

Virtual  
Collection  
Registry

Software

# Virtual Collection Registry

## Citing and sharing virtual research data collections

A scientific publication typically includes a bibliography based on persistent references. Ideally any data sets used in the reported research are referred to with persistent references as well. Persistence of references to data has become a topic of growing importance for researchers, as more and more research data are becoming available online for reuse. Furthermore, the need for advanced ways to version, group, and share data - both within and beyond the boundaries of a single research organisation - has become paramount. To meet and support such needs, new concepts and tools have been developed.

A **Virtual Collection (VC)** is a coherent set of links to a set of digital resources (e.g., annotated text, video fragments, interview data, etc.) that can be accessed and cited for future use. The links can originate from different archives, hence the term 'virtual'. The individual resources referred to in a VC may have been generated by different researchers and teams, and usually, they are maintained at the repositories of multiple organisations. For discovery and management purposes, a VC is described by its creator with a set of metadata with preservation of the original access permissions. When a VC is registered in a **Virtual Collection Registry (VCR)**, it is assigned a Persistent Identifier (PID) so that it can be cited and referred to as any other data collection.

The VCR developed in SSHOC will support researchers in arranging and re-using existing resources and collections for new purposes, avoiding the need to create new datasets/corpora for every project.

The VCR was originally created by CLARIN as a service to enable researchers, lecturers, and students to easily collect, group and share virtual collections of language materials and is further developed in SSHOC to also support users working with other data types.

## How does the Virtual Collection Registry work?

The Virtual Collection Registry offers researchers user-friendly support for referencing research data that may be heterogeneous and hosted at different repositories. After authentication, users can create and publish VCs that will remain linked to their identity, whereas unauthenticated users can only browse and search published VCs.

The VCR also offers an API for tight integration with, for instance, repository software. Such integration allows users to create and extend VCs directly from the repository user interface and avoids the use of copy and paste when creating VCs.

Currently, CLARIN operates a Virtual Collection Registry that allows the creation and management of VCs through a user-friendly graphical interface and API. The VCR is one of the thematic services registered at the EOSC portal and therefore available also for disciplinary communities beyond SSH.

## Benefits

- The Virtual Collection Registry (VCR) helps researchers to create persistently identifiable Virtual Collections (VCs) for citing/referencing heterogeneous and distributed research data.
- The VCR enables researchers to access and reuse large samples of data generated by others in the context of a specific research project.
- The VCR facilitates the preparation of sets of resources for specific workflows and makes it possible to share and to work collaboratively on distributed datasets.
- The VCR personal workspace allows researchers to keep track of the steps taken in their research process and data selections used.

## More Information

CLARIN material

→ Information page: [What is a Virtual Collection?](#)

→ User manual: [Virtual Collection Registry](#)

SSHOC material

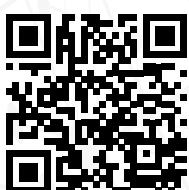
→ Factsheets on GitHub: [Detailed VCR user guide](#)

## Access the VCR

[Access the VCR directly](#)

[Access via the SSH Open Marketplace](#)

[Access via the EOSC Portal](#)



Data Management



Virtual Collection Registry



Software