Version 1.0

List of the authors

Short title	SWHAP
Full title	Software Heritage Acquisition Process
	Roberto Di Cosmo, Software Heritage, Inria and University of Paris ⟨roberto@dicosmo.org⟩ Carlo Montangero, Dept. of Computer Science, University of Pisa ⟨carlo@montangero.eu⟩ Guido Scatena, Dept. of Computer Science, University of Pisa ⟨guido.scatena@unipi.it⟩
Date	November 25, 2021
Contact	Roberto Di Cosmo ⟨roberto@dicosmo.org⟩

Abstract

The presentation of acquired source code of landmark legacy software is particularly important. This document presents the first version of the integration of WikiMedia and ScinceStories.io presentation into SWHAP, the Software Heritage Acquisition Process: a protocol for the collection and preservation of software of historical and scientific relevance. SWHAP results from a fruitful collaboration of the University of Pisa with Software Heritage in this area of research, under the auspices of UNESCO, and has been validated on a selection of software source code produced in the Pisa area over the past 50 years.

Acknowledments TBD

Contents

1) Introduction

Create a SWH-story

The suggested process to document the recovered source code with a story in the SWH-stories website https://swh.stories.k2.services/stories/ has two phases: - Collect, where the Presentation designer collects the images, videos, documents to be published. For each item, he also gathers in a suitably structered *inventory* in the Workbench the information nedeed to insure that once uploaded in the appropriate Wikimedia data base the item is satisfactorily self documenting. The Presentation designer should also take care that the item can be granted a pubblic domain license, as required by Wikimedia policies. - Publish, where the Web engineer, in this case the Wikimedia expert, uploads the items in such a way that they are best presented exploiting the STORIES SERVICE at [http://stage.stories.k2.services/publisher/].

The story will be available at [(https://swh.stories.k2.services/stories/){.underline}]. Since the story is dinamically constructed at each access, the above process can be freely iterated, to add new elements or to correct the presented information.

TBD : Choose between inserting the sentence *The process is described in detail in ???* or elaborate here the description above. The presentation process, abstract view {#sec:processabs}

TBD : Presentation designer might be abstract, and instantiated as SWH-stories designer in this version.

2) Presenting in SWH-stories

This way of presenting the recovered source code is inspired by the https://sciencestories.io website.

TDB : decide between an *embedded* or an *external* approach. In the embedded case we insert here a description as abstract as possible of the process to create a SWH-story, and see how to put other information in the rest of the document. In the external case, we refer to a new document, based on Morane and Kat's report, completed with a section related to the support in the adjourned template of SWHAPPE.

2.1) Resources in the process

- SWHAP Repository
- Wikidata
- Wikimedia Commons
- Software Heritage
- Publisher Workspace (?)

2.2) SWHAP Repository structure

Cover Page

Cover page TODO:

Story Inventory

The story invetory is the central worikng point of the presentation process of SWHAP. It consists in a document and a set of folders. The document serves ad catalogue of collected items and guides the process of tranfer acquired matherials to wiki*.

Presentation Moments

The story of a software, as presented by ScienceStories.io, is divided in "moments".

People Moments

as a software story is also a story about people ... This section, and the realtive folders, provides informatins and storage for the images, videos, documents, etc. related to the people involved in the software project.

The people story moments fall in two classes, with a dedicated sub-folder each:

- *media_gallery*, for images and videos, and
- *library*, for documents.

An entry should be added to the StoryInventory file for each added item, following the pattern offered in the dedicated section.

Software Moments

This section, and the relative folders, provides storage for the images, videos, documents, etc. related to the software.

The software story moments fall in four classes, with a dedicated sub-folder each:

- **code_listing**, for PDF versions of the source code, annotated with permalinks to the SWH archive,
- **media_gallery**, for images of the original source code recovered from non digital documents,
- **library** for links to documents related to the software, and
- **videos** for related videos.

An entry should be added to the StoryInventory file for each added item, following the pattern offered in the dedicated section.

Wikidata Entities in this section we annotate the entities to be created

2.3) Roles in the process

Collector

The curator is in charge of retirve alle the mandatory information abot the collected matherials, as, for instance, the authors, the dates and the copytright.

Curator

The curator is in charge of recustructing the story behind the software, about people and places. Moreover, the curator is in charge of selecting and creating a presentable objects for acquired matherials. For instance he select and crop the acquired images, he create the annotaded pdf version of relevant pieces of code listng, with links to the actual code archived in Software Heritage by the SWHID.

Presentation designer and Web engineer*

Though most of the presentations of the archived software will be on line, the abilities to design the contents of a presentation should be considered separately from the technical ones. For instance, in the case of the SWH-stories, the presentation designer should be competent in the topic addressed by the code, to be able to search and select the items to be inserted in the story. On the other side, the web engineer should be acknowledgeable of Wikimedia and the other tools involved in creating the stories.

Visitor

The visitor can navigate to https://swh.stories.k2.services/ where he can find the collection of software acquired and presented by Software Heritage.

3) A walkthrough on a running example