

Core Components for CLARIN metadata

CMDI task force
CLARIN Annual Conference Bazaar 2020

What are core components, and why do we need them?

What are core components, **and why do we need them?**

Currently:

- ❑ it is **hard** for both metadata creators and metadata modelers to choose **which components and profiles to use**
- ❑ CMDI best practices are not always implemented: varying **quality**
- ❑ **vocabulary** is not uniform and hard to align
- ❑ CMDI is **not well integrated** into the broader research metadata world

What are core components, and why do we need them?

- ❑ Family of **components and profiles**
- ❑ Implementation of **CMDI best practices**
- ❑ Designed for **good discoverability and presentation** in the VLO
- ❑ Linking to **widely used vocabularies**
- ❑ In essence, still **ordinary components and profiles**

What are core components, **and why do we need them?**

With core components:

- ❑ metadata creators are offered a set of **“go-to” profiles** which in most cases should meet their needs
- ❑ there is **reduced need for custom modelling**
- ❑ there are better opportunities for **metadata exploitation**
- ❑ CMDI is better positioned to be **integrated into the broader world of research metadata**

What distinguishes a core component from other components?

- ❑ Technically, core components are **not much different** from other components
- ❑ They do form a curated 'family' with **ensured high quality & interoperability**
- ❑ The power of the core components lies in their **combined use in profiles for specific use cases**.
- ❑ CLARIN infrastructure developers will be able to pay special attention to the **'performance' of these profiles** in metadata infrastructure components such as the VLO

Which core components based profiles exist/are planned?

An overview can be found via  tinyurl.com/cmd-core

- ❑ Current use case based profiles:
 - ❑ Virtual Collection Registry (export from a database)
 - ❑ DDI (conversion from another format)
- ❑ Plans for:
 - ❑ Datacite (for exchange to/from native Datacite)
 - ❑ Historical documents (thematic use case; conversion from TEI)
 - ❑ Learner corpora (thematic use case)
 - ❑ Lexical/conceptual resources (thematic use case)
 - ❑ QUEST (resource type centric)
 - ❑ Software tools/services/components/workflows (resource type centric)

Do I need to switch to core components?

It is **not required** to switch to core components, and you can safely choose to continue using CMDI metadata in the way you have.

Core components are **implemented ‘on top of’ the CMD framework**, which remains the basis for metadata support and interoperability within CLARIN.

How much effort is it to 'switch' to core components?

Obviously depends on the specific use case :)

- ❑ (Relatively) low effort if
 - ❑ You are using a repository solution that offers CMDI export for you
 - ❑ CMDI is **generated by export or conversion** from another format or representation (**and not too large**)
- ❑ Possibly a bit less trivial if
 - ❑ You store 'native' CMDI based on a **custom profile**

Who can make new core components and/or profiles?

- ❑ Members of the **CMDI task force** currently develop and maintain the core components.
- ❑ The task force also decides **which new or existing components and profiles get included** in the 'core component' group and get the 'recommended' status.
- ❑ Anyone with interest in contributing to this task is **welcome to join!**
- ❑ The task force tries to work with **use case specific experts**
- ❑ There will likely be a **procedure to suggest changes** to existing components or **request new ones**.

Can I combine core components with other components in the same profile?

Yes you can, nothing stops you from combining core components with other components!

Good example: custom components that you need to represent **information that is specific to your use case**. Doing this might mean that you will still get many of the benefits of using an 'official' core components profile.

I have found some core components that look useful on the whole. However, some of the elements don't quite fit my resources. How to proceed?

- ❑ The components are designed to be broadly applicable and therefore often **allow for more information** than might be applicable in a single use case.
- ❑ This does **not** mean that all fields within a component are **mandatory** or even necessarily **recommended**. By design, the core components **“accept” relatively sparse information**.
- ❑ In case the availability of certain elements is problematic at the metadata creation stage, for example in the case of manual metadata creation, we suggest hiding the ‘extraneous’ elements in the **editing environment**.