

CC/WD 36001:2019

Standardization documents – Metanorma – Document metamodel

THE CALENDARING AND SCHEDULING CONSORTIUM

TC VCARD

Ronald Tse **AUTHOR**
Nick Nicholas **AUTHOR**

CALCONNECT STANDARD

WORKING DRAFT

WARNING FOR DRAFTS

This document is not a CalConnect Standard. It is distributed for review and comment, and is subject to change without notice and may not be referred to as a Standard. Recipients of this draft are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

© 2019 The Calendaring and Scheduling Consortium, Inc.

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from the address below.

The Calendaring and Scheduling Consortium, Inc.
4390 Chaffin Lane
McKinleyville
California 95519
United States of America

copyright@calconnect.org
www.calconnect.org

CONTENTS

Foreword

Introduction

Influences

1. Scope

2. Normative references

3. Terms and definitions

4. Structure

5. Standards Document model

6. Metadata and bibliographic information models

7. Section models

8. References Section

9. Terms Section and Definition Section

10. Inline elements

11. Block

12. Data Types

Appendix A(normative) ISO Standard Document Model Profile

A.1. General

A.2. Document, Section

A.3. Block

A.4. Bibliography

Appendix B(normative) CSD Standard Document Model Profile

B.1. General

B.2. Document, Section

B.3. Block

B.4. Inline Elements

B.5. Bibliography

Bibliography

FOREWORD

The Calendaring and Scheduling Consortium (“CalConnect”) is a global non-profit organization with the aim to facilitate interoperability of collaborative technologies and tools through open standards.

CalConnect works closely with international and regional partners, of which the full list is available on our website (<https://www.calconnect.org/about/liaisons-and-relationships>).

The procedures used to develop this document and those intended for its further maintenance are described in the CalConnect Directives.

In particular the different approval criteria needed for the different types of CalConnect documents should be noted. This document was drafted in accordance with the editorial rules of the CalConnect Directives.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. CalConnect shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be provided in the Introduction.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

This document was prepared by Technical Committee *VCARD*.

INTRODUCTION

The process of generating standards documents is problematic, and becoming more so with the increased work burden on standards organisations. ISO is no exception to this issue, and concerns have been raised over both the efficiency and the effectiveness of the workflows around generating standards in those bodies and elsewhere.

- The rigorous and detailed directives around document structure articulated in [ISO/IEC DIR 2:2018](#) are being manually validated by the editorial teams, with mixed results.
- Editing of documents substantially revolves around Microsoft Word toolsets, with all the problems that conflating document structure and presentation introduce. The editorial process has to spend significant amounts of time on formatting rather than document content, and is vulnerable to error. Opportunities to automate document presentation independently of content (as is routinely done in the HTML world with CSS) are not being taken up.
- Existing exchange mechanisms for documents have been claimed to focus on document appearance, rather than document structure, and so do not alleviate the formatting burden on the editorial process.
- Different standards have completely different workflows, with different inputs: anyone submitting the same standards text to more than one standard body has to replicate a significant amount of work.

The Metanorma document metamodel provides data structures that enable an end-to-end standardization document production process. The process centers around the Metanorma document model, which is intended to encompass the core similarities between standards bodies' versions of standards documents; particularities of different standards bodies are kept to a minimum, and expressed as local extensions. The process generates a semantic representation of the standards document expressed in the abstract document model, based on non-WYSIWYG, lightly marked-up text-based input. A further step then generates different output formats from the same document model expression, treating them as the application of styling rather than as document editing.

This approach has the following benefits:

- The approach enables an end-to-end, author-to-publishing workflow, with well-managed artefacts along each stage, that editorial bodies can exercise appropriate control over. This becomes all the easier as each artefact until the very end is text rather than binary: that allows the document management infrastructure around computer source code (such as version control) to be used to manage standards document production.
- Much of the tedium and unpredictability around formatting documents to align with presentation requirements is automated, since the document model is semantic, and abstracted away from presentation. Formatting is applied only at the very end of the

process. This includes not only formatting details such as margin sizes and font choices, but also auto-numbering of sections and tables, cross-referencing, automated generation of references, and generating tables of contents.

- The structural and content requirements around standards documents can be captured in the document model, and documents can therefore be subject to formal validation against the model. That means that a Metanorma-based process can provide substantial feedback to authors on the extent to which they comply with the rules expressed in specifications like [ISO/IEC DIR 2:2018](#) – including rules around wording (e.g. where requirements can occur), as well as document structure.
- The document model is substantially the same across standards, as is the non-WYSIWYG text input that it is derived from. That means that text can be repurposed for different standards bodies with minimal effort: it is treated mostly as a swap-out of stylesheets, with only minor adjustments needed to content.

Influences

The Metanorma document model has drawn on several existing models, with a view to reflecting the expressiveness of existing standards, but also to align with the capability of established document production tools.

The standards models consulted most closely for Metanorma are those for ISO standards [ISO/IEC DIR 2:2018](#) and IETF standards ([RFC 7749](#), [RFC 7991](#)).

1. SCOPE

This document provides the Metanorma reference model for standards documents. Implementation of the reference model and serialisation formats out of scope of this document.

The mapping of the reference model to output formats such as HTML, and the application of organisation-specific styling to those outputs, is out of scope of this document.

This document discusses processes around customising the reference model for different standards, and describes customisation for the ISO and CSD standards as example profiles in [Appendix A](#) and [Appendix B](#). Prescribing profiles of the document model for particular standards, however, is out of scope of this document.

The modelling of bibliographies and bibliographic citations is part of the reference model, but is the subject of a separate proposal, [ISO 6900--](#), and is not discussed in detail in this document.

The modelling of text markup at the block and inline level, and of changes to documents is part of the reference model, but is the subject of a separate proposal, [basicdoc](#), and is not discussed in detail in this document.

2. NORMATIVE REFERENCES

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

- ISO 639 (all parts), *Codes for the representation of names of languages*
- ISO 8601, *Data elements and interchange formats – Information interchange – Representation of dates and times*
- ISO/IEC 10118 (all parts), *Information technology – Security techniques*
- ISO/IEC 14888 (all parts), *Information technology – Security techniques*
- ISO 15924, *Information and documentation – Codes for the representation of names of scripts*
- ISO 6900:–, *Information and documentation – Bibliographic reference model and serialization*
- CC 36010,¹ *Lightweight document – Document metamodel*
- IETF RFC 3986, *Uniform Resource Identifier (URI): Generic Syntax*

3. TERMS AND DEFINITIONS

For the purposes of this document, the following terms and definitions apply.

3.1. standardization body

standards body

body that has recognized activities in standardization

[SOURCE: 901-03-03]

3.2. standardization document

standards document

document, established by consensus and approved by a recognized *standardization body* (Clause 3.1), that provides, for common and repeated use, rules, guidelines or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context

[SOURCE: 901-02-02]

3.3. class

structure containing a description of an entity in terms of its components

3.4. subclass

class (Clause 3.3) which inherits from another class its component descriptions, and optionally adds to them its own component descriptions

3.5. document metamodel

model

formal specification of the structure of a document in terms of its components and their arrangement, expressed through *classes* (Clause 3.3)

3.6. document model

model

formal specification of the structure of a document in terms of its components and their arrangement, expressed through *classes* (Clause 3.3)

3.7. standards model

document model (Clause 3.6) for *standards* (Clause 3.2)

3.8. paragraph

subdivision of running text, normally run on throughout, that is separated from text before and after by a change of line and stands below any chapters or sections (Clause 3.11)

[SOURCE: ISO 5127:2017, Clause 3.5.8.07]

3.9. block

paragraph (Clause 3.8)-level grouping of text

3.10. inline Element

grouping of text that can be contained within a *paragraph* (Clause 3.8), including plain strings

3.11. section

hierarchical subdivision of a document, consisting of one or more *blocks* (Clause 3.9), and/or one or more sections

3.12. identifier

a character, or group of characters, used to identify or name an item of data and possibly to indicate certain properties of that item

[SOURCE: 714-21-07]

3.13. term

expression with a particular meaning that is part of the specialized vocabulary of a field

4. STRUCTURE

The Metanorma document model is expressed hierarchically.

- At its base is a *Basic Document* model ([CC 36010](#)), which expresses the structure of generic documents.
- The *Basic Document* model is specialised as a *Standards Document* structure, conveying the structure particular to standards documents as a class.
- The Standards Document model is further specialised into models specific to particular standards.
- Specific standards models can form the basis in turn of other standards models. For example, the Chinese National Standards model, as described in [GB/T 1.1-2009](#), is structurally modelled after ISO documents, and is explicitly described as a localisation of the 2004 version of [ISO/IEC DIR 2:2004](#). So the Chinese Standards model is best expressed as a specialisation of the ISO Standards model.

Specialisation of a model consists of:

- Adding classes to a base model.
- Changing attributes of a base model class. This is not restricted to adding attributes, as is the case in typical entity subclassing; it can also include removing attributes from a class, changing their obligation and cardinality, and changing their type, including changing enumerations. Attributes can be overruled at any level; for example, standards-specific models routinely enhance the bibliographic model at the base of the hierarchy.
- For reasons of clarity, renaming classes and attributes is avoided in specialisation.

The classes involved in the document model are of three classes:

- Sections ([basicsection](#), [Clause 7](#))
- Blocks (paragraph-level groupings of text) ([basicblock](#), [Clause 11](#))
- Inline elements (groupings of text smaller than a paragraph, including plain strings) ([inline](#))

In the Metanorma model, the classes are in a strict hierarchical relation: documents consist of sections, which consist of blocks, which consist of inline elements. Sections can be nested within sections (e.g. clauses and subclauses); blocks can be nested within blocks (e.g. nested lists). Although the model does not currently provide for it, inline elements can also be embedded within other inline elements (e.g. bold + italics). However, sections are not expected to be siblings of blocks, nor blocks of inline elements. For example, a list (block) is not expected to occur next to inline text within a paragraph. For that reason, paragraphs cannot contain other block elements, such as lists or tables.

NOTE This constraint is inherited from lightweight text markup, and is not present in XML-based schemas like HTML, TEI-C, or DocBook. It helps to simplify the document model significantly, and while it sacrifices some expressive potential, the difference is minor, particularly with regards to the rendering of paragraphs.

The *Basic Document* model contains the bulk of the blocks and inline elements in the Metanorma model; that model is documented separately ([CC 36010](#)), and is not discussed further here. The Standards Model and the standard-specific models mostly deal with the organisation of sections within a standards document. However, because Metanorma inherits from the Basic Document model, specialisation in Metanorma can customise Basic Document classes, as well as classes specific to Metanorma.

The Metanorma document model depends on the [ISO 690](#) bibliography model for its expression of bibliographic references. The specific bibliography model instantiation and serialisation it uses is the subject of a separate proposal, [ISO 6900:--](#), and is not discussed further here.

5. STANDARDS DOCUMENT MODEL

The *Standards Document* model extends the *Basic Document* modelling of the document by requiring the following specific types of section:

- An optional *boilerplate* section, intended to appear at the front of the document, and consisting of text addressing *copyright*, *license*, *legal*, and *feedback* concerns (each modelled as Standard Hierarchical Sections: [Clause 7](#)).
- Zero or more optional *preface* sections (Standard Hierarchical Sections: [Clause 7](#)).
- One or more *sections* (Standard Hierarchical Sections).
- Zero or more *annexes* (Standard Hierarchical Sections).
- Zero or more *bibliographies* (Standard References Sections: [Clause 8](#)).

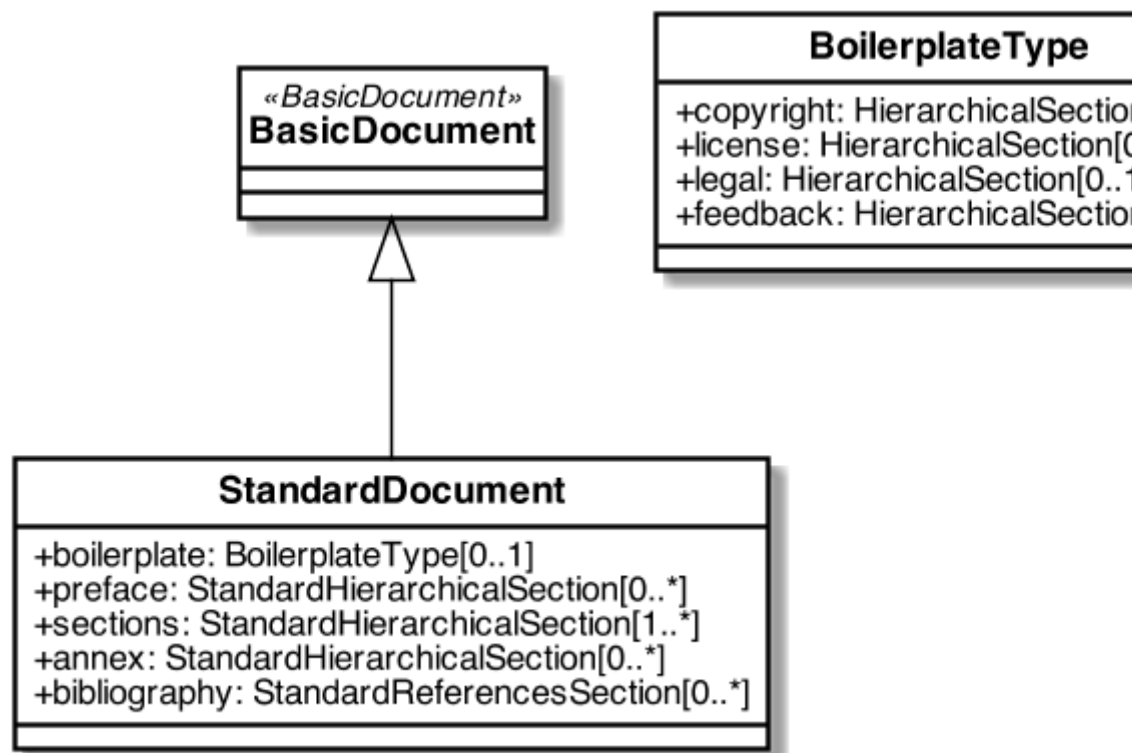


Figure 1 – Standards Document model: Document

6. METADATA AND BIBLIOGRAPHIC INFORMATION MODELS

Bibliographic information appears in two places in Metanorma, following the Basic Document model: document citations under Standard References Sections: [Clause 8](#), and document metadata (`bibdata`) as a root element of the document.

The modelling of `bibitem` (document citations) follows the *BibliographicItem* class in [ISO 6900](#)--, and readers are referred to that specification. The *BibliographicItem* class is intended to capture document citations, and to be applicable to any document type, without any further specialisation; that is because a document can cite documents of any type. The modelling of citations is not intended to vary between different Standards Developing Organizations, let alone between standards documents and other types of document.

The modelling of `bibdata` (document metadata) likewise follows the *BibliographicItem* class in [ISO 6900](#)--. The `bibdata` class inherits from `Relaton` the ability to extend metadata with metadata specific to a document class; this information is modelled in `BibDataExtensionType`, and is contained in `bibdata/ext`. The only extension point modelled as generically applicable in Metanorma model is the document type, which is populated in the generic instance with the single value “standard”. It is assumed that particular specialisations of the document model will substitute their own enumerations of particular subclasses of document, which will be more granular.

7. SECTION MODELS

The *Standards Document* model extends the *Basic Document* modelling of sections by specialising the Basic Section class as follows:

- All sections in a *Standards Document* model are instances of the Standard Section class, which is a subclass of Basic Section.
- The Standard Section class indicates its status as normative or informative explicitly, through the `status` attribute. (This attribute has been added because it is overtly specified in [ISO/IEC DIR 2:2018](#), but the normative status of individual sections in a standard is pertinent to any standards document.)
- Sections in a standards document fall into two categories: Clause Sections, which are numbered, and form part of the main flow of the document, and Content sections (Standard Hierarchical Sections), which are optionally not numbered, and which lie outside the main flow of the document.
- Clause Sections are by default hierarchical (but see [Clause 9](#)). The subclauses of Clauses are themselves Clause Sections; they continue the hierarchical numbering of Clause Sections at deeper hierarchical levels. The subclauses of Content sections are Standard Sections, and need not be numbered.
- Annexes and Clause Sections are modelled in the same way in the *Standards Document* model. However they are expected to be numbered differently, and to be rendered as separate sequences.

Although it is difficult to capture in a container-based model, the *Standards Document* model requires “hanging paragraphs” to be avoided: a Clause section in particular should contain either blocks or subsections, but not both. This is a requirement in [ISO/IEC DIR 2:2018](#), but the need to reference text in clauses unambiguously by clause number makes the avoidance of hanging paragraphs desirable for standards documents in general. The requirement is less pressing for Content sections (which are unnumbered), and Annexes (as exemplified in [ISO/IEC DIR 2:2018](#) itself.)

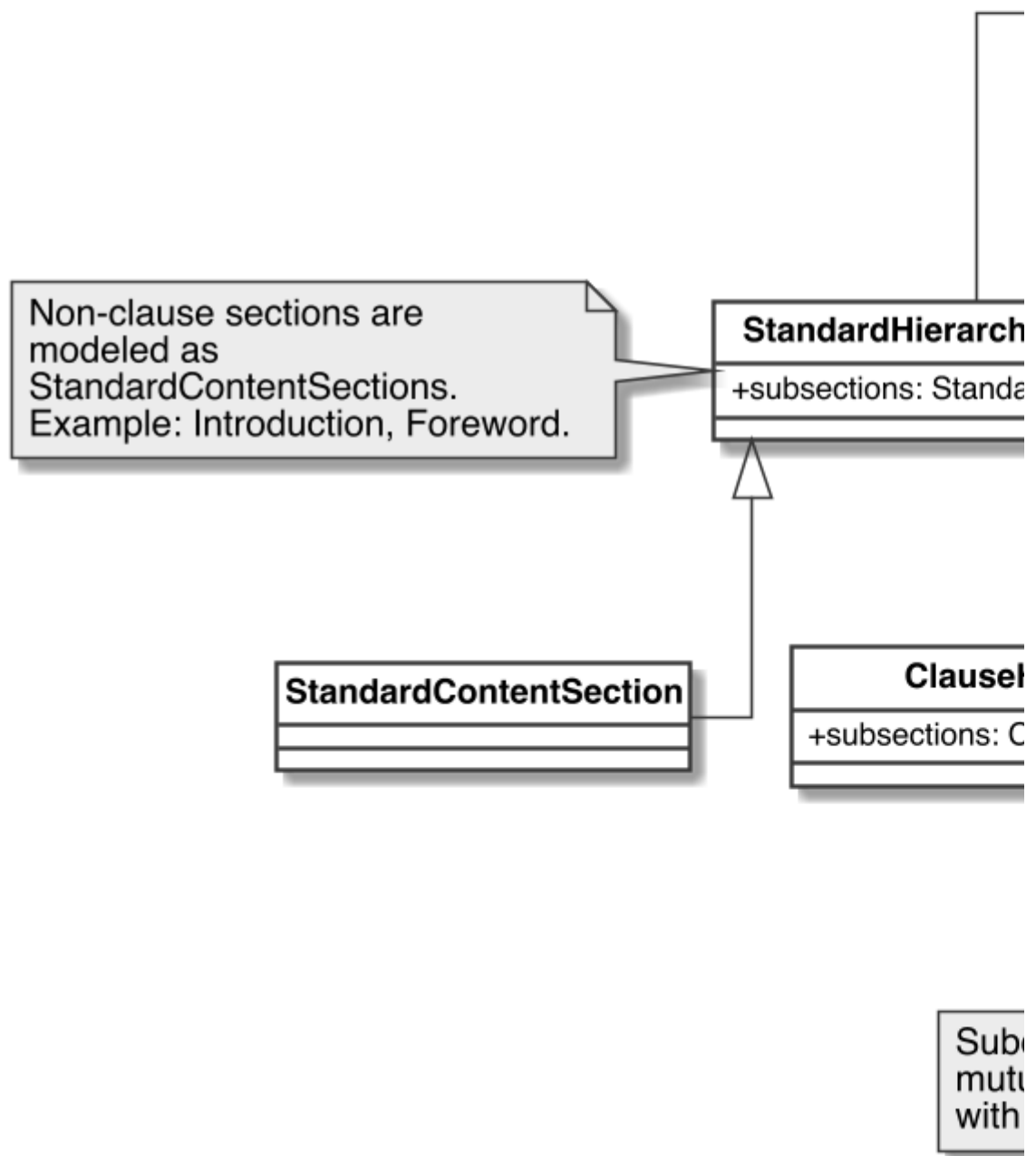


Figure 2 – Standards Document model: Section

8. REFERENCES SECTION

Standard References Sections are a subclass of Standard Section. Like the References Sections of the *Basic Document* model ([basicsection](#)), they are leaf nodes which contain zero or more bibliographical references (as modelled in [ISO 6900:--](#)), along with any prefatory text.

9. TERMS SECTION AND DEFINITION SECTION

The *Standards Document* model models two further subclasses of Clause Sections:

- *Definition Sections* consist of one or more Definition lists (see [CC 36010](#)), and are used to define symbols and abbreviations used in the remainder of the document. They can also be used as glossaries, with simple definitions, in contrast to the more elaborate definitions given in Terms Sections.
- *Terms Sections* give elaborated definitions of terms used in the document; they are intended to align to the structure of terms in [ISO/IEC DIR 2:2018](#). Term Sections consist of zero or more Terms; *Terms* in turn are composed of the following elements:

<code>id</code>	An optional identifier for the term, to be used in cross-references.
<code>preferred</code>	One or more names under which the term being defined is canonically known.
<code>admitted</code>	Zero or more names which are acceptable synonyms for the term being defined.
<code>deprecates</code>	Zero or more names which are deprecated synonyms for the term being defined.
<code>domain</code>	An optional semantic domain for the term being defined, in case the term is ambiguous between several semantic domains.
<code>definition</code>	The definition of the term applied in the current document.
<code>note</code>	Zero or more notes about the term.
<code>example`</code>	Zero or more examples of how the term is to be used.
<code>source</code>	Zero or more bibliographical sources for the term. These include the <code>origin</code> of the term, which is its bibliographical citation (as defined in ISO 6900:--); the <code>status</code> of the definition (whether <i>identical</i> to the definition given in the origin cited, or <i>modified</i>); and, if the definition is modified, a description of the <code>modification</code> to the definition applied for this document.

Both types of section can include prefatory text; in the case of Definition Sections, each Definition list can be preceded by prefatory text. Both types of section are leaf nodes, and cannot contain subclauses of their own: nested Terms or Definition Sections are modelled as Clause Sections with subclauses, terminating in Definition and Terms clauses.

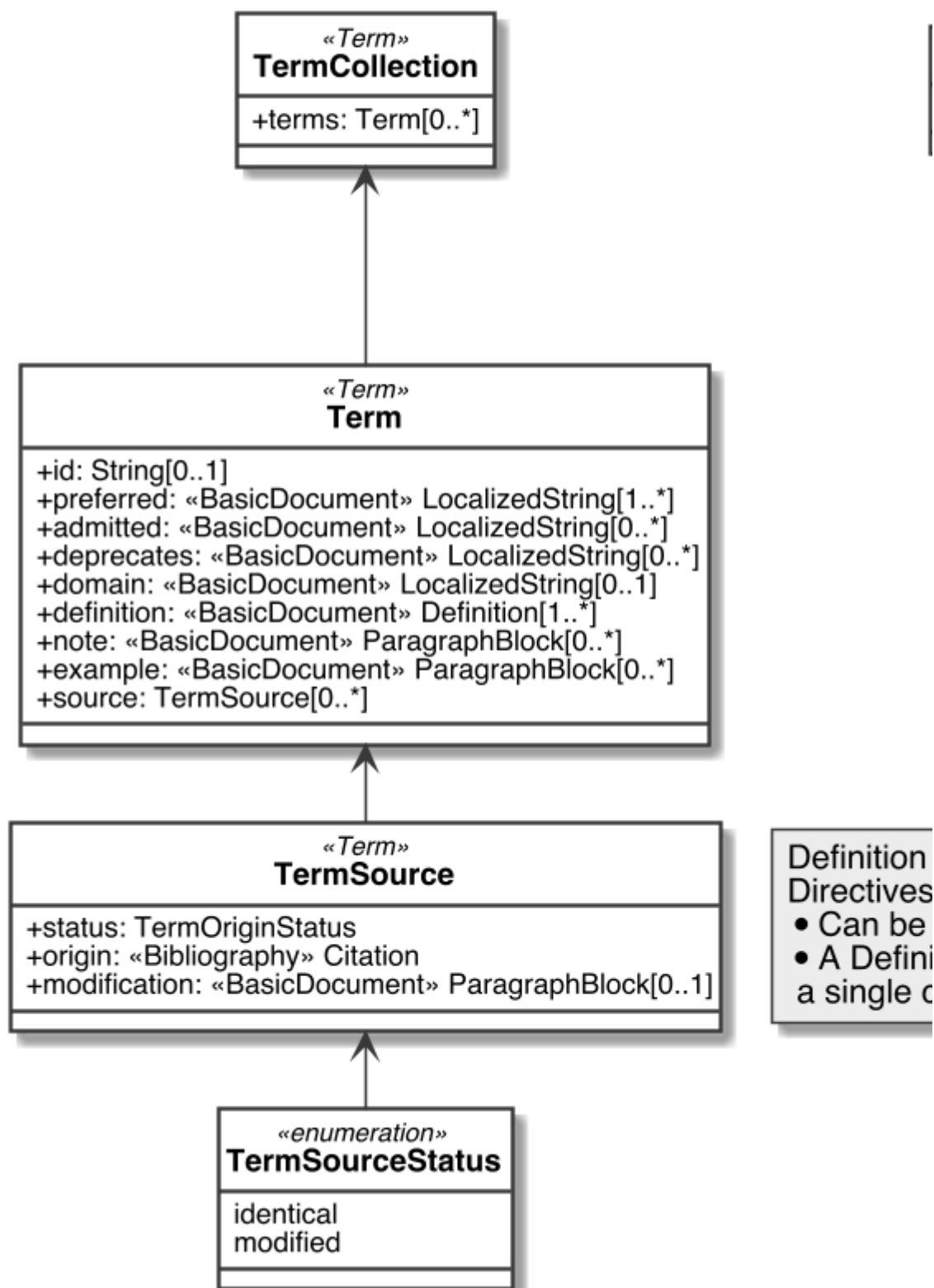


Figure 3 – Standards Document model: Term Section and Definition Section

10. INLINE ELEMENTS

There is no difference in the modelling of inline element models between the *Basic Document* model and the *Standards Document* model. Specialisations and profiles of the Standards Document model may specialise the modelling of inline elements.

11. BLOCK

There is no difference in the modelling of blocks between the *Basic Document* model and the *Standards Document* model. Specialisations and profiles of the Standards Document model may specialise the modelling of blocks.

12. DATA TYPES

There is no difference in the modelling of data types between the *Basic Document* model and the *Standards Document* models. Specialisations and profiles of the Standards Document model may specialise the modelling of data types.

APPENDIX A (NORMATIVE) ISO STANDARD DOCUMENT MODEL PROFILE

A.1. General

The *ISO Standard Document* model is given here as an exemplar of the application of Metanorma modelling to a particular standard.

A.2. Document, Section

The *ISO Standard Document* model extends the *Standards Document* modelling of the document by requiring the following specific types of section:

- A *Preface*, consisting of a mandatory *Foreword* (a Basic Section: [basicsection](#)), and an optional *Introduction* (a ClauseSection: [Clause 7](#), since an ISO Introduction can optionally have numbered subsections).
- One or more Clause Sections ([Clause 7](#)).
- Zero or more Annexes ([Clause 7](#)).

Annexes in the *ISO Standard Document* can optionally contain one or more *Appendixes*, which can be thought of as annexes of the annex section.

NOTE Appendixes are not defined in [ISO/IEC DIR 2:2018](#), but they appear in the companion standard [ISO/IEC DIR 1](#). For example, Annex SL of [ISO/IEC DIR 1](#) includes Appendix 1, Appendix 2, and Appendix 3; Appendix 2 “High level structure, identical core text, common terms and core definitions” itself contains an Introduction and 10 clauses, including a Terms clause.

The *ISO Standard Document* model document also includes zero or more `termSources` elements, giving the citations for the documents that the Terms Section draws on for its definitions and repertoire of terms.

The *ISO Standard Document* model document uses ISO-specific extensions to the [ISO 6900](#)-- metadata about a document ([Appendix A.4](#)).

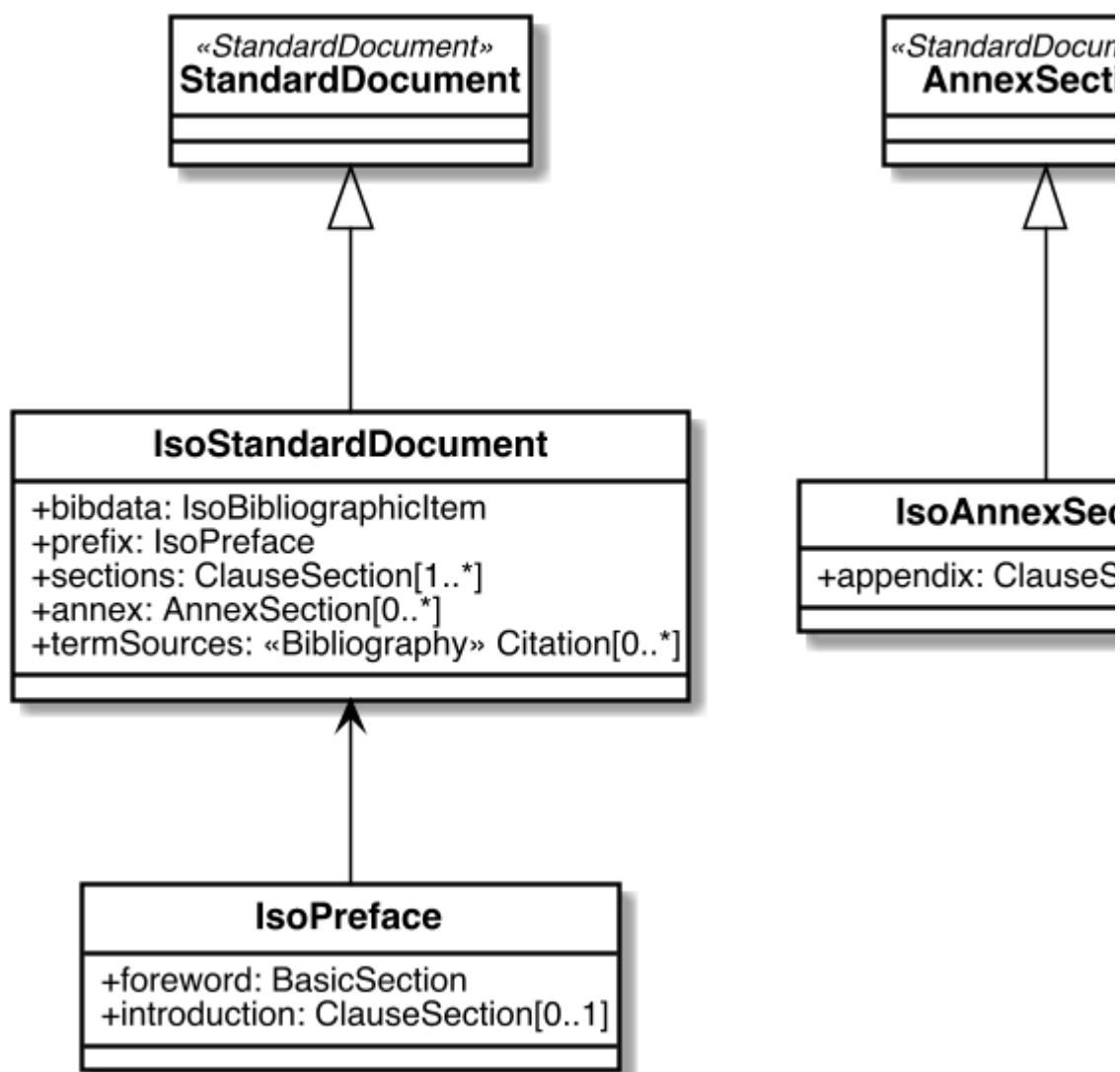


Figure A.1 – ISO Standard Document model: Document

A.3. Block

The *ISO Standard Document* model extends the *Standards Document* modelling of blocks for one block definition:

- The types that an Admonition can belong to are overridden to belong to the following set prescribed in [ISO/IEC DIR 2:2018](#): *Danger, Caution, Warning, Important, Safety Precaution*.

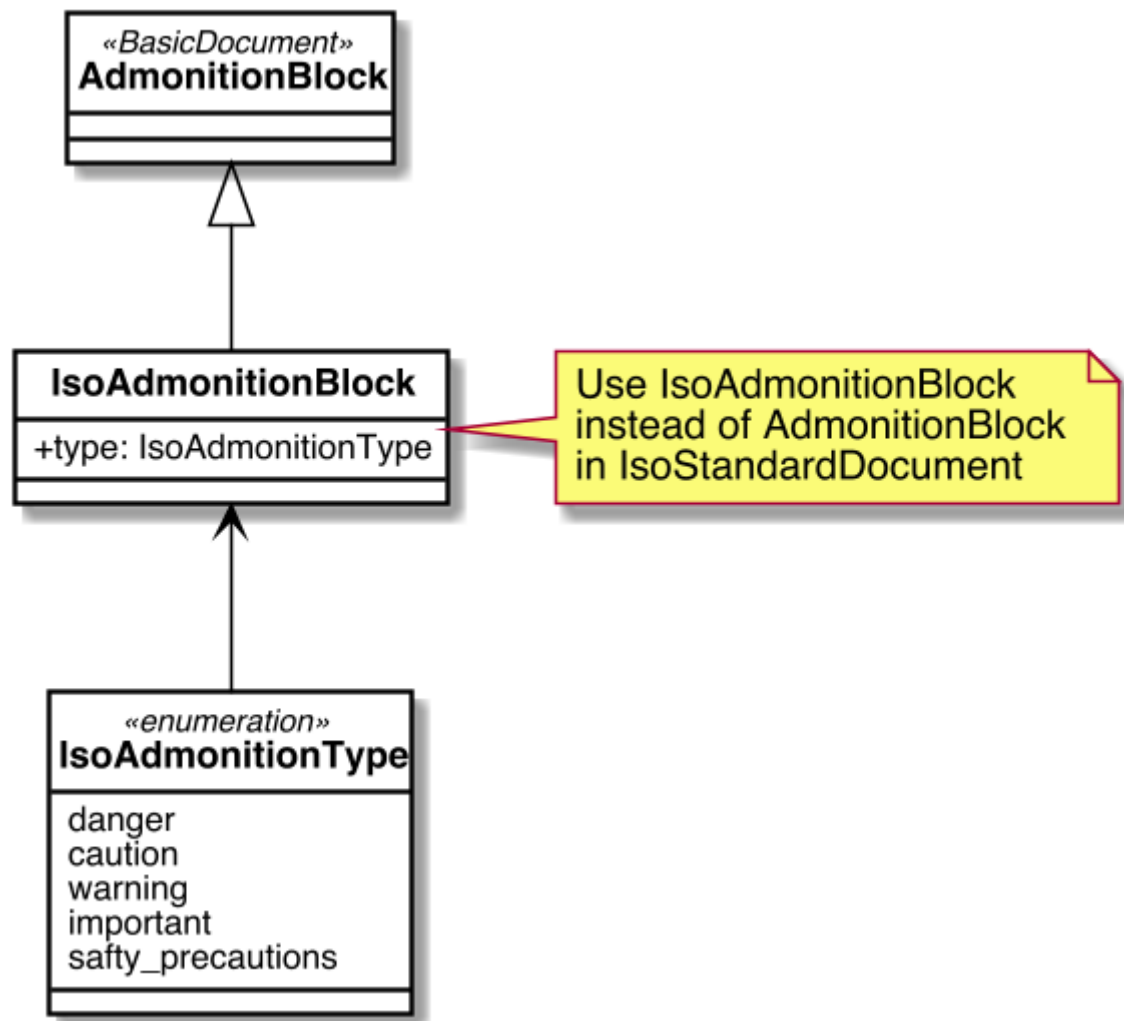


Figure A.2 – ISO Standard Document model: Block

A.4. Bibliography

We note here the *ISO Standard Document* overrides of the *ISO 6900:--* bibliographic model; although the contents of *ISO 6900:--* are out of scope of this document, its customisation for a specific standard are in scope.

ISO 6900:-- is intended as a highly generic model, and custom bibliographic elements would impede the reuse of bibliographic entries between standards. However, it is useful for the metadata describing the current document (bibdata: *CC 36010*) to be enriched to provide metadata relevant to the standard class.

The bibliographic modelling of ISO documents is enriched compared to *ISO 6900:--* as follows:

- The document identifier is refined to contain three optional parts: a project number, a part number, and a technical committee document number.

- The document title is broken down into an optional Introductory component, a mandatory Main component, and an optional Part component.
- The document type is constrained to be one of the following: *International Standard*, *Technical Specification*, *Technical Report*, *Publicly Available Specification*, *International Workshop Agreement*, *Guide*.
- The document status is broken down into an ISO Document Stage code ([ISO International harmonized stage codes](#)), an ISO Document Substage code ([ISO International harmonized stage codes](#)), and an Iteration number.
- The editorial group responsible for the document is defined with four components: a mandatory Technical Committee, an optional Subcommittee, an optional Workgroup, and an optional Secretariat. The Technical Committee, Subcommittee and Workgroup, in turn, are instances of the ISO Subgroup class, defined as having a name, an optional type, and an optional number.
- The document has one or more International Classification for Standards classifications [ISO International Classification for Standards](#). The ICS classification is defined as a code, and an optional text explaining the code.
- The document has zero or more source URIs.
- The document has an optional abstract.
- The document has an optional “All Parts” boolean attribute, indicating whether the bibliographic description applies across all Part documents under the same project number (which makes it a multi-document description).

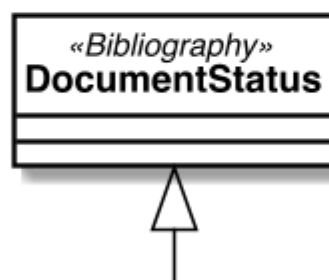




Figure A.3 – ISO Standard Document model: Bibliography

APPENDIX B (NORMATIVE) CSD STANDARD DOCUMENT MODEL PROFILE

B.1. General

The *CSD Standard Document* model is given here as an exemplar of the application of Metanorma modelling to a particular standard (standards published by the Calendaring and Scheduling Consortium).

B.2. Document, Section

The *CSD Standard Document* model extends the *Standards Document* modelling of the document by requiring the following specific types of section:

- An optional *Foreword* (a Standards Hierarchical Section: [Clause 7](#)).
- An optional *Introduction* (a Standards Hierarchical Section: [Clause 7](#)).
- One or more Clause Sections ([Clause 7](#)).
- Zero or more Annexes ([Clause 7](#)).

B.3. Block

The *CSD Standard Document* model extends the *Standards Document* modelling of blocks for one block definition:

- The definition of Figure is altered to allow Ascii art as an option for its content, as a string (`pre`). As with Sourcecode ([basicsourcecode](#)), the string should be treated as preformatted text, with whitespace treated as significant.

B.4. Inline Elements

The *CSD Standard Document* model extends the *Standards Document* modelling of inline elements by one element:

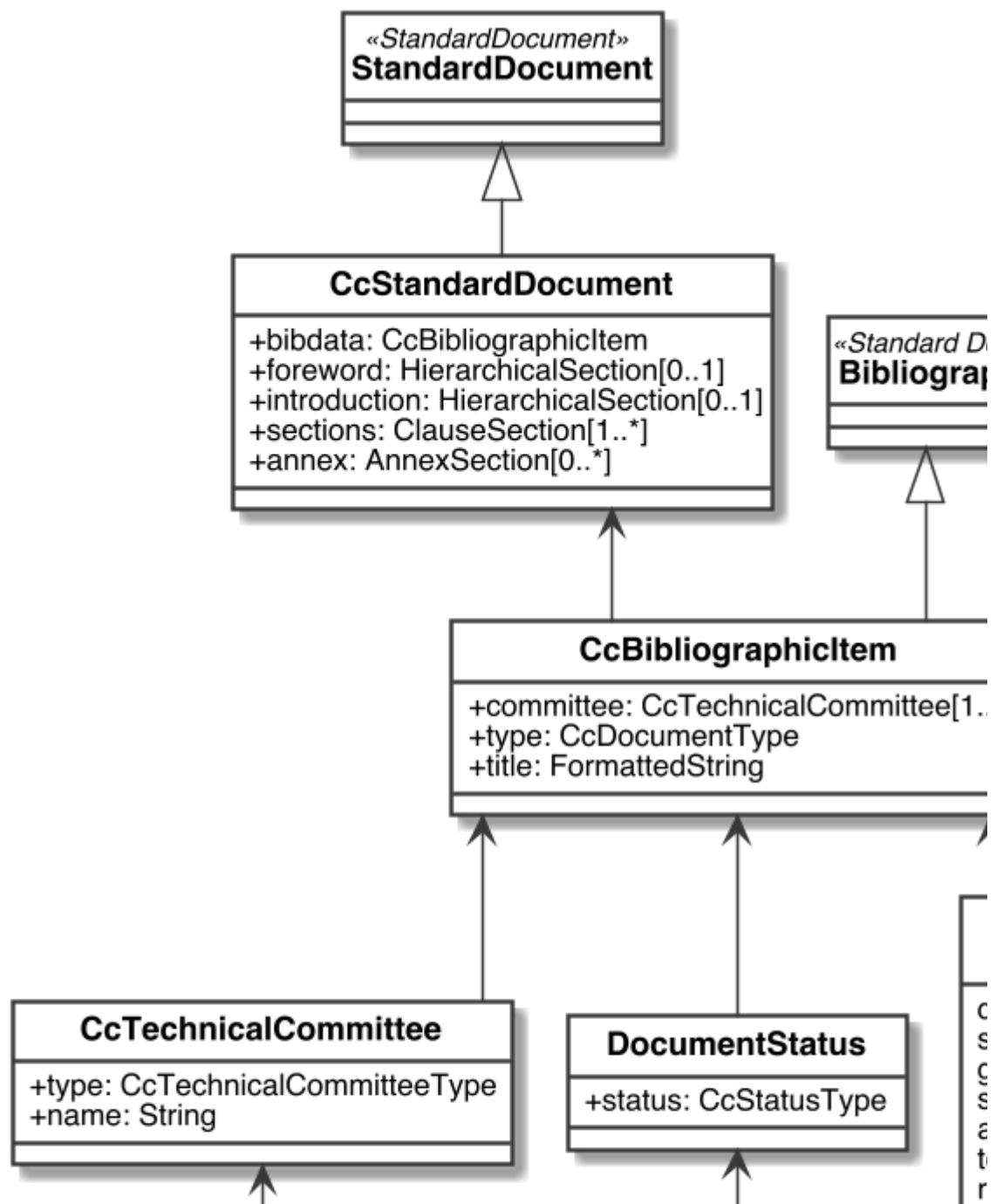
- The `keyword` element is added as a subclass of Text Element ([textelements](#)). The element is used to represent code keywords, as with `code` in HTML, and should be rendered in monospace.

B.5. Bibliography

The provisos which apply to ISO bibliographic entries ([Appendix A.4](#)) also apply here.

The bibliographic modelling of CSD documents is enriched compared to [ISO 6900](#)-- as follows:

- The document type is constrained to be one of the following: *Code, Presentation, Proposal, Standard, Report*.
- The document status is constrained to be one of the following: *Proposal, Working Draft, Committee Draft, Final Draft, Published, Withdrawn*.
- The editorial group responsible for the document is defined with a single components: a mandatory Technical Committee. The Technical Committee has a type of either *Technical* or *Provisional*.



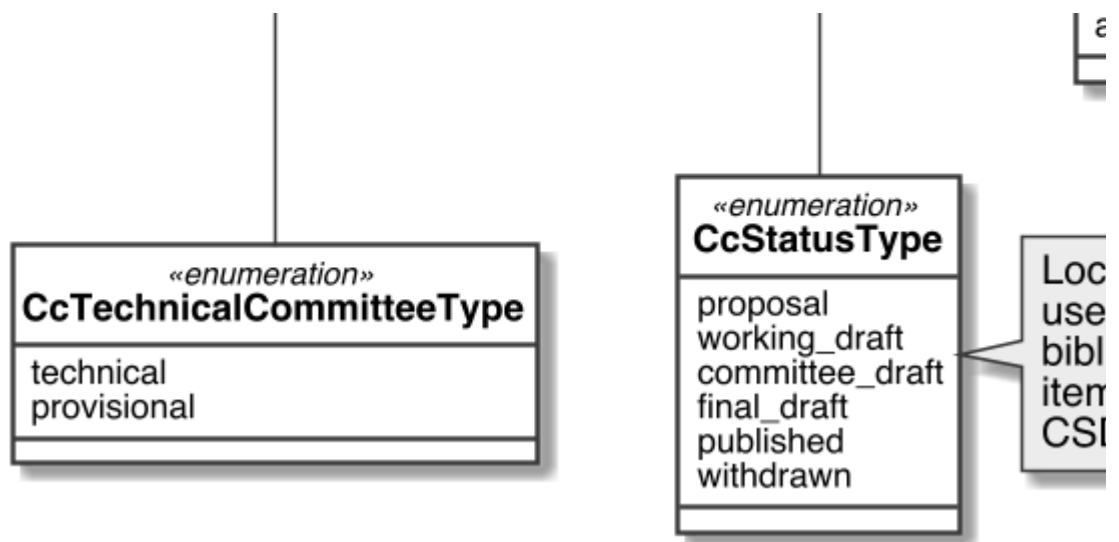


Figure B.1 – CSD Standard Document model

BIBLIOGRAPHY

- [1] ISO 690, *Information and documentation – Guidelines for bibliographic references and citations to information resources*
- [2] ISO 5127:2017, *Information and documentation – Foundation and vocabulary*
- [3] ISO/IEC DIR 1, *Procedures for the technical work*
- [4] ISO/IEC DIR 2:2004, *Principles and rules for the structure and drafting of ISO and IEC documents*
- [5] ISO/IEC DIR 2:2018, *Principles and rules for the structure and drafting of ISO and IEC documents*
- [6] Chinese Standard GB/T 1.1-2009, *标准化工作导则 第1部分 标准化文件的结构和起草规则*
- [7] IETF RFC 7749, *The “xml2rfc” Version 2 Vocabulary*
- [8] IETF RFC 7991, *The “xml2rfc” Version 3 Vocabulary*
- [9] ISO International Classification for Standards, *ISO International Classification for Standards (ICS)*, <https://www.iso.org/publication/PUB100033.html>
- [10] ISO International harmonized stage codes, *International harmonized stage codes*, <https://www.iso.org/stage-codes.html>

¹In draft

