Fine-Grained Semantic Annotations of Standards Content

Aim
We want to enrich standards with additional semantic information (semantic tagging) at content level (instead of document level).

Needs
- any span of text in standards should allow being annotated with multiple self-defined pieces of information (for example as key value pairs)
- this multi-tagging (“tagging” not necessarily in the sense of adding XML tags; it might be implemented as attributes on existing elements, as additional inline tags like std/std-id-group, as standoff annotations that rely on persistent IDs in the content, ...) should be linked to a metamodel like a taxonomy, ontology, database, controlled vocabulary, ... (analogous: @vocab, @vocab-term, @vocab-identifier, @vocab-term-identifier)
- our external classification system has the following characteristics that needs to be considered:
  - key value pairs, usually single-valued but multi-valued keys should be possible
  - hierarchical relationships
- a single annotation container (which might be a named-content element) should also accept multiple properties (analogous to multi-tagging described above). For example:
  - processing-method-property: the value of the tag was set manually or automatically
  - pre-post-processing-property: the value was set by a standardization committee (preprocessing) or was determined by postprocessing
- while we expect that in most cases the text extents to be annotated will properly nest, there may be the need to annotate text that stretches, for example, across paragraph boundaries, covering only parts of the affected paragraphs. For our use cases, this option to create overlapping annotations is an expendable requirement but others may need it.

Examples from IEC 60079-14
<p>Low-pressure sodium lamps shall not be transported unprotected through a hazardous area owing to the risk of ignition due to free sodium from a broken lamp.</p>
...
<p>Cables employ materials which could exhibit “coldflow” characteristics. Coldflow in cables can be described as the movement of the cable sheath under the compressive forces created by the displacement of seals in cable glands where the compressive force applied by the seal is greater than the resistance of the cable sheath to deformation.</p>

What we want
We would like to annotate this example with descriptions of the content and its context. For example, we have a classification system with the following key value pairs related to the example above:
- StandardizationSubject = {household appliance, electrical installation, switching device, control module, electrical equipement, luminaire, cable etc.}
- ProductLifecycle = {design, election, erection, delivery, maintenance, transport etc.}
- AtmosphericConditions = {damp, corrosive, explosive, normal, not explosive etc.}
- Motive = {health, safety, capability, environment, explanation etc.}
For every key-value pair we want to set the processing-method-property and the pre-post-processing-property as well.
Solution approach
We don't anticipate a specific solution, but one approach might be that the element subj-group is made available within all content elements.

The following code shows a possible tagging of the IEC 60079 paragraphs with the subj-group element:

```xml
<p>Low-pressure sodium lamps shall not be transported unprotected through a hazardous area owing to the risk of ignition due to free sodium from a broken lamp.

<subj-group id="DKE1.0" originator="DKE" vocab="DKE-Classification"
xml:lang="en" subj-group-type="Standardization for Electrical, Electronic & Information Technologies"
vocab-identifier="https://www.dke.de/ClassificationSystem">

<subject id="DKE1.0-1"
vocab-term-identifier="https://www.dke.de/ClassificationSystem/StandardizationSubject">
luminaire</subject>

<subject id="DKE1.0-2"
vocab-term-identifier="https://www.dke.de/ClassificationSystem/ProductLifecycle">
transport</subject>

<subject id="DKE1.0-3"
vocab-term-identifier="https://www.dke.de/ClassificationSystem/AtmosphericConditions">
explosive</subject>

<subject id="DKE1.0-4"
vocab-term-identifier="https://www.dke.de/ClassificationSystem/Motive">
safety</subject>

</subj-group>
</p>
```

...Cables employ materials which could exhibit “coldflow” characteristics. Coldflow in cables can be described as the movement of the cable sheath under the compressive forces created by the displacement of seals in cable glands where the compressive force applied by the seal is greater than the resistance of the cable sheath to deformation.

```
<subj-group id="DKE2.0" originator="DKE" vocab="DKE-Classification"
xml:lang="en" subj-group-type="Standardization for Electrical, Electronic & Information Technologies"
vocab-identifier="https://www.dke.de/ClassificationSystem">

```
Markup of requirements

Aim

We want to mark requirements in standards texts. For requirements our multi-tagging approach is not sufficient. A requirement is a complex type with multiple information and relations. We need a practical and efficient way of requirement tagging.

Needs

- any text in standards should allow to be marked as a requirement
- every requirement should be able to be marked with these types of information
  - a version and an id
  - the requirement type (shall, should, may, must, must) (not mandatory, but desirable for fast processing)
    - the specific type of requirements should be extensible and not limited to fixed keywords
  - semantic information (see multi-tagging proposal)
  - relations
    - between requirements e.g. groups of requirements (e.g. thematically coherent), sequence of requirements (e.g. sequence of examination), alternatives of requirements
    - between other text types e.g. example or graphics related to a requirement

Other open questions (low priority)

- To mark up granularity: A requirement consists of specific linguistic components (conditions, subject, object, actions, constraint of action). Should these components also be taggable?
- requirements which are in special text formatting sections such as tables or lists (not as contiguous paragraphs but with the information spread across table columns or list items, for example). Should this also be taggable?
- How to handle non-atomic requirements: e.g. a requirement over two sentences or two different requirements in one sentence?

Examples from IEC 60079-14
Example of a simple requirement
Simple apparatus shall be clearly identifiable by durable labelling.

What we want
We would like to mark this example as a single requirement with the following information:

ID = 1
Version = 1.0
Type = shall

Semantic Information:
- Standardization Subject = Simple apparatus

Relations: None

Example of alternative requirements
Electrical equipment should, as far as is reasonably practicable, be located in non-hazardous areas. Where it is not possible to do this, it should be located in an area where an explosive atmosphere is least likely to occur.

What we want
We would like to mark this example as two requirements with the following information:

Requirement 1:
ID = 2
Version = 1.0
Type = should

Semantic Information:
- Standardization Subject = Electrical equipment
- Atmospheric Conditions = not explosive

Relations: requirement 2 is an alternative of requirement 1

Requirement 2:
ID = 3
Version = 1.0
Example of requirements that belong together
Electrical equipment and materials shall be installed and used within their electrical ratings for power, voltage, current, frequency, duty and such other characteristics where non-conformity might jeopardize the safety of the installation.
In particular, care shall be taken to ensure that the voltage and frequency are appropriate to the supply system with which the equipment is used and that the temperature classification has been established for the correct voltage, frequency and other parameters.

What we want
We would like to mark this example as 2 requirements with the following information:

Requirement 1:
ID = 4
Version = 1.0
Type = shall
Semantic Information: (analogous to ex. above)
Relations: belongs to group 1

Requirement 2:
ID = 5
Version = 1.0
Type = shall
Semantic Information: (analogous to ex. above)
Relations: belongs to group 1