

SSHOC Considerations for the Vocabulary Platforms: CESSDA

Taina Jääskeläinen, Finnish Social
Science Data Archive

SSHOC Workshop
6 November, 2020



Contents

- **CESSDA:**
 - **Vocabularies used**
 - **Vocabulary platform requirements**
 - **Tools currently used**



CESSDA vocabularies



- CESSDA uses vocabularies to:
 - Harmonise metadata across organisations
 - Improve metadata quality and understandability
 - Increase findability of metadata and relevance of search results
- Two subject vocabularies:
 - Thesaurus ELSST (keywords, large, over 3500 concepts)
 - CESSDA Topic Classification (broad categorisation of topics)



CESSDA vocabularies

- Vocabularies providing broad categorisations of research methods
 - Created by DDI Alliance for the data documentation standard
 - Translated by CESSDA member or associated organisations
 - Examples: Mode of Collection, Unit of Analysis, Sampling Procedure
- Vocabularies used for CESSDA Data Catalogue
 - May be used e.g. for filter cleaning
 - Examples: CDC Publisher Names, CESSDA Persistent Identifier Types

Vocabulary platform requirements

- Governed access, specified by roles, agencies and languages
- Online access, no installations
- Simultaneous use from multiple locations
- Vocabulary management: creation, translation, updating, publishing, validity checks
- Publication and discovery platform for general users
- Export/download in required formats

Vocabulary platform requirements: information needs

- Multilingual concepts and definitions of concepts
- Hierarchy of concepts (broader, narrower)
- For the thesaurus: also related concepts, multilingual synonyms and different types of notes
- For DDI vocabularies: code value – the identifier across languages, not a stable concept identifier
- CESSDA organisations producing metadata from varied disciplines would benefit from having concepts specific to a research domain within a broader controlled vocabulary

Vocabulary platform requirements

- Human-readable version history or an easy way of producing clear and specific version history, specified by language
- CESSDA's purpose for using controlled vocabularies is to harmonise metadata and increase findability but without clear and specific version history:
 - Organisations will not update their legacy metadata → the main purpose of using controlled vocabularies is defeated
 - Vocabulary translating organisations will not be able to maintain their language up-to-date if not know what has changed in the source language

Vocabulary tools currently in use in CESSDA



- **ELSST thesaurus:**
 - VocBench3 as the editor
 - SKOSMOS for publication platform, download and API
 - To be launched 16 November
 - elsst.cessda.eu
- **Other, smaller vocabularies:**
 - CESSDA Vocabulary Service both as editor and publication platform
 - vocabularies.cessda.eu



Vocabulary platform requirements: technical

- SKOS support
- The thesaurus requires more SKOS features than CESSDA Vocabulary Service platform produces, therefore VocBench and SKOSMOS as platforms
- Versioned APIs
- Resolvable URIs
- CESSDA requirements for tools and services (e.g. software quality and maturity)

<https://docs.tech.CESSDA.eu/services/requirements.html>



Thank you for your attention!

Join our community



<https://www.sshopencloud.eu>



@SSHOpenCloud



info@shopencloud.eu



/in/shopencloud

