

6.9 Domain

6.9.1 Definition



The [Domain](#) data category is used to identify the domain of content.

This data category addresses various challenges:

- Often domain related information in content does exist, e.g. keywords in the HTML `meta` element. The [Domain](#) data category addresses this by providing a mechanism to point to this information.
- There are many flat or structured lists of domain related values, keywords, key phrases, classification codes, ontologies. The [Domain](#) data category does not propose a given list; rather it provides a mapping mechanism to associate values in content with workflow specific values needed for processing domain information.

6.9.2 Implementation



The [Domain](#) data category can be expressed only with global rules. The information applies to the textual content of the element, *including* child elements, but *excluding* attributes. There is no default.

GLOBAL: The **domainRule** element contains the following:

- A required **selector** attribute. It contains an XPath expression which selects the nodes to which this rule applies.
- A required **domainPointer** attribute that contains a relative XPath expression pointing to a node that holds the domain information.
- An optional **domainMapping** attribute that contains a comma separated list of mappings between values in the content and workflow specific values. The values may contain spaces; in that case they **MUST** be delimited by quotation marks.

Example xxx: The **domainRule** element

The **domainRule** element expresses that the content of the HTML `body` element is in the domain expressed by the HTML `meta` element with the name attribute, value `DC.Subject`. The **domainPointer** attribute points to that `meta` element.

```
<its:rules
  xmlns:its="http://www.w3.org/2005/11/its" version="2.0">
  <its:domainRule selector="/html/body"
domainPointer="/html/head/meta[@name='DC.subject']/@content"/>
</its:rules> [Source file: tbd]
```

Example xxx: The **domainRule** element associating existing values with workflow specific values

The **domainRule** element expresses that the content of the HTML `body` element is in the domain expressed by associated values. The **domainPointer** attribute points to the values in the source content. The **domainMapping** attribute contains the comma separated list of mappings. In the example, `automotive` is available in the source content, and `auto` is part of the workflow.

```
<its:rules
  xmlns:its="http://www.w3.org/2005/11/its" version="2.0">
```

```
<its:domainRule selector="/html/body"
domainPointer="/html/head/meta[@name='DC.subject']/@content"
domainMapping="automotive auto, medical medicine, 'criminal law' law,
'property law' law/>

</its:rules> [Source file: tbd]
```

Note:

In source content, if available, it is recommended to use [dublin core subject](#) as the metadata term for domain information. In HTML, this can be achieved via a meta element with the name="DC.subject" attribute.

In the area of machine translation (e.g. machine translation systems or systems harvesting content for machine translation training), there is no agreed upon set of value sets for domain. Nevertheless it is recommended to use a small set of values both in source content and within workflows, to foster interoperability. If larger value sets are needed (e.g. detailed terms in the law or medical domain), mappings to the smaller value set needed for interoperability should be provided. An example would be a **domainMapping** attribute for generalizing the law domain: domainMapping="'criminal law' law, 'property law' law, 'contract law' law".