6 The SDP and Metadata

6.1 Introduction

This chapter introduces the metadata for the Schedule Data Profile. Metadata is used to describe one or more SDP Documents that form the collection of a schedule version (including all revisions and route depot versions) or schedule version revision.

6.2 What is SDP Metadata?

Metadata is often defined as "data about data". It summarizes the "who, what, when, where, why and how" of the data set. Metadata helps people find data that is appropriate for their use. SDP metadata will help[1]:

- preserve the data history so that it can be re-used or adapted,
- assess the age and character of your data set
- provide a place for agencies to document extensions to their data sets for internal or special projects and applications
- instill data accountability by requiring you to state what you know about the data and realizing what you don't, but should, know about your data
- limit data liability by explicitly designating the effective and administrative limits of use of the data.
- monitor data development by regular review of the process steps completed and recorded within the metadata
- access the lineage and content of the data production process

In theory, metadata is a "best practice". In practice, metadata is time consuming and tedious. The geospatial industry has learned over the years that metadata is an important tool in order to effectively manage and re-use data resources. To this end, an SDP Document submitted for use should be accompanied by a metadata document. A data repository that stores a SDP Document should use the metadata document enable Data Consumers to discover the data resources available at the site. To aid in the collection of metadata, the data repository, when it registers the SDP Document, should support the documentation and importation of metadata components of the SDP Document submissions.

6.3 Why SDP Metadata

The TSDEA Data Repository will serve as a portal in which consumers of schedule data may find resources submitted by downstate NY regional Operators. This portal environment requires a "registry" of information about the quality, fitness for use, dissemination policies and interfaces supported by each data set organization. In theory, the TSDEA will support multiple transit agency schedules; it may support several schedule versions and revisions for a single transit operator. Managing these data sets will provide information on the SDP Document submission's identity, status, extensions, customizations, data quality and other information needed to describe the purpose, fitness for use and distribution policies. Management will be

---

[1] List adapted from "Why bother with Metadata", http://www.fgdc.gov/metadata/metadata-business-case

accomplished through a SDP Metadata XML Schema submission and metadata application programming interface (API).

6.4 SDP Metadata XML Schema

6.4.1 Overview

The SDP Metadata XML Schema is composed of several "packages". These packages are based on best practices identified by several metadata standards including

- ASTM E2468-05 - Standard Practice for Metadata to Support Archived Data Management Systems
- IEEE 1489/1488 ITS Data Dictionary and Message Set
- FGDC Content Standard for Digital Geospatial Metadata
- ISO 11179 – Metadata Registries

The ASTM references seven categories of information that should be included in a metadata. They include:

1. Identification Information
2. Data Quality Information
3. Spatial Data Organization Information
4. Spatial Reference Information
5. Entity and Attribute Information
6. Distribution Information
7. Reference Information

Many of the elements from the ASTM reference are incorporated in the requirement description for the SDP Metadata.

6.4.2 SDP Metadata Requirements

This section describes the requirements that drive the SDP Metadata XML Schema. The needs are listed in Table 1.

Table 1: Requirement Description for SDP Metadata

#	Category	Requirements
1	Identification Information	• SDP shall contain identification information about the submission and standards used to define an operator's set of Schedule and Related Data. The information should include: o Originator/submitter (name, telephone and email) o Original registration date o Approval staff person (name, signature, date) o Publication date o List of SDP Files covered by this metadata (only one Schedule Version per metadata) o SDP Schema version used to produce SDP files o SDP Metadata Schema version

#	Category	Requirements
		○ Online linkages (e.g., maps, pdf timetables)
2	Description and Time Period of Content	• Abstract (general description of SDP files) • Schedule Version ○ revision • Activation date / deactivation date (or 12-12-9999) • Data sources (e.g., Scheduling Application, RTIF, including version numbers of applications or file descriptions)
3	Status	• Phase: registered, Levels 1-3 • Date entered phase • Estimated date for next schedule version update • Update frequency: schedule change frequency (biweekly, quarterly, semi-annually) • Processing and Change Logs ○ Tests conducted • Type, who, what, why, activation date, disposition of updated record(s), ○ Revision History • Lineage (actual processes used to convert data) ○ Transformation processes needed to convert data ○ Date recorded ○ change start date, change stop date
4	Schedule Calendar	• Calendar • List of holidays and special days • Schedule Calendar Days
5	Data Quality Information	• Procedures (instructions) ○ Process of loading data set(s) ○ Transformations required to pass testing (levels 1-3) • List of Exceptions or Constraints ○ Exception descriptions ○ Date recorded
6	Special Conventions	• Schedule day (based on up to a 36 hours clock which may include a start time from the day before and end at a time the next day) • Naming convention for indexes (may be included in XML schema annotation)
7	Code Set Extensions	• Based on codes supported by SDP XML Schema.
8	Spatial Dataset	• Description of each location reference with respect to its measurement quality and datum or coordinate system (e.g., GPS, map coordinates using SP-NY), e.g., x-coordinate/y-coordinate pair is New York State Plane. Accuracy of measurement or map base may include its own metadata. ○ For example, the New York State Plane may reference the State GIS metadata link: http://www.nysgis.state.ny.us/gisdata/metadata/nysogs.sta

#	Category	Requirements
		tepln.html o Default is NAD '83 and UTM 18 o Default spherical coordinates units is decimal degrees with 6 decimal places • Linear reference units (e.g., feet, meters, decimeters; include precision, resolution and accuracy for measurement) o Default is feet
9	Distribution Information	• List of available interfaces • Contact Information o Name, telephone, email, URL

When the TSDEA provides a capability to document and store the metadata, many of the requirements may be automatically or manually entered into a metadata document. The schedule calendar requirement was separated from the metadata and will be submitted as a separate document. Guidance information on the schedule calendar XML schema may be found in the SDP Guidance Part 2, Chapter 9.

6.4.3 Metadata XML Schema Model

The Metadata XML Schema is described in this section. The high level SDP Metadata schema is depicted in Figure 1: High Level SDP Metadata XML Schema. The depiction includes the categories described in the Requirements Description. Each node of the schema includes the details related to the required information.
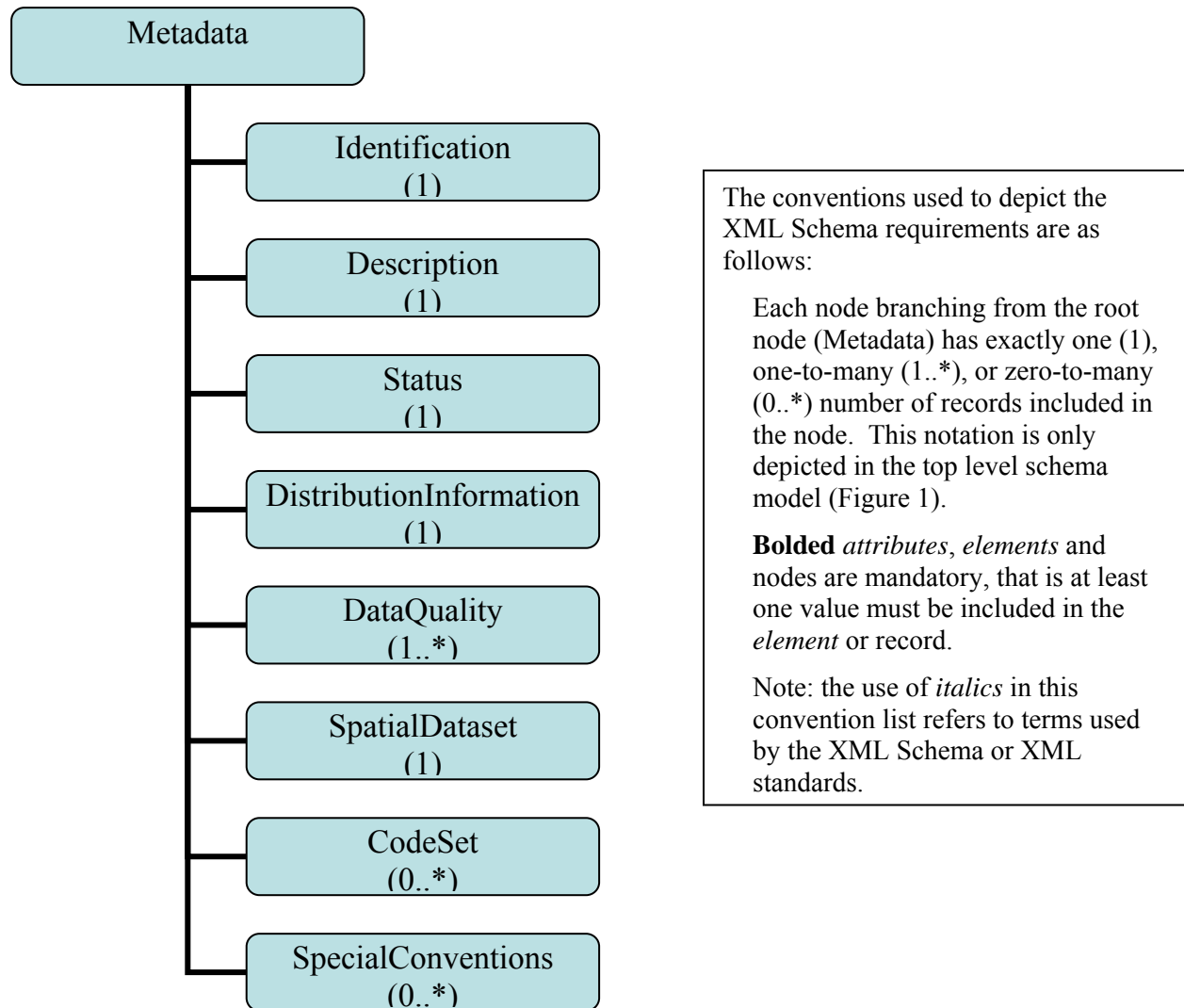
The conventions used to depict the XML Schema requirements are as follows:

Each node branching from the root node (Metadata) has exactly one (1), one-to-many (1..*), or zero-to-many (0..*) number of records included in the node. This notation is only depicted in the top level schema model (Figure 1).

Bolded *attributes*, *elements* and nodes are mandatory, that is at least one value must be included in the *element* or record.

Note: the use of *italics* in this convention list refers to terms used by the XML Schema or XML standards.

Figure 1: High Level SDP Metadata XML Schema

Using the XMLSpy software (see Appendix A for description of the notation), the high level SDP Metadata XML Schema is depicted in Figure 2. The Metadata document attribute group is depicted in Figure 3. Figures 4-11 describe the high level elements in the schema. The schema is composed of three files, similar to the SDP XML Schema. The files include:

- Main Schema Document: SDP_Metadata_XML_Schema_V0.1.xsd
- Complex Type Descriptions: SDPM_common_V0.1.xsd
- Simple Type Descriptions: SDPM_domain_V0.1.xsd

These documents may be found on http://www.consystec.com/tsdea/rstwg/docs.html.

Although, a white paper describing the requirements was circulated during the TSDEA Project, the schema and requirements have not yet been implemented.
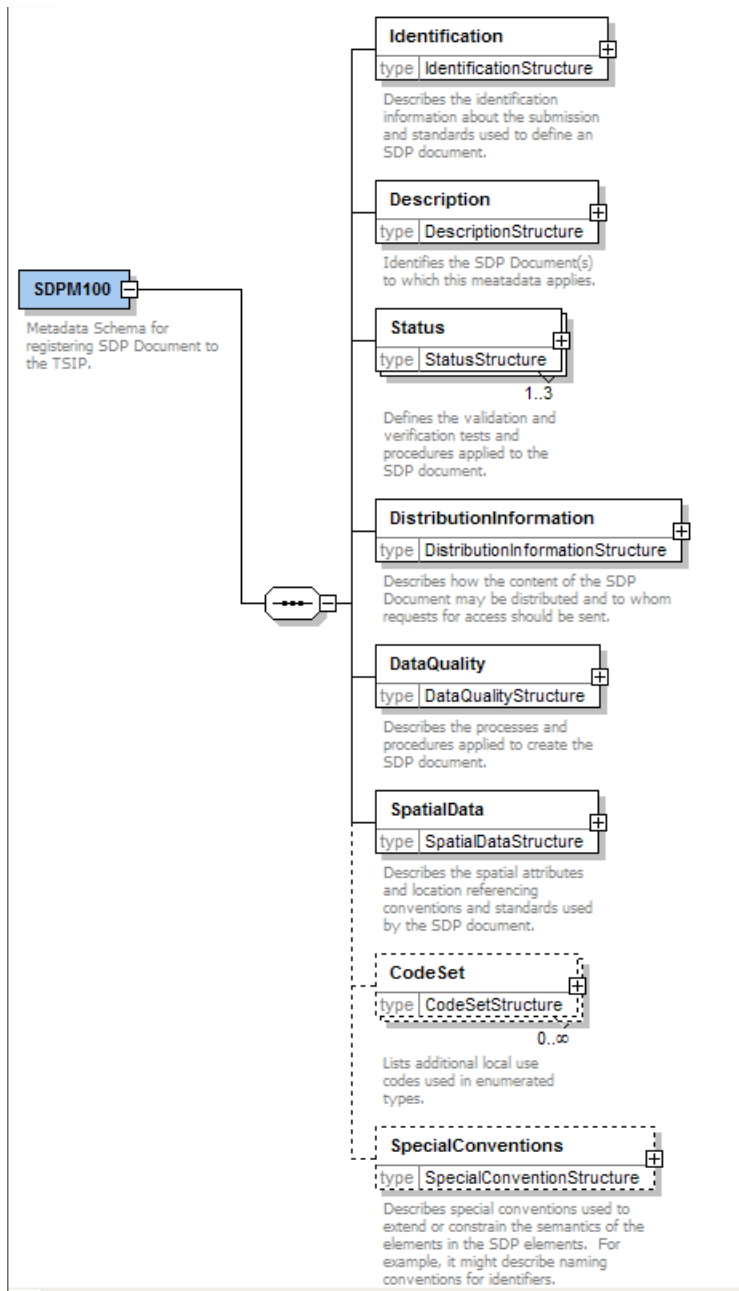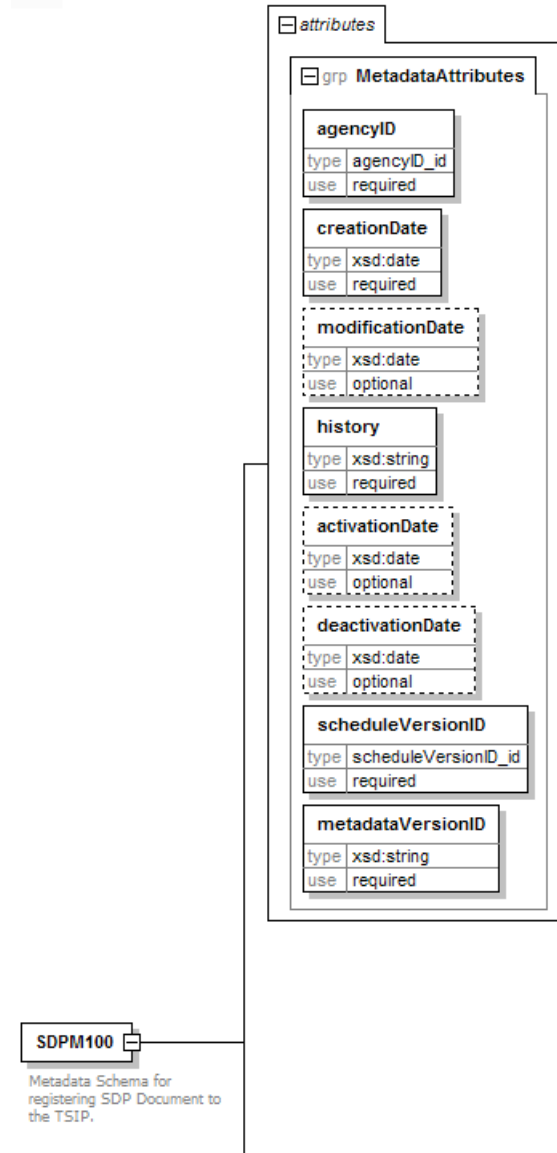
Figure 2: Metadata SDP XML Schema Model (from XMLSpy)
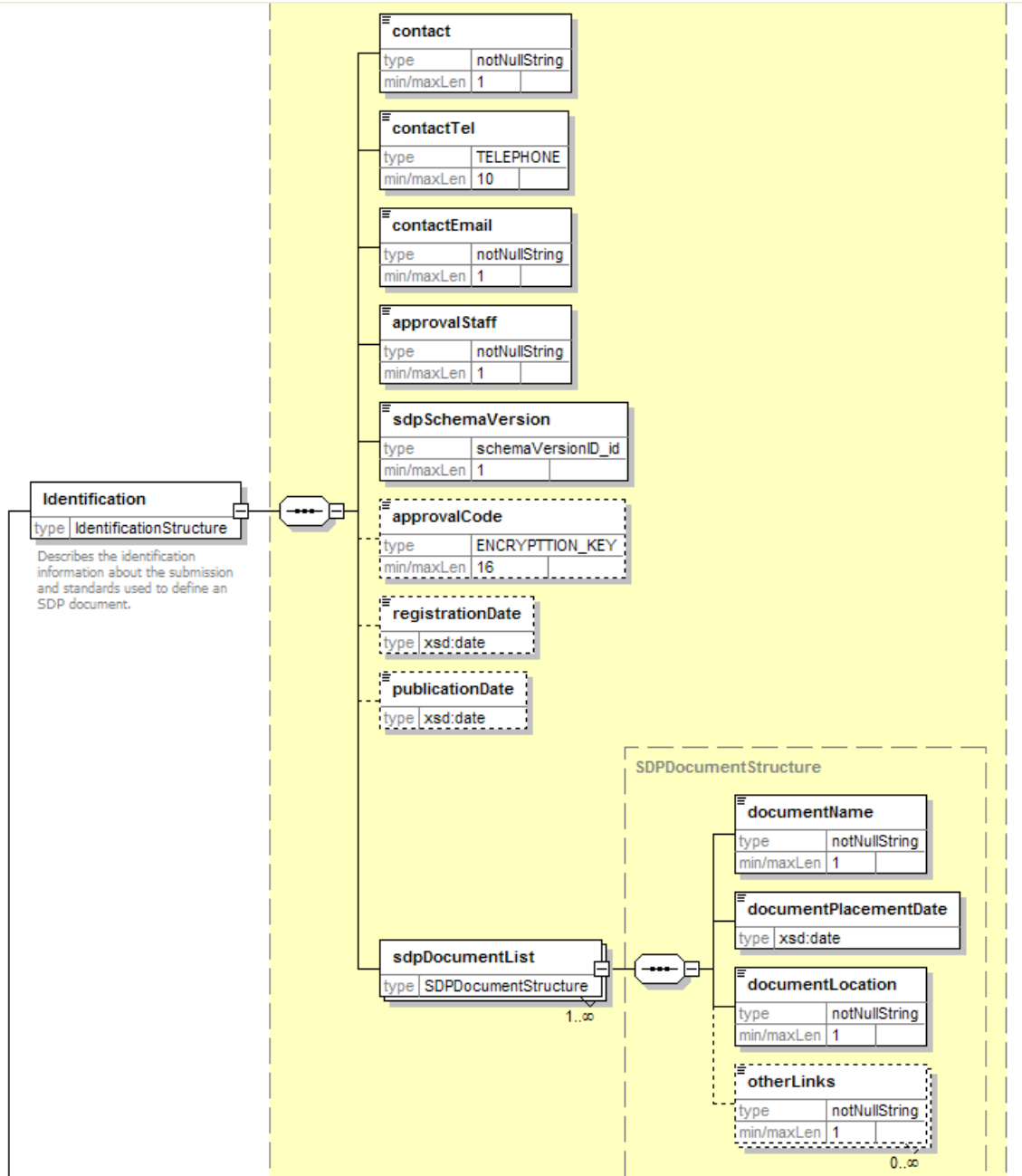
Figure 3: SDP Metadata Attribute Group

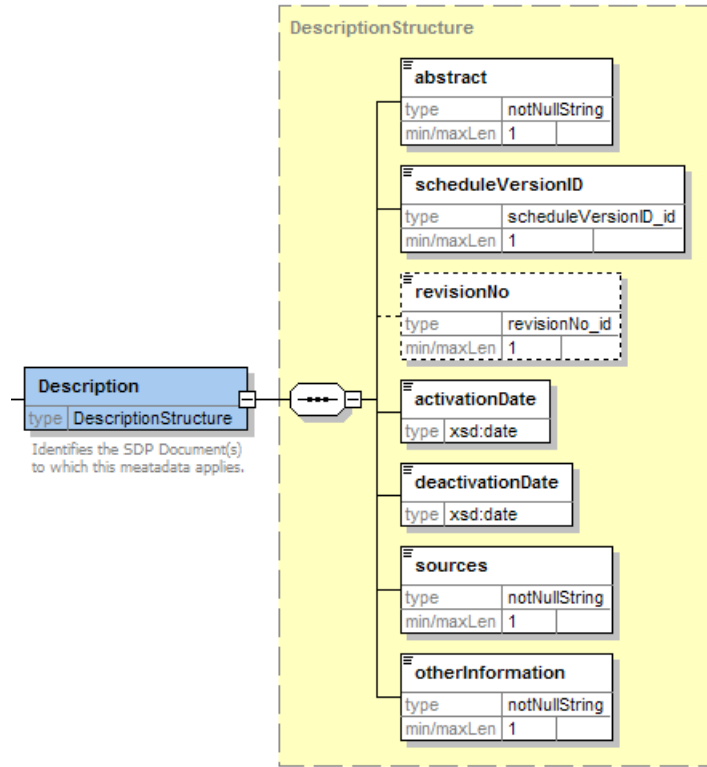**Figure 4: SDP Metadata XML Schema Fragment of Identification Element**

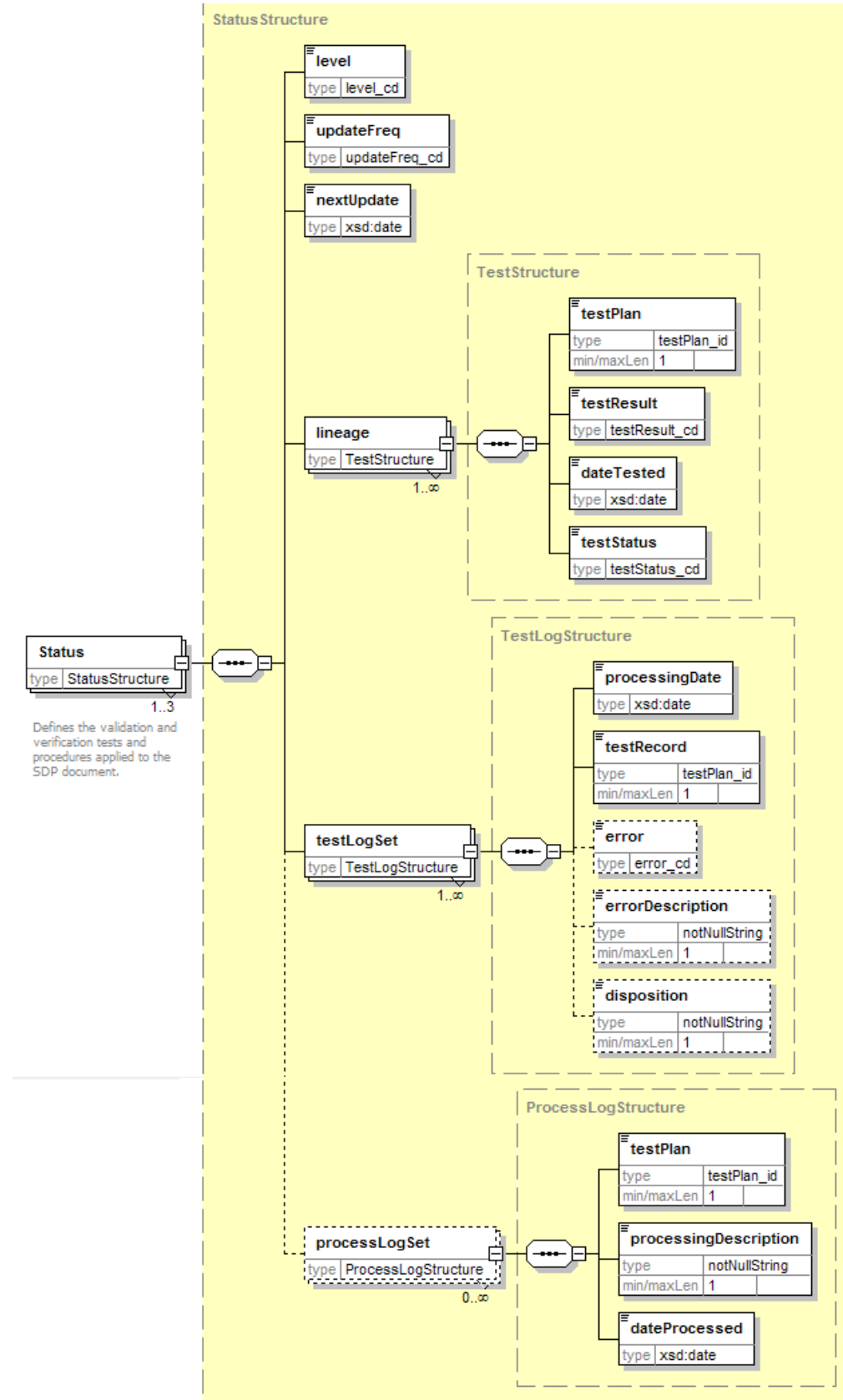**Figure 5: SDP Metadata XML Schema Fragment of Description Element**

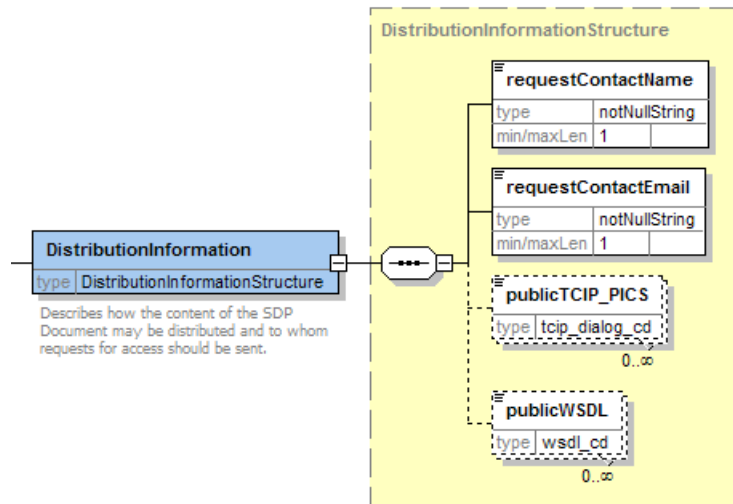**Figure 6: SDP Metadata XML Schema Fragment of Status Element**

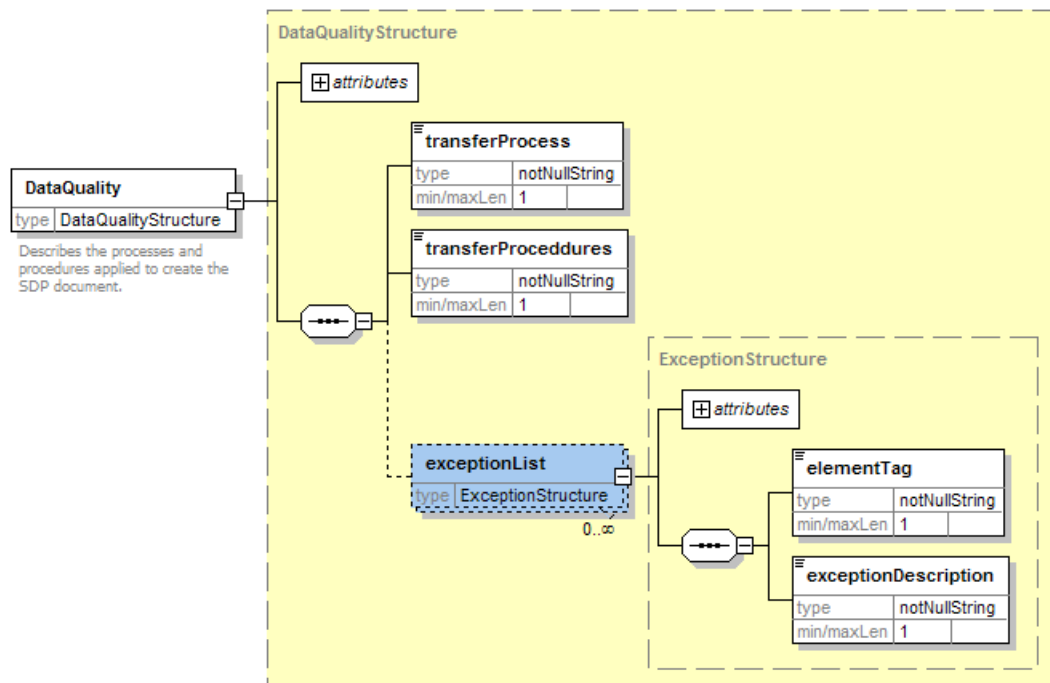**Figure 7: SDP Metadata XML Schema Fragment of Distribution Information Element**

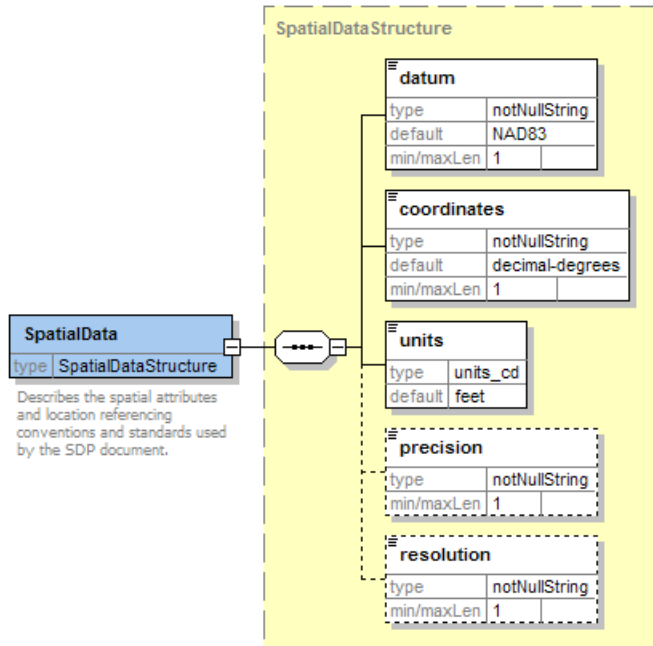**Figure 8: SDP Metadata XML Schema Fragment of Data Quality Element**

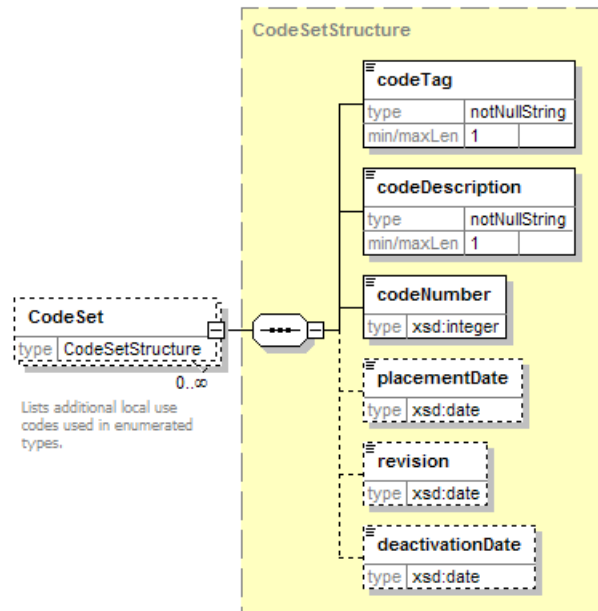**Figure 9: SDP Metadata XML Schema Fragment of Spatial Data Element**

Figure 10: SDP Metadata XML Schema Fragment of Code Set Element
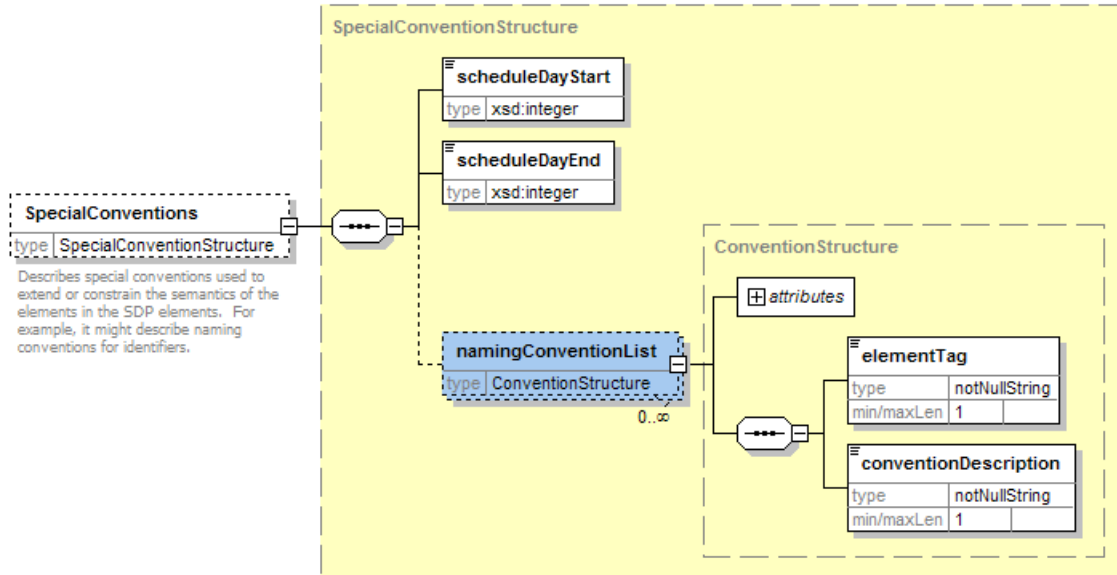
Figure 11: SDP Metadata XML Schema Fragment of Special Conventions Element

6.5 Appendix A: XMLSpy Schema Notation

The XML Schema notation, as extracted from the XMLSpy application, is used to describe the organization and format of the SDP XML Schema. The Schema is based on a hierarchical organization where parent nodes or elements may contain child elements (which may in turn be a parent element to child elements). The XML Schema format and document instance are based on the standard notation of an XML Schema and instance document.

Figure 12 below illustrates the different levels of the XML Schema and key notation, using Transit Facility as the example. In addition, the figure shows the type description for each element. A type reference may have one of the following prefixes or suffixes:

- Prefix of "xsd" asserts the type is native to the XML standard
- Suffix of "_id" implies the type is defined as an SDP identifier domain
- Suffix of "_cd" implies an enumerated code type.

A "Structure" in the type name implies that the element is a complex type. An element also includes the constraint on the number of times it is allowed. An element enclosed by a dotted lined box indicates that the element is optional. Elements that may be repeated will include a notation of the minimum and maximum (e.g., $0..\infty$) under the right hand corner of the element enclosure. "plantComponentList" is an example of an element that is optional, but may be repeated. One element is required when the element is enclosed with a solid line (and does not contain a min-max value).
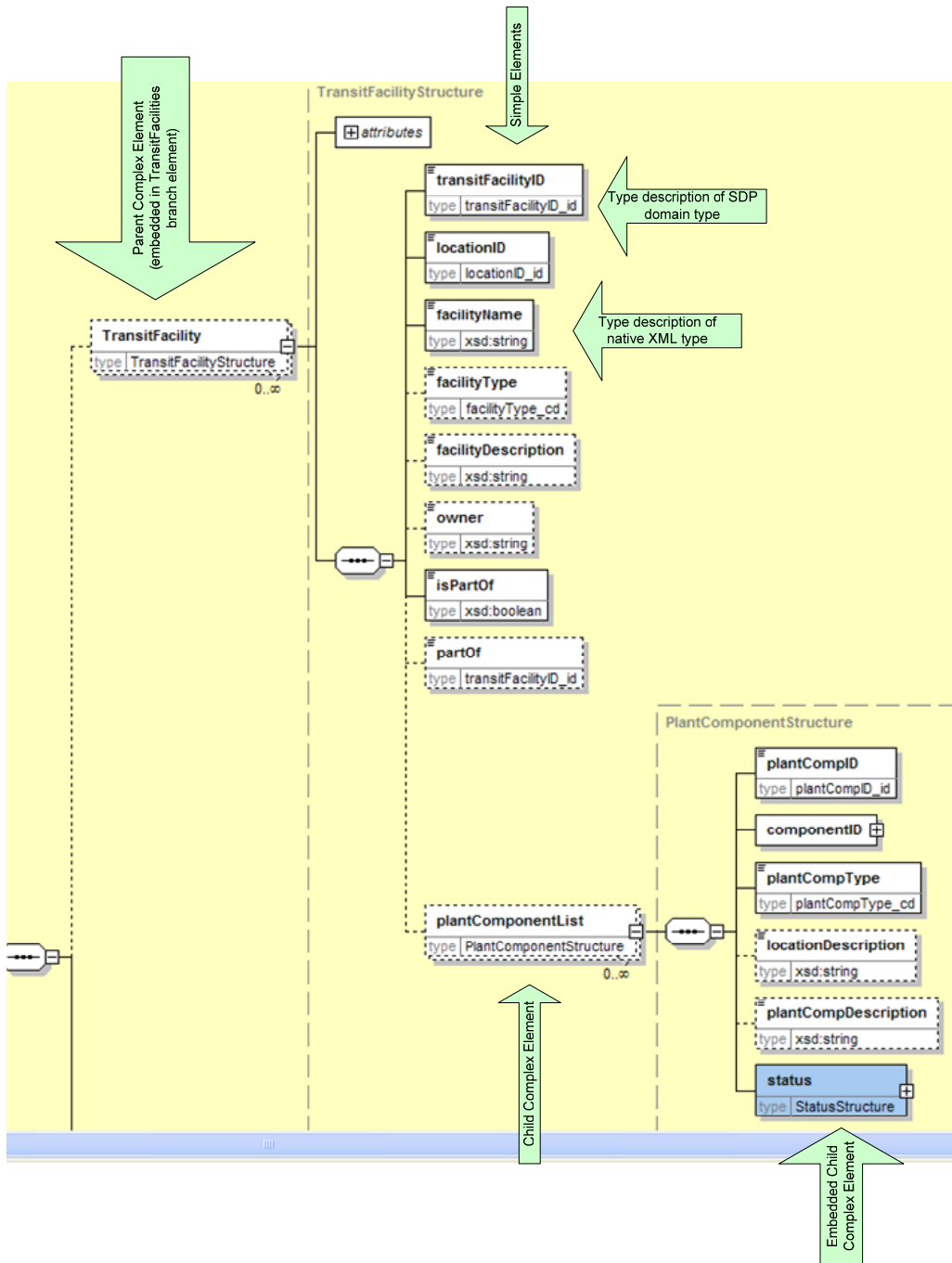
Figure 12: Example of the XMLSpy Diagram Notation