FAIR Data in Medical Research

Incorporating the FAIR Principles in the Research Data Life Cycle

Kersloot, M.G.

Publication date
2022

Chapter 8

The International Society for the Study of Vascular Anomalies (ISSVA) Ontology

Philip van Damme*
Martijn G. Kersloot*
Bruna dos Santos Vieira
Leo Schultze Kool
Ronald Cornet

* These authors contributed equally

Submitted

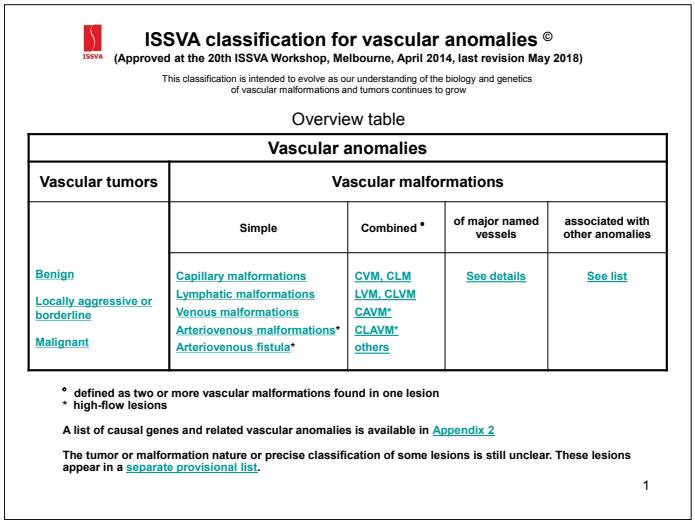
Abstract

The International Society for the Study of Vascular Anomalies (ISSVA) provides a classification for vascular anomalies that enables specialists to unambiguously classify diagnoses. This classification is only available in PDF format and is not machine-readable, nor does it provide unique identifiers that allow for structured registration. In this paper, we describe the process of transforming the ISSVA classification into an ontology. We also describe the structure of this ontology, as well as two applications of the ontology using examples from the domain of rare disease research. We used the expertise of an ontology expert and clinician during the development process. We semi-automatically added mappings to relevant external ontologies using automated ontology matching systems and manual assessment by experts. The ISSVA ontology should contribute to making data for vascular anomaly research more Findable, Accessible, Interoperable, and Reusable (FAIR). The ontology is available at https://bioportal.bioontology.org/ontologies/ISSVA.

Introduction

Vascular anomalies represent a broad spectrum of common and rare disorders of blood vessel growth, ranging from a simple "birthmark" to life-threatening entities, which affect mainly infants, children, and young adults [245, 246]. To support diagnosis, management, and further research of vascular anomalies, Mulliken and Glowacki created a comprehensive classification system for vascular anomalies [247]. This classification system, published in 1982, distinguishes between vascular tumors (formerly referred to as *hemangiomas*) and vascular malformations.

The International Society for the Study of Vascular Anomalies (ISSVA), adopted this classification in 1996. ISSVA, founded in 1992, is the society for specialists of various medical disciplines involved in the treatment of patients afflicted with vascular anomalies [246]. During international ISSVA workshops, the classification is revised by specialists, as the understanding of the biology and genetics of vascular anomalies continues to grow [248]. The latest revision of the classification was published in 2018 [248] and is available as an interactive PDF (Figure 8.1).



ISSVA classification for vascular anomalies ©

(Approved at the 20th ISSVA Workshop, Melbourne, April 2014, last revision May 2018)

This classification is intended to evolve as our understanding of the biology and genetics of vascular malformations and tumors continues to grow

Overview table

Vascular anomalies				
Vascular tumors	**Vascular malformations**			
	Simple	**Combined °**	**of major named vessels**	**associated with other anomalies**
Benign Locally aggressive or borderline Malignant	Capillary malformations Lymphatic malformations Venous malformations Arteriovenous malformations* Arteriovenous fistula*	CVM, CLM LVM, CLVM CAVM* CLAVM* others	See details	See list

° defined as two or more vascular malformations found in one lesion
* high-flow lesions

A list of causal genes and related vascular anomalies is available in Appendix 2

The tumor or malformation nature or precise classification of some lesions is still unclear. These lesions appear in a separate provisional list.

1

Figure 8.1: The first page of the ISSVA classification interactive PDF

The interactive PDF contains hyperlinks to several sections of the classification and allows specialists to easily find anomalies by their type (e.g., a *Congenital hemangioma* is a *Benign vascular tumor*) or their causal gene (e.g., the gene *VEGFC* is associated with *Primary hereditary lymphedema*). While the PDF allows specialists to find and unambiguously classify diagnoses, it does not allow for structured registration of these diagnoses using unique identifiers, nor implementation in software systems, such as Electronic Health Records (EHRs) and Electronic Data Capture (EDC) systems. To make the data for vascular anomaly research more Findable, Accessible, Interoperable, and Reusable (FAIR) [18], it is important that these diagnoses are registered in a structured manner.

A majority of the vascular anomalies present in the ISSVA classification are rare diseases. In line with the FAIR principles, the European Platform on Rare Disease Registration recognizes the importance of structured registration and therefore strongly recommends the use of the Orphanet nomenclature of rare diseases for registering diagnoses in their Common Data Elements (CDEs) [203]. Each disease in this nomenclature is attributed a unique identifier, an ORPHAcode [249]. The ORPHAcodes, together with descriptions of the diseases and relationships with associated genes, are available in a machine-readable format in the Orphanet Rare Disease Ontology (ORDO). ORDO is the preferred ontology in the European Union for the rare disease domain and mappings exist between this ontology and other biomedical ontologies, such as MeSH, SNOMED CT, UMLS, and MedDRA [249].

ORDO only contains rare diseases, and therefore only vascular anomalies that are considered rare. The ISSVA classification on the other hand also covers non-rare vascular anomalies, such as capillary malformations. Hence, not all classes present in the ISSVA classification have a corresponding concept in ORDO and not all the relationships between anomalies or genes present in the ISSVA classification are available in ORDO. In addition, ISSVA is an internationally recognized classification, whereas ORDO is mostly being used in Europe. Therefore, we identified the need for a new ontology based on the ISSVA classification. Creating such an ontology ensures that the classes present in the classification are assigned a Uniform Resource Identifier (URI) and are available in a machine-readable format.

In this paper, we describe the efforts we took to transform the ISSVA classification into an ontology. In addition, we describe how the classes in the classification are mapped to existing biomedical ontology concepts, to increase the interoperability of the new ontology. We also describe the structure of the ontology and its current and planned applications.

The ISSVA ontology

The ISSVA ontology describes vascular anomalies, the causal genes of such anomalies, and related vessels. The ontology is available at https://bioportal.bioontology.org/ontologies/ISSVA in the Web Ontology Language (OWL) format. At the time of writing the ISSVA ontology contains 202 concepts and 1126 axioms. The ontology is organized as a monohierarchy that has three main branches: *Gene*, *Vascular anomaly*, and *Vessel*. Figure 8.2 shows an overview of the hierarchy. The *Gene* branch contains all genes that are associated with one or more vascular anomalies. The *Vascular anomaly* branch contains all vascular anomalies that are either a vascular malformation or a vascular tumor. The *Vessel* branch contains all vessels that can be affected by a vascular anomaly. Gene-disease associations are captured as an annotation property using *sio:gene-disease-association* that links a gene to its associated vascular anomaly concept(s) in the ontology (Figure 8.3.1). Each vascular anomaly can be a disorder of one or more vessel types, which is captured using an object property *affects vessel type* in vascular anomaly concepts. For example, *Capillary malformation* affects vessels of the type *Capillary vessels* (Figure 8.3.2).

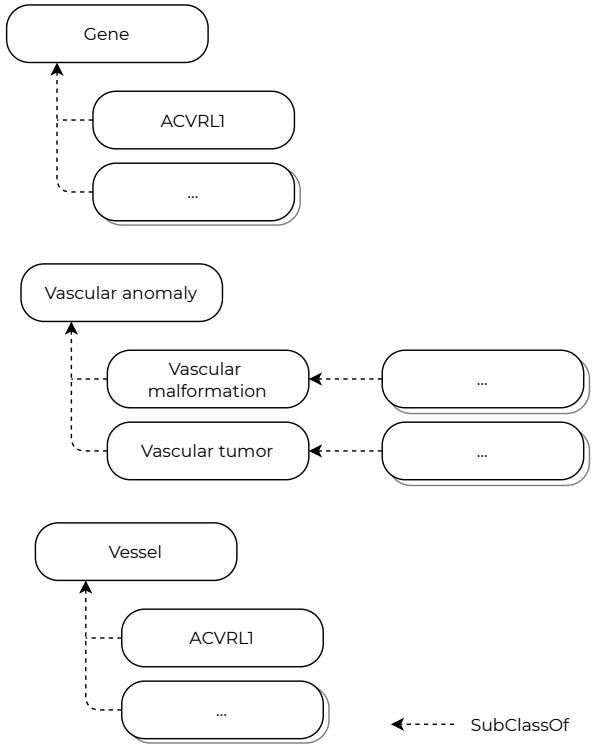
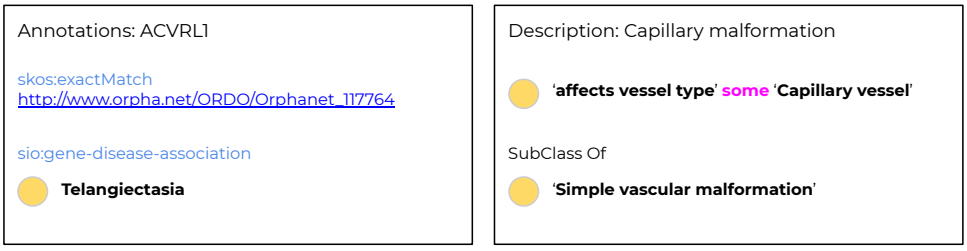


Figure 8.2: Overview of the ISSVA ontology hierarchy.



(1) Mapping and property annotations

(2) Object property 'affects vessel type'

Figure 8.3: Examples of the mapping annotations, property annotations, and object properties in the ISSVA ontology.

All concepts contain English labels (*rdfs:label*). Alternative labels (*skos:altLabel*) and descriptions (*rdfs:comment*) are present if applicable.

Mapping annotations

122 out of 194 ISSVA ontology concepts were matched to equivalent concepts in ORDO, the National Cancer Institute thesaurus (NCIt) [250], SNOMED Clinical Terms (SNOMED CT) [70], the Human Phenotype Ontology (HPO) [251], and/or the HUGO Gene Nomenclature (HGNC) [252] (see Table 8.1). This set of ontologies was chosen based on input from an ISSVA board member. Every gene has at least one mapping to ORDO and/or HGNC. Mappings are captured as *skos:exactMatch* annotations (Figure 8.3.1). A concept in the ISSVA ontology can be matched to multiple concepts in the other ontologies, and each mapping contains the machine-readable URI of the matched concept.

Table 8.1: Number of mappings to external ontologies sorted by ISSVA hierarchy

	Concepts	Matched	HGNC	HPO	NCIt	ORDO	SNOMED CT
Gene	41	40	40	0	0	38	0
Vascular malformation	111	49	0	7	17	37	25
Vascular tumor	36	28	0	5	20	12	18
Vessel	5	4	0	0	4	0	4

Versioning

The ISSVA ontology uses the Semantic Versioning system to version new releases of the ontology [253]. Since the ontology is based on the ISSVA classification, major version updates indicate that the ontology was updated based on changes to the ISSVA classification. New versions are uploaded to the ISSVA ontology GitHub repository [254] and BioPortal [255]. In case there are changes in the content of the ontology, a domain expert is consulted to approve the changes.

Methods

The development of the ISSVA ontology can be divided into four phases: 1) add concepts and relationships to the ontology, 2) semi-automatically add mappings to external ontologies, 3) validate the ontology and its mappings, and 4) publish the ontology.

Add concepts and relationships to the ontology

All classes present in the ISSVA classification [248] (i.e., all vascular anomalies, genes, and vessel types) were represented as concepts in the ontology. The hierarchy of the classification was adopted as the hierarchy of the *Vascular anomaly* branch in the ontology. The naming conventions for biomedical ontologies [256] were followed and, as such, the

labels used in the classification were used in their singular form in the ontology (e.g., 'vessel', as opposed to 'vessels'). In addition, the abbreviations that were used in the labels were expanded to their full form according to the abbreviation list in the classification (e.g., *CM of CM-AVM* is expanded to *Capillary malformation of Capillary malformation-arteriovenous malformation*). The original label with abbreviations was subsequently added as an alternative label (skos:altLabel). The labels of residual categories were expanded and added as alternative labels to the parent of the category (e.g., *Others*, a child of *Benign vascular tumors*, was expanded to *Other benign vascular tumor* and added as an alternative label for *Benign vascular tumors*). Lastly, we automatically generated unique identifiers for every concept in the ontology. The structure of our identifiers is based on the OBO Foundry ID Policy [257]: *ISSVA* is used as a prefix (IDSPACE), followed by an underscore (_) and three ascending digits (LOCALID). For example, the identifier for *Vascular anomaly* is http://purl.bioontology.org/ontology/ISSVA/ISSVA_144 and can be shortened according to the OBO Foundry ID Policy as ISSVA:144. By using RDF prefixes this can be shortened as ISSVA:ISSVA_144 (with http://purl.bioontology.org/ontology/ISSVA/ as prefix).

Semi-automatically add mappings to external ontologies

Reusing content of existing ontologies is a best practice when developing new ontologies and should improve interoperability [258]. ORDO only covers rare diseases and not all vascular anomalies in the classification are rare. To ensure interoperability between ISSVA and other relevant ontologies, we semi-automatically added mappings to ORDO, NCIt, SNOMED CT, and HPO. To do so, we first automatically obtained an initial set of pairwise mappings between ISSVA and each aforementioned ontology. Subsequently, experts manually assessed and validated this initial set (Subsection 8). We used three state-of-the-art ontology matching systems to generate this initial set of mappings, namely, AgreementMakerLight 2.0 [259], LogMap 2.0 [260], and FCA-Map [261]. We ran the systems using their default parameters and combined the resulting mappings for each pair of ontologies (ISSVA - external ontology). Duplicate mappings were removed in the case of two systems returning the same mapping. Additionally, we manually added mappings between the genes in the ISSVA ontology and HGNC/ORDO because the matching systems did not return any mappings for the gene concepts.

Validate the ontology and its mappings

The initial set of mappings was assessed by two medical informaticians (PvD and MK) and one data steward (BdSV). Mappings were classified as correct, incorrect, or in need of review (Table 8.2). A correct mapping is a mapping that is both syntactically and semantically correct: a mapping to an ontology concept with a label (preferred or alternative) equal to the ISSVA label (syntactically) and in a *Finding* or *Disorder* hierarchy (semantically). An incorrect mapping is a mapping to a concept with a different label or a different hierarchy. When the label was similar, but not identical, to the label in ISSVA or when questions about the hierarchical position of the concept arose, the mapping

was classified as in need of review. Mappings in need of review were presented to a domain expert and ISSVA representative (LSK). The domain expert assessed if the mapping should be added to the ontology, or if the mapping should be deleted.

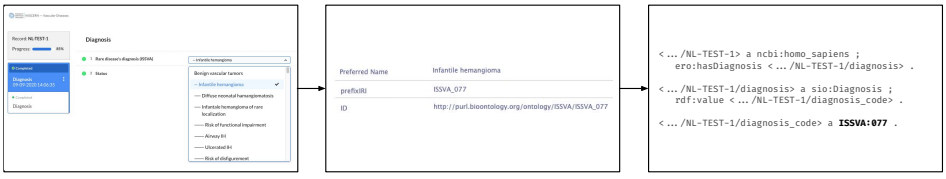
Table 8.2: Examples of mappings classified as correct, incorrect, or in need of review, with their associated ontology and hierarchical position.

	ISSVA term	Matched concept	Ontology	Hierarchy
Correct	Sturge-Weber syndrome	Sturge-Weber syndrome	SNOMED CT	Clinical finding
Incorrect	Other locally aggressive or borderline vascular tumors	NPI - Any Other Aggressive or Agitated Behaviors	NCIt	Activity
Needs review	Arteriovenous malformation in capillary malformation-arteriovenous malformation	Capillary malformation-arteriovenous malformation	ORDO	Clinical entity

After the reviewed mappings were added to the ontology, an expert in the field of ontologies (RC) reviewed the structure of the ontology. Based on their feedback, we revised the ontology.

Publish the ontology

We published the ontology in BioPortal, a repository for biomedical ontologies [262]. BioPortal automatically assigns Persistent Uniform Resource Locators (PURLs) to the concepts in the ontology and provides a web interface to browse the ontology. Moreover, the repository also exposes an Application Programming Interface (API) to search for concepts or to annotate clinical text, which allows software providers to implement the ontology in their systems.



(1) Capture diagnosis (2) Get associated PURL (3) Generate machine-readable data

Figure 8.4: The steps from capturing a diagnosis in the Registry of Vascular Anomalies to making it machine-readable

Application of the ISSVA ontology

The ISSVA ontology can be used in settings where the structured recording of a human-readable and machine-readable diagnosis is preferred. Here, we describe two applications: data collection and annotation and data integration and analysis.

Data collection and annotation

The ontology is currently implemented in an EDC system to support FAIR data collection and exposure for the Registry of Vascular Anomalies [63, 263]. In the registry, based on the previously mentioned European Platform on Rare Disease Registration CDEs, the diagnosis of the patient is registered using the ISSVA classification (Figure 8.4.1). The diagnosis is saved using the unique ISSVA concept identifier (e.g., *ISSVA_077* for *Infantile hemangioma* with http://purl.bioontology.org/ontology/ISSVA/**ISSVA_077** as URL). This identifier is then used to get the PURL associated with the machine-readable version of the diagnosis present in the ontology (Figure 8.4.2). Finally, the data, among which the diagnosis, are transformed into a machine-readable format (Figure 8.4.3), which is then hosted in a repository. The machine-readable format is then ready for re-use and can be queried using SPARQL queries.

Data integration and analysis

Information silos, i.e., systems that do not exchange data with other similar systems, are common in healthcare and, for example, rare disease research [264,265]. Ontologies play a role in the exchange of data between distributed data sources as they unambiguously conceptualize a domain. Using an ontology to annotate data in different sources, therefore, eliminates any potential semantic heterogeneities between those sources [266]. The ability to integrate data from different sources is crucial for data analysis. ISSVA advocates the exchange of data about vascular anomalies by distributing its classification. Consequently, the ISSVA ontology enables integrating and analyzing distributed data sources that contain data about vascular anomalies. A possible future implementation of the ISSVA ontology for data analysis lies in the development of a virtual platform by the European Joint Programme on Rare Diseases (EJP RD) [267]. This platform will describe rare disease resources and will utilize relevant ontologies.

Discussion

In this paper, we described the incentive, structure, development process, and applications of the ISSVA ontology. We also described how we semi-automatically added mappings to existing ontologies using ontology matching systems and expert evaluation. The ISSVA ontology should support the registration of vascular anomalies according to the FAIR data principles.

Strengths and limitations

We can identify several strengths and limitations of our efforts in transforming the ISSVA classification into an ontology. First of all, a major strength is the improved interoperability of the ISSVA classification now that it is also available as a machine-readable ontology. Another strength is that we highly valued the interoperability of the ISSVA ontology with other relevant ontologies, and, therefore, added equivalence mappings to ORDO, SNOMED CT, NCIt, HPO, and HGNC. Another strength is that we consulted an ontology expert and clinician during the development of the ontology.

However, as we relied on the knowledge of only one ontology expert and one clinician, we consider this also a limitation of our engineering approach. Ideally, we would have involved multiple experts for evaluating the mappings and assessing our design decisions. Additionally, we were often bound to the content of the ISSVA classification in designing the ontology. This led to trade-offs between, for example, reusing content from existing ontologies or adding new content to the ISSVA ontology with mappings to equivalent concepts in other ontologies.

Design decisions

Throughout the development process of the ontology, various (evidence-based) design decisions were made. We followed the naming conventions for biomedical ontologies [256] and used the common OBO Foundry ID Policy [257]. We adopted the hierachy of the ISSVA classification in our ontology. Based on the feedback from domain experts, we did not include the residual categories from the classification, but included the category names as alternative labels for the branch associated with the category. Since reusing (concepts of) existing ontologies is a best practice in ontology development and supports semantic interoperability [258], we added mappings to other ontologies. We decided that importing large ontologies into the ISSVA ontology was undesirable. The ISSVA ontology only has around 200 concepts, and importing external ontologies would therefore make our ontology unnecessarily large and complex. Lastly, as vascular anomaly research is a rapidly evolving field, we decided to add dedicated gene concepts to the ontology. In that way, we are able to more easily maintain new versions of the ontology, without being dependent on the content of external ontologies.

Mappings to other ontologies

The matching systems that we used to generate our initial set of mappings (AgreementMakerLight 2.0, LogMap 2.0, and FCA-Map) are participants of the disease and phenotype track in the Ontology Alignment Evaluation Initiative (OAEI) [268]. We used three different systems that apply different ontology matching techniques to increase the number of mappings the systems could expose. Our approach of combining automated matching systems with the manual assessment of mappings by medical informaticians and domain experts is known as a best practice [268]. As ontologies are dynamic and change over time, using automated matching improves the feasibility of adding and maintaining

mappings in future versions of the ISSVA ontology. Due to the small number of concepts in the ontology, we could detect that the matching systems did not return mappings for the gene concepts. Hence, we manually mapped all genes to ORDO and HGNC. Mappings to HGNC could not be generated automatically at all because there was no officially supported and maintained version of HGNC available in the OWL format.

We noticed that some vascular anomalies are modeled as both *Body structure*s and *Clinical finding*s in SNOMED CT. For example, the ISSVA term *Angiokeratoma* was automatically mapped to SNOMED CT concepts *Angiokeratoma (morphologic abnormality)* and *Angiokeratoma of skin (disorder)*. Since the ISSVA classification is focused on registering diagnoses, we included the mappings that were marked as a disorder (*Clinical finding*) as *skos:exactMatch* and the mappings marked as morphologic abnormality (*Body structure*) as *skos:closeMatch*.

Implications for practice

Prior to our work, vascular anomaly diagnoses were registered in a human-readable manner following the free-text descriptions of classes in the ISSVA classification. The ISSVA ontology ensures these diagnoses can be registered in a machine-readable manner and thus increases the interoperability of clinical studies and registries that collect data for vascular anomaly research. Since the prevalence of most vascular anomalies is low, data are often distributed over various sources. Machine-readable data enables the integration of data across these sources, increasing the sample size and thus power of the data, in order to perform valid analyses.

Future work

Future research should focus on the evaluation of the increased interoperability of clinical studies and registries and on showing the value of machine-readable diagnosis registration. In addition, future work should also focus on the implementation of the ontology in different settings. Different implementations of the ontology might require different views on the ontology as well, and possibilities for showing different views or subsets of the ontology should be explored.

Conclusion

In this paper we introduced the ISSVA ontology, a monohierarchical ontology based on the ISSVA classification with mappings to existing biomedical ontologies. The ontology is available at https://bioportal.bioontology.org/ontologies/ISSVA and is currently used in a clinical patient registry to capture machine-readable diagnoses of vascular anomaly patients. We believe that the ontology will increase the interoperability between clinical studies and registries and, therefore, will contribute to FAIRer vascular anomaly research.