

Software Heritage

The universal source code archive

Morane Guenpeter

Metadata specialist, Software Heritage
Crossminer, University of L'Aquila

`morane@softwareheritage.org`

June 11th, 2019



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

- 
- 1 Software Heritage
 - 2 SWH data model and persistent identifiers
 - 3 Demo



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Collect, preserve and share the source code of all the software

Preserving our heritage, enabling better software and better science for all



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Collect, preserve and share the source code of all the software

Preserving our heritage, enabling better software and better science for all

Reference catalog



find and reference **all** the
source code



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Collect, preserve and share the source code of all the software

Preserving our heritage, enabling better software and better science for all

Reference catalog



find and reference **all** the
source code

Universal archive



preserve **all** the source
code



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Collect, preserve and share the source code of all the software

Preserving our heritage, enabling better software and better science for all

Reference catalog



find and reference **all** the
source code

Universal archive



preserve **all** the source
code

Research infrastructure



enable analysis of **all** the
source code

Cultural Heritage



Industry



Research



Education



Software Heritage

Cultural Heritage



Industry



Research



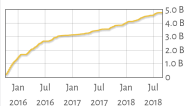
Education



Software Heritage

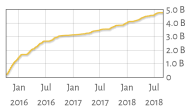
Source files

5,335,334,261



Commits

1,191,059,588



Projects

86,555,394



Growing Support

Raising awareness: landmark agreement, 3/4/2017; grand opening, 7/6/2018

Inria
INVENTEURS DU MONDE NUMÉRIQUE



Sharing the vision



Morane Gruenpeter

Sponsoring our work

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Platinum sponsors



Microsoft



SOCIÉTÉ GÉNÉRALE

Golden sponsor



MINISTÈRE
DE L'ÉCONOMIE, DE LA RECHERCHE
ET DE LA RECHERCHE
DIGITALE

Silver sponsors



MINISTÈRE
DE L'ÉCONOMIE, DE LA RECHERCHE
ET DE LA RECHERCHE
DIGITALE



www.softwareheritage.org

June 11th, 2019

4 / 8

- 
- 1 Software Heritage
 - 2 SWH data model and persistent identifiers
 - 3 Demo

An important distinction: DIOs vs. IDOs

The term “Digital Object Identifier” is construed as “digital identifier of an object,” rather than “identifier of a digital object”

Norman Paskin. 2010



An important distinction: DIOs vs. IDOs

The term “Digital Object Identifier” is construed as “digital identifier of an object,” rather than “identifier of a digital object”
Norman Paskin. 2010

DIO (Digital Identifier of an Object)

digital identifiers for (potentially) **non digital objects**

- epistemic complexity (manifestations, versions, locations, etc.)
- need an **authority** to ensure persistence and uniqueness

An important distinction: DIOs vs. IDOs

The term “Digital Object Identifier” is construed as “digital identifier of an object,” rather than “identifier of a digital object”
Norman Paskin. 2010

DIO (Digital Identifier of an Object)

digital identifiers for (potentially) **non digital objects**

- epistemic complexity (manifestations, versions, locations, etc.)
- need an **authority** to ensure persistence and uniqueness

IDO (Identifier of a Digital Object)

digital identifiers (only) for **digital objects**

- can provide both **integrity** and **no middle man**
- broadly used in modern software development (git, etc.)

An important distinction: DIOs vs. IDOs

The term “Digital Object Identifier” is construed as “digital identifier of an object,” rather than “identifier of a digital object”
Norman Paskin. 2010

DIO (Digital Identifier of an Object)

digital identifiers for (potentially) **non digital objects**

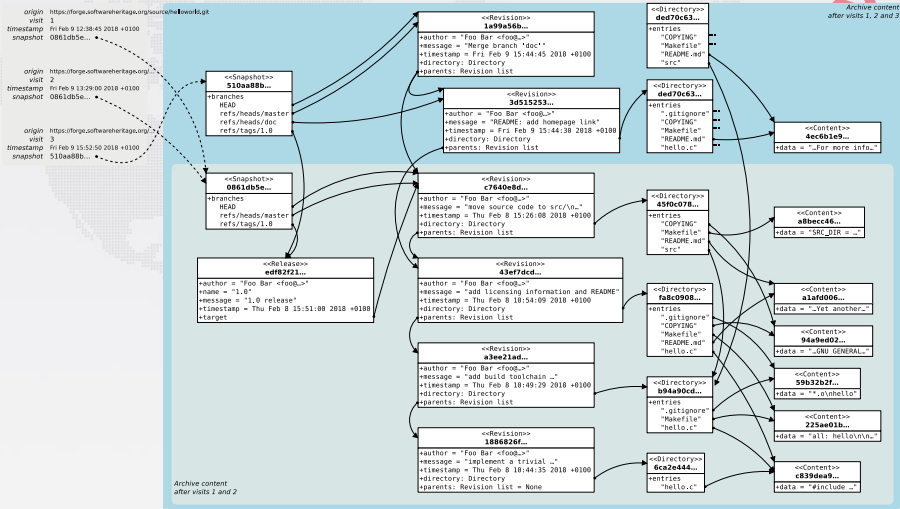
- epistemic complexity (manifestations, versions, locations, etc.)
- need an **authority** to ensure persistence and uniqueness

IDO (Identifier of a Digital Object)

digital identifiers (only) for **digital objects**

- can provide both **integrity** and **no middle man**
- broadly used in modern software development (git, etc.)

Software Heritage provides intrinsic **IDOs** for all archived artefacts



The Software Heritage IDO schema (see <http://bit.ly/swhpids>)

swh:1:**cnt**:94a9ed024d3859793618152ea559a168bbcbb5e2

full text of the GPL3 license

The Software Heritage IDO schema (see <http://bit.ly/swhpids>)

swh:1:**cnt**:94a9ed024d3859793618152ea559a168bbcbb5e2

full text of the GPL3 license

swh:1:**dir**:d198bc9d7a6bcf6db04f476d29314f157507d505

Darktable source code

The Software Heritage IDO schema (see <http://bit.ly/swhpids>)

swh:1:**cnt**:94a9ed024d3859793618152ea559a168bbcbb5e2 full text of the GPL3 license

swh:1:**dir**:d198bc9d7a6bcf6db04f476d29314f157507d505 Darktable source code

swh:1:**rev**:309cf2674ee7a0749978cf8265ab91a60aea0f7d

a **revision** in the development history of Darktable

The Software Heritage IDO schema (see <http://bit.ly/swhpids>)

swh:1:**cnt**:94a9ed024d3859793618152ea559a168bbcbb5e2 full text of the GPL3 license

swh:1:**dir**:d198bc9d7a6bcf6db04f476d29314f157507d505 Darktable source code

swh:1:**rev**:309cf2674ee7a0749978cf8265ab91a60aea0f7d

a **revision** in the development history of Darktable

swh:1:**rel**:22ece559cc7cc2364edc5e5593d63ae8bd229f9f

release 2.3.0 of Darktable, dated 24 December 2016

The Software Heritage IDO schema (see <http://bit.ly/swhpids>)

swh:1:**cnt**:94a9ed024d3859793618152ea559a168bbcbb5e2 full text of the GPL3 license

swh:1:**dir**:d198bc9d7a6bcf6db04f476d29314f157507d505 Darktable source code

swh:1:**rev**:309cf2674ee7a0749978cf8265ab91a60aea0f7d

a **revision** in the development history of Darktable

swh:1:**rel**:22ece559cc7cc2364edc5e5593d63ae8bd229f9f

release 2.3.0 of Darktable, dated 24 December 2016

swh:1:**snp**:c7c108084bc0bf3d81436bf980b46e98bd338453

a **snapshot** of the entire Darktable repository (4 May 2017, GitHub)

The Software Heritage IDO schema (see <http://bit.ly/swhpids>)

swh:1:**cnt**:94a9ed024d3859793618152ea559a168bbcbb5e2 full text of the GPL3 license

swh:1:**dir**:d198bc9d7a6bcf6db04f476d29314f157507d505 Darktable source code

swh:1:**rev**:309cf2674ee7a0749978cf8265ab91a60aea0f7d

a **revision** in the development history of Darktable

swh:1:**rel**:22ece559cc7cc2364edc5e5593d63ae8bd229f9f

release 2.3.0 of Darktable, dated 24 December 2016

swh:1:**snp**:c7c108084bc0bf3d81436bf980b46e98bd338453

a **snapshot** of the entire Darktable repository (4 May 2017, GitHub)

Current resolvers: archive.softwareheritage.org and n2t.org

- 
- 1 Software Heritage
 - 2 SWH data model and persistent identifiers
 - 3 Demo



Demo Time!

<https://archive.softwareheritage.org>

- search by repository url
- search by metadata
- save code now