dimensions information, like an artifact's width, height, and depth. It then filters out results that do not have a working image. Before selecting an image, the application tries to remove duplicate images from other keywords. The application then randomly selects a result from the pruned list. This result has all the information needed to display an image in relative size to the other images in the collage interface for you.

Filter results Q&A

- Are there many artifacts without dimensions data?
 - Yes, there are many, but not exactly known how many, artifacts are missing dimensions.
- Why are the results randomized?
 - This is a creative liberty the team took. Results are randomized to promote the feeling of wandering and discovery. In a previous test, each word returned the same image and duplicate words would use subsequent images from the results list. While this is a clear system in writing, it is not evident in the final interface presented to you. Further, it was noted by the advisors of the Activating Smithsonian's Open Access grant that the same results made for a more static experience. The team's aspirations are to bring the spirit of going to the museum to you through this interface. Therefore, the act of wonder and discovery took priority over presenting images in a more hierarchical and analytical manner.
- What other attributes are commonly added to artifacts?
 - You can see a full list of attributes over at [Smithsonian's Open Access API Developer Documentation](#) page.
- Why are there duplicate images?
 - While the application tries to remove duplicates, it does this by using the unique identifier provided by Open Access. Each museum part of the Smithsonian has their own identification nomenclature. As a result, there are certain artifacts that exist in more than one place in the Open Access database.
- What does “randomly selects” mean?

- Every browser has a random function that when called returns a value between zero and one. This value is then multiplied against how many results the application can choose from. The new value represents the index of the result in the list. Thus a result is chosen. While not truly random, each browser's implementation is different and this method provides a simple way to introduce the idea of chance to interactions.

5. Place image

Entering the final step, the keyword now has an associated result connected to it. With this result, the application fetches the image and places it in the collage interface at a scale relative to its dimensions based on the other artifacts' dimensions in the collage. In addition, every image is wired up to be clickable to display its' result data. Lastly, when an image does not have a description, it is processed through [Microsoft's Image Description](#) service. This description is added as the alternate text (see [alt text](#)) to the image for people experiencing the application through [screen readers](#). The result of this step adds one image to the collage.

Image retrieval Q&A

- How does image placement work?
 - The placement of each image is incremental in a loose spiral pattern. It is inspired by many two-dimensional packing algorithms and generative design systems. This particular feature was developed by [Hiro Yamane](#).
- How many images in total can be displayed?
 - The total number of images that can be displayed is based on the device and browser you use. In most cases, using a computer versus a smartphone will allow you to have more images on the page.
- Can I select a different image?
 - Unfortunately at this time there isn't a button to cycle through an image that is already displayed. This is because it would create issues with the layout of the whole collage. That being said, you can always delete a word and retype the same word to try to get a new one to appear.
- Are there images related to one another?
 - Smithsonian's Open Access API does not have any sense of relationship between images. However, the life of a keyword creates a list of results that match a word. These results, in a way, are related; though they do not show up in the interface at this time. This is the extent of relatedness in the project.
 - While Open Access does not have a sense of related images, there are many other sets of collections (digital and otherwise). Some of these sets do keep track of different kinds of relationships. A list the team found in their research are:
 - [Google Image Search](#)
 - [Pinterest](#)
 - [LINEAGE](#)
 - [Grand Palais](#)
 - [Artsy](#)

- Why does it take a long time to display the collage?
 - There are many requests that happen all at the same time. The query and image loading for any given resource can take a few seconds. This, compounded by hundreds or thousands of keywords, can create long loading times. For this project, a fast internet connection will improve the wait time.
- If this process occurs for each keyword, how come the images do not display progressively?
 - Due to limitations in the library used to maintain state, [React.js](#), it is difficult to both keep track of which items are loaded and the order in which they are laid out. For this reason, the application leaves a blank space before displaying the collage in its entirety.

Broader Context

This 5 step process is what happens in the background after you stop typing. It happens hundreds and sometimes thousands of times a second to every person that uses the site. This is a complex orchestration of technical transformation. This type of transformation is the throughline in new media artist Jono Brandel's work. This transformation also has an academic name coined by Lev Manovich: transcoding. Jono wrote about this kind of transformation and how it is an essential piece to making creative expression with technology in his thesis: [*Charting Artistic Rigor: How three new media artists advance Wassily Kandinsky's theories on abstract painting and harness technology as a form of artistic expression*](#).

Even more questions with answers

The life of a keyword has questions related to each of the 5 critical steps to realizing Writing with Open Access. However, in publishing the project there have been additional questions raised. This section is where those questions live:

- What typeface is the project using?
 - [Baton and Baton](#) Turbo by [Fatype](#)
- Is this project's code open sourced?
 - Writing with Open Access is a personal project initiated by Jono Brandel. There are subsequent iterations of this project's code in development and at this time it is not openly available to read or use.
- How many images are in the Smithsonian's collection?
 - Some of this is outlined here: <https://www.si.edu/dashboard/national-collections>
- How many of those are in Open Access?
 - Some of this is outlined here: <https://www.si.edu/dashboard/virtual-smithsonian>
- How many artifacts in Open Access have dimensions data?
 - At the time of writing, this is not known by the project team at Writing with Open Access.
- How many artifacts in Open Access have digital images?
 - Some of this is outlined here: <https://www.si.edu/dashboard/virtual-smithsonian>
- How many artifacts in Open Access have been exhibited in a museum?

- At the time of writing, this is not known by the project team at Writing with Open Access.
- How does the search ranking work for the Open Access API?

External Channels

Hopefully this information has provided clear grounds for how and where decisions are made and operate. This by no means is meant to answer every question. Quite the contrary, it is meant to give framing so that when questions arise you know who to ask or where to seek further information. To that end, here are the main external channels if you have further questions:

- Contact the Open Access team for questions regarding the collection: openaccess@si.edu
- Contact Google Cloud Support for questions regarding language analysis: <https://cloud.google.com/support-hub>
- Contact Microsoft Azure Support for questions around image descriptions: <https://azure.microsoft.com/en-us/support/options/>
- Contact the makers of this project for everything else: inquiries@curaturae.com