

Data Management Plan

This project will primarily result in three forms of data: computer source code, machine learning models, and documentation.

In keeping with the STUDIO's commitment to open source development, our prototype software tools, models, associated project source code, workflows, documentation, and use cases will be posted on the STUDIO's GitHub site. This will be released under the MIT open source media license or CC BY 4.0 Creative Commons license as applicable to allow for the largest possible application. These will remain available on GitHub for a period of not less than five years, and we anticipate them remaining there for the foreseeable future. We will also host a mirror of the source code, documentation, and models on the STUDIO's website to protect against the event that GitHub is no longer available for hosting.

Our primary data set, the Teenie Harris Archive, consists of images and metadata that are publically available through the Carnegie Museum of Art. CMOA is the custodian of these data sets and will continue to maintain the canonical copy of this data. Throughout the project, our use of the Teenie Harris Archive will generate annotations in the Open Annotation (<http://www.openannotation.org/spec/core/>) standard that augment that collection. We have agreed to provide these annotations at the completion of the project to the Carnegie Museum of Art for potential integration into their permanent collection. Additionally, through the use of our tools, these annotations can be regenerated as needed.

(It is possible that IIIF will replace Open Annotation with W3C Web Annotations (<https://www.w3.org/TR/annotation-model/>) before this project can be completed. In that case, our annotations would use the Web Annotation Model.