

CultuurLINK: Connecting Cultural Heritage

CultuurLINK is a Web application to link vocabularies. The application has been developed to support the cultural heritage community with the alignment of their vocabularies, and is available at cultuurlink.beeldengeluid.nl. Collection administrators can use CultuurLINK to link their terminology resources to large thesauri that are authoritative for the community. The service is in particular interesting for smaller cultural organizations. Based on their current sources, they can now gain access to external data that enrich their own collections by, for example, additional background information or multilingual descriptions.

CultuurLINK has already been used by the NIOD Institute for War, Holocaust and Genocide Studies to link their terminology with subject terms to a large audio visual thesaurus (GTAA) from the Netherlands Institute for Sound and Vision institute. The resulting links connect, among other things, a collection of Dutch historic news reels with photographic collections of the NIOD. An application developed on top of this data can now suggest related photographs to someone watching a news item, [Linked Open Images](#).

Also, the linking of vocabularies creates opportunities for new services and applications for museums. [An initial prototype](#) demonstrates how links to DBpedia expose the collection of the Rijksmuseum in new and broader perspectives. For instance, the collection can now be searched in multiple languages, and connects 'La Lechera' to Vermeer's Milkmaid. The Rijksmuseum staff actively uses CultuurLINK to link their terminology resources to international databases.

Existing alignment tools

While several tools exist to perform the alignment of datasets, we experienced that they are difficult to apply in practice. As a result, people end up creating ad hoc solutions to find links, one for each use case. [SILK](#) partially overcomes these limitations as it allows the user to manually construct the workflow to find links. With the graphical editor of SILK the user constructs this workflow out of building blocks. SILK provides powerful building blocks, including a variety of algorithms to compare string values. However, constructing a workflow for a specific case is not straightforward. First, the user has to be familiar with the structure of the data, and needs to know how to create the corresponding SPARQL-like path expressions to select the right content. Second, SILK assumes a linear process, where the user first constructs the appropriate workflow and afterwards inspects and evaluates the results. In practice, the construction of the appropriate workflow requires several iterations, where the intermediate results help to understand how to improve or change the workflow.

[Research prototype Amalgame](#) has been developed with the intent to better support this interactive process of vocabulary alignment. With Amalgame, the user constructs a workflow as in SILK, but in this case the user can inspect intermediate results at each step. This allows the user to quickly try out different strategies. The downside is that Amalgame did not provide all the powerful blocks of SILK.

CultuurLINK combines the support for an interactive alignment process as provided by Amalgame with the powerful match techniques found in SILK. CultuurLINK has been deployed as an open service for the Dutch cultural heritage community. The service fits within the national strategy for digital heritage to develop a digital infrastructure that connects collections from all over the Netherlands.

Using CultuurLINK

CultuurLINK is used by first uploading a SKOS vocabulary. Next, the user selects the *Hub* target vocabulary to link with. This Hub contains large vocabularies that are authoritative in its domain. Currently, CultuurLINK contains four vocabularies in the Hub: the audiovisual thesaurus from the Netherlands Institute for Sound & Vision, the Dutch registry of species from Naturalis, the heritage thesaurus of the Dutch Heritage Agency and the Art and Architecture thesaurus from the Getty Institute.

The screenshot displays the CultuurLINK interface. At the top, there are navigation tabs: 'Upload source', 'Select target', 'Edit strategy', and 'Export'. The 'Edit strategy' tab is active, showing a graphical editor with various blocks connected by lines. The blocks include 'Custom' (with 'niod' as a source), 'GTAA', 'Property (object)' (with 'skos.inScheme' and 'gtaa.OnderwerpenBenG'), 'String match' (with 'skos.prefLabel' and 'skos.altLabel'), 'String fuzzy match' (with 'skos.prefLabel' and 'rdfs.label'), 'Overlap mappings', and 'Ambiguity'. Below the editor, a table shows the results of the strategy, with 138 matches. The table has columns for 'Id', 'skos:prefLabel', 'skos:altLabel', and 'skos:broader'.

Id	skos:prefLabel	skos:altLabel	skos:broader
1	niod-terms.Aandelen	Aandelen	
	gtaa.217661	waardepapieren	aandelen, effecten, obligaties, opties
			documenten
2	niod-terms.Aanplakblijetten	Aanplakblijetten	
	gtaa.216030	affiches	posters, aanplakblijetten
			publicaties
3	niod-terms.Moordaanslagen	Moordaanslagen	
	gtaa.218996	moorden	moordaanslagen
			criminaliteit

Given a source and target vocabulary, the next step is to model a unique link strategy. The graphical strategy editor shown in the figure above supports the user in this process. The user builds the link strategy step-by-step out of basic building blocks. CultuurLINK contains blocks to filter vocabularies, match concepts by comparing their labels, filter matches using structural properties and partition the result sets for analysis. Using Spinque's search engine as the backend, the output of each step is efficiently computed to provide the user direct access to the (intermediate) results. Using the result visualization panel, the user may inspect the results, manually evaluate links and decide which step to take next. When the link strategy is completed, the established links between the thesauri are exported as SKOS triples. The

definition of the link strategy provides the provenance of the links. The strategy definition can be exported and uploaded later on to re-run the strategy, for example on an updated dataset.