Software Heritage
Legacy software archival - an archeology rescue mission

Morane Gruenpeter
Software Heritage
morane@softwareheritage.org

June 21th, 2020
Outline

1. Software Heritage Acquisition Process (SWHAP)

2. The Software Stories new project
We are at a turning point

Looking at the past

- a lot of old software misplaced, lost, or behind barriers, but...
- most founding fathers are still here, and willing to share
- urgent to collect their knowledge

Only a few years left.
We are at a turning point

Looking at the past

- a lot of old software misplaced, lost, or behind barriers, but...
- most founding fathers are still here, and willing to share
- urgent to collect their knowledge

Only a few years left.

Looking at the future

- software development and use skyrockets: more programmers, and more code!
- essential to provide a universal platform for all the future software source code

Every year that goes by makes the problem worse.
We are at a turning point

Looking at the past
- a lot of old software misplaced, lost, or behind barriers, but...
- most founding fathers are still here, and willing to share
- urgent to collect their knowledge

Only a few years left.

Looking at the future
- software development and use skyrocket: more programmers, and more code!
- essential to provide a universal platform for all the future software source code

Every year that goes by makes the problem worse.

it is urgent to take action!
Paris Call on Software Source Code

“[We call to] support efforts to gather and preserve the artifacts and narratives of the history of computing, while the earlier creators are still alive”
Paris Call on Software Source Code

“[We call to] support efforts to gather and preserve the artifacts and narratives of the history of computing, while the earlier creators are still alive”
Paris Call on Software Source Code

“[We call to] support efforts to gather and preserve the artifacts and narratives of the history of computing, while the earlier creators are still alive”

- **Rescue** Legacy Software from different media
  - physical
  - digital
    - legacy / unsupported
    - recent / supported

- **Curate** the code
  - reconstructing the development history
  - collecting metadata

- And **illustrate** with dedicated presentations
SWHAP: A variety of situations

**Code support**
- physical (Softi, TAUmus)
- digital, legacy (TAUmus)
- digital, recent (CMM, SPA, OrbFit)

**Authors availability:**
- from next door, CMM
- to next block, OrbFit
- to far away, SPA
- to passed away, Softi
SWHAP: Four phases workflow to streamline the effort

Collector -> Deposit Engineer

Collect

Journal/Catalogue

DepositedSC

Curator

Journal/Catalogue

CuratedSC

Archive Engineer

Archive

Journal/Catalogue

ArchivedSC

[Present]

Presentation Designer

[Web Designer]

Journal/Catalogue

PresentedSC

[ Warehouse ]

Depository

Curated Source Code Deposit

SWH

Wikies
Collect: find and gather the source code and related materials

- **Warehouse** physical location
- **Depository** virtual warehouse
- **Workbench (virtual)** to support the process activities
- **Curated source code deposit** to be saved and referenced in SWH
Collect: find and gather the source code and related materials

Warehouse physical location

Depository virtual warehouse

Workbench (virtual) to support the process activities

Curated source code deposit to be saved and referenced in SWH
Curate: analyze, cleanup and structure the raw materials

How to curate?

- Rebuilt the development history:
  - ascertain versions,
  - production dates,
  - contributors for historical accuracy

- Collect descriptive metadata
  - project description
  - purpose
  - contextual information

- Procure legal authorization
  - Identify the owners,
  - obtain the authorizations to publish the artifacts

- Organize metadata to be consumed for the semantic web
  - recommend vocabulary: CodeMeta
Where to archive

- **Software Heritage** for the source code
- **Wikimedia** for images or videos,
- open access repositories for research articles,
- **Wikidata** for software descriptions and properties, etc.
The control code for the music synthesizer TAU2

FORTRAN II and TAUmus command language

Istituto di Elaborazione dell’Informazione CNR

Group led by the late M° P. Grossi

Le Sacre du Printemps (ABSTRACT)
SWHAP@PISA: Capturing metadata in branch master

CMM

a Customisable Memory Manager

This repository contains the CMM Development History.
The original files are stored in the Depository containing the raw materials and the browsable source.
Information on the acquisition of this code can be found in the CMM-Workbench repository.
This repository was created with the support of the Software Heritage Acquisition Process Pilot Enabler.
SWHAP@PISA: Recreating development history in branch SourceCode
Browse the archive

https://github.com/Unipisa/CMM

30 September 2020, 22:19 UTC

Branches (2)

Commit Date

<table>
<thead>
<tr>
<th>Revision</th>
<th>Author</th>
<th>Date</th>
<th>Message</th>
<th>Commit Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>c0f32b6</td>
<td>Giuseppe Attardi</td>
<td>02 March 1998, 23:00 UTC</td>
<td>1.0 - Contributors mentioned in Changelog - Giuseppe Attardi @attar...</td>
<td>11 December 2019, 14:35 UTC</td>
</tr>
<tr>
<td>55778ad</td>
<td>Giuseppe Attardi</td>
<td>15 May 1997, 06:24 UTC</td>
<td>1.8 - Contributors mentioned in Changelog - Giuseppe Attardi @attar...</td>
<td>11 December 2019, 14:35 UTC</td>
</tr>
<tr>
<td>cfa2967</td>
<td>Giuseppe Attardi</td>
<td>24 November 1996, 18:13 UTC</td>
<td>1.7 - Contributors mentioned in Changelog - Giuseppe Attardi @attar...</td>
<td>11 December 2019, 14:35 UTC</td>
</tr>
<tr>
<td>b2be54b</td>
<td>Giuseppe Attardi</td>
<td>06 June 1996, 09:07 UTC</td>
<td>1.6 - Contributors mentioned in Changelog - Giuseppe Attardi @attar...</td>
<td>11 December 2019, 14:35 UTC</td>
</tr>
<tr>
<td>b41693b8</td>
<td>Giuseppe Attardi</td>
<td>02 November 1995, 15:11 UTC</td>
<td>1.5 - Contributors mentioned in Changelog - Giuseppe Attardi @attar...</td>
<td>11 December 2019, 14:35 UTC</td>
</tr>
<tr>
<td>b5c4b31</td>
<td>Giuseppe Attardi</td>
<td>03 October 1995, 10:43 UTC</td>
<td>1.4 - Contributors mentioned in Changelog - Giuseppe Attardi @attar...</td>
<td>11 December 2019, 14:35 UTC</td>
</tr>
<tr>
<td>29d75258</td>
<td>Giuseppe Attardi</td>
<td>07 November 1994, 16:36 UTC</td>
<td>1.3 - Contributors mentioned in Changelog - Giuseppe Attardi @attar...</td>
<td>11 December 2019, 14:35 UTC</td>
</tr>
<tr>
<td>ce186db6</td>
<td>Giuseppe Attardi</td>
<td>27 October 1994, 11:26 UTC</td>
<td>1.1 - Contributors mentioned in Changelog - Giuseppe Attardi @attar...</td>
<td>11 December 2019, 14:35 UTC</td>
</tr>
</tbody>
</table>
SWHAP@PISA: Referencing source code on SWH

In this online visualization of the objects present in the Software Heritage archive, pre-commit hooks based on SoftHive Heritage persisted Identifiers (SWHIDs) must be used instead of copying and pasting the URL from the address bar of the browser (as there is no guarantee the current URI scheme will remain the same over time).

Select below a type of object currently browsed in order to display its associated SWHID and permalink.

- revision
- content
- snapshot

*This text is only a visual representation of the SWH database.*
Outline

1. Software Heritage Acquisition Process (SWHAP)

2. The Software Stories new project
Software Stories

Goal

- Create dedicated presentations of the curated materials
- Showcase legacy software in a virtual exposition
- see https://sciencestories.io example

Actors

- SWH team: Roberto and Morane (project management)
- Science-stories team: Kat and Kenneth (design and implementation)
- Pisa team: Carlo, Guido and Laura (testing and guides)
The activity plan

Stage 1: Design & Prototype implementation

- Design the data model
- Define a “formal language” for storyboards of SWH stories
- Define a high-level architectural design for the “render engine”
- Develop rendering engine
- Test prototype

Stage 2: Design a meta-process for the curation of stories as part of the SWHAP

- Develop the process for creation and maintenance of SWH stories
- Create a curator guide to assist curators in the process of legacy software preservation
Completing the process