

The background of the image is a photograph of a white bookshelf filled with numerous white and light-colored binders or folders. The binders are arranged in neat rows on the shelves, creating a sense of order and storage. The lighting is soft, highlighting the texture of the binders and the structure of the shelf.

SOFTWARE HERITAGE ARCHIVE



Diego Villanueva Berros

Pelayo Reguera García

Rubén del Rey Álvarez

A close-up portrait of Roberto Di Cosmo, a middle-aged man with a receding hairline, light blue eyes, and a grey beard. He is wearing a dark checkered shirt. The background is a soft, out-of-focus grey.

ROBERTO DI COSMO





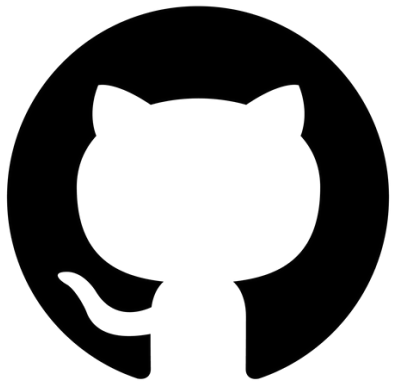
Software Heritage



GitLab

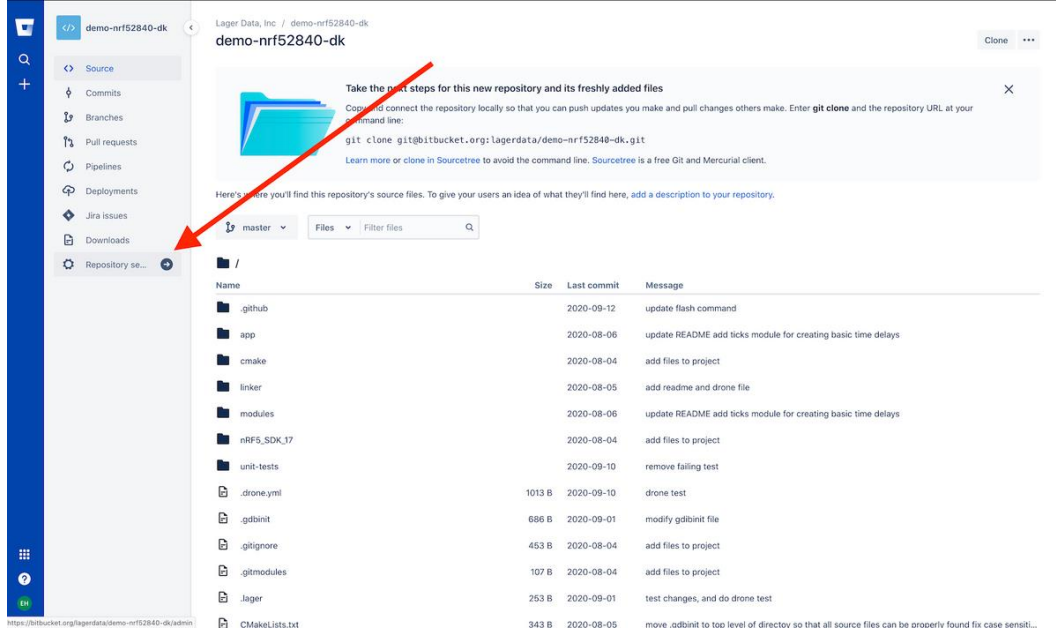
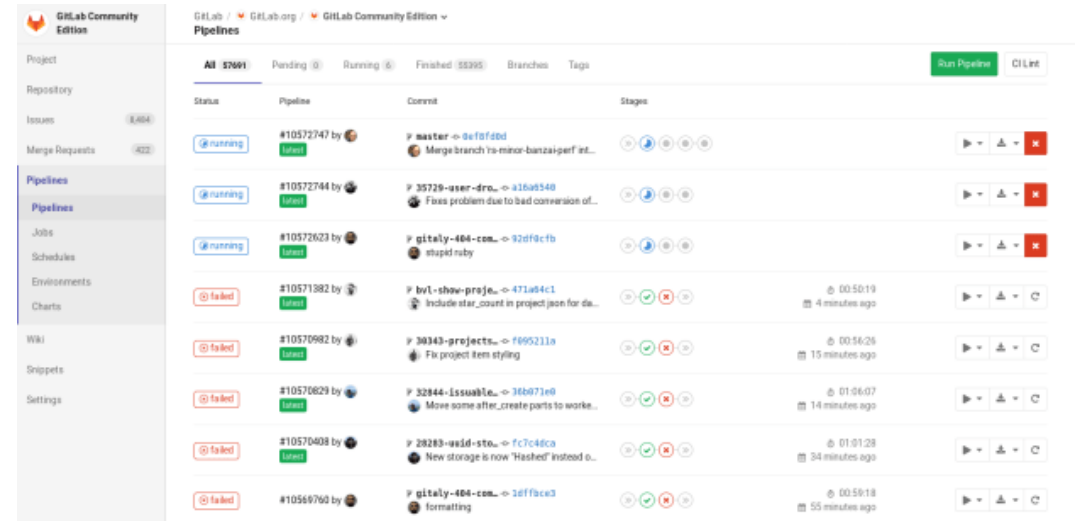
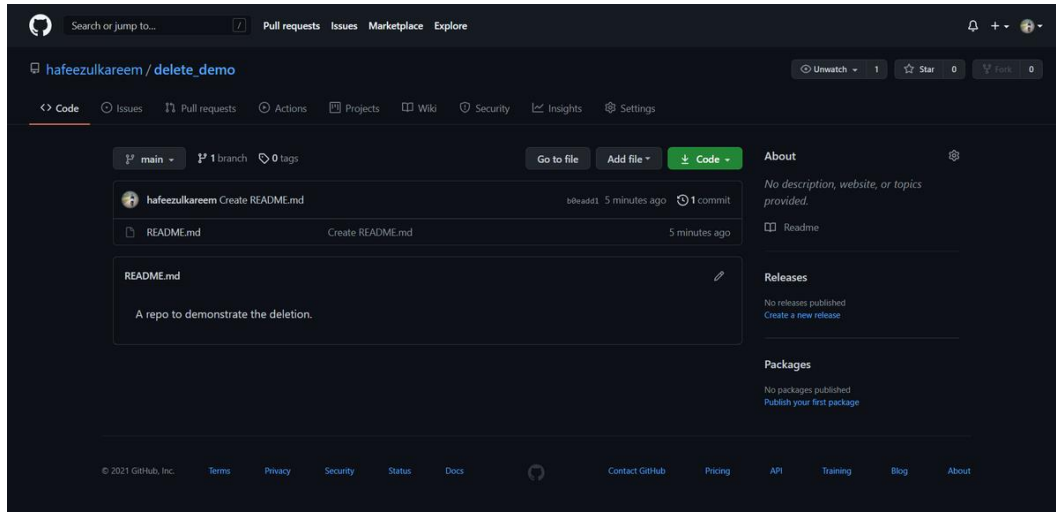


Bitbucket



GitHub



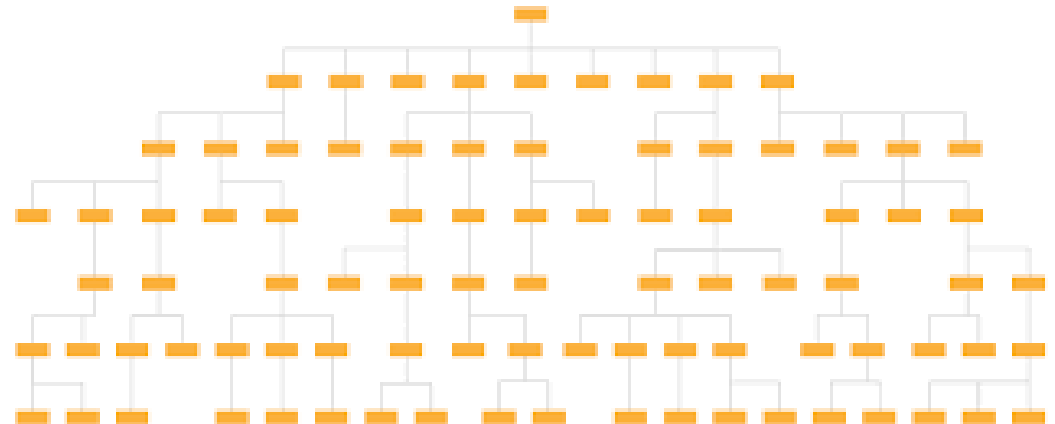
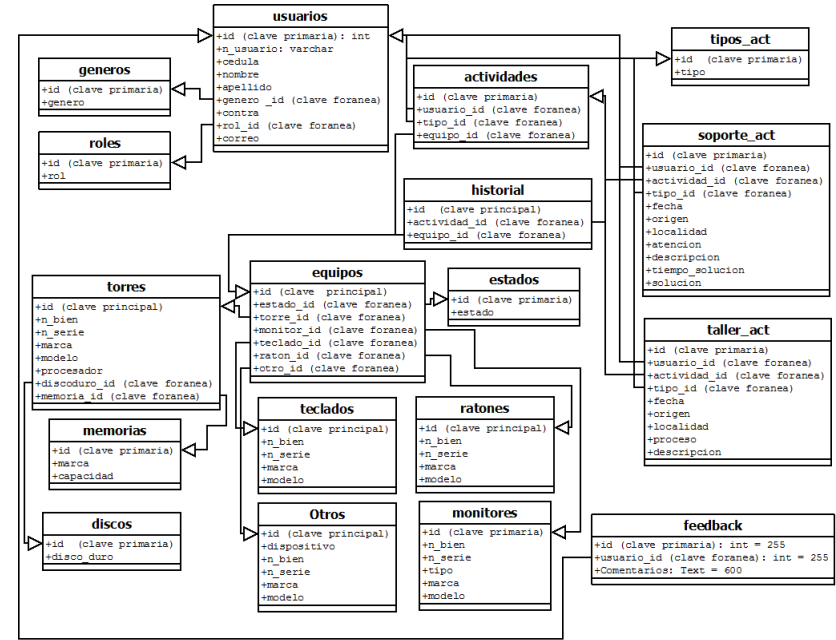




HOW TO STORE ALL THIS DATA

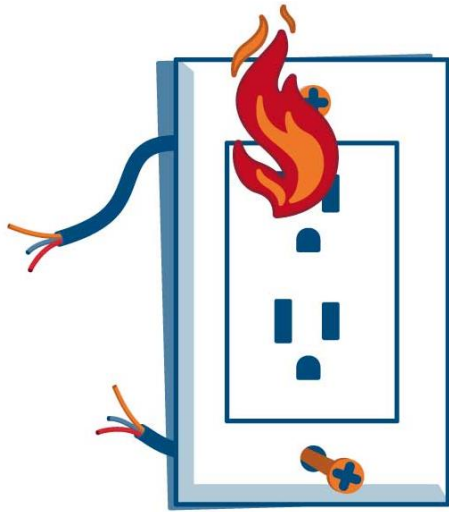


WHAT DATABASE MODEL TO USE?

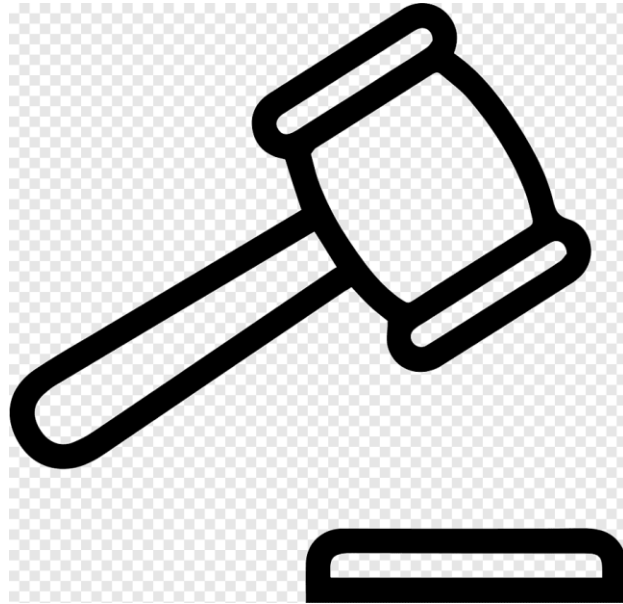


Problems that could arise and how to face them

Natural disasters



Modification in the legislation



HOW OFTEN TO STORE CONTENT AND HOW TO DO IT?



Periodically each three or four months





A formal agreement between the organization and the stakeholder



Visiting save.softwareheritage.org and uploading it yourself

Features

 Search Downloads Save code now Add forge now Help

You can contribute to extend the content of the Software Heritage archive by submitting an origin save request. To do so, fill the required info in the form below:

Origin type

Origin url

git



Submit

Help

Browse save requests

A "Save code now" request takes the following parameters:

- **Origin type:** the type of version control system the software origin is using. Currently, the supported types are:

- **git**, for origins using Git
- **hg**, for origins using Mercurial
- **svn**, for origins using Subversion
- **cvs**, for origins using CVS
- **bzr**, for origins using Bazaar

- **Origin url:** the url of the remote repository for the software origin.

In order to avoid saving errors from Software Heritage, you should provide the clone/checkout url as given by the provider hosting the software origin.

It can easily be found in the web interface used to browse the software origin.

For instance, if you want to save a **git** origin into the archive, you should check that the command `$ git clone <origin_url>` does not return an error before submitting a request.

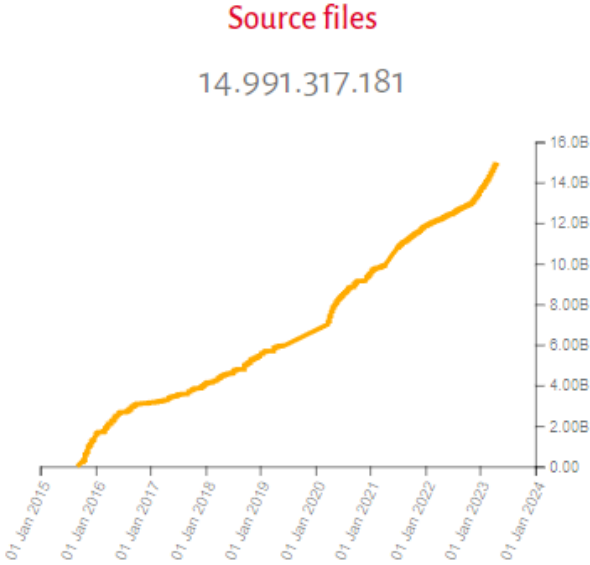
Once submitted, your save request can either be:

- **accepted:** a visit to the provided origin will then be scheduled by Software Heritage in order to load its content into the archive as soon as possible
- **rejected:** the provided origin url is blacklisted and no visit will be scheduled
- put in **pending** state: a manual review will then be performed in order to determine if the origin can be safely loaded or not into the archive

Once a save request has been accepted, you can follow its current status in the [submitted save requests list](#).

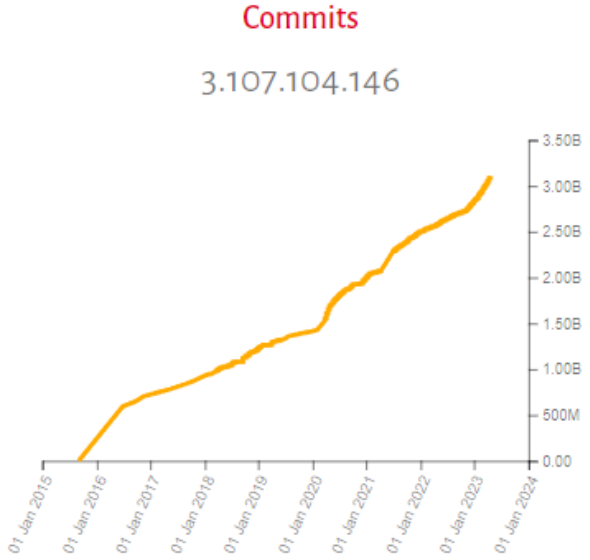
If you submitted requests while **authenticated**, you will be able to only display your own requests.

SOFTWARE HERITAGE ID



Directories

12.162.066.440



Intrinsic VS Extrinsic

Needs a standard



Needs a central authority



What needs an ID?

- Files
- Directories
- Commits





Software is the key

to access information and
a fundamental part of
human heritage

We are building the universal software archive



Collect
Preserve
Share

We **collect** and **preserve** software in source code form, because software embodies our technical and scientific knowledge and humanity cannot afford the risk of losing it.

Software is a precious part of our cultural heritage. We curate and make accessible all the software we collect, because only by **sharing** it we can guarantee its preservation in the very long term.

[Browse the archive](#)

[Discover our mission](#)

Search



- ☒ only show origins visited at least once
- ☒ filter out origins with no archived content
- ☐ search in metadata (instead of URL)

visit type:

any

- any
- bzr
- cran
- cvs
- deb
- deposit
- ftp
- git
- hg
- maven
- nixguix
- npm
- opam
- pypi
- svn
- tar

... or check out the

API

Overview

The long term goal of the Software Heritage initiative is to **collect** all publicly available software in source code form together with its development history, replicate it in order to ensure its **preservation**, and **share** it with everyone who needs it. The Software Heritage archive is growing over time as we crawl new source code from software projects and development forges.

Content

A significant amount of source code has already been ingested in the Software Heritage archive, see the [Archive Changelog](#) for notable changes to the archive over time. It currently includes the following software origins.

Regular crawling

These software origins get continuously discovered and archived using the [listers](#) implemented by Software Heritage.



2,166,787 origins



17,674 origins



22,054 origins



132,107 origins



7,955 origins



172,278,853 origins



3,997,686 origins



30,945 origins



354 origins



Features

Search

Downloads

Save code now

Add forge now

Help

google

☒ only show origins visited at least once

☒ filter out origins with no archived content

☐ search in metadata (instead of URL)

visit type: any

Origin type	Origin url	Archiving status
Checking	https://muckibude.googlecode.com/hg/	Checking
Checking	https://404forum-testclone.googlecode.com/hg/	Checking
Checking	https://github.com/sanghkaang/google-interview-university	Checking
Checking	https://github.com/VictorMiBe/Google-Maps-API	Checking
Checking	http://osgitrail.googlecode.com/svn/	Checking
Checking	https://github.com/davionxiang/google-interview-university	Checking
Checking	https://new-attitude-cms.googlecode.com	Checking
Checking	http://vs2012.googlecode.com/svn/	Checking
Checking	http://start-menu2.googlecode.com/svn/	Checking
Checking	http://gyropode-insarennnes-4eii.googlecode.com/svn/	Checking
Checking	https://github.com/mlssky/google-gson	Checking
Checking	https://bitbucket.org/pandinka/google-hashcode-2016.git	Checking
Checking	https://loverain9999-google-authenticator.googlecode.com	Checking
Checking	https://teodororosana-aaaaaaaaa.googlecode.com/hg/	Checking
Checking	https://github.com/xxz7694/googleads-admob-examples	Checking
Checking	http://cameronpurchase.googlecode.com/svn/	Checking
Checking	https://harryhao-myschedule.googlecode.com/hg/	Checking

QUESTIONS?

