Table of Contents

Figures

Chapter 2: SDP XML Schema Model Overview

In This Chapter

- Learn about the organizational structure of the Schedule Data Profile.
- Learn the concepts included in each part of the SDP XML Schema Model.
- Identify the mandatory high level elements of the SDP XML Schema.
- Review SDP XML Schema notation.

Schedule Data Profile
The Schedule Data Profile (SDP) is a specification that describes operator generated schedule and related data. It is a business semantics specification that describes schedule information, specifically each data element and its relationship to scheduling data concepts, and preserves the referential integrity of these data concepts.

The SDP will be based on recognized information technology (IT) standards such as Extensible Markup Language (XML) and XML Schema, as well as standards and best practices in the IT and transit industries.

XML Schema Defined

An XML Schema is based on the e**X**tensible **M**arkup **L**anguage, XML, a general-purpose markup language.[1] There are many tutorials and descriptions of the language and schema standards that are helpful in understanding the SDP. A list of XML-related resources is included in Appendix C: Resources.

Purpose of the Schedule Data Profile (SDP) XML Schema

The SDP XML Schema is a structured description of the data concepts needed to describe key schedule-related data involved in transit service provision. The SDP XML Schema defines and constrains the format of an operating agency's schedule data and the relationships among the pieces of data. The industry-adopted XML and XML Schema standards, upon which the SDP was developed, provide a formal syntax that describes the "schedule" data requirements.

SDP XML Document

When schedule-related data is formatted into a file, following the rules of the SDP XML Schema, the file is called a SDP XML Document. This Guidance Document refers to a file of schedule data formatted using the SDP XML Schema as a "SDP Document" or "SDP XML Document."

Because formal XML standards have been adopted and implemented extensively, there are editing and validation tools readily available as off-the-shelf software that support native data extraction, loading, accessing, and storing (ELAS). In addition, many database management systems support these same ELAS functions.

Scope of the SDP XML Schema Model

The Schedule Data Profile consists of scheduling and related data that help describe a schedule. The scheduling elements are constrained to those elements needed for public information dissemination (e.g., timetable and trip planning functions), regional coordination and schedule planning activities. In particular, operator and track management functions are ***excluded*** from the current representation of regional schedule exchange. However, data concepts related to transit facilities and their location information are included in the scope of the SDP model.

[1] Link to resources with description of XML and XML Schema

Additionally, the model attempts to capture key metadata information on data quality, ownership, custodianship, lineage and currency. Metadata is discussed in greater detail in *Part 3, Programmer's Manual*.

Typology Used to Partition the SDP Conceptual Reference Data Model (CRDM)

Many different pieces of data and information are needed to accurately describe transit schedules and their relationship to service. In addition, some of the relationships between those pieces of data can be complex to describe. Partitioning a domain enables the organization of key ideas into smaller, less complex concepts. The SDP CRDM is partitioned to help users of the SDP better understand the data and data relationships. The partitioning of the SDP CRDM is based on the following categories, given that the data are organized by organization:

- Definition: General Agency Information
- Network: Transportation Network
- Place: Transit Feature Gazetteer and Transit Facilities
- Temporal: Service Provision

Places reference real-world locations, networks are made up of places, temporal services traverse the networks. Due to the complexity of transit "places," the "Place" category was divided into two—Facilities and Features (other than Facilities).

A fifth category (though not included in the SDP model) is the transportation network, a representation of the real-world network referenced by transit network and services. These categories are illustrated in Figure 2-1 below.

The partitioning helps distinguish different types of information the SDP must support. Building these separate blocks of information will eventually facilitate maintenance of the SDP model. Transit agencies update different types of information at different frequencies; additionally, various data sets require maintenance at different times. Partitioning the SDP Model into discrete logical categories identify the strong and weak relationships among the various scheduling elements. Elements within the same category are strongly related, and data in another category may be weakly related. Separating the elements also helps segment data that may be maintained at various times of the year. Stops may not be updated at the same cycle as service schedules. As such, building separate "branches" of information enables agencies to submit separate documents for different types of information.

Description of SDP Categories

The categories used to partition the SDP are further described below.

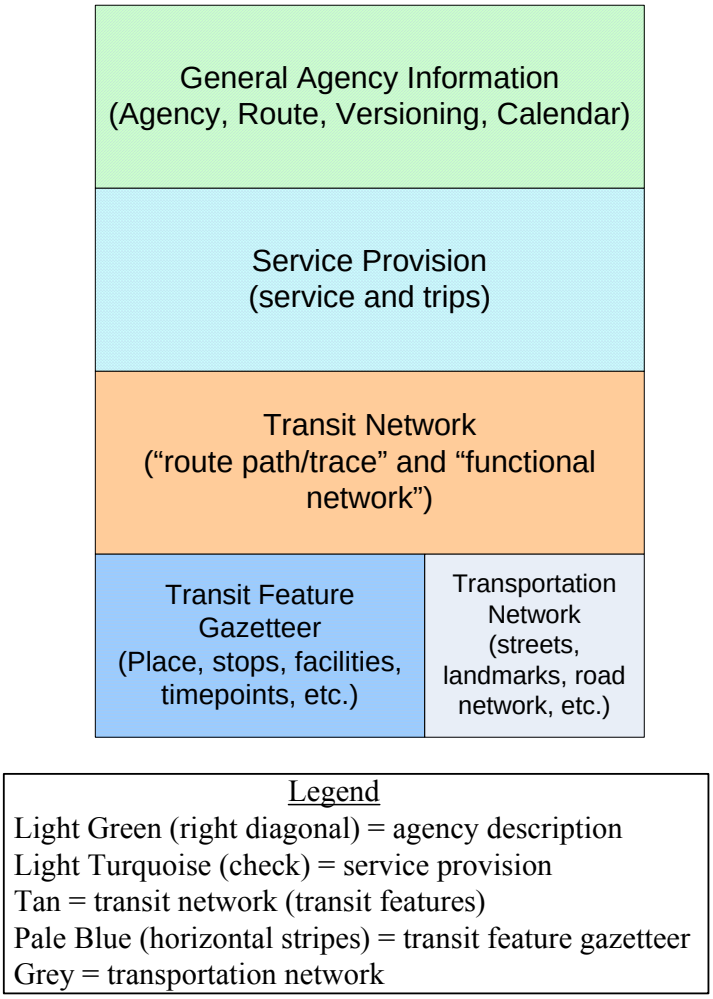


Figure 2-1: Partition of SDP Domain

General Agency Information: This category describes general information related to the agency that registers schedule-related data and the actual data that are registered. Included in this category is information on the agency, schedule version and revisions, and the organization that submitted the data or is referred to by the data set content. The SDP XML Schema stores this information in the AgencyRegistration branch.

Service Provision: This category provides information on the service provided by an agency or organization unit of an agency. The scheduling elements include information on trips and trip times, scheduling notes, and bus assignment schedules. The SDP XML Schema stores this information in the Service branch.

Transit Network: This category describes the route path traversed by transit service. The network is composed of transit paths called patterns. Events occur on each pattern, and patterns may be composed of segments, often called route segments or timepoint intervals (TPI). The SDP XML Schema stores this information in the TransitNetwork branch.

Transit Feature Gazetteer: This category defines places and their locations. A valuable addition is the incorporation of a "Location Table" which aggregates spatial references. The table simplifies data maintenance and enables the linking of places to equivalent locations (which may be described using a different location referencing system). For example, Penn Station Long Island Rail Road (LIRR) is the equivalent location as Penn Station Amtrak, New Jersey Transit, New York City Transit Lines A, C, E. Other transit features such as timepoints and transfer locations are also included in this category. The SDP XML Schema stores this information in the TransitGazetteer branch.

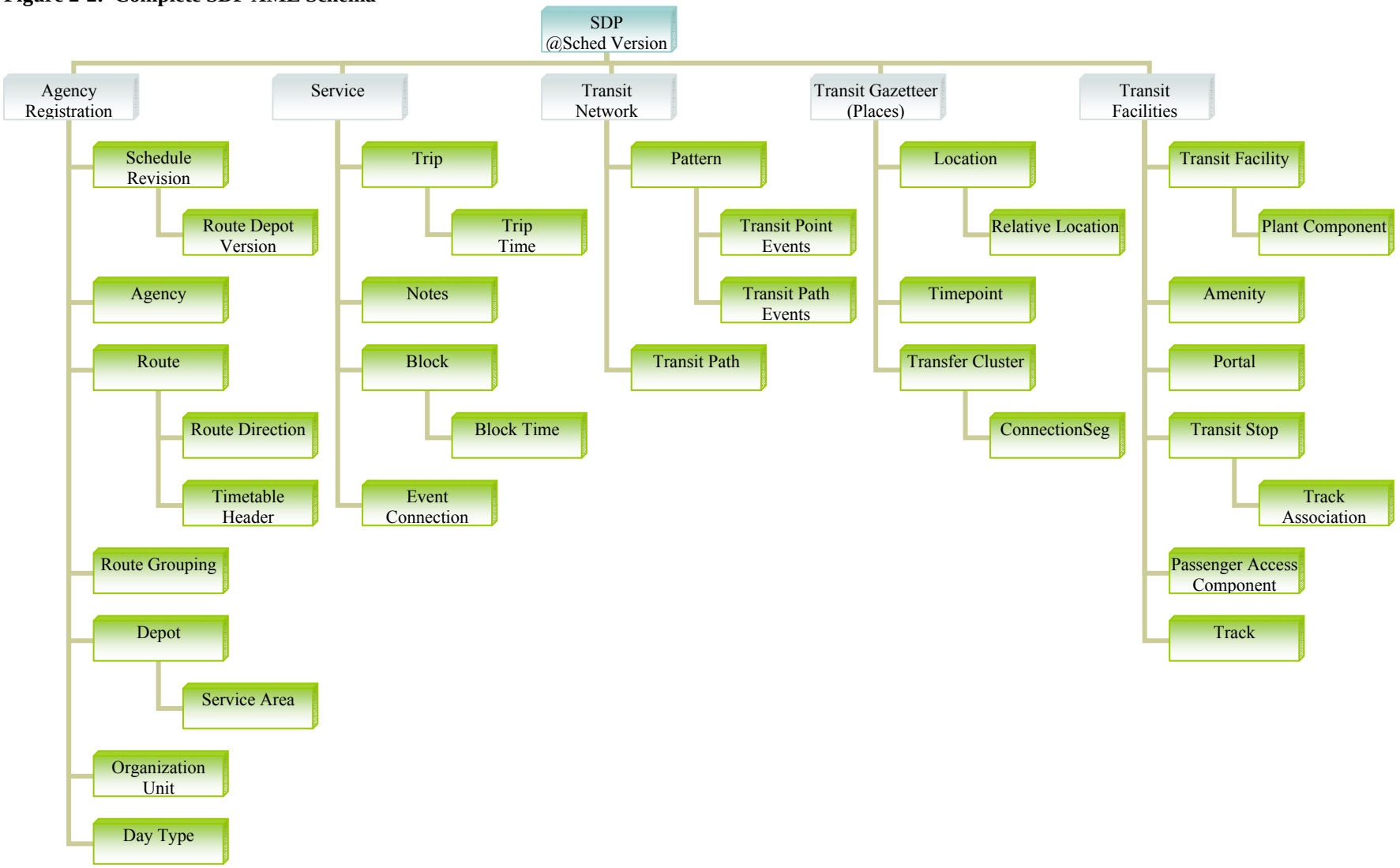
Transit Facilities: This category includes Transit Facilities and plant components that are included in or related to a transit facility. For example, parking lot, boarding area (platform), track, entrance, and elevator are plant components of a train station. The category includes the

asset inventories related to transit facilities, that is, transit stops, amenities, portals, passenger access elements and tracks (associated with a transit facility). Due to the complex nature of key New York City transit facilities like Port Authority Bus Terminal, Jamaica Station, Grand Central Terminal, and Pennsylvania Station in New York, the model enables a facility to be part of another facility. The SDP XML Schema stores this information in the TransitFacilities branch.

SDP Categories Drive SDP XML Schema Organization

The partitioning of schedule-related data concepts drives the organization of the SDP XML Schema. The five areas related to the model partitions drove the categories used to separate the SDP schema into branches, accounting for all the scheduling elements described in the SDP Functional Requirements document. Each element in this SDP XML Schema model will be described in detail in subsequent chapters. As the abstract partitioning of the schedule-related data was transformed into an XML schema, some of the category names changed slightly. For example, Agency Registration is used in the schema while General Agency Information is used in the model above.

Figure 2-2: Complete SDP XML Schema



Mandatory SDP XML Schema Elements

The SDP XML Schema is designed to meet many existing and future needs. As a result, the Schema contains both mandatory and optional SDP XML schema elements. The SDP must meet the schedule information requirements of several downstream applications such as Trip Planning, Timetable, and Ad Hoc Schedule Planning[2], as well as support the existing upstream requirements of the data providers. Many data requirements were also included in the Conceptual Reference Data Model to allow extending the SDP to future downstream applications or to support current Operator practices. To support the diversity of requirements, there are many optional elements in the SDP Schema.

The high level mandatory (complex type) elements that are required by every validated SDP document are depicted in Figure 2-3. When implementing the SDP, the business rules described in each chapter must be carefully reviewed, because some additional elements from a different level of the SDP XML Schema may be required in the SDP XML Document.

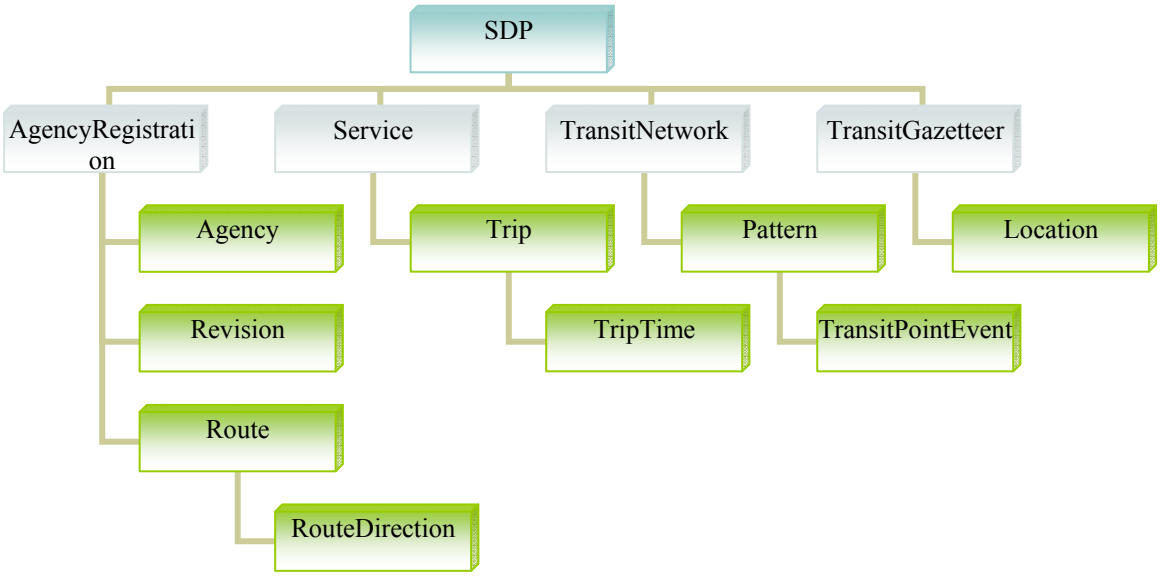


Figure 2-3: Mandatory SDP XML Schema Elements

XML 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 2-4 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

[2]The high level requirements for these applications may be found in the Appendices of the TSDEA Concept of Operations document.

- Suffix of "_id" implies the type is defined as an SDP identifier domain
- Suffix of "_cd" implies an enumerated code type.

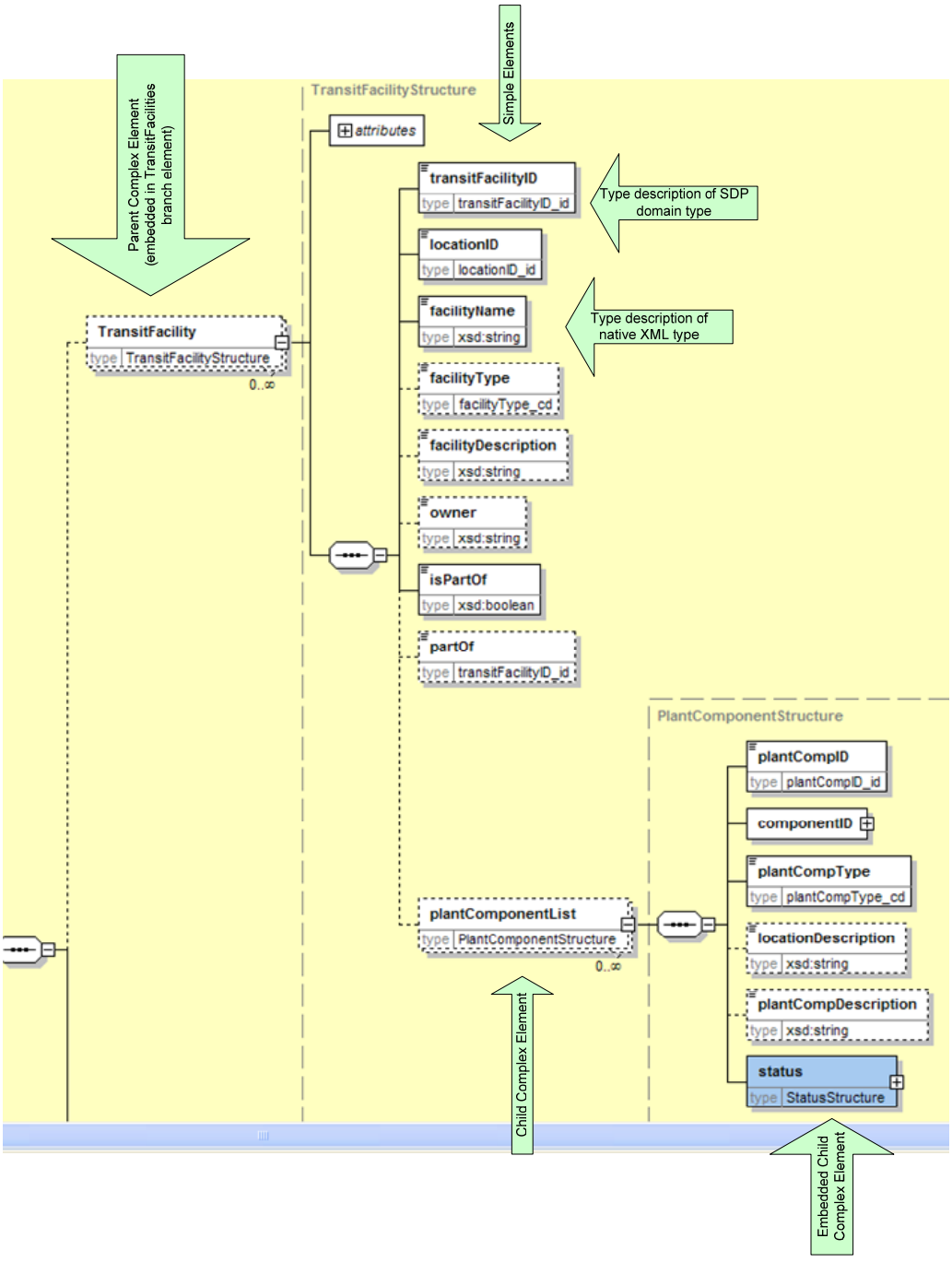


Figure 2-4: Example of the XMLSpy Diagram Notation

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..∞) 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).