Context

Currently, at the Flemish government's Environment Department, business analyses culminate in the creation of distributed, non-machine-readable documents such as Word files or wiki pages.

When developing a business application, the reference data present in business terms lists needs to be converted into machine-readable formats, enumerated values in java code or tables in relational databases.

For publication, the enumerated values and reference data tables are transformed into SKOS (Simple Knowledge Organization System) concepts.

Need

To optimize the governance of reference data, the focus should be on adopting machine-readable formats and web standards as outcomes from business analysis. This enables seamless integration with business applications and easy sharing through web publication. The reusability of machine-readable lists in both business applications and web publication enhances data consistency, facilitates collaboration, and drives operational efficiency.

Figure 1: governance work groups oversee and guide the data governance processes.

Figure 2: business analysts and business architects producing the desired outcomes from their analysis.

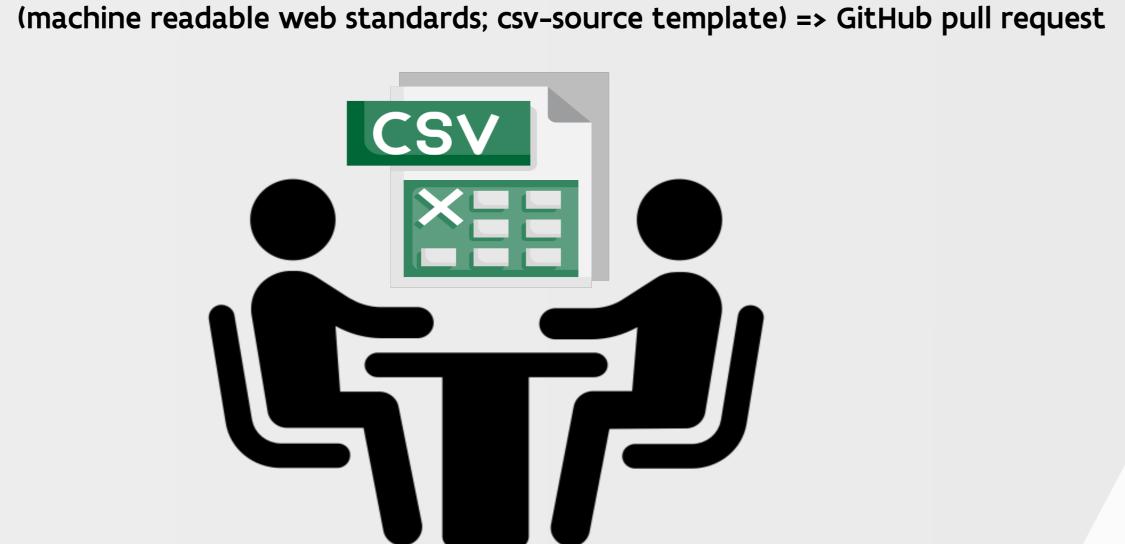
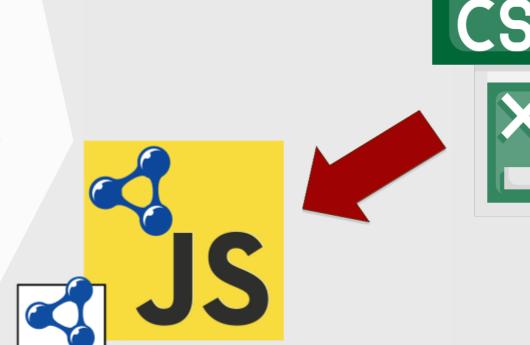


Figure 3: csv to rdf-distributions + generate metadata + shacl validation Local nodejs scripts

of reference data.

Task



Scan code for conceptscheme overview

• Implementing a set of user-friendly tools for managing reference

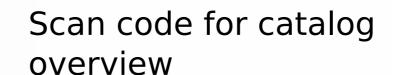
data, accommodating individuals with limited technical background.

Establishing a transparent system for managing different versions

Ensuring machine-readable distributions, versioned code

dependencies, and adherence to web standards.







View it on GitHub



Streamlining Data Governance: A Solution for Managing Reference Data

Figure 6 a,b,c:

a) jar dependencies via code repository (jfrog)

b) Conceptschemes and DCAT metadata as linked

data via subjectpages/sparql endpoint (virtuoso/

c) findable via metadata on metadata Vlaanderen

Publication

netkernel)

Ufora

Linked Data & Solid 2023

Conclusion

In conclusion, the proposed system is suitable and user-friendly for managing simple lists and controlled vocabularies with limited interdependencies and a small number of concepts. It efficiently handles straightforward reference data requirements.

However, for more complex lists, particularly those categorized as master

data, such as chemical substance lists that necessitate a step-by-step protocol, it may be beneficial to develop specific business applications to assist in the governance process. These applications can provide dedicated functionalities and features tailored to the unique complexities of managing and governing intricate reference data, enhancing efficiency and accuracy in the governance process.

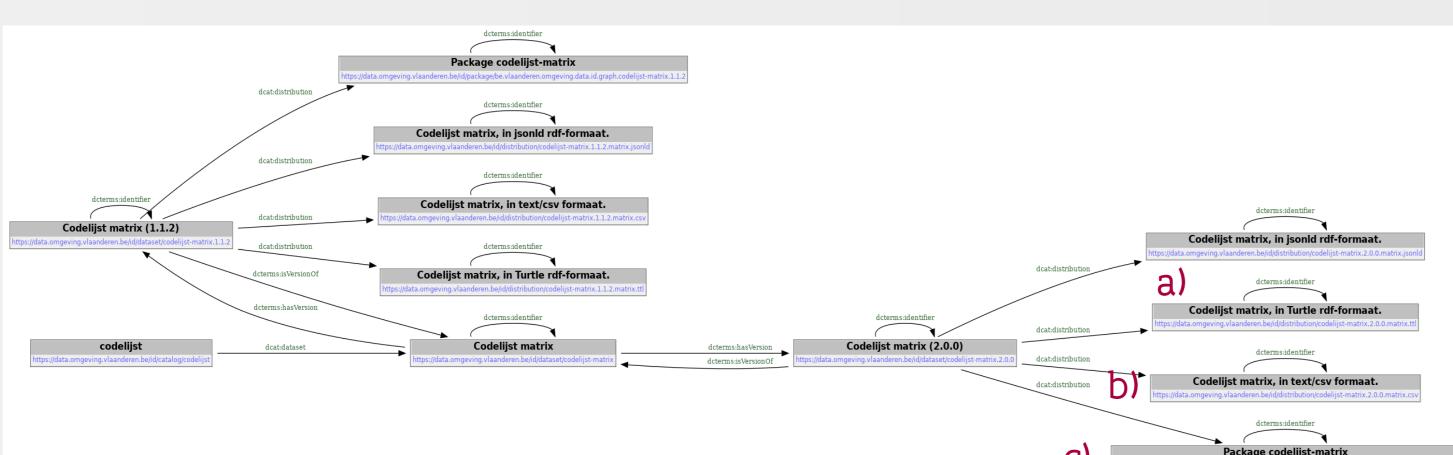


Figure 7 a, b, c:
Usage
a) lod publication, etl-processes
b) etl-processes
c) java code





Bamboo build server.

La Company of the Com

Solut

The proposed solution for this

project involves utilizing GitHub

Solution

version control software to manage the reference data. The source of the reference data is a CSV file, which undergoes transformation, using javascript, into multiple RDF distributions.

Additionally, a JAR dependency is created, containing these distributions.

To ensure proper documentation, metadata for the versioned distributions is generated in DCAT (Data Catalog Vocabulary) format, which is added to the version control system and also packaged as a jar dependency.

Once there is consensus on the content of the reference data, all these components are released as a version in a Maven build on a

Figure 4: push result to GitHub

Geert Van Haute

Environment Departement, Flemish Governement

Vlaamse overheid

Metadata Vlaanderen