

Arches Lingo: An Application to Build and Use Authority Data

Part III: Modelling Authority Data

George Bruseker

The Challenge

Data Models and Lingo

The foundation for supporting the editorial and curatorial requirements of authority data management and for building the software functionalities to create an ergonomic and reliable data management environment is the data model and the derivatives it supports. **Right choices here are crucial for everything that follows!**

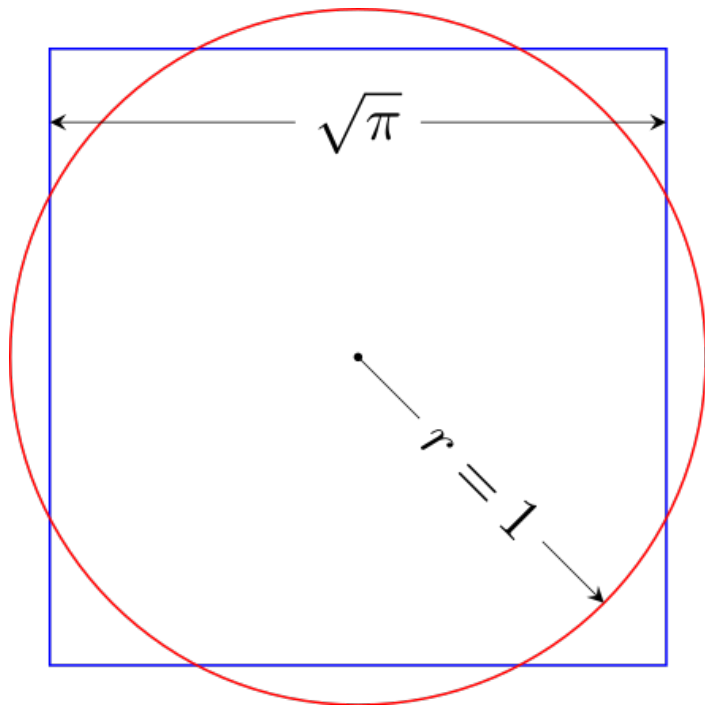


Arches and Data Structures

- Empty
 - Designed for Born Semantic Data
- Causes a Decision
 - What representation strategy do you want to take?



The Task and Challenge



Construct a

- well documented and reusable semantic data model
- supporting the management of authority and vocabulary data in a manner compliant with the highest standards in the field (ISO 25964)
- maintaining backward compatibility with well known and used existing semantic models

that

- applies best practice and knowledge in 2024 of semantic data modelling
- aligns with broader strategies in cultural heritage semantic data modelling world

The world to semantically model and to practically support

Model

- Vocabularies
- Concepts
- Provenance / History
- Contribution

Support

- Production of Vocabularies
- Consumption / Application of Vocabularies
- The lifecycle of Vocabularies

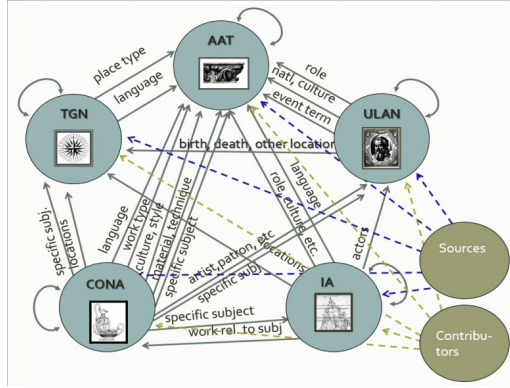


Arches Lingo Internal Semantic Data Model Candidates!

Name	Pro	Con
None	Free hand!	Zero control; bad practice
SKOS	Widespread adoption format for implementation	Does not have most of expressivity to handle best practice.
Linked.Art	Actively developed, links to wider CRM standard, rich enough to handle much of the good practice, a Getty supported and maintained standard which aligns with GD strategy, competencies and tooling.	Does not have all the expressivity to handle best practice.
Linked.Art+	Adopt the Linked.Art where possible, deviate where necessary in order to match best practice requirements from the community.	Does not conform 100% to Linked.Art

A Learning Collaboration with GVP

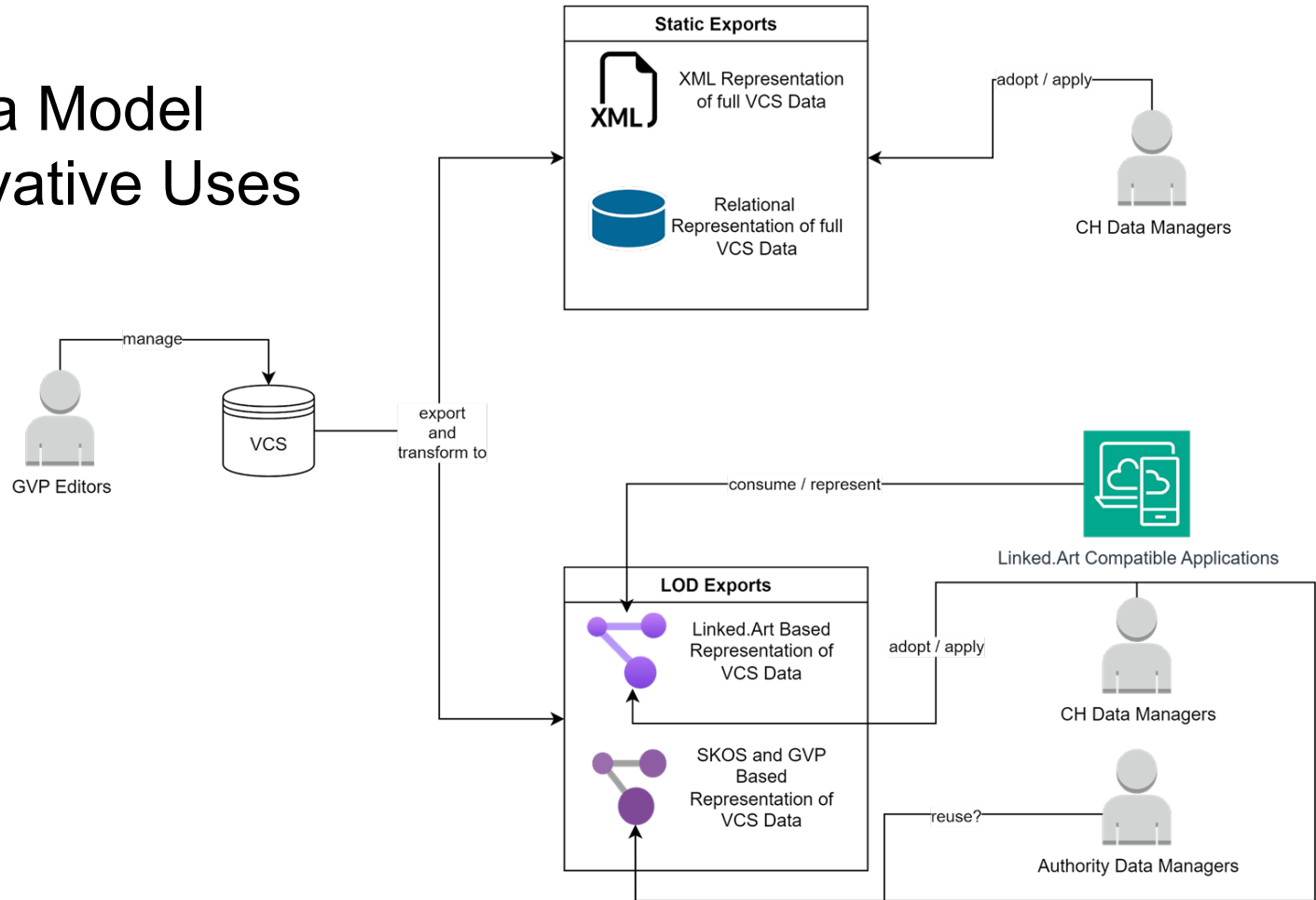
Arches Lingo and Getty Vocabulary Project



Collaborate with and learn from world recognized experts in authority data management to develop a system that is informed by industry standard best practices, aiming to more closely align with the requirements of a project as large and complex as GVP.



GVP Data Model And Derivative Uses Picture

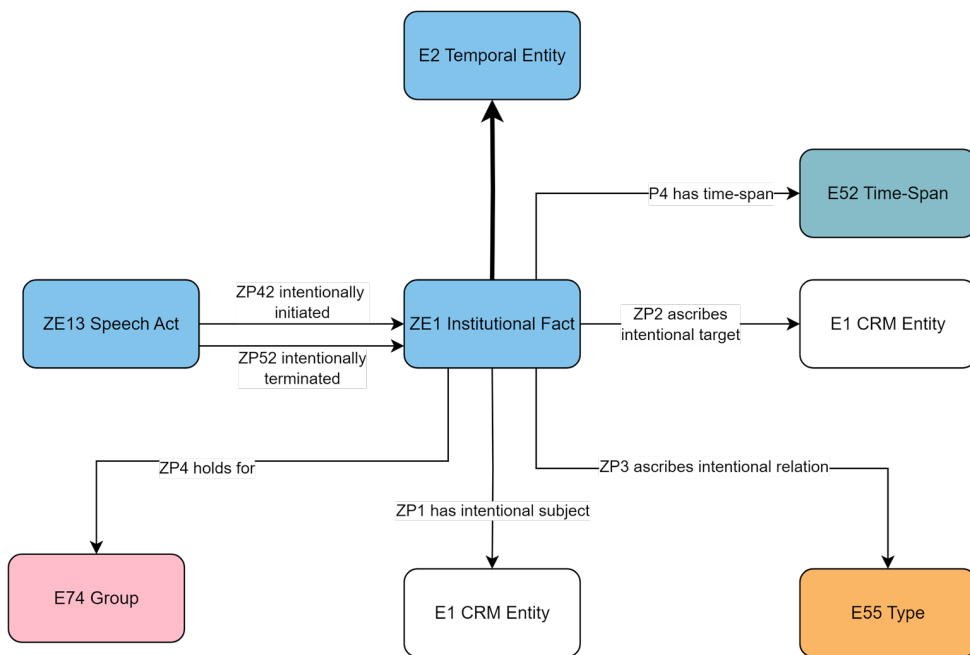


Challenges to Semantically Representing the Richness of VCS / ISO 25964

- Different Demands on Data Shapes from Different Data Models
- Term history-temporality / status / function
- Relation history history-temporality / status / function
- Inter-concept relations
- Provenance Data

The Proposed Solution

Building a Modelling Bridge: Linked.Art+



Core data modelling pattern uses CRMaaa which introduces a core ontological structure, “Institutional Fact” to document mutable social facts such as

- Appellative Status
- Hierarchical Status
- Associative Relation Statuses

While respecting and extending the core modelling patterns of Linked.Art.

Thus it is possible to have the expressive richness of VCS including its rich set of specialized properties & concordance with Linked.Art.

Term History

lambruscatura

Source: <http://vocab.getty.edu/aat/term/1000002693-it>

Subject (14) Predicate Object All Download in: [JSON](#) | [JSONLD](#) | [RDF](#) | [N3/Turtle](#) | [N-Triples](#)

Inference Explicit only

Statements in which the resource exists as a subject.

Predicate	Object
rdf:type	xl:Label
rdfs:comment	Medieval term for wainscoting
gvp:term	lambruscatura@it
gvp:displayOrder	9
gvp:estStart	0900
gvp:estEnd	1700
xl:literalForm	lambruscatura@it
gvp:historicFlag	http://vocab.getty.edu/historic/historic
gvp:termType	http://vocab.getty.edu/term/type/UsedForTerm
gvp:termPOS	http://vocab.getty.edu/term/POS/Noun
gvp:contributorNonPreferred	aat_contrib:10000000
gvp:sourcePreferred	aat_source:2000018341
dc:identifier	1000002693
dcterms:language	aat:300388474

There is a concept:

<http://vocab.getty.edu/aat/300002691>

Wainscoting

It has a label ‘lambruscatura’

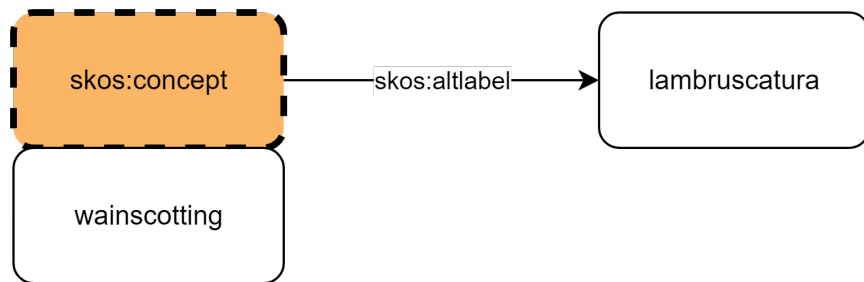
This label was used from 900-1700

This usage is historic (now)

This label is in Italian

This label was attributed to the concept by an institution using a source

Term History Representation: Skos



Representative Capacity:

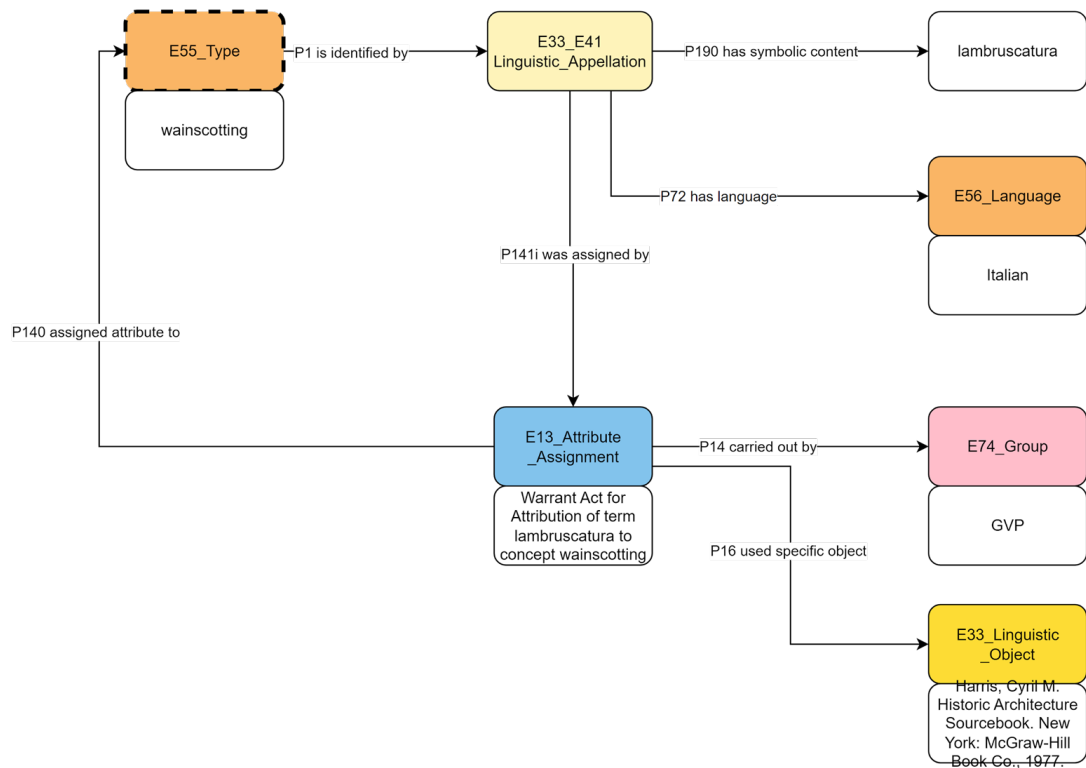
label
Label preference

Conclusion:

good application format

bad documentation
format

Term History Representation: LinkedArt



Representative Capacity:

Label
Label language
Label Source
Label contributor

Conclusion:

much richer but still
missing important
contextual information

Representative Capacity:

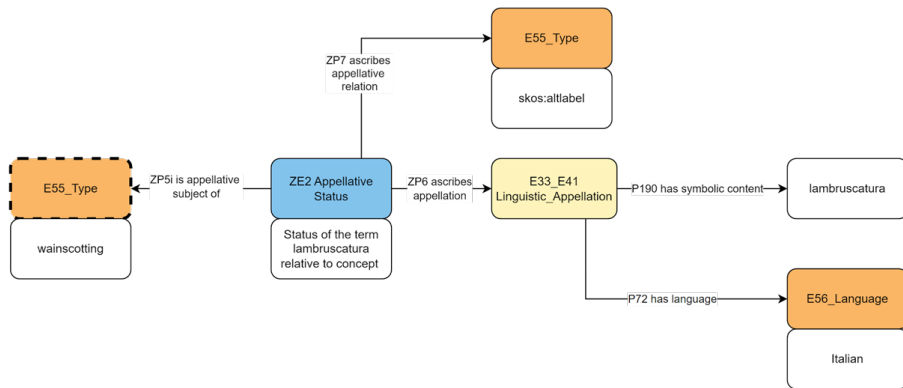


Conclusion:

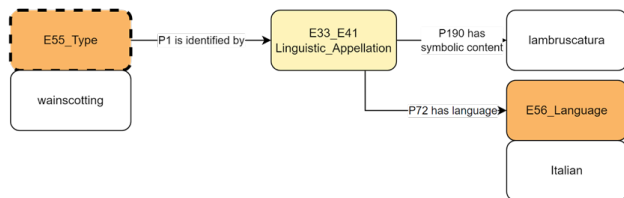
Complete?
Reduces back to LA and
skos, so these can also be
produced programmatically

One “+” Model to Support 3 (and more)

Linked.Art+



Linked.Art



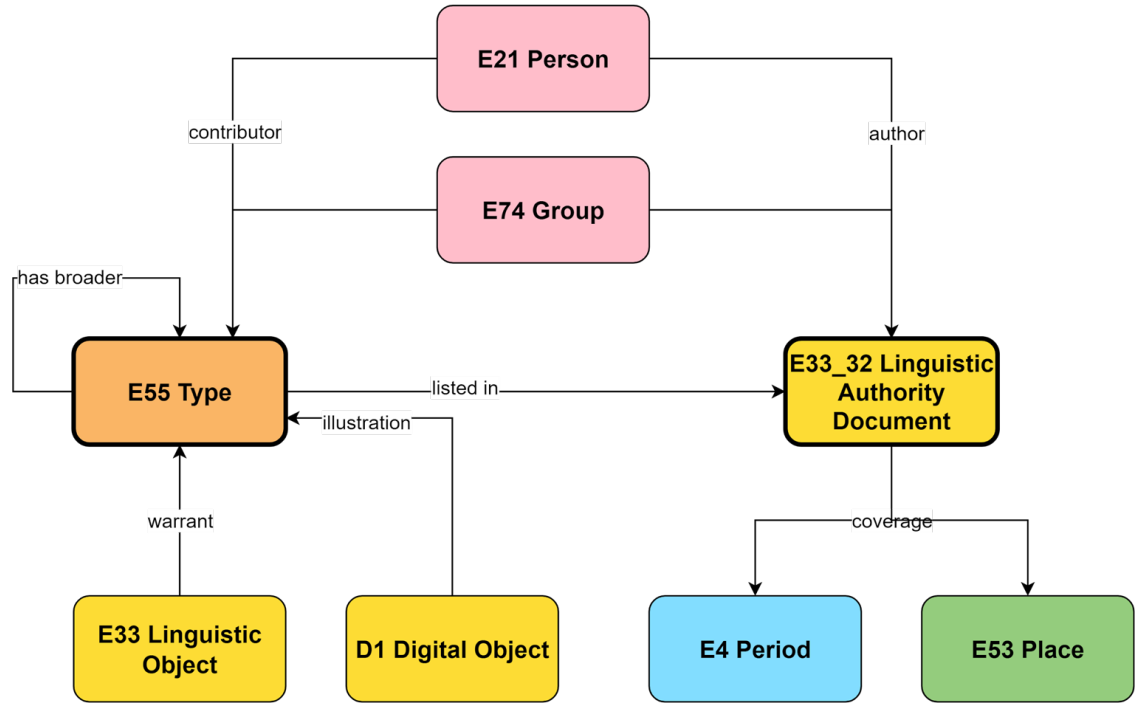
SKOS



Simple
1:1
Transforms

Outcomes

Data Models



Name	Semantic Class
Concept	E55 Type
Scheme	E33_E32 Linguistic Authority Document
Person_RDM_System	E21 Person
Group_RDM_System	E74 Group
Place_RDM_System	E53 Place
Texual_Work_RDM_System	E33 Linguistic Object
Digital_Object_RDM_System	D1 Digital Object
Period_RDM_System	E4 Period

Concept

Derivative Generators

Generators

Attributes

KeyField	RDM.1_Concept
Identifier	RDM.1
Name	Concept
Description	This is the core model of the Lingo semantic model project. The Concept model is where the majority of data curation and management is intended to occur. The concept model allows for the documentation of formally defined concepts in the sense of ISO 25964. It enables the temporally qualified and provenanced documentation of the labels/terms associated, associative and matching relations and hierarchical relations. Each concept is documented in relation to a Scheme as either a simple member or as a top concept. An instance of Concept's overall provenance information can be documented and it can be linked to digital asset resources that illustrate it.

[CAT.1] Names and Identifiers

Identifier	Name	CRM Path	Expected Value Type	Expected Collection	Expected Model	Derivatives
RDMF.12	URI	->p1->E42[RDMF.11_1]->p190->rdfs:literal	String			Generators
RDMF.13	URI Type	->p1->E42[RDMF.11_1]->p2->E55[RDMF.13_1]	Concept			Generators
SRDF.524	Label	->zp5i->ZE2[SRDF.518_1]->zp6->E41[SRDF.524_1]	Collection	Name (Simple)		Generators
SRDF.525	Label Type	->zp5i->ZE2[SRDF.518_1]->zp7->E55[SRDF.525_1]	Concept			Generators
SRDF.522	Label Status	->zp5i->ZE2[SRDF.518_1]->p2->E55[SRDF.522_1]	Concept			Generators
SRDF.523	Label Status Metatype	->zp5i->ZE2[SRDF.518_1]-p2-E55[SRDF.522_1]->p2->E55[SRDF.523_1]	Concept			Generators
SRDF.526	Label Temporal Validity	->zp5i->ZE2[SRDF.518_1]->p4->E52[SRDF.526_1]	Collection	Timespan (Simple)		Generators
RDMF.57	Label Warrant Assertion Event	zp5i->ZE2[SRDF.518_1]->p14ii->E13[RDMF.57_1]	Collection	Warrant Assertion Event (Simple)		Generators

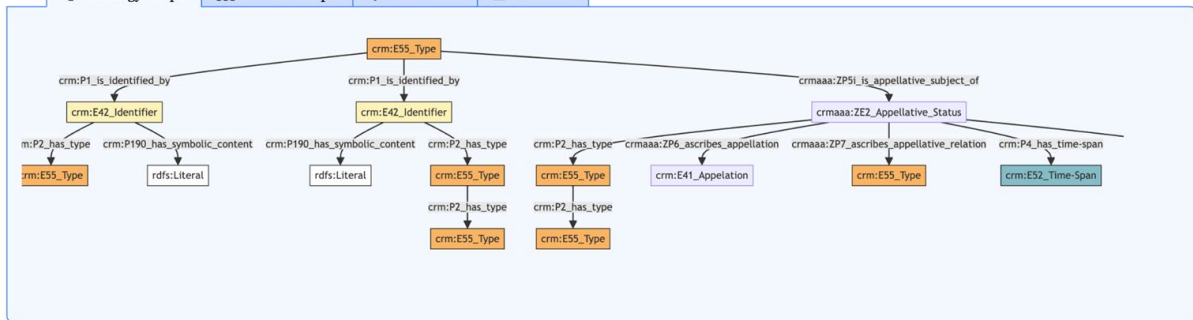
Rich, Reusable Semantic Documentation

Ontology Graph

Instance Graph

Turtle RDF

JSON-LD

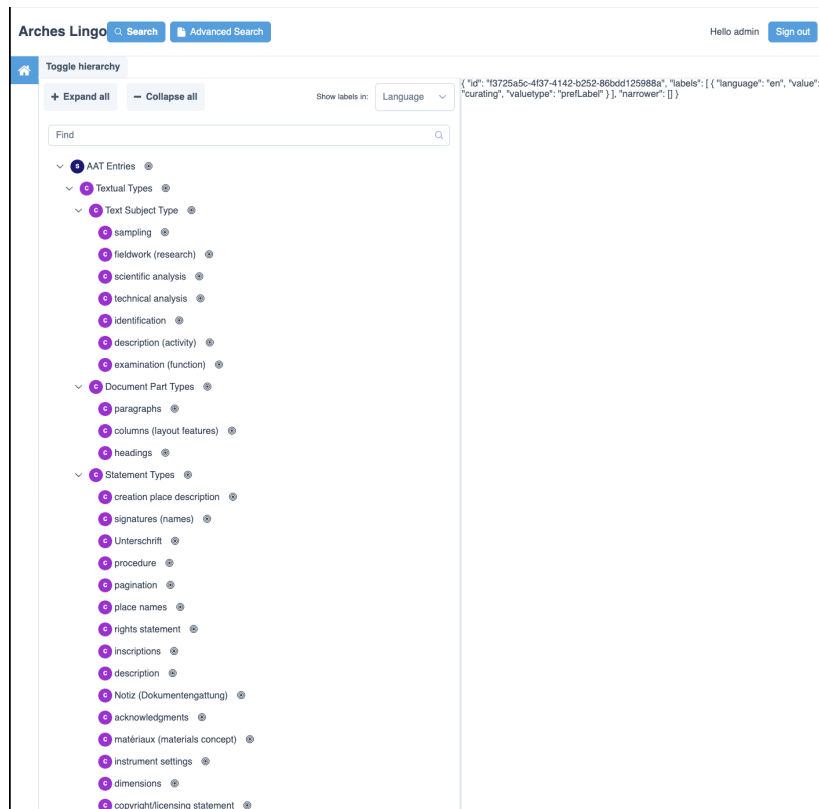


Implementation

Sandbox: <https://rdm.dev.fargeo.com>

Github:

<https://github.com/archesproject/arches-lingo>



Questions?

george@takin.solutions