

Software Heritage

The Great Library of Source Code

David Douard & Antoine R. Dumont

Software Engineer - Software Heritage
info@softwareheritage.org

27 juin 2019

PSES - Choisy-le-Roy



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Software is everywhere



Software is everywhere



Software embodies a growing part of...

... our scientific, *technical* and Cultural Heritage!

Source Code: *executable* and *human readable* knowledge



“The source code for a work means the preferred form of the work for making modifications to it.”

GPL Licence





“The source code for a work means the preferred form of the work for making modifications to it.”

GPL Licence

Hello World



"The source code for a work means the preferred form of the work for making modifications to it."

GPL Licence

Hello World

Program (excerpt of binary)

```
4004e6: 55
4004e7: 48 89 e5
4004ea: bf 84 05 40 00
4004ef: b8 00 00 00 00
4004f4: e8 c7 fe ff ff
4004f9: 90
4004fa: 5d
4004fb: c3
```



"The source code for a work means the preferred form of the work for making modifications to it."

GPL Licence

Hello World

Program (excerpt of binary)

```
4004e6: 55
4004e7: 48 89 e5
4004ea: bf 84 05 40 00
4004ef: b8 00 00 00 00
4004f4: e8 c7 fe ff ff
4004f9: 90
4004fa: 5d
4004fb: c3
```

Program (source code)

```
/* Hello World program */

#include<stdio.h>

void main()
{
    printf("Hello World");
}
```

Apollo 11 (~60.000 lines), 1969



"When I first got into it, nobody knew what it was that we were doing. It was like the Wild West."

Margaret Hamilton



~ 50 years, a lightning fast growth

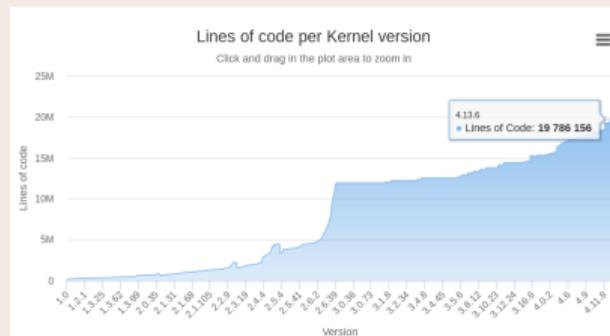
Apollo 11 (~60.000 lines), 1969



"When I first got into it, nobody knew what it was that we were doing. It was like the Wild West."

Margaret Hamilton

Linux Kernel (in your pockets!)



~ 50 years, a lightning fast growth

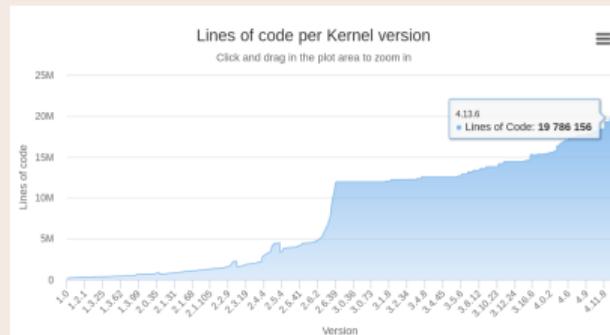
Apollo 11 (~60.000 lines), 1969



"When I first got into it, nobody knew what it was that we were doing. It was like the Wild West."

Margaret Hamilton

Linux Kernel (in your pockets!)



Harold Abelson, Structure and Interpretation of Computer Programs

(1985)

"Programs must be written for people to read, and only incidentally for machines to execute."

~ 50 years, a lightning fast growth

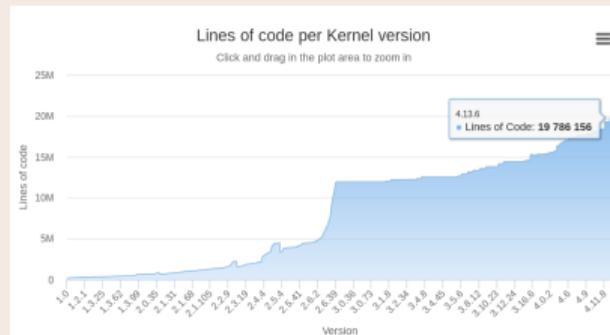
Apollo 11 (~60.000 lines), 1969



"When I first got into it, nobody knew what it was that we were doing. It was like the Wild West."

Margaret Hamilton

Linux Kernel (in your pockets!)



Harold Abelson, Structure and Interpretation of Computer Programs

(1985)

"Programs must be written for people to read, and only incidentally for machines to execute."

Len Shustek, Computer History Museum

(2006)

"Source code provides a view into the mind of the designer."



A word cloud of terms related to software fragility, including: damage, disaster, malicious, deletion, obsolete, attack, dependencies, dangling, wear, corruption, encryption, format, reference, storage, media, aging, and tear.

Like all digital information, FOSS is fragile

- inconsiderate and/or malicious code loss (e.g., Code Spaces)
- business-driven code loss (e.g., Gitorious, Google Code)
- for obsolete code: physical media decay (data rot)



Like all digital information, FOSS is fragile

- inconsiderate and/or malicious code loss (e.g., Code Spaces)
- business-driven code loss (e.g., Gitorious, Google Code)
- for obsolete code: physical media decay (data rot)

Where is the archive...

where do we go if (a repository on) GitHub or GitLab.com goes away?

We are at a turning point

Preserve the past

Only a few years left to recover the history of software technology

We are at a turning point

Preserve the past

Only a few years left to recover the history of software technology

Improve the future

We need a **universal** platform for all the future software source code



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Our mission

Collect, **preserve** and **share** the *source code* of *all the software* that is publicly available.

Past, present and future

Preserving the past, enhancing the present, preparing the future.

Core principles

Cultural Heritage



Industry



Research



Education



Software Heritage

Cultural Heritage



Industry



Research



Education



Software Heritage

Open approach

- 100% Free and Open Source Software
- transparency

In for the long haul

- replication
- non profit

Archiving goals

Targets: VCS repositories & source code releases (e.g., tarballs)

We DO archive

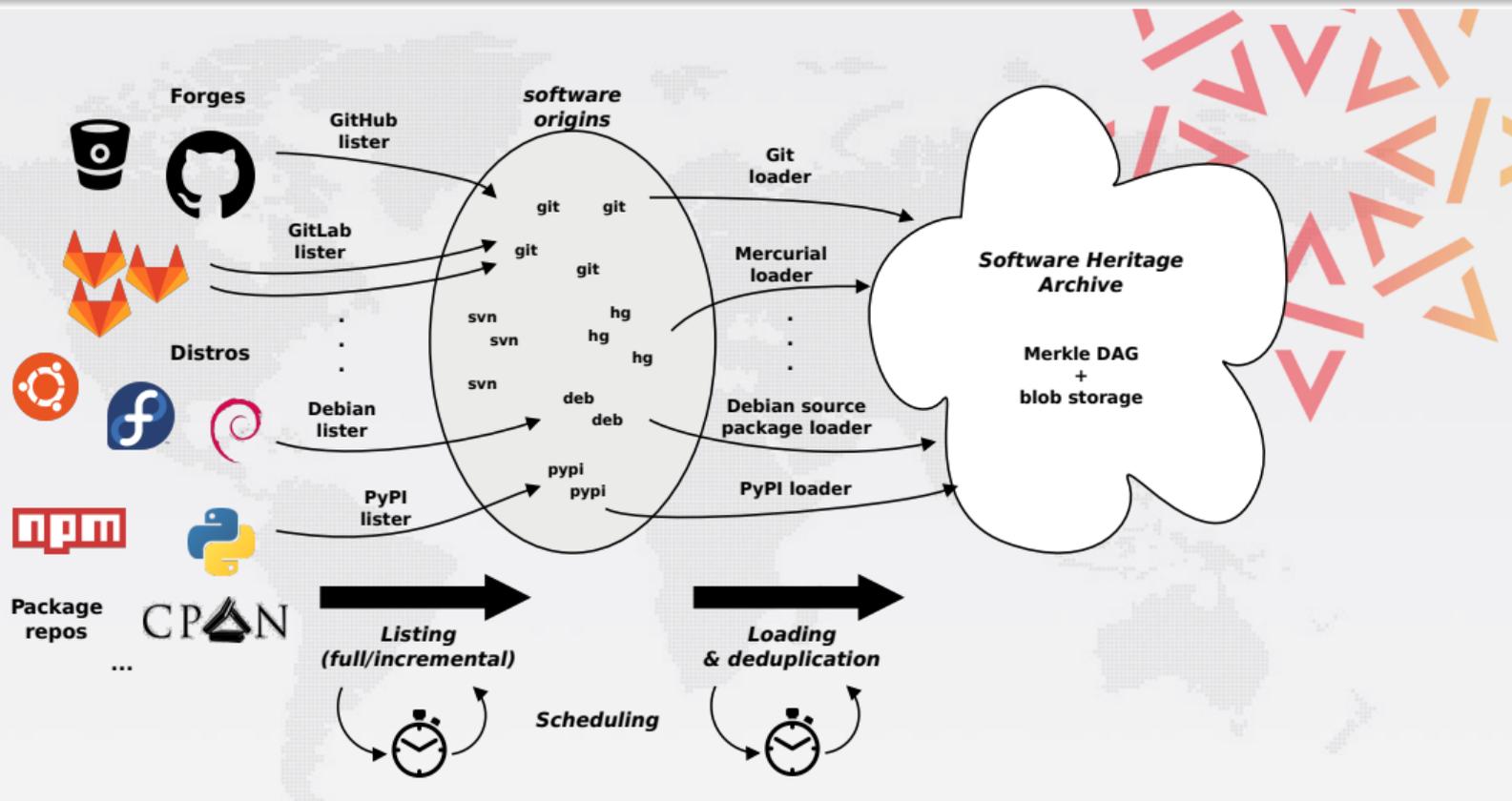
- file **content** (= blobs)
- **revisions** (= commits), with full metadata
- **releases** (= tags), ditto
- where (**origin**) & when (**visit**) we found any of the above

... in a VCS-/archive-agnostic **canonical data model**

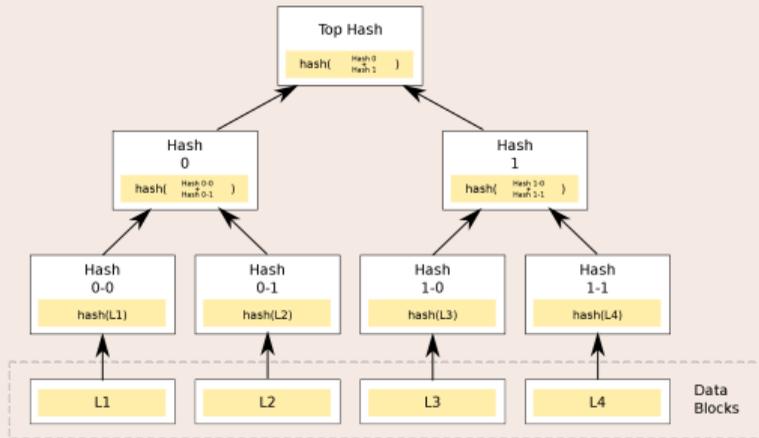
We DON'T archive

- homepages, wikis
- BTS/issues/code reviews/etc.
- mailing lists

Long term vision: play our part in a *"semantic wikipedia of software"*



Merkle tree (R. C. Merkle, Crypto 1979)

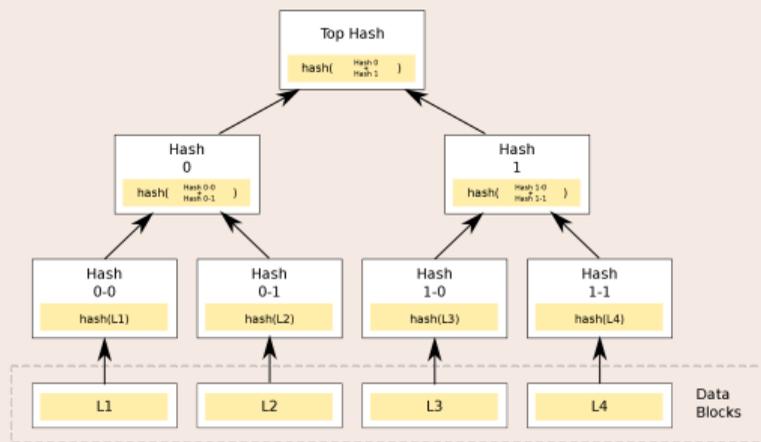


Combination of

- tree
- hash function

Merkle trees

Merkle tree (R. C. Merkle, Crypto 1979)



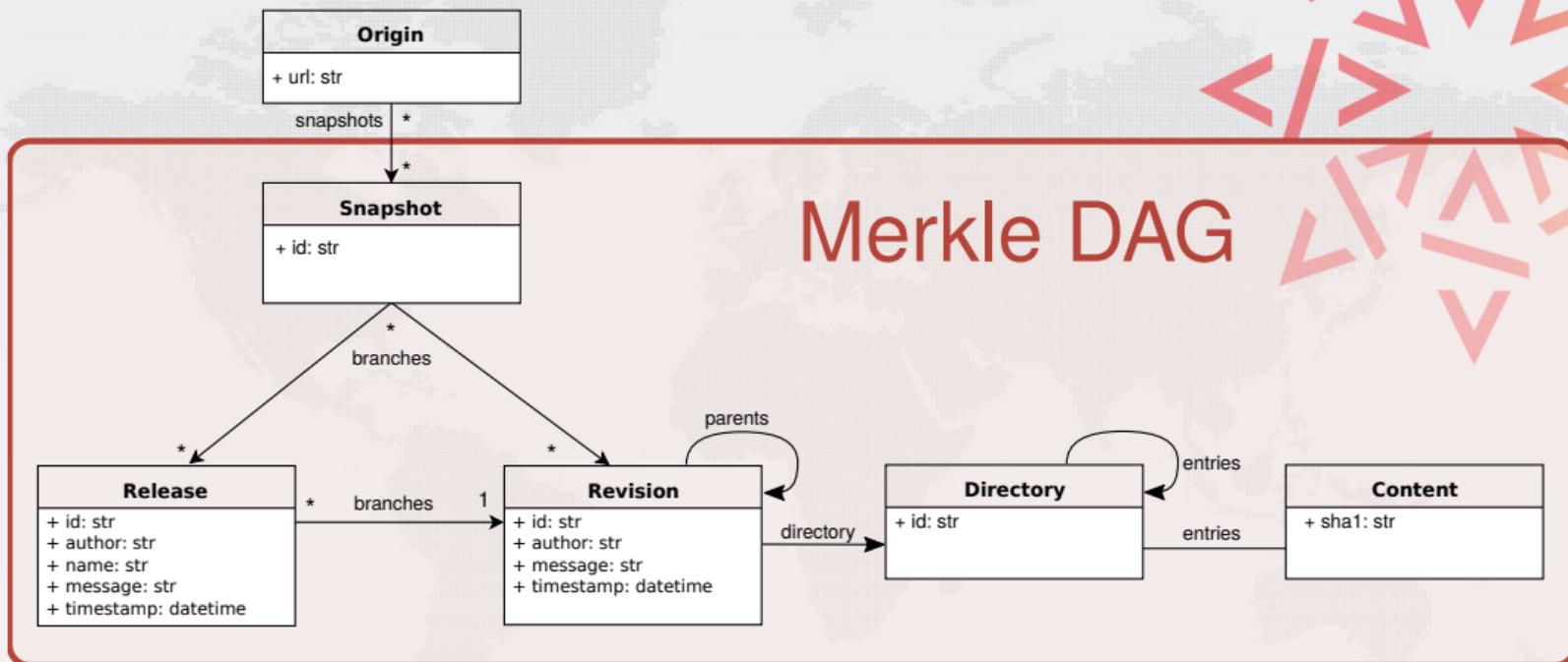
Combination of

- tree
- hash function

Classical cryptographic construction

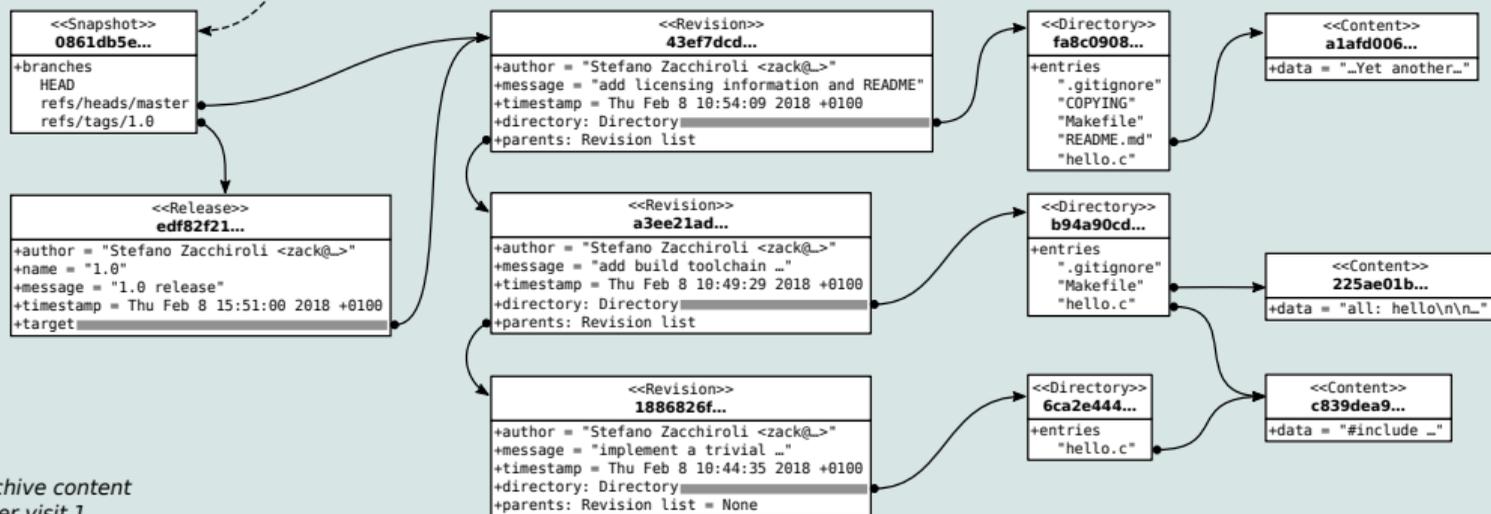
- fast, parallel signature of large data structures
- widely used (e.g., Git, blockchains, IPFS, ...)
- built-in deduplication

The archive: a (giant) Merkle DAG



The archive: a (giant) Merkle DAG

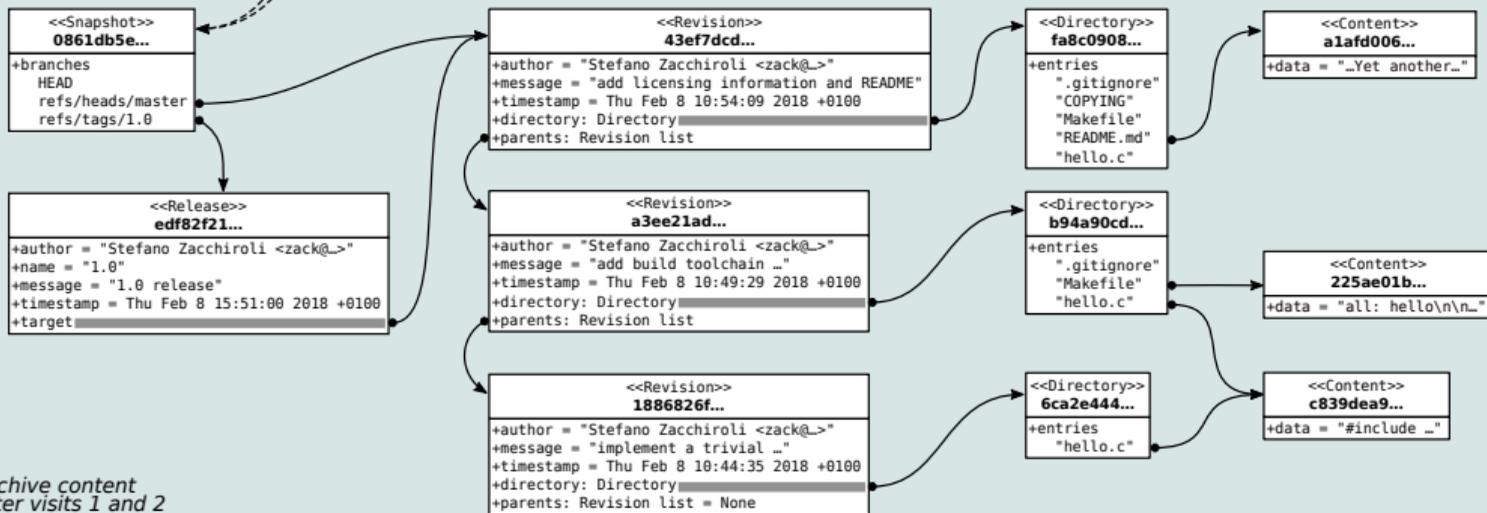
origin https://forge.softwareheritage.org/source/helloworld.git
visit 1
snapshot 0861db5e...
timestamp Fri Feb 9 12:38:45 2018 +0100



Archive content
after visit 1

The archive: a (giant) Merkle DAG

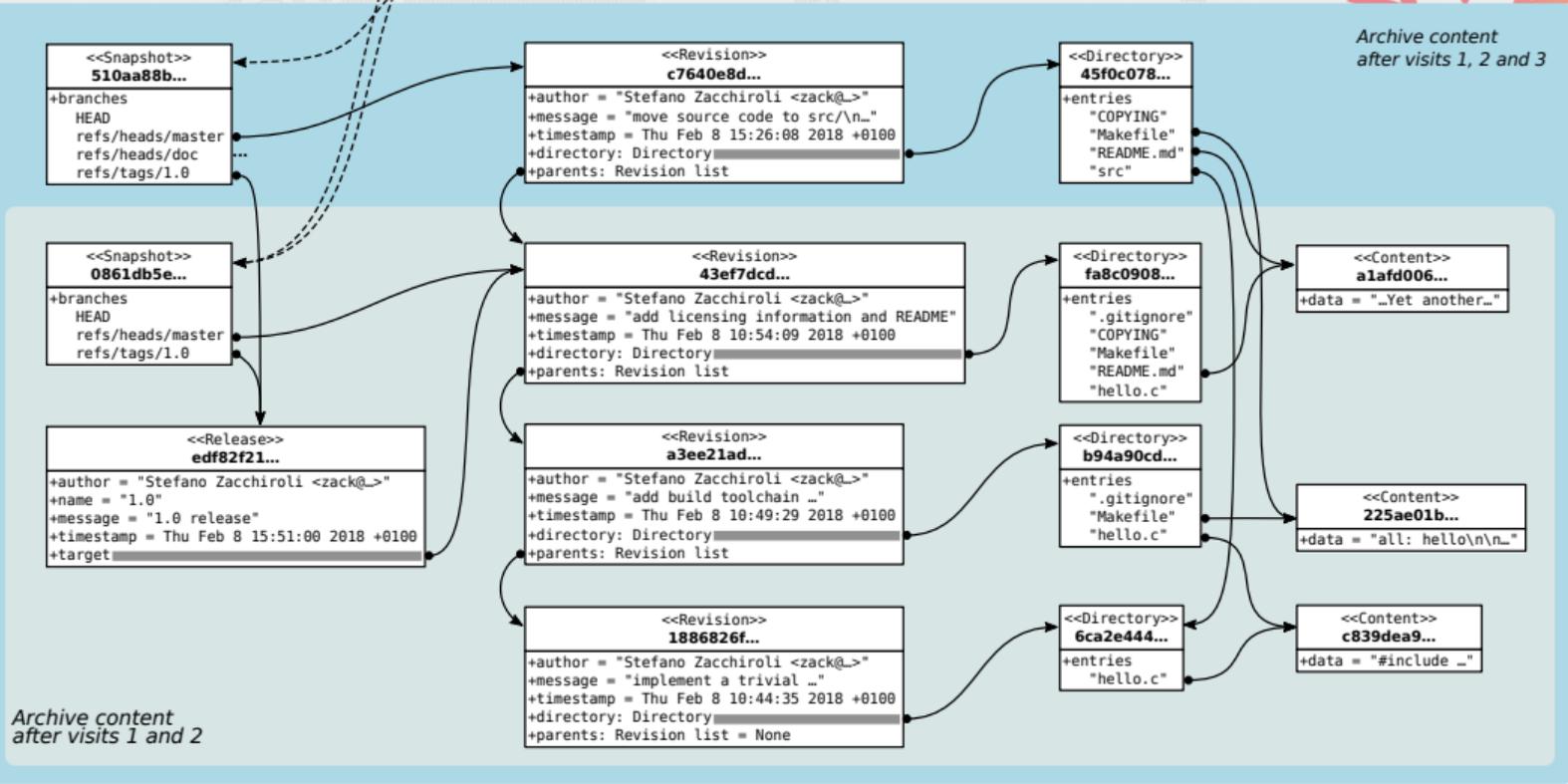
origin	visit	snapshot	timestamp
https://forge.softwareheritage.org/source/helloworld.git	1	0861db5e...	Fri Feb 9 12:38:45 2018 +0100
https://forge.softwareheritage.org/source/helloworld.git	2	0861db5e...	Fri Feb 9 13:29:00 2018 +0100



Archive content
after visits 1 and 2

The archive: a (giant) Merkle DAG

origin	visit	snapshot	timestamp
https://forge.softwareheritage.org/source/helloworld.git	1	0861db5e...	Fri Feb 9 12:38:45 2018 +0100
https://forge.softwareheritage.org/source/helloworld.git	2	0861db5e...	Fri Feb 9 13:29:00 2018 +0100
https://forge.softwareheritage.org/source/helloworld.git	3	510aa88b...	Fri Feb 9 15:52:50 2018 +0100

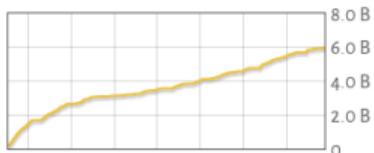


Archive content after visits 1 and 2

Archive content after visits 1, 2 and 3

Source files

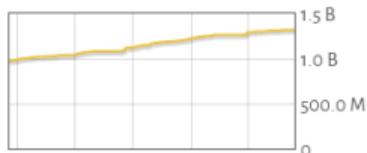
6,006,503,960



Jan Jul Jan Jul Jan Jul Jan
2016 2016 2017 2017 2018 2018 2019

Commits

1,326,776,432



Apr Jul Oct Jan Apr
2018 2018 2018 2019 2019

Projects

89,301,694



Apr Jul Oct Jan Apr
2018 2018 2018 2019 2019

GitHub

debian



GitLab

Google code

GITORIOUS

GNU

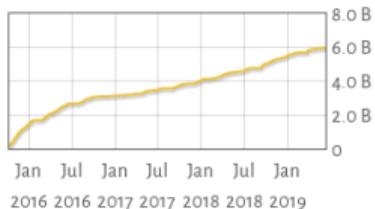
HAL
archives-ouvertes.fr

Inria
inventeurs du monde numérique

python™
Package Index

Source files

6,006,503,960



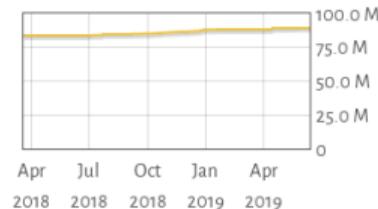
Commits

1,326,776,432



Projects

89,301,694



GitHub

debian



GitLab

Google code

GITORIOUS

GNU

HAL
archives-ouvertes.fr

Inria
inventeurs du monde numérique

python
Package Index

- 410 TB (uncompressed) blobs, 20 B nodes, 280 B edges
- The *richest* public source code archive, ... and growing daily!

Browser-based interface to browse the Software Heritage archive

<https://archive.softwareheritage.org/browse/>

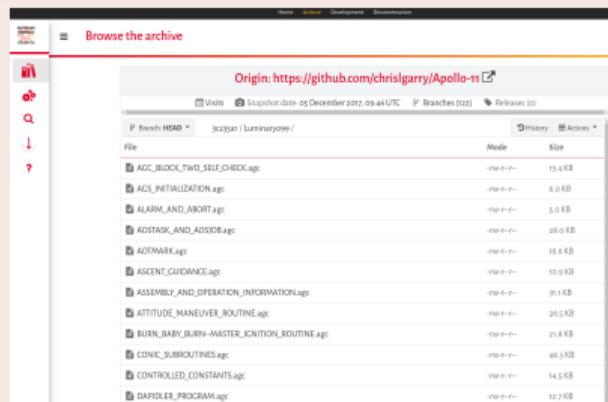
Features

- all **REST API features**, but good looking :-)
 - browsing: snapshots → revisions → directories → contents ...
 - access to metadata and crawling information
- **origin search**, as full text indexing of origin URLs
- bulk **download**, via integration with the Vault

Margaret Hamilton today



The Apollo 11 source code in SWH



The screenshot shows the SWH interface for the repository `sc3b51 / Luninary099 /`. The origin is `https://github.com/chrisgarry/Apollo-11`. The interface includes a search bar, a commit history table, and a file list.

File	Mode	Size
<code>ADC_BLOCK_TWO_SELF_CHECK.agc</code>	mp+r--	15.4 KB
<code>ACS_INITIALIZATION.agc</code>	mp+r--	6.0 KB
<code>ALARM_AND_ABORT.agc</code>	mp+r--	5.0 KB
<code>ADSTAR_AND_ADSDS.agc</code>	mp+r--	28.0 KB
<code>ADTMARK.agc</code>	mp+r--	16.6 KB
<code>ASCENT_GUIDANCE.agc</code>	mp+r--	10.9 KB
<code>ASSEMBLY_AND_OPERATION_INFORMATION.agc</code>	mp+r--	30.1 KB
<code>ATTITUDE_MANEUVER_ROUTINE.agc</code>	mp+r--	26.5 KB
<code>BURN_BABY_BURN-MASTER_KC110N_ROUTINE.agc</code>	mp+r--	21.8 KB
<code>CDMIC_SUBROUTINES.agc</code>	mp+r--	46.5 KB
<code>CONTROLLED_CONSTANTS.agc</code>	mp+r--	14.5 KB
<code>DAPFLER_PROGRAM.agc</code>	mp+r--	12.7 KB

Some pointers

- Entry point
- Burn, baby, burn!

Visiting the archive: the Quake 3 source code

John Carmack



The Quake 3 source code in SWH

A screenshot of the Software Heritage (SWH) archive interface. The page title is "Browse the archive". The origin is listed as "https://github.com/id-Software/Quake-III-Arena". The current branch is "HEAD" with commit hash "c8f0702". A table lists the files and directories in the archive:

File	Mode	Size
code	d-----	
common	d-----	
icc	d-----	
libs	d-----	
q3asm	d-----	
q3map	d-----	
q3radant	d-----	
ui	d-----	
COPYING.txt	-rwxr-xr-x	14.8 KB
README.txt	-rwxr-xr-x	8.8 KB

Below the table, there is a link for "README.txt" and a red heading "Quake III Arena GPL source release".

Some pointers

- Entry point
- What the f...



Alpha version

Search Software Heritage origins to browse

Origin type	Origin browse url	Visit status
git	/browse/origin/git/url/https://github.com/emacsimize/org-admin/	✓
git	/browse/origin/git/url/https://github.com/emacsattic/org-babel-plugins/	✓
git	/browse/origin/git/url/https://github.com/giorgos-pontikakis/emacs.d/	✓
git	/browse/origin/git/url/https://github.com/emacsmirror/org-oddmuse/	✓
git	/browse/origin/git/url/https://github.com/ajitgeorge/emacs.d/	✓
git	/browse/origin/git/url/https://github.com/emacsmirror/wn-org/	✓
git	/browse/origin/git/url/https://github.com/wizardxbl/EmacsOrg/	✓
git	/browse/origin/git/url/https://github.com/byorgey/noah-emacs/	✓
git	/browse/origin/git/url/https://github.com/jorgenschaefer/emacs-ixio/	✓
git	/browse/origin/git/url/https://github.com/visionnoob/.emacs.d_org-init/	✓
git	/browse/origin/git/url/https://github.com/timyguo/dotemacs.org/	✓
git	/browse/origin/git/url/https://github.com/jorgeavelar/dotemacs/	✓
git	/browse/origin/git/url/https://github.com/emacsmirror/emacswiki.org/	✗
git	/browse/origin/git/url/https://github.com/emacsmirror/org-extension/	✓
git	/browse/origin/git/url/https://github.com/adversary-org/emacs-themes/	✓





SWH origin: <https://github.com/hylang/hy>

Go to origin

SWH origin visits

Visits overview

- Total number of visits: 17
- Last full visit: ✓ 02 March 2018, 20:03 UTC
- First full visit: ✓ 04 August 2015, 22:26 UTC
- Last visit: ✓ 02 March 2018, 20:03 UTC

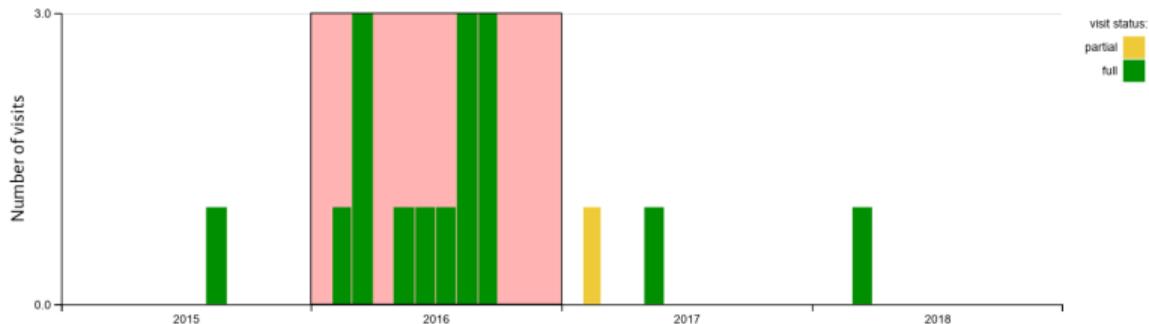
Visits history

Show full visits with different snapshots

Show all full visits

Show all visits

Timeline



Calendar



Calendar

2016

January

Su	Mo	Tu	We	Th	Fr	Sa
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

May

Su	Mo	Tu	We	Th	Fr	Sa
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

September

Su	Mo	Tu	We	Th	Fr	Sa
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

February

Su	Mo	Tu	We	Th	Fr	Sa
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29					

June

Su	Mo	Tu	We	Th	Fr	Sa
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	

October

Su	Mo	Tu	We	Th	Fr	Sa
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

March

Su	Mo	Tu	We	Th	Fr	Sa
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

July

Su	Mo	Tu	We	Th	Fr	Sa
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

November

Su	Mo	Tu	We	Th	Fr	Sa
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30			

April

Su	Mo	Tu	We	Th	Fr	Sa
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30						

August

Su	Mo	Tu	We	Th	Fr	Sa
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

December

Su	Mo	Tu	We	Th	Fr	Sa
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

List

✓ 22 February 2016, 16:56 UTC

✓ 23 May 2016, 11:22 UTC

✓ 16 August 2016, 22:40 UTC

✓ 03 March 2016, 17:25 UTC

✓ 07 June 2016, 04:42 UTC

✓ 29 August 2016, 19:38 UTC

✓ 19 March 2016, 01:53 UTC

✓ 26 July 2016, 14:59 UTC

✓ 07 September 2016, 09:19 UTC

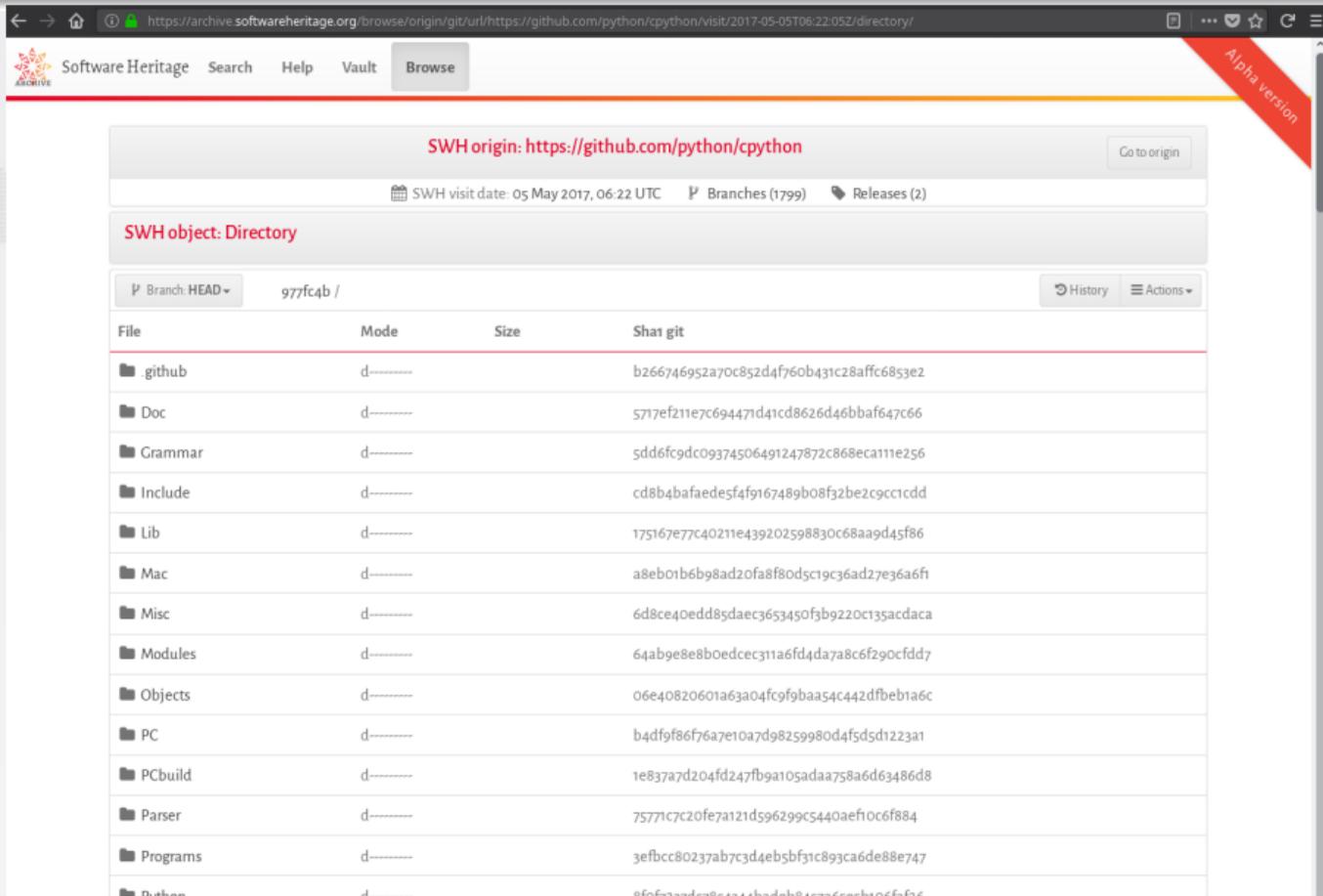
✓ 29 March 2016, 06:26 UTC

✓ 13 August 2016, 03:45 UTC

✓ 14 September 2016, 11:04 UTC



Web UI — directory browsing



The screenshot shows the Software Heritage web interface for a directory. At the top, the browser address bar shows the URL: `https://archive.softwareheritage.org/browse/origin/git/url/https://github.com/python/cpython/visit/2017-05-05T06:22:05Z/directory/`. The page header includes the Software Heritage logo and navigation links: Search, Help, Vault, and Browse. A red banner in the top right corner says "Alpha version".

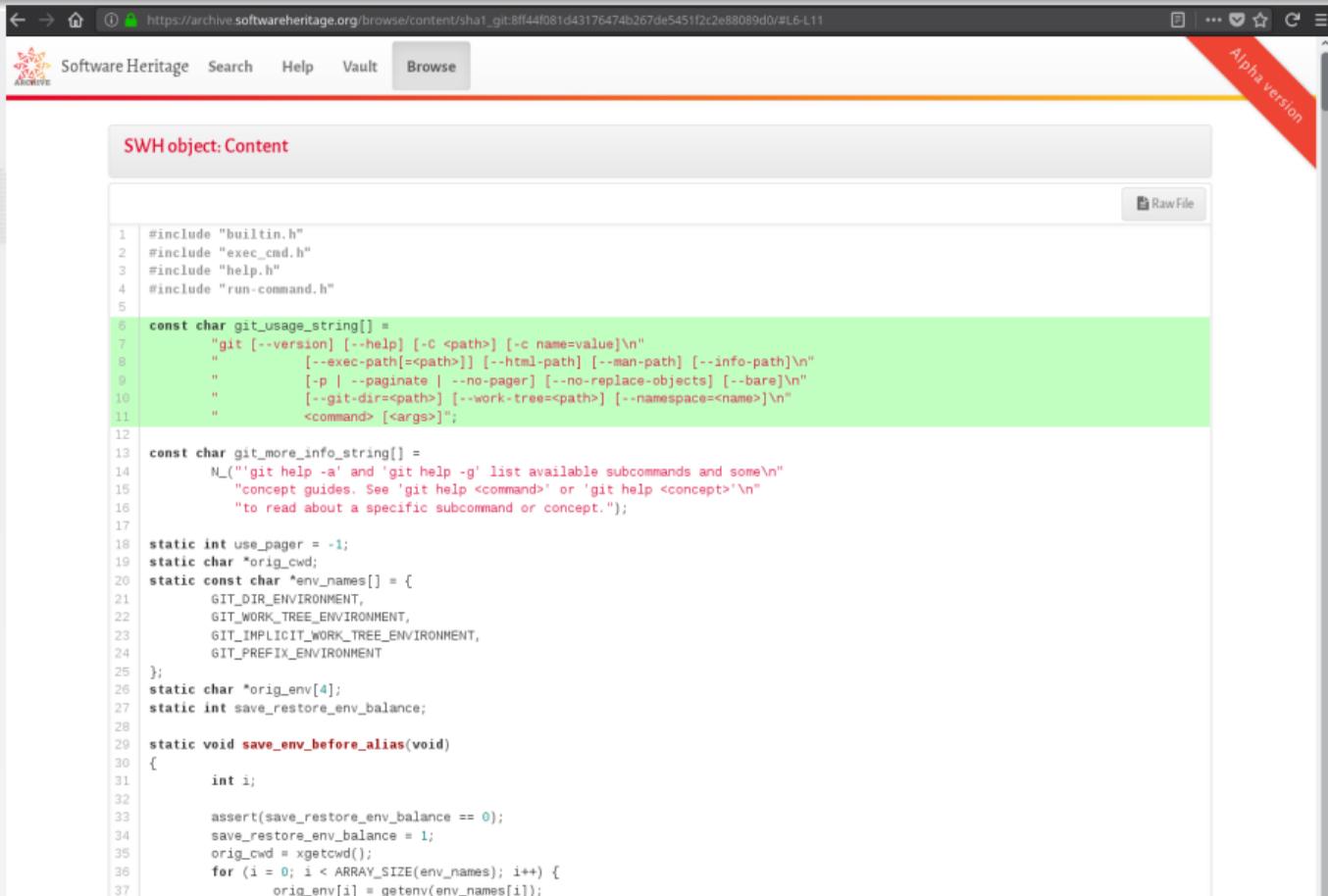
The main content area displays the SWH origin: `https://github.com/python/cpython` with a "Go to origin" button. Below this, it shows the SWH visit date: "05 May 2017, 06:22 UTC", and the number of branches (1799) and releases (2).

The SWH object is identified as a "Directory". The current branch is "HEAD" with commit hash "977fc4b". There are "History" and "Actions" buttons.

A table lists the directory contents with columns for File, Mode, Size, and Sha1 git:

File	Mode	Size	Sha1 git
github	d-----		b266746952a70c852d4f760b431c28affc6853e2
Doc	d-----		5717ef211e7c694471d41cd8626d46bbaf647c66
Grammar	d-----		5dd6fc9dc09374506491247872c868eca111e256
Include	d-----		cd8b4bafaedesf4f9167489b08f32be2c9cc1cdd
Lib	d-----		175167e77c40211e439202598830c68aa9d45f86
Mac	d-----		a8eb01b6b98ad20fa8f80d5c19c36ad27e36a6f1
Misc	d-----		6d8ce40edd85daec3653450f3b9220c135acdaca
Modules	d-----		64ab9e8e8b0edcec311a6fd4da7a8c6f29ocfdd7
Objects	d-----		06e40820601a63a04fc9f9baa54c442dfbeb1a6c
PC	d-----		b4df9f86f76a7e10a7d98259980d4f5d5d1223a1
PCbuild	d-----		1e837a7d204fd247fb9a105adaa758a6d63486d8
Parser	d-----		7571c7c20fe7a121d596299c5440aef10c6f884
Programs	d-----		3efbcc80237ab7c3d4eb5bf31c893ca6de88e747
Python	d-----		8f6f237dc28c2244badeb84c7a6e8cb106fa96

Web UI — syntax highlighting and selection



Software Heritage Search Help Vault Browse

Alpha version

SWH object: Content

Raw File

```
1 #include "builtin.h"
2 #include "exec_cmd.h"
3 #include "help.h"
4 #include "run-command.h"
5
6 const char git_usage_string[] =
7     "git [--version] [--help] [-C <path>] [-c name=value]\n"
8     "  [--exec-path[=<path>]] [--html-path] [--man-path] [--info-path]\n"
9     "  [-p | --paginate | --no-pager] [--no-replace-objects] [--bare]\n"
10    "  [--git-dir=<path>] [--work-tree=<path>] [--namespace=<name>]\n"
11    "  <command> [<args>]";
12
13 const char git_more_info_string[] =
14     N_("'git help -a' and 'git help -g' list available subcommands and some\n"
15     "  concept guides. See 'git help <command>' or 'git help <concept>'\n"
16     "  to read about a specific subcommand or concept.");
17
18 static int use_pager = -1;
19 static char *orig_cwd;
20 static const char *env_names[] = {
21     GIT_DIR_ENVIRONMENT,
22     GIT_WORK_TREE_ENVIRONMENT,
23     GIT_IMPLICIT_WORK_TREE_ENVIRONMENT,
24     GIT_PREFIX_ENVIRONMENT
25 };
26 static char *orig_env[4];
27 static int save_restore_env_balance;
28
29 static void save_env_before_alias(void)
30 {
31     int i;
32
33     assert(save_restore_env_balance == 0);
34     save_restore_env_balance = 1;
35     orig_cwd = xgetcwd();
36     for (i = 0; i < ARRAY_SIZE(env_names); i++) {
37         orig_env[i] = getenv(env_names[i]);
```

Software Heritage Search Help Vault Browse

Alpha version

SWH object: Revision

Revision [f1b94134a4b879bc55c3dacdb496690c8ebdc03f](#) authored by Vikram Fugro on 11 March 2016, 12:16 UTC

gstdecode: support alloc'ing vlc pictures with padding

Allocate the output vlc pictures with dimensions padded, as requested by the decoder (for alignments). This further increases the chances of direct rendering.

Signed-off-by: Jean-Baptiste Kempf <jb@videolan.org>

1 parent [6c813cb](#)

Files Changes

Showing 6 changed files with 109 additions and 53 deletions (6 / 6 diffs computed) [Compute all diffs](#)

modules/codecs/gstreamer/gstdecode.c [Unified](#) [Side-by-side](#) [View file](#)

```
@@ -179,6 +179,7 @@
179 179     VLC_UNUSED( p_ele );
180 180     decoder_t *p_dec = p_data;
181 181     decoder_sys_t *p_sys = p_dec->p_sys;
182 +   GstVideoAlignment align;
182 183
183 184     msg_Info( p_dec, "got new caps %s", gst_caps_to_string( p_caps ));
184 185
@@ -189,8 +190,9 @@
189 190     }
190 191
191 192     gst_vlc_dec_ensure_empty_queue( p_dec );
192 -
193 -   return gst_vlc_set_vout_fmt( &p_sys->vinfo, p_caps, p_dec );
193 +   gst_video_alignment_reset( &align );
194 +
195 +   return gst_vlc_set_vout_fmt( &p_sys->vinfo, &align, p_caps, p_dec );
196 196 }
```

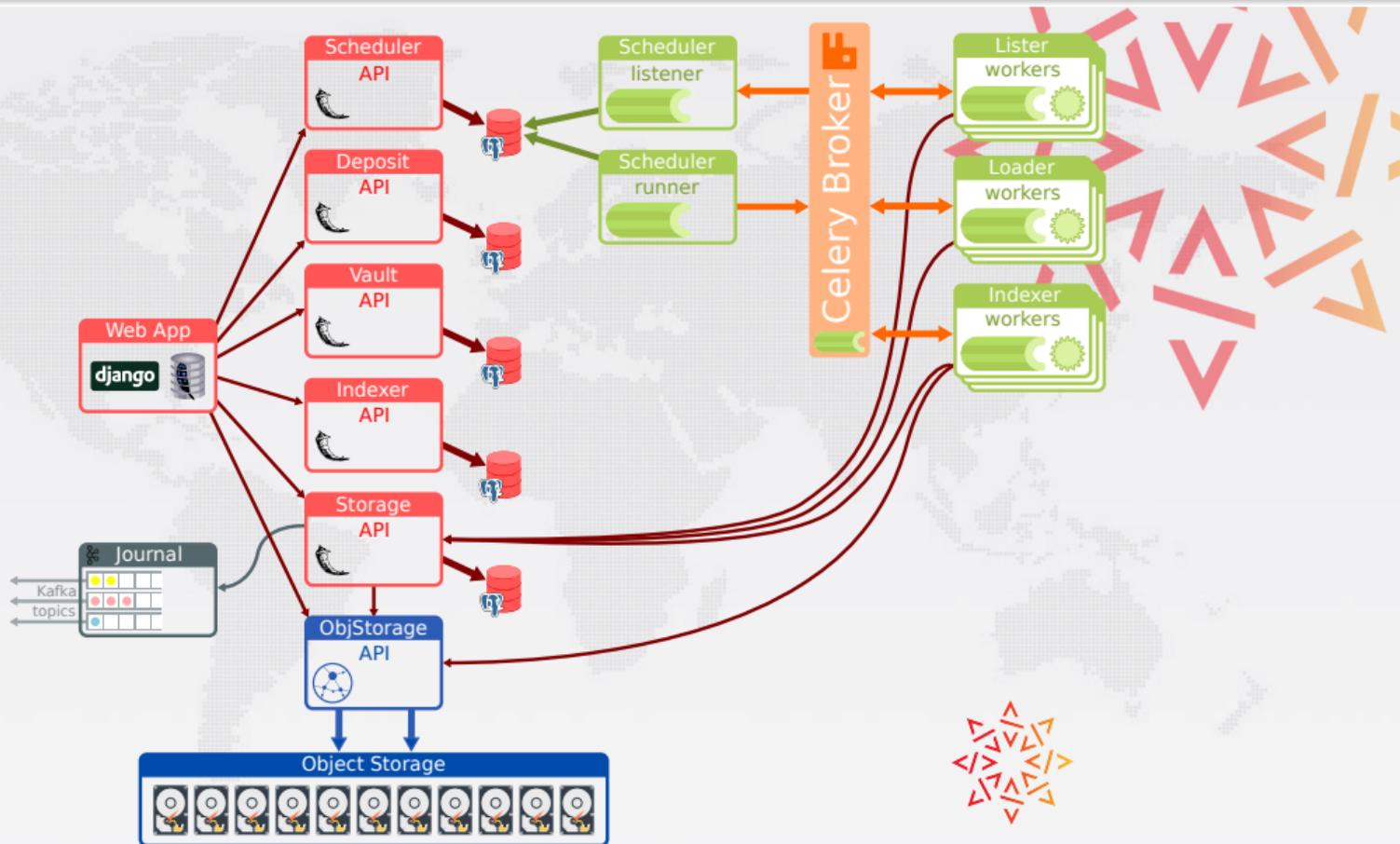
3rd party

- Debian, Puppet, Ceph
- PostgreSQL for metadata storage, with barman & pglogical
- Celery (RabbitMQ backend) for task scheduling
- Python3 and psycopg2 for the backend
- Django, Bootstrap, D3.js for Web stuff

in house

- *ad hoc* object storage (to avoid imposing tech to mirrors)
- data model implementation, listers, loaders, scheduler
- ~60 Git repositories (~20 Python packages, ~30 Puppet modules)
- ~60 kSLOC Python / ~12 kSLOC SQL / ~4 kSLOC Puppet
- licence choice: GPLv3 (backend) / AGPLv3 (frontend)

Deployment architecture



in house

- 2x hypervisors with ~20 VMs
- 1x high performance database server
- 2x dedicated storage server using
- 2x high density storage array (60 * 6TB => 300TB usable each)
- 3x nodes for a kafka+elasticsearch cluster

on Azure

- full object storage mirror
- full mirror of the database containing the graph
- workers for content indexing
- workers for download bundle preparation

classic FOSS development

- language: English
- development mailing list
<https://sympa.inria.fr/sympa/info/swh-devel>
- IRC
#swh-devel / FreeNode
- Forge
<https://forge.softwareheritage.org>
- Git, tasks, code review, etc.

for more information

<https://www.softwareheritage.org/community/developers/>

Yes, now you can!

Wayback machine for source code

retrieve the source code as it was

Open science

deposit scientific software (via HAL)

Reference catalog

use intrinsic identifiers for software

Universal knowledge base

store the knowledge about source code

And much, much more is in store!

With your help?

The Software Heritage Foundation

- independent
- long term mission
- multistakeholder

The community

- academia: Open Access, research
- industry: better software
- cultural heritage: **all** the software history

The mirror network

- resilience
- biodiversity

“Let us save what remains: not by vaults and locks which fence them from the public eye and use in consigning them to the waste of time, but by such a multiplication of copies, as shall place them beyond the reach of accident.”

Thomas Jefferson



Software Heritage

www.softwareheritage.org

@swheritage

Everybody is concerned, everybody can help build

The Library of Alexandria of code



- recover the past
- structure the future

A CERN for Software



- build better software
 - for industry
 - for society as a whole

Q: do you archive *only* Free Software?

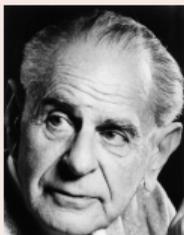
- We only crawl origins *meant* to host source code (e.g., forges)
- Most (~90%) of what we *actually* retrieve is textual content

Our goal

Archive **the entire Free Software Commons**

- Large parts of what we retrieve is *already* Free Software, today
- Most of the rest *will become* Free Software in the long term
 - e.g., at copyright expiration

Reproducibility is the key

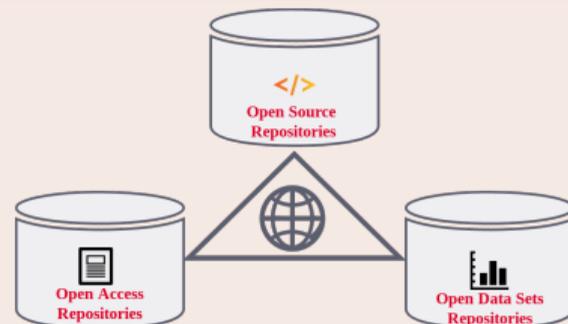


non-reproducible single occurrences are of no significance to science

Karl Popper, The Logic of Scientific Discovery, 1934

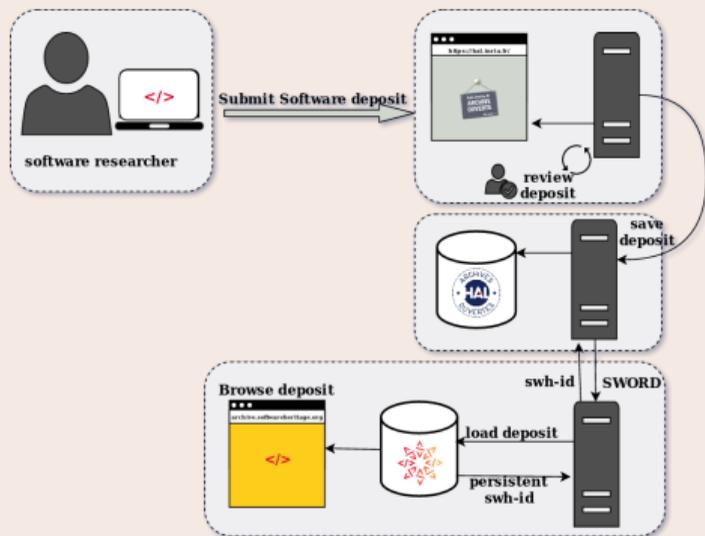
the software source code is special

- It **embodies the logic** of the data transformation.
- It **must** be included in strategies for *scientific knowledge preservation*.
- knowing the **exact version** of the software used is essential for reproducibility.



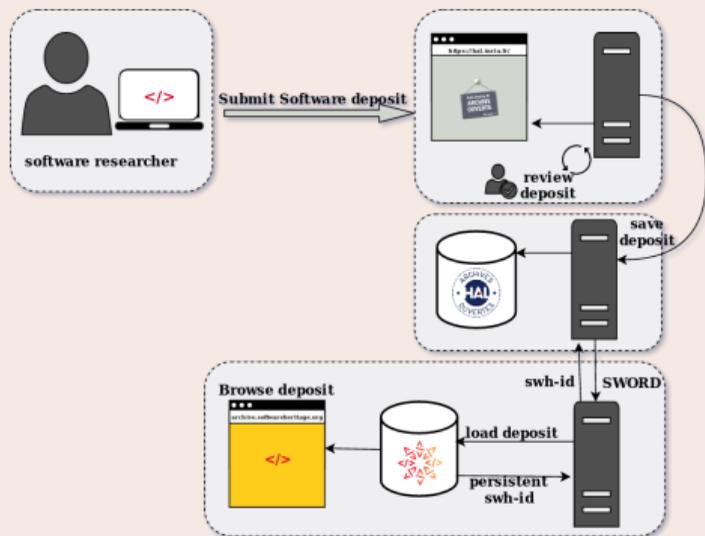
Deposit software in HAL

<http://hal.inria.fr/hal-01738741>



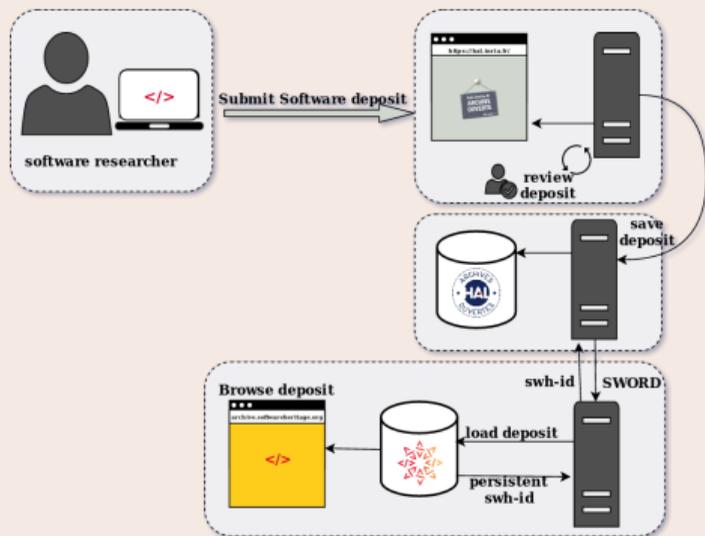
Deposit software in HAL

<http://hal.inria.fr/hal-01738741>



Generic mechanism:

- SWORD based
- review process
- versioning



Generic mechanism:

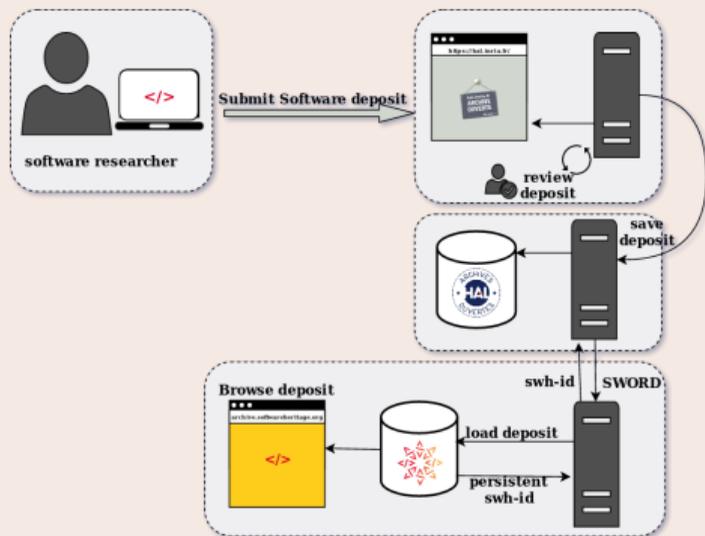
- SWORD based
- review process
- versioning

How to do it:

- **today:** deposit .zip or .tar.gz file (*guide*)
- **tomorrow:**
 - provide *SWH id* and metadata
 - include *metadata file* for automatic metadata extraction
 - ...

Deposit software in HAL

<http://hal.inria.fr/hal-01738741>



Generic mechanism:

- SWORD based
- review process
- versioning

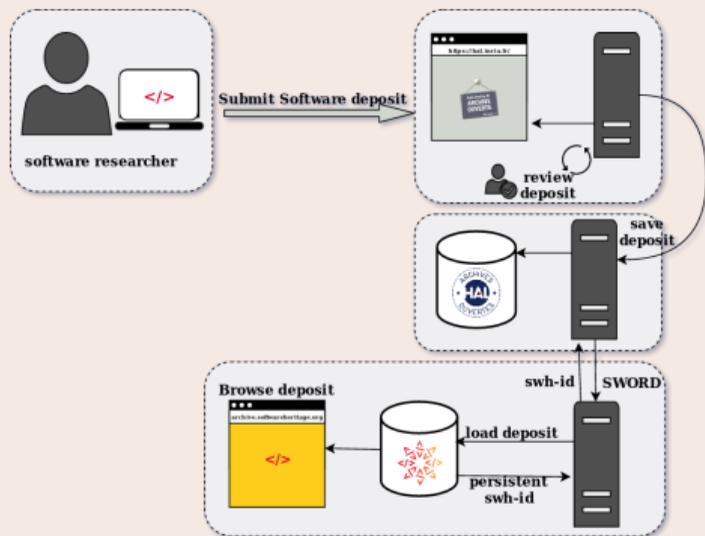
How to do it:

- **today:** deposit .zip or .tar.gz file (*guide*)
- **tomorrow:**
 - provide *SWH id* and metadata
 - include *metadata file* for automatic metadata extraction
 - ...

September 2018: **open to all** on <https://hal.archives-ouvertes.fr/>

Deposit software in HAL

<http://hal.inria.fr/hal-01738741>



Generic mechanism:

- SWORD based
- review process
- versioning

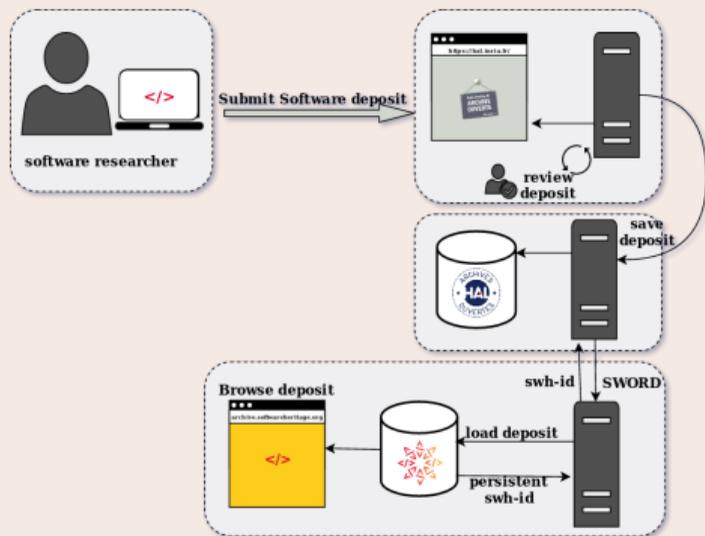
How to do it:

- **today:** deposit .zip or .tar.gz file (*guide*)
- **tomorrow:**
 - provide *SWH id* and metadata
 - include *metadata file* for automatic metadata extraction
 - ...

September 2018: **open to all** on <https://hal.archives-ouvertes.fr/>

Deposit software in HAL

<http://hal.inria.fr/hal-01738741>



Generic mechanism:

- SWORD based
- review process
- versioning

How to do it:

- **today:** deposit .zip or .tar.gz file (*guide*)
- **tomorrow:**
 - provide *SWH id* and metadata
 - include *metadata file* for automatic metadata extraction
 - ...

September 2018: **open to all** on <https://hal.archives-ouvertes.fr/>