



Aalto University
School of Science
and Technology

DataSuomi

Metadata editor and smart search portal for
Open datasets

School of Science and Technology
Department of Media Technology

University of Helsinki
Department of Computer Science



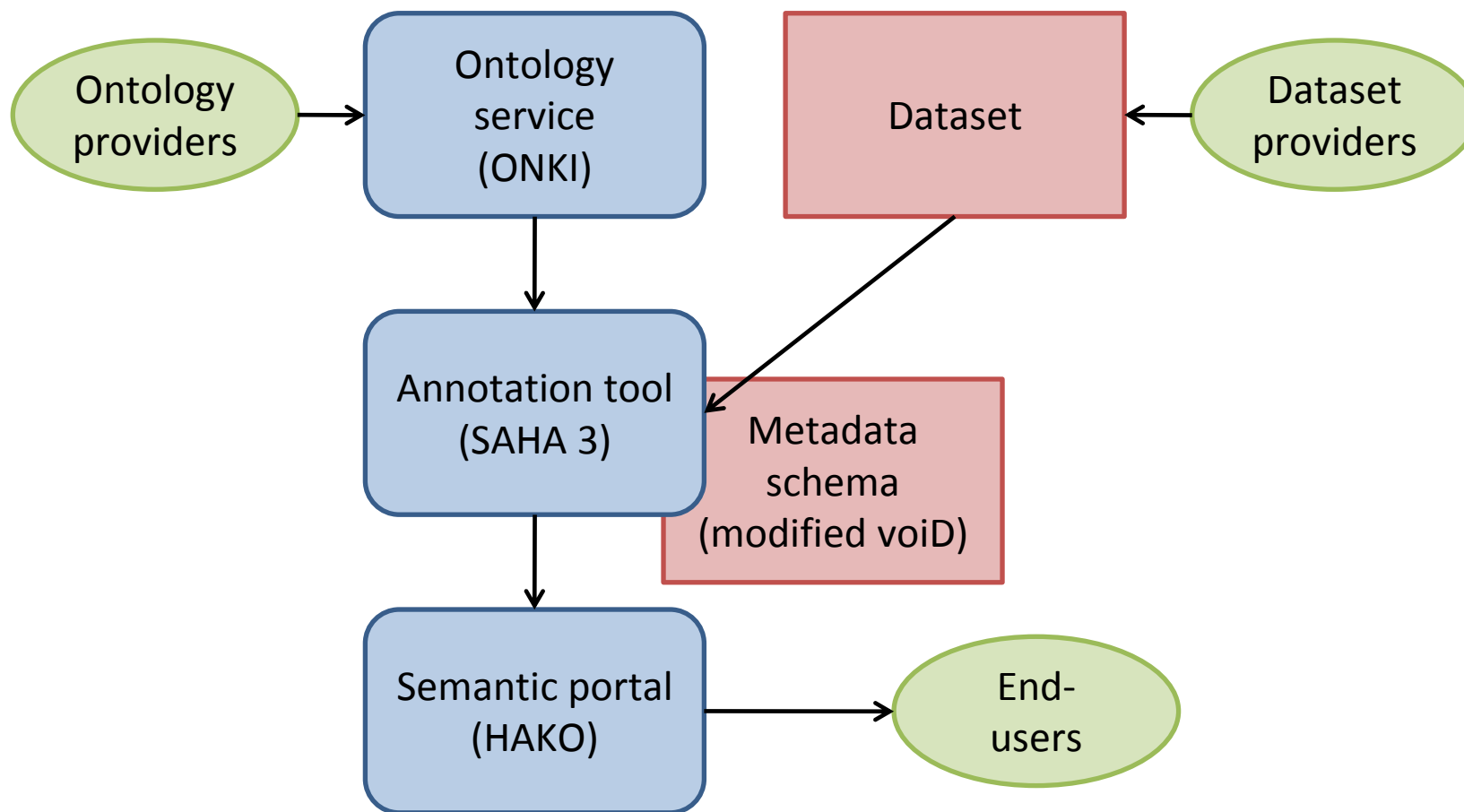
UNIVERSITY OF HELSINKI



Introduction to DataSuomi

- DataSuomi is a tool for editing metadata on and searching for open datasets
- There are two main components to DataSuomi
 - SAHA 3 is used for creating and editing metadata on datasets
 - HAKO portal allows for the publishing of the annotated datasets in a searchable, view-based portal on the web
- The metadata schema used in the annotations is a modified version of the Vocabulary of Interlinked Datasets
 - Allows also non-RDF datasets

Overview of the publication process 1/2

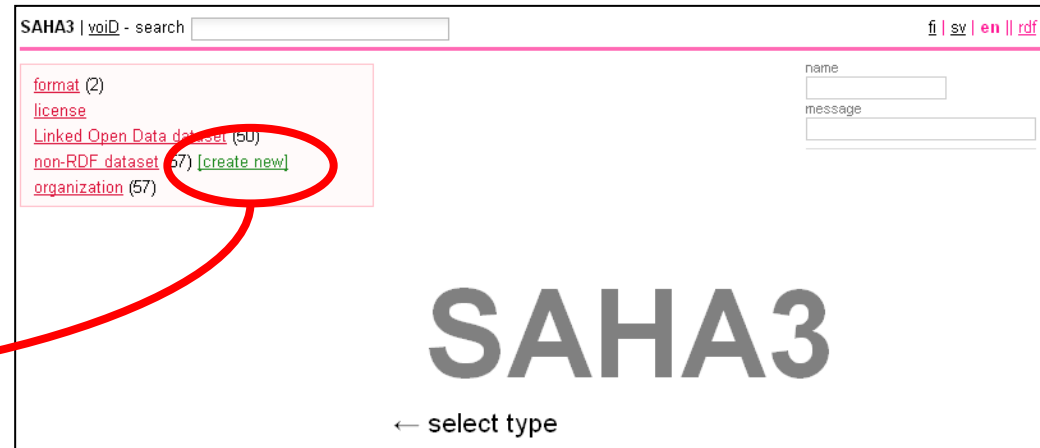


Overview of the publication process 2/2

- First, a dataset is published
- Metadata about the dataset recorded using an annotation tool, SAHA 3 in this case, according to a metadata schema, modified void
- Interoperability between annotations is achieved through the use of shared ontologies provided by an ontology service
- Finally, the metadata about the datasets is published in a smart portal, view-based search engine HAKO in this case

Instructions on how to annotate and publish a dataset using DataSuomi

- First, choose from the left either a Linked Open Data dataset or a non-RDF dataset



Instructions on how to annotate and publish a dataset using DataSuomi

- A new dataset annotation has been created
- There are two kinds of values that can be input
 - Literals, where you can write free text
 - References, that refer to RDF resources defined somewhere else
 - Reference fields feature auto-completion (see next slide)

SAHA3 | [voID](#) - search

[fi](#) | [sv](#) | [en](#) | [rdf](#)

<http://seco.tkk.fi/saha3/u82f36fa4-14d1-4dda-aa7a-79bb024275cf>

name

message

Linked Open Data dataset: u82f36fa4-14d1-4dda-aa7a-79bb024275cf

[\[view\]](#) | [\[rdf\]](#) | [\[config\]](#) | [\[remove\]](#)

title	add new literal <input type="text"/> <input type="button" value="add"/>
creator The person or organization responsible for the dataset	select reference <input type="text"/>
homepage	add new literal <input type="text"/> <input type="button" value="add"/>
subject key concepts related to the dataset	select reference <input type="text"/>
description	add new literal <input type="text"/> <input type="button" value="add"/>
is published under license	select reference <input type="text"/>
language of the dataset	select reference <input type="text"/>
format of the dataset The format in which the dataset is available (e.g. RDF, PDF, Word document etc).	select reference <input type="text"/>
has data dump at Announcement of an RDF dump of the dataset.	add new literal <input type="text"/> <input type="button" value="add"/>
has a SPARQL endpoint at	add new literal <input type="text"/> <input type="button" value="add"/>
has an URI look-up endpoint at	add new literal <input type="text"/>

Instructions on how to annotate and publish a dataset using DataSuomi

- For example, one of the subjects of BBC Music dataset is music, which can be found by writing 'mus' in the subject field and choosing music from the drop-down menu

SAHA3 | [voiD](#) - search [fi](#) | [sv](#) | [en](#) | [rdf](#)

<http://seco.tkk.fi/saha3/u261a6ff1-e519-4e00-b0a0-66576fba1f4d> name
message

Linked Open Data dataset: BBC Music
[\[view\]](#) | [\[rdf\]](#) | [\[config\]](#) | [\[remove\]](#)

title	add new literal <input type="text"/> <input type="button" value="add"/> [remove] (fi) BBC Music
creator The person or organization responsible for the dataset	select reference <input type="text"/> <input type="button" value="v"/> [remove] BBC [edit]
homepage	add new literal <input type="text"/> <input type="button" value="add"/> [remove] http://www.bbc.co.uk/music
subject key concepts related to the dataset	select reference <input type="text"/> <input type="button" value="v"/> <input type="button" value="create a new instance quick inline"/> results from koko ontology (showing top 15 of 48 results) music music music academies musical analysis musical colleges musical exercise musical instruments musical notation musicals musicals musical talent music awards music business music classes music clubs
description	<input type="text"/> artist features, BBC music blogs. Largely
is published under license	<input type="text"/>

Instructions on how to annotate and publish a dataset using DataSuomi

- For the creator, format and license fields you can click the down-arrow on the field and choose “inline”
- A new annotation frame opens where you can describe the reference directly

SAHA3 | voiD - search [fi](#) | [sv](#) | [en](#) | [rdf](#)

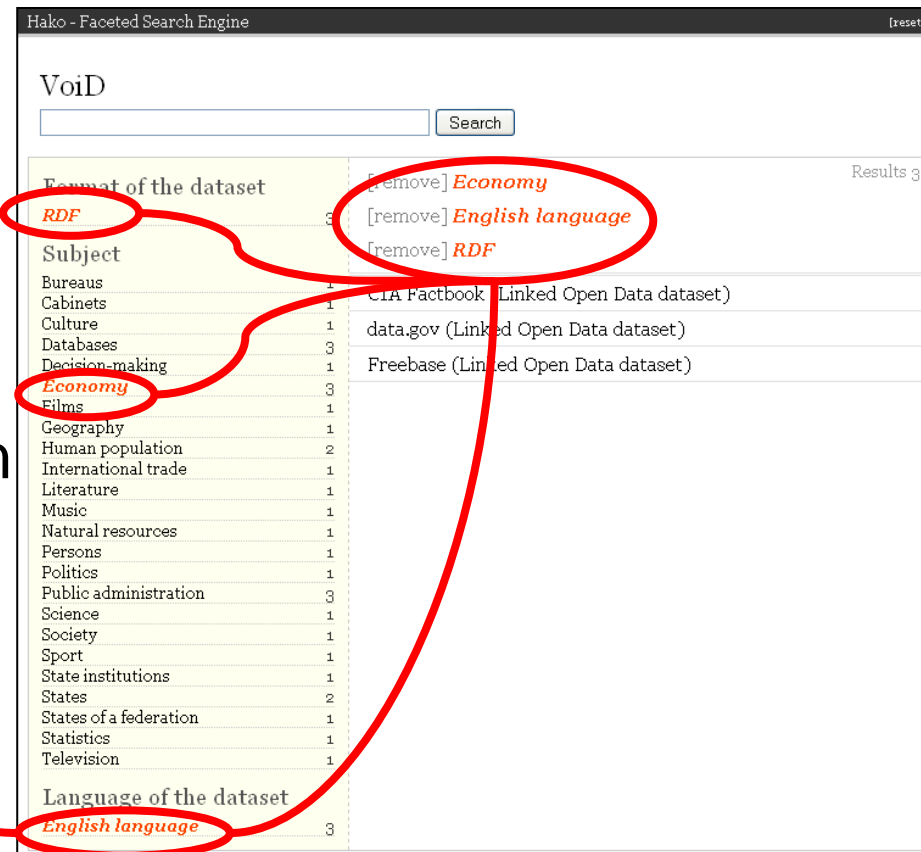
<http://seco.tkk.fi/saha3/u261a6ff1-e519-4e00-b0a0-6f576fba1f4d> name
message

Linked Open Data dataset: BBC Music
[\[view\]](#) | [\[rdf\]](#) | [\[config\]](#) | [\[remove\]](#)

title	add new literal en <input type="text"/> <input type="button" value="add"/> [remove] (fi) BBC Music
creator The person or organization responsible for the dataset	select reference <input type="text"/> <input type="button" value="[close]"/> title add new literal <input type="text"/> <input type="button" value="add"/> homepage add new literal <input type="text"/> <input type="button" value="add"/> description add new literal <input type="text"/> <input type="button" value="add"/> [remove] BBC [edit] [remove] u98266856-03e1-472c-a515-91e34b7011ae [edit]
homepage	add new literal <input type="text"/> <input type="button" value="add"/> [remove] http://www.bbc.co.uk/music
subject key concepts related to the dataset	select reference <input type="text"/> [remove] art music [edit] [remove] blogs [edit] [remove] entertainers [edit] [remove] jazz [edit] [remove] music [edit] [remove] news [edit] [remove] popular music [edit] [remove] rock'n'roll [edit] [remove] äänilevyt [edit]
description	add new literal en <input type="text"/> <input type="button" value="add"/>

Instructions on how to annotate and publish a dataset using DataSuomi

- HAKO enables the searching of datasets
- Free text search field is in the upper left corner
- You can also choose facets from the left to narrow down your search



Links

- DataSuomi
 - SAHA 3: <http://demo.seco.tkk.fi/saha3sandbox/voiD/index.shtml>
 - HAKO: <http://demo.seco.tkk.fi/saha3sandbox/voiD/hako.shtml>
- Research article:
<http://www.seco.tkk.fi/publications/submitted/frosterus-hyvonen-saha-void-2010.pdf>