

# ESTABLISHING AN OPEN-SOURCE PACKAGE "ARCHIVE"

**Euan Cochrane**

*Yale University Library  
USA*

*euan.cochrane@yale.edu  
0000-0001-9772-9743*

**Rafael Gieschke**

*University of Freiburg  
Germany*

*rafael.gieschke@rz.uni-freiburg.de  
0000-0002-2778-4218*

(Extended) Abstract - Many Linux-based operating systems use a package management system that enables users to install a wide range of applications using one or more simple workflows, without needing to find and download the applications from their original publishers. The package management systems also resolve dependencies for users by finding and installing any dependent-applications or "packages" that are needed in order to run the application that the user is trying to install. These workflows greatly improve the experience of working with the operating systems and save a great deal of time for the end-users. When setting up the Emulation as a Service Infrastructure (EaaSI) platform to work with Linux-based operating systems we have encountered a number of issues when working with the operating systems' built-in Package Management Systems (PKMS). The PKMSs usually include a list of servers that host the packages that the PKMS can install for the user. We have found that often those lists are out-of-date and point to servers that no longer exist. In some cases, it is impossible or nearly impossible to find alternative servers that are still actively serving the packages, and where they are available the speed/bandwidth is often much slower/limited for packages for older operating system versions than it was when the operating systems were current. Even when an alternative server can be found that is still actively serving the packages for a legacy operating system, the average end-user can often struggle to understand how to point the PKMS at a custom server as this often requires editing relatively obscure configuration files within the Operating System.

To address these issues the EaaSI program of work and Yale University Library have launched a spin-off project to create a central "archive"<sup>1</sup> of compiled, open-source software packages. The "archive" will host copies of packages for many versions of legacy operating systems and make them freely available to any users, whether working with EaaSI or using the legacy operating systems in other contexts. The archive is hosted directly on a public S3-compatible bucket without an additional frontend server to improve availability, reduce maintainability, allow for practically limitless scaling, and enable users to easily explore and clone its contents (or parts thereof). In addition, the EaaSI software is being updated to enable instance administrators to configure the platform to automatically re-route connections being made to the package servers configured as defaults in legacy Linux-based operating systems and dynamically and seamlessly re-map the connections to the appropriate locations in the new package "archive". This is achieved either by automatically manipulating the DKMS's configuration files or by transparently routing network requests to the originally configured domain to an emulated version of the DKMS's server in a virtual network running in EaaSI.

Unlike Software Heritage [1], the world's preeminent archive of software source code, this project aims to primarily maintain an archive of compiled binaries and not the application source code. While there are some operating systems that use a PKMS that dynamically compiles from source code

<sup>1</sup> This term is used very loosely in this context

when installing applications, these are rare<sup>2</sup> and we are comfortable with the possibility of overlapping with the work of the software heritage team in this limited area.

There is a wide scope of future work for this project. In spite of being comfortable mildly overlapping on the work of the Software Heritage team, we are also interested in potentially creating reproducible builds of the packages in our “archive” within the EaaSI platform, using source code Sourced from Software Heritage. Doing the compilation from source within the EaaSI platform would provide an audit trail to further strengthen the trustworthiness of the packages provided in our “archive”.

Another future extension will be enabling sourcing packages from multiple sources, so that users can use one URL for any version of the same operating system, even if the packages were originally collected from different origin servers by our project. Here, a useful addition could be to seamlessly provide provenance information in S3 for individual packages, i.e., individual files in a directory sourced from multiple different origins.

In the regular use and development of repositories used by PKMSs the specific versions of the packages supplied by the repository are regularly updated and changed. IT is conceivable that users of our archive would need access to a specific version of a package, something that might not be possible if we only included the last version made available in the original repository. Therefore, a further (but much more elaborate) future extension could be to include different versions of the same package, a service offered for Debian at <https://snapshot.debian.org/>.

Further developments could include archiving the repositories of other package managers used in non-operating system contexts, e.g., PKMSs used for acquiring libraries for programming languages, for example., the Python Package Index (PyPI), Node.js’ npm, or various Docker registries (e.g., DockerHub, GitHub Container Registry). Here, a particular focus could be on providing time-travel functionality (as offered by [snapshot.debian.org](https://snapshot.debian.org/)) as software projects often do not fully specify their required library versions (e.g., by only specifying a minimum version instead of an exact version) and do not work anymore when newer versions of their required libraries become available at a later time and are selected by the package manager. Here, a bigger focus would probably be on the integration into the system than on duplicating the archive.

In this poster, we will provide a visual overview of the plans outlined in this extended abstract and hope

---

<sup>2</sup> Many of these PKMS will still serve the automatic build instructions as a special source

to use it to raise awareness of the “archive” in order to ensure extensive use of it once it is available.

**Keywords** – emulation, software preservation, open-source

**Conference Topics** – Sustainability: Real and Imagined; Immersive Information

## 1. REFERENCES

- [1] Software Heritage. Software heritage team. 2023. <https://www.softwareheritage.org/> (accessed 9/3/2023)

package as well that would not fall under the scope of Software Heritage.