

## Appendix 8.

# Outcomes Assessment

*Marcela Mora and William Ulate*

Initially we chose four different open-source annotations tools to assess how they satisfy the annotation needs of Botanical users (Hypothesis.is, Digilib, Annotorious and RERUM). While evaluating the tools we realized that only Hypothes.is and Annotorious are readily applicable interactive tools while Digilib and RERUM are annotation servers that support and facilitate annotations.

We found three additional free source interactive annotation tools with potential desirable functionalities requested by the botanists interviewed. These were (1) Pundit, (2) Recogito, and (3) VGG image annotator.

Below we provide a short summary for the annotations Tools we evaluated. The tools are listed alphabetically.

**Annotorious** (<https://annotorious.github.io/>) is an image annotation tool. The original tool has been discontinued since we evaluated it. By April of 2020 this tool has been rewritten using JavaScript technology, it uses SVG rather than canvas for drawing, and is built right on top of the #RecogitoJS core and adheres to WC3 Web annotation standard unlike its predecessor.

It was adapted to work well for annotating any HTML-formatted text also. It properly handles overlaps and crossing tag boundaries

On early July 2020, the polygon-drawing tool was made available in the Annotorious OpenSeaDragon plugin and later that month, editable polygons were made available in Annotorious and a new feature was available to configure a controlled vocabulary for tagging with Annotorious and RecogitoJS.

It was also announced that new code releases for Recogito and Annotorious (and its plugin version in OpenSeaDragon) were then available on GitHub and NPM. Apart from bug fixes, these changes added polygon drawing, tag vocabulary and autocomplete. Also new changes were included in the API with a Tutorial and demos on how to build a completely customized user interface.

**Digilib** (<https://robcast.github.io/digilib/>) is an image server and web client for viewing images with an “annotator” plugin for annotations on images under the Lesser General Public License, jointly developed by the Max Planck Institute for the History of Science, the Bibliotheca Hertziana, the University of Bern and others.

**Hypothes.is** (<https://web.hypothes.is/>), originally launched in July 2011 by founder and CEO Dan Whaley, is web-based tool that allows users to highlight and/or annotate online text and pdfs. This tool can be accessed using a Chrome extension, a bookmarklet, or the via.hypothes.is proxy server. The “Highlight” option acts like a color highlighter over the text. The “Annotate” option allows users to add notes, comments, images, links, and tags in the right side margin that opens next to the selected text.

**Pundit Annotator Pro** (<https://thepund.it/>): an advanced tool for web annotation that allows users to create semantic annotations (built by one or more triples) using text fragments, web pages, Cultural Heritage Objects or Linked Data entities. This tool is intended for scholars and researchers. All the data is stored in an Annotations Server.

**Recogito** (<https://recogito.pelagios.org>) a software platform that facilitates annotation of text and images. Through both automatic annotation and manual annotation by users, the software links uploaded files to geographic data and facilitates the sharing and downloading of this data in various formats. The software is freely available for download through GitHub, and a version is also hosted online. In the online version, users have a private workspace as well as the ability to share documents among a group or publicly.

**RERUM** is a product of the Center for Digital Humanities at Saint Louis University and is both, an entity annotation ecosystem and repository for annotated corpora. RERUM is not an interface for creating annotations but rather storing them. The annotations they store always include asserted ownership and transaction metadata so consumers can evaluate trustworthiness and relevance.

**VGG Image Annotator [VIA]** (<http://www.robots.ox.ac.uk/~vgg/software/via/>) is a simple and standalone manual annotation software for image, audio and video developed at the Department of Engineering Science of the University of Oxford. VIA runs in a web browser and does not require any installation or setup. The complete VIA software fits in a single self-contained HTML page of size less than 400 Kilobyte that runs as an offline application in most modern web browsers. It is based solely on HTML, JavaScript and CSS (no dependency on external libraries) and is intended for both academic projects and commercial applications.

### **Qualitative comparison within existing tools**

We provide a qualitative comparison within some of the most recent annotation tools, following the features requested by botanists in Appendix 6, where an extensive table with the requirements is broken down into specific sub requirements for easier assessment. A check mark was used to indicate if the annotation tool exhibited the requirement at the time of the assessment, an X mark if the tool lacked the requirement. If not applicable or not sure that the annotation tool exhibited the requirement, then a N/A or a question mark was used respectively.

**Table 2. Set up complexity of the different annotation tools.**

Annotation Tool	Setup Complexity
<b>Annotorious</b>	It requires technical skills to install it such as basic knowledge of html & css. It is a JS library that allows a webpage to use any repository to store the annotations.
<b>Digilib</b>	It requires the installation of Docker or the installation of (1) git, (2)Java JDK and (3) Maven
<b>Hypothes.is</b>	Runs in a web browser and does not require any installation or set up. It needs to be added in Chrome or Firefox as an extension. It is designed to work with existing online content
<b><u>Pund.it (Pro)</u></b>	Runs in a web browser and does not require any installation or set up. It needs to be added in a browser as an extension. It can be signed in with Google and Facebook or a Pundit account. The program presents some bugs and apparently lack of maintenance and sometimes it doesn't allow the user to log in
<b>Recogito</b>	Runs in a web browser and does not require any installation or set up. It has a 10 min tutorial in six languages: <a href="https://recogito.pelagios.org/help/tutorial">https://recogito.pelagios.org/help/tutorial</a>
<b>RERUM</b>	N/A. It is a Repository that facilitates annotation but it is not a per se interactive annotation tool.
<b>VIA image annotator</b>	VIA does not require any additional software installation and runs solely in a web browser. Therefore, this application can run seamlessly in many platforms like Windows, MacOS and Linux and does not demand any technical knowledge from its users

**Table 3. Pros and cons of interactive annotation tools (annotators)**

Annotation Tool	Pros	Cons
<b>Annotorious</b>	<ul style="list-style-type: none"> <li>• Allows to tag images easily</li> </ul>	<ul style="list-style-type: none"> <li>☒ user needs to know html &amp; css</li> <li>☒ Annotations are for images only, not texts</li> <li>☒ Requires knowledge of JavaScript</li> </ul>
<b>Hypothes.is</b>	<ul style="list-style-type: none"> <li>• its source code is hosted in GitHub.</li> <li>• Has browser extensions for both Chrome and Firefox, which makes annotating web pages even easier.</li> <li>• Searchable stream of annotations, replies, and email notifications</li> <li>• Users can work in private groups and then publish, keep annotations private, or make it public from scratch.</li> <li>• Unique URLs for each annotation</li> <li>• Users can link directly to an annotation.</li> <li>• User can sort by tags</li> <li>• keyboard shortcuts for creating annotations and highlights accessible to people with disabilities</li> <li>• Any pdf can be annotated, activating the hypothesis browser</li> </ul>	<ul style="list-style-type: none"> <li>• Annotations are for text only, not images</li> <li>• no semantic annotation</li> </ul>
<b><u>Pund.it (Pro)</u></b>	<ul style="list-style-type: none"> <li>• Users are able to log in using social accounts</li> <li>• Supports the use of semantic annotation</li> <li>• Last update was on Dec 2019</li> </ul>	<ul style="list-style-type: none"> <li>• It has some bugs. It does not always work</li> <li>• Assumes users' knowledge of technical terminology such as key concepts behind the Semantic Web such as the RDF triple, recognize the subject, predicate and object in a sentence (parsing), and a familiarity with prominent Linked Data sources such as Freebase and</li> </ul>

	<ul style="list-style-type: none"> <li>• Unique URLs for each annotation</li> <li>• Allows to export and reuse your annotation</li> <li>• User can annotate text and images</li> </ul>	DBpedia are required in order to make proper use of Pundit
<b>Recogito</b>	<ul style="list-style-type: none"> <li>• Supports the use of semantic annotation to link and explore historical place information</li> <li>• User can annotate text and images</li> <li>• Allows to connect between people (names), places or events</li> <li>• Geotagging features</li> <li>• Supports both additional image standards (such as IIIF) and text standards (TEI export).</li> </ul>	<ul style="list-style-type: none"> <li>• It does not really annotate a website. Sources are imported through an upload feature and managed by the users themselves. It uses notebooks where an image or a text can be uploaded.</li> </ul>
<b>VIA image annotator</b>	<ul style="list-style-type: none"> <li>• Do not depend on any external library</li> <li>• Easy to use and does not require complex installation</li> <li>• Attributes can be defined with different anchor and input types (i.e. text, checkbox, radio, select &amp; image)</li> </ul>	<ul style="list-style-type: none"> <li>• Annotations are for images and videos only, not texts</li> </ul>

## Conclusions and Future work

We explored different open source annotators to evaluate if these tools fulfill the annotation needs of botanists. We installed the selected tools (when necessary) and evaluated them based on different functional aspects.

We concluded that there is no single tool that comprehensively complies with the annotation needs of botanists. However, the combination of the collective strengths of each tool satisfy many of the basic needs of the botanical community. In cases where OCR and pdf texts are always visible and available in a digital library, we recommend the use of Hypothes.is. Since most of the literature in the Botanicus digital library is in the form of images, we consider that the best software that can be used to annotate such images are VIA and Recogito. One thing in common with the Humanities and biodiversity sciences is the importance of places and geographic entities. In the biodiversity sciences, the geographic distribution of a species is vital for the understanding of its ecology, evolution, and conservation. Recogito is particularly well suited for annotating documents with references to gazetteers, and supports the annotation of both texts and images (i.e., digitized maps).

Although many historical biodiversity documents have been digitized, searching or retrieving their content is challenging. Many manuscripts refer to outdated taxonomic classifications and historical place names that are written in multiple languages. Semantically annotating the content (structured annotations from the named entities in texts) is an important step for text mining and would further facilitate biodiversity research.

We hope that our project raises the awareness of annotation tools already available and optimizes the use of existing and future resources to take advantage of the functionality already available in open software tools, to minimize the need for costly and time-consuming competing developments. Our vision aims to recognize the strengths each different external third-party tool and promote development of the needed functionality while providing pluggable interfaces into a repository such as RERUM. We hope that by making available these annotation capabilities in biodiversity digital libraries such as Botanicus or BHL (Biodiversity Heritage Library) will entice the community to participate actively in expanding biodiversity knowledge.

In the near future, we intend to write a proposal to implement our ideas following the real use cases presented and working closely with the community to allow emerging new ideas and requests to guide any development.