

Software Heritage

a year of work, evolution and progress

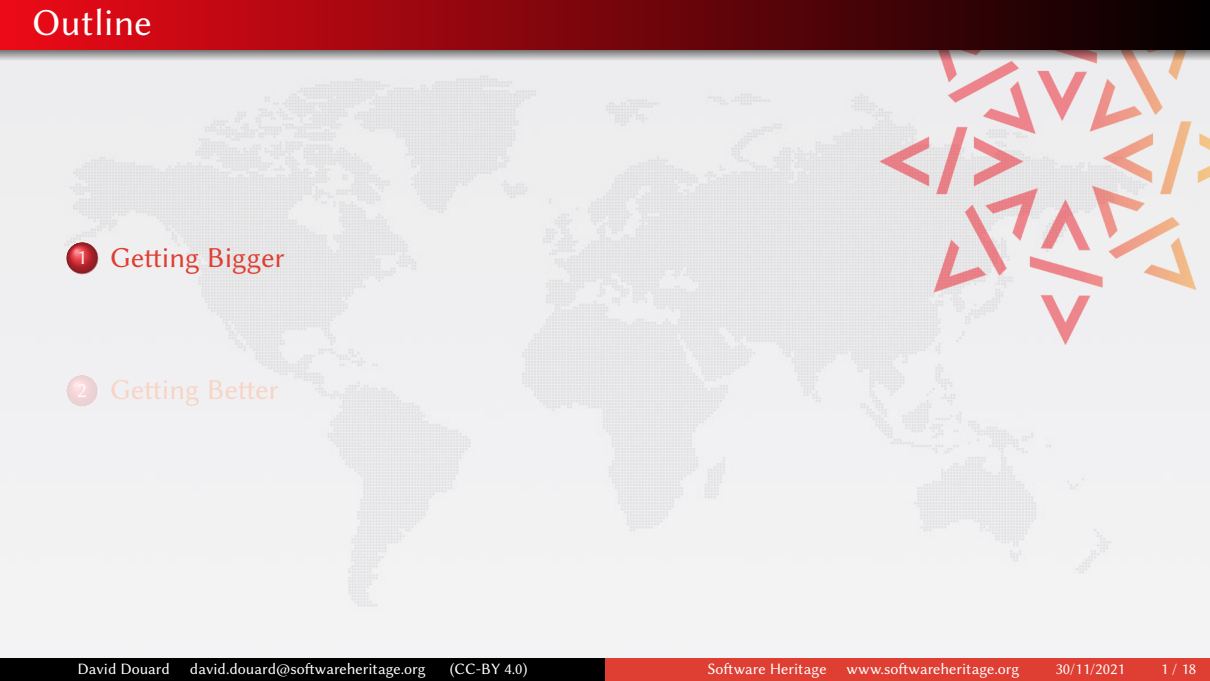
David Douard
Inria

30 November 2021
Inria



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE



1 Getting Bigger

2 Getting Better

SWH packages

- 4 new 'swh-' packages
- more than 2k git revisions
- 25 active contributors

from: (2020-12-31)

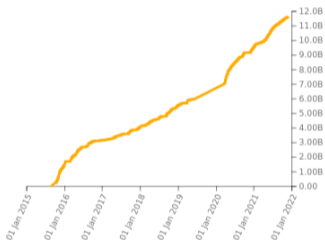
- python: 124445
- javascript: 8006
- java: 4379

to: (2021-11-29)

- python: 146903
- javascript: 9238
- java: 5915

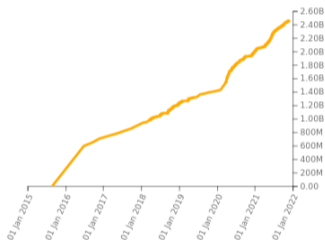
Source files

11 625 317 027



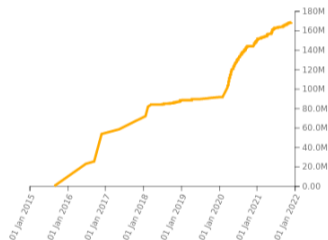
Commits

2 463 695 836

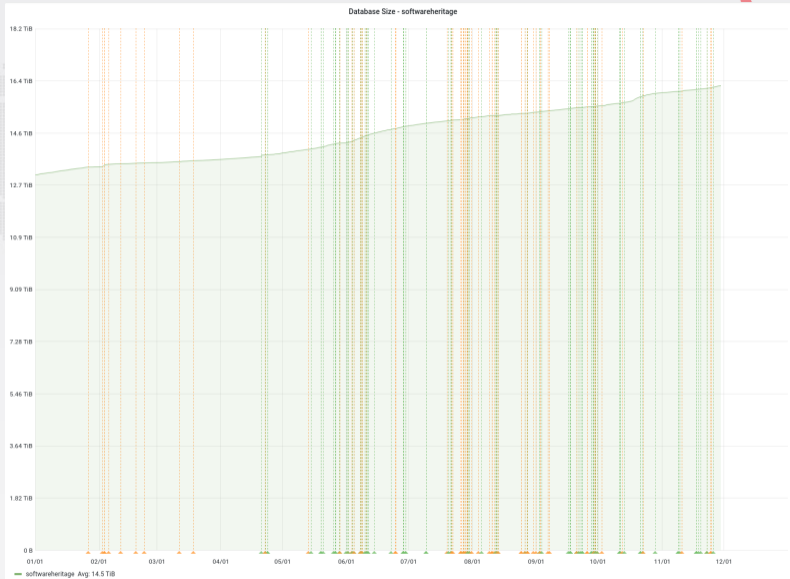


Projects

168 431 599



Database





1 Getting Bigger

2 Getting Better

New objstorage in development

- Design by Loic Dachary (Easter-Eggs) & al.
- Ceph-based
- Massive scale up

New storage backend to be deployed

- Design by Valentin Lorentz (SWH) & al.
- Large scale benchmarking on Grid5000 by Vincent Sellier (SWH)
- Massive (concurrent-write) scale up

- On-going large scale experiments at ENEA

Improved vault design

- Better stability
- Better performances



Many of improvements on the Deposit

- Integration with InvenioRDM/Zenodo in progress (Cottage Labs)
- Allow metadata-only deposits
- Better UI

Better Save-Code-Now behavior

- Improved architecture design by Antoine R. Dumont (SWH) & al.
- Improved reporting to the user of the status of a save code now request
- Much better latency

New Software Heritage Filesystem

- Design by Thibault Allançon & al.
- User-space POSIX filesystem to browse the Software Heritage Archive

Better search capabilities and performances

- New Elasticsearch-based search engine
- Support for searching in matadata

- Design by Nicolas Dandrimont (SWH) & al.
- Better overall "scheduling efficiency"

- New Extrinsic Metadata data model
- New APIs to retrieve metadata

- Ongoing work on improving the technical documentation
- Ongoing work on providing user documentation
- Development of a FAQ page

Better overall stability

- Improved global uptime of the infrastructure
- Reduced maintenance cost

- Much better Mercurial loader (Raphaël Gomez & al., Octobus)
- Better Subversion loader,
- Better CVS loader (tefan Sperling,
- New tuleap lister (Boris Baldassari, Castalia)
- New Sourceforge lister (Raphaël Gomez, Octobus)
- New Opam lister and loader (Leo Andres, OCamlPro)
- New Maven lister (Boris Baldassari, Castalia)
- New bazaar loader in progress (Raphaël Gomez, Octobus)

Thank you

- Thank you all for your involvement in the SWH project
- Stay tuned for the 2022 Roadmap (next january)