The CIEE Web Database Document Type Definition

A. Joseph Koshy,
jkosh@freebsd.org

Version: 0.3
15-Jun-2002
Preface

The CIEE WEB DATABASE project aims to make information collected about government aided schools in the Indian state of Karnataka available on the world wide web.

Information collected by this project about each school is recorded in the system in a structured manner. This structure is formally specified in this Document Type Definition.

Typographical Conventions

This document follows a few simple typographical conventions:

- Important terms and words are highlighted using a **bold font**.
- Emphasis is given to certain points in the text by typesetting phrases *using italics*.
- Computer output is typeset in a *monospaced font*.

Copyright

This work is being made available under a “BSD-style” license. Please see chapter 14 for the text of the license and copyright notice.
CONTENTS

6 Student Information 43
  6.1 DTD . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 43
    6.1.1 Information collected per form . . . . . . . . . . . . . . . . 44
    6.1.2 Student Results . . . . . . . . . . . . . . . . . . . . . . . . 45

7 Staff Information 47
  7.1 DTD . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 48
    7.1.1 Staff General Information . . . . . . . . . . . . . . . . . . . 48
    7.1.2 Staff Attendance . . . . . . . . . . . . . . . . . . . . . . . . 50
    7.1.3 Teachers Opinion . . . . . . . . . . . . . . . . . . . . . . . . 50
    7.1.4 Incentive Schemes . . . . . . . . . . . . . . . . . . . . . . . 50
    7.1.5 Administrative Work . . . . . . . . . . . . . . . . . . . . . . 51

8 School Administration 53
  8.1 DTD . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 54
    8.1.1 Visitors . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 54
    8.1.2 Data on SBC/AY/BCS . . . . . . . . . . . . . . . . . . . . . . 55
    8.1.3 Samudayadatta Shala . . . . . . . . . . . . . . . . . . . . . . 57

9 Infrastructure 59
  9.1 Building Ownership . . . . . . . . . . . . . . . . . . . . . . . . . . 61
  9.2 Verandah / Corridor . . . . . . . . . . . . . . . . . . . . . . . . . . 62
  9.3 Rooms . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 62
    9.3.1 Room information . . . . . . . . . . . . . . . . . . . . . . . . 62
  9.4 Water . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 63
  9.5 Toilet . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 65
  9.6 Compound Wall . . . . . . . . . . . . . . . . . . . . . . . . . . . . 66
  9.7 Flooring . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 66
  9.8 Electricity . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 67
  9.9 Furniture . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 68
  9.10 Teaching Material . . . . . . . . . . . . . . . . . . . . . . . . . . . 69
  9.11 Playground . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 70
  9.12 Sports Equipment . . . . . . . . . . . . . . . . . . . . . . . . . . . 70
  9.13 Library . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 72
  9.14 Medical Facilities . . . . . . . . . . . . . . . . . . . . . . . . . . . 73
  9.15 Location of classes . . . . . . . . . . . . . . . . . . . . . . . . . . 74
  9.16 Maintenance . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 75
  9.17 Environment . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 76
  9.18 Development . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 77
  9.19 Previous Complaints . . . . . . . . . . . . . . . . . . . . . . . . . 79

10 Other Information 81
  10.1 DTD . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 82
    10.1.1 Local Group Activities . . . . . . . . . . . . . . . . . . . . . 82
    10.1.2 Government Programmes . . . . . . . . . . . . . . . . . . . . 83
    10.1.3 Extra Curricular Activities . . . . . . . . . . . . . . . . . . . 84
    10.1.4 Education Support Activities . . . . . . . . . . . . . . . . . . 85

11 Neighbours 87
  11.1 DTD . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 87
CONTENTS

12 Improvement Opinions 89
  12.1 DTD ............................ 90
    12.1.1 Improvement Opinion Table 90
    12.1.2 Past Initiatives ................. 91
    12.1.3 Childrens Input ................. 92

13 School Pictures 93
  13.1 DTD ............................ 93
    13.1.1 Known Picture formats .......... 94
    13.1.2 Photographer ................. 95
  13.2 Sample .......................... 95

14 Copyright and License 97

II Appendices 99
A Glossary 101
B An XML Primer 103
  B.1 What is XML? ...................... 103
  B.2 What is a DTD? ................. 103
C Rationale 105
  C.1 Data description using SGML .... 106
  C.2 Design Decisions ................. 107
    C.2.1 Design: One File Per School .... 107
    C.2.2 Design: Unicode encoding .... 107
    C.2.3 Design: Formal file structure expressed in XML 107
    C.2.4 Design: Data Entry Tool .... 108
  C.3 Alternative approaches ............ 108
    C.3.1 Alternate: Use a database .... 108
    C.3.2 Alternate: Use plain text files 108

Index 111
  C.4 Chunk Index ...................... 111
  C.5 Identifier Index ................. 113

Colophon 117
List of Tables

2.1 Attributes of element phone ........................................ 12
2.2 Helpers .......................................................... 15
2.3 Attributes of element result ....................................... 17
2.4 Distances .................................................................. 18
2.5 Dimension units ....................................................... 18
2.6 Quality judgements .................................................. 19
2.7 Infrastructure constructed by ...................................... 19
2.8 Medium of instruction ............................................... 20
2.9 Text Field Attributes ............................................... 21
2.10 Premises ownership ................................................ 22
2.11 Roles .................................................................. 22
2.12 Attributes of element form ........................................ 26
3.1 Attributes of element samudayadatta-shale ................. 57
3.2 Attributes of element room-info .................................. 63
3.3 Attributes of element water-source .............................. 64
3.4 Water sources .......................................................... 64
3.5 Kinds of water storage .............................................. 65
3.6 Attributes of element toilet ........................................ 66
3.7 Attributes of element compound-wall ......................... 66
3.8 Kinds of compound walls .......................................... 66
3.9 Attributes of element teaching-material ...................... 67
3.10 Kinds of flooring ..................................................... 67
3.11 Attributes of element has-electricity .......................... 68
3.12 Attributes of element furniture ................................. 68
3.13 Kinds of furniture .................................................. 68
3.14 Attributes of element sports-equipment ..................... 69
3.15 Kinds of teaching materials ....................................... 70
3.16 Attributes of element sports-equipment ..................... 71
3.17 Kinds of sports equipment ........................................ 71
<table>
<thead>
<tr>
<th>Table Number</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.18</td>
<td>Attributes of element library</td>
<td>72</td>
</tr>
<tr>
<td>9.19</td>
<td>Attributes of element library-item</td>
<td>72</td>
</tr>
<tr>
<td>9.20</td>
<td>Kinds of library items</td>
<td>73</td>
</tr>
<tr>
<td>9.21</td>
<td>Attributes of element medical</td>
<td>73</td>
</tr>
<tr>
<td>9.22</td>
<td>Attributes of element health-checkup</td>
<td>73</td>
</tr>
<tr>
<td>9.23</td>
<td>List of health checking organizations</td>
<td>74</td>
</tr>
<tr>
<td>9.24</td>
<td>Attributes of element class-location</td>
<td>74</td>
</tr>
<tr>
<td>9.25</td>
<td>Locations of classes</td>
<td>75</td>
</tr>
<tr>
<td>9.26</td>
<td>Teaching styles</td>
<td>75</td>
</tr>
<tr>
<td>9.27</td>
<td>Attributes of element maintenance</td>
<td>76</td>
</tr>
<tr>
<td>9.28</td>
<td>Attributes of element hazard</td>
<td>76</td>
</tr>
<tr>
<td>9.29</td>
<td>Kinds of hazards</td>
<td>77</td>
</tr>
<tr>
<td>9.30</td>
<td>Causes of hazards</td>
<td>77</td>
</tr>
<tr>
<td>9.31</td>
<td>Attributes of element development-item</td>
<td>78</td>
</tr>
<tr>
<td>9.32</td>
<td>Kinds of development items</td>
<td>78</td>
</tr>
<tr>
<td>9.33</td>
<td>Attributes of element additional-classrooms</td>
<td>79</td>
</tr>
<tr>
<td>9.34</td>
<td>Attributes of element complaint</td>
<td>80</td>
</tr>
<tr>
<td>10.1</td>
<td>Government Schemes for Schools</td>
<td>84</td>
</tr>
<tr>
<td>10.2</td>
<td>Organizers for Extra Curricular activities</td>
<td>85</td>
</tr>
<tr>
<td>13.1</td>
<td>Attributes of element school-picture</td>
<td>94</td>
</tr>
<tr>
<td>13.2</td>
<td>Views of the school</td>
<td>94</td>
</tr>
<tr>
<td>13.3</td>
<td>Recognized image formats</td>
<td>95</td>
</tr>
</tbody>
</table>
List of Examples

1.1 Naming the DTD using SYSTEM . . . . . . . . . . . . . . . . . . . . 3
1.2 DOCTYPE declaration using the PUBLIC keyword . . . . . . . . . . . 3
1.3 Generating typeset documentation . . . . . . . . . . . . . . . . . . . 4
1.4 Extracting the DTD and sample data . . . . . . . . . . . . . . . . . . 4
1.5 Validating sample data . . . . . . . . . . . . . . . . . . . . . . . . . 5
Chapter 1

Introduction

The aim of the CIEE WEB DATABASE project is to develop the software infrastructure allowing information about government funded schools in the Indian state of Karnataka collected by the CIEE project to be published on the World Wide Web.

The project stores surveyed information about schools in a structured form. This document formally specifies the structure of data collected.

We develop a model of the structure of the stored school data in the form of an XML Document Type Definition (DTD). Using literate programming techniques (see section and section), the text of this document also serves as the source for the actual DTD. Parallely, we also clarify the expected file structure by examining a sample XML file containing school data.

1.1 Intended Audience

This document is intended to be read by:

- people seeking to understand how school data is being stored
- programmers implementing or maintaining the CIEE WEB DATABASE project
- people wishing to learn from, or to reuse this work in other contexts

1.1.1 Pre-requisite knowledge

To make the most of this document, you should be familiar with:

- basic XML concepts, like well-formedness and validation,
- XML document syntax,
- XML DTD specifications

Appendix has a short introduction to the necessary XML concepts. It also has pointers to further resources for XML.

Experience on using a UNIX-like operating system will help you feel comfortable with some of the operational examples presented.

1 The status of the CIEE WEB DATABASE project can be found at http://people.freebsd.org/~jkoshy/projects/ciee-db/.
2 The CIEE project home pages are present at http://www.ciee.net/
1.2 Structure of this document

This document is structured as follows:

- Subsequent sections of this chapter expand on the usage of the DTD (section 1.5) during the processing of a school data file, and on the way the DTD should be enhanced (section 1.6).
- Chapter 2 on page 9 gives an overview of the DTD. The DTD is described side-by-side with sample data of a fictional school.
- Chapters 3–14 delve into the details of the DTD. These chapters are intended to be used as reference material.
- In the appendices, we have a glossary of terms (appendix A), a short tutorial (appendix refxml-tutorial) on XML, some background on the design (appendix C) and an index (appendix F).

1.3 Reading this document

If you are unfamiliar with XML and structural markup, you can start by taking a look at Appendix B. There are pointers there to a number of XML related resources available on the Internet.

Appendix C on page 105 gives some background information about the CIEE WEB DATABASE project. It briefly covers the design decisions that led to the formulation of this DTD. Readers new to the project may wish to read this material first.

Section 1.4 describes the terminology and special use of language in this document.

Section 1.5 describes the steps in validating a file containing school information against this DTD.

This document is both the formal DTD specification and the explanation of the formal specification. Literate programming tools have been used to fuse the two. Section 1.6 describes the process of extracting the DTD from this document. It also describes the steps to be taken care of when modifying this document.

Each major section of the printed CIEE form is described in a chapter with the same name in this document. You can also refer to the relevant sections directly; the appropriate links to related information are present in the text. Using the online version of this book is particularly convenient, since accessing linked-in information is very convenient.

At the end of the book, an index of concepts and identifiers.

1.4 Terminology

A discussion on the terminology (TODO).

XXX should this be merged with or converted into a glossary?

1.5 Using the DTD

Given a text file that contains school survey data for one school, a validating XML parser can check for the well-formedness of the data by checking that the document
conforms to the structure specified in the DTD. Further, some validating XML parsers can use the information supplied in the DTD to supply default values of attributes that were not explicitly specified in the school survey document.

For this to happen, a couple of points need to be taken care of.

- The school survey data document needs to name the CIEE DB DTD in a DOC-TYPE line, as in Example 1.1 (see also section 2.1):

```
<!DOCTYPE school-information SYSTEM "cieedbd.dtd">
```

The formal public identifier for this DTD is the string: "-//Joseph Koshy//DTD CIEE School Information Database v0.3//EN". Thus a school data file can also name the DTD using a DOCTYPE line as in Example 1.2.

```
<!DOCTYPE school-information PUBLIC "-//Joseph Koshy//DTD CIEE School Information Database v0.3//EN" "cieedbd.dtd" >
```

Using a PUBLIC declaration is more flexible as it allows one to avoid hardcoding the system location of the DTD specification in the school data files. However, not all XML parsers today support the PUBLIC keyword.

- The DTD file needs to be made accessible to the XML parser. The CIEE DTD is kept in a file called cieedbd.dtd. The XML parser needs to know how to locate this file.

How to do this depends on the XML parser involved and is out of the scope of this document.

## 1.6 Modifying the DTD

This document and the formal DTD it describes are both derived from a single source, namely, cieedbd-dtd.nw. This file contains the text for the description that you are reading now, interspersed with NOWEB directives embedding the text of the formal DTD. The text of description uses \LaTeX\ directives for typesetting and document structuring, so the contents of cieedbd-dtd.nw thus comprise of a mixture of \LaTeX\ commands and NOWEB processed sections.

From this file, the NOWEB tools generate:

- an annotated \LaTeX\ file cieedbd-dtd.tex that contains the description along with the DTD syntax appropriately typeset
- the plain text of the DTD for use with XML tools.
- the plain text of the sample school data
CHAPTER 1. INTRODUCTION

1.6.1 Generating typeset documentation

The steps involved in converting the NOWEB source to printable (and viewable) documentation are as follows (see Example 1.3).

Example 1.3 Generating typeset documentation

```
$ noweave -index -delay cieedb-dtd.nw > cieedb-dtd.tex
$ pdflatex cieedb-dtd.tex
```

1. The NOWEB tool noweave is used to convert the cieedb-dtd.nw file to annotated \texttt{\LaTeX} source cieedb-dtd.tex.

2. The \texttt{pdflatex} program then converts the \texttt{\LaTeX} source to PDF (the Adobe Portable Document Format) form.

1.6.2 Extracting the DTD and Sample data

The DTD and sample school data present in the cieedb-dtd.nw source can be extracted using \texttt{notangle} (see Example 1.4).

Example 1.4 Extracting the DTD and sample data

```
$ notangle -Rcieedb.dtd cieedb-dtd.nw > cieedb.dtd
$ notangle -Rsample.xml cieedb-dtd.nw > sample.xml
```

The \texttt{notangle} command requires a root chunk to be specified using its \texttt{-R} option. The file cieedb-dtd.nw contains two root chunks:

- cieedb.dtd, being the root chunk of the XML DTD, and,
- sample.xml, being the root chunk of the sample data

Given the name of a root chunk, the \texttt{notangle} tool will extract all the relevant NOWEB sections in the correct inclusion order.

1.6.3 Verification

Sample school data is provided along with the DTD and its explanation, for the following reasons:

- The sample data provides a concrete example of how the DTD would translate to real-world use
- The sample data and the XML DTD are meant to be kept synchronized with each other and to serve as a sanity check when developing or maintaining the DTD

The sample data can be validated against the DTD using any validating XML parser. Example 1.5 shows the use of Richard Tobin’s \texttt{RXP} parser for XML validating the extracted sample.xml against the DTD.
1.6. MODIFYING THE DTD

Example 1.5 Validating sample data

$ rxp -asV sample.xml

If validation is successful, the command will complete silently without flagging any errors.

3 These examples were generated under FreeBSD. The process would be similar for most POSIX compatible operating systems.
4 The home page for the RXP parser is at http://www.cogsci.ed.ac.uk/~richard/rxp.html
Part I

The DTD
Chapter 2

Overview

We present a brief overview of the school data file format by examining some sample data. The structure of the file is briefly examined, and the formal DTD specification that models this structure is presented.

A sample data file is examined in the next section. The layout of the DTD specification itself is described in section 2.3.

2.1 Data file organization

Let us start by looking at a sample school data file named by the chunk \(\langle \text{sample.xml} \rangle\).

\[
\langle \text{sample.doctype-declaration} \rangle \equiv \langle \text{sample:school-information} \rangle
\]

All school data files start with the standard XML declaration, namely,

\[
?xml \text{ version='1.0' encoding='UTF-8' standalone='no'}?\rangle
\]

This various parts of this declaration inform an XML aware processor that

- **version** the version of the XML standard being followed is “1.0”,
- **standalone** that this XML document will require to be validated against and external document type definition
- **encoding** that the encoding used in this document is “UTF-8”. UTF-8 encoding allows Unicode text to be represented in a way that is backwards compatible with regular ASCII.

The next line would be a document type declaration informing the XML aware tool about the DTD defining the document structure.

\[
\langle \text{sample:doctype-declaration} \rangle \equiv
\langle!DOCTYPE \text{ school-information PUBLIC "-//Joseph Koshy//DTD CIEE School Information Database v0.3//EN" \"cieedb.dtd\rangle
\]
CHAPTER 2. OVERVIEW

This declaration identifies the DTD as being present in system file name cieedb.dtd and alternately named by the formal public identifier -//Joseph Koshy//DTD CIEE School Information Database v0.3//EN.

The top-level structural element which contains all the information about a school is `<school-information>`.

```
<sample:school-information>
  <sample:form-info>
  </sample:form-info>
  <sample:community-information>
  </sample:community-information>
  <sample:school-general-info>
  </sample:school-general-info>
  <sample:student-info>
  </sample:student-info>
  <sample:staff-info>
  </sample:staff-info>
  <sample:admin-info>
  </sample:admin-info>
  <sample:infrastructure>
  </sample:infrastructure>
  <sample:other-info>
  </sample:other-info>
  <sample:neighbours>
  </sample:neighbours>
  <sample:improvement-opinions>
  </sample:improvement-opinions>
  <sample:pictures>
  </sample:pictures>
</sample:school-information>
```

These top-level sections in the school data file correspond to the major sections of the paper form being used to collect data from the school.

- `<sample:form-info>` holds records of updates to the school data file
- `<sample:community-information>` contains information about the community that the school is situated in
- `<sample:school-general-info>` contains information about the school that is relatively static, like addresses and school numbers
- `<sample:student-info>` records information about the children of the school
- `<sample:staff-info>` records information about the staff of the school
- `<sample:admin-info>` has information on how the school is being administered.
- `<sample:infrastructure>` records the infrastructure present in this school.
- `<sample:other-info>` tracks extracurricular activities and section
- `<sample:neighbours>` has information on neighbouring educational institutions of repute.
- `<sample:improvement-opinions>` is a record of improvement suggestions collected from various interested parties
- `<sample:pictures>` is the last section of the data file with some photographs of the premises.

Each of these major sections of the school data file are further examined in subsequent chapters.
2.2. BASIC DTD ELEMENTS

2.1.1 Top level DTD structure

In the DTD, we model this structure as a top-level school-information element with eleven mandatory child elements.

```xml
<!ELEMENT school-information (form-list, community-information, school-general-information, student-information, staff-information, school-administration, infrastructure, other-information, neighbour-information, school-improvement-opinions, school-pictures)>
<!ATTLIST school-information
id ID #IMPLIED>
```

All eleven sub-elements have to be present, in the order specified. Some of these sections may be empty, depending on how much is known about the school, but the section’s top-level element itself is required.

2.2 Basic DTD Elements

In this section we will examine some of the basic elements that are the building blocks for the school information DTD.

2.2.1 Standard DTD attributes

We would like all the DTD elements to possess some “standard” attributes. Currently, we only define one: an id attribute, which is marked as optional.

```xml
<!ATTLIST school-information
id ID #IMPLIED>
```

An id attribute that is marked optional is useful since it allows us to name individual elements of the school data file and refer to these from other places. For an example of such usage of the id attribute, the results element (section 2.2.12) can name an issue element (section 2.2.10) element by its ID.

2.2.2 Academic years

Some elements are used to hold data that changes every year. We use an academic-year attribute on such element to record the year associated with the data recorded by the element.

This attribute is mandatory for elements that are defined to use it. Finally, we restrict the list of years to a limited set via an XML enumeration, to allow for stricter validation.

```xml
<!ATTLIST school-information
```
2.2.3 Representing Telephone numbers

While telephone numbers can be represented as strings of numerals, it is useful to keep comments alongside with them. Phones are sometimes person-to-person phones, or may be physically located in a central place like the village headman’s house. Some phones may be manned only during certain times of the day.

We represent a telephone number with a `phone` structural element. This element groups the actual number (the `phone-number` attribute) with an optional comment (Table 2.1). The comment element is useful in documenting restrictions or comments about the use of the phone: some phones may be person-to-person phones, others may have definite times when they are attended to.

```
<phone (comment?)>
<ATTLIST phone (

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>phone-number</td>
<td>text of the phone number</td>
</tr>
</tbody>
</table>
```

Table 2.1 Attributes of element `phone`

The structure of data recorded in the `phone-number` attribute is not laid down in this DTD. Checks for validity of the phone number need to be done in the processing application.

A `phone-list` is a list of `phone` elements. This element would be used where we would expect more than one phone number.

```
<phone-list (phone*)>
<ATTLIST phone-list (dtd: standard attributes) >
```

2.2.4 Addresses

A postal address is modelled using an `address` element. This element has many optional text fields that hold the parts of the actual postal address, along with an optional list of phone numbers associated with the address.

```
<address (additional-text?, village-or-place?, gram-panchayat | ward ?, taluk?, district?, pin?, phone-list?)>
<ATTLIST address (dtd: standard attributes) >
```
2.2. BASIC DTD ELEMENTS

The structure of the address element is as follows:

- The village-or-place, taluk, and district elements are common elements of rural Indian postal addresses, corresponding to the administrative units.
  - village-or-place holds the the village name or equivalent.
  - taluk holds the name of the Taluk (sub-division of a district).
  - district holds the district name.
- Addresses can have either a gram-panchayat name or a ward number. This is reflected in the DTD declaration above, which allows only one or the other of gram-panchayat and ward. Generally speaking urban areas are divided into wards while rural areas are grouped into gram-panchayats.
- The pin element holds the pin-code for the village.
- The phone-list element holds telephone numbers associated with this address.
- The additional-text element is a catch-all element that is used for all other parts of the address that do not fit into the broad scheme described above.

Each of village-or-place, gram-panchayat, ward, taluk, district, and pin hold plain text. Formally they are specified as:

```xml
<!ELEMENT village-or-place (#PCDATA)>
<!ATTLIST village-or-place (dtd: standard attributes) (dtd: standard text field attributes)>

<!ELEMENT gram-panchayat (#PCDATA)>
<!ATTLIST gram-panchayat (dtd: standard attributes) (dtd: standard text field attributes)>

<!ELEMENT ward (#PCDATA)>
<!ATTLIST ward (dtd: standard attributes) (dtd: standard text field attributes)>

<!ELEMENT taluk (#PCDATA)>
<!ATTLIST taluk (dtd: standard attributes) (dtd: standard text field attributes)>

<!ELEMENT district (#PCDATA)>
<!ATTLIST district (dtd: standard attributes) (dtd: standard text field attributes)>

<!ELEMENT pin (#PCDATA)>
<!ATTLIST pin (dtd: standard attributes) (dtd: standard text field attributes)>```
Sample Data

A sample address that corresponds to the declaration above could be:

```xml
<address>
  <additional-text>32 Murugesh Palya</additional-text>
  <village-or-place>Inaimathihalli</village-or-place>
  <gram-panchayat>I. G. Gram Panchayat</gram-panchayat>
  <taluk>Nelamangala Taluk</taluk>
  <district>Bangalore</district>
  <pin>5000001</pin>
  <phone-list>
    <phone phone-number="0801112222">;
      <comment>phone belongs to village headman</comment>
    </phone>
    <phone phone-number="0801113333" />
  </phone-list>
</address>
```

### 2.2.5 Helpers

In a few places information is collected on people who assist in the daily tasks of running a school (for example, maintaining water supplies and cleaning the premises). We list the possible helpers that we would like to distinguish in our school data in Table 2.2.

```
| head-master | teachers | students | paid-help | community | nobody | other |
```
### 2.2. BASIC DTD ELEMENTS

#### 2.2.6 Names

Names are required at many places in the form. These are modelled by a `name` element.

\[
\langle \text{dtd: name declaration} \rangle \equiv \langle \text{!ELEMENT name (#PCDATA)} \rangle \\
\langle \text{!ATTLIST name} \rangle \\
\langle \text{!dtd: standard attributes} \rangle \\
\langle \text{!dtd: standard text field attributes} \rangle 
\]

A `name` element contains plain text. It has the standard attributes associated with textual elements.

#### 2.2.7 Yes/No Answers

In some places of the school data form, we need to record “yes”, “no”, and “unspecified” answers. This is modelled by using an optional attribute on an element accepting an enumeration that is a choice of two symbols, `yes` and `no`. The absence of the optional attribute means “unknown” or “unspecified”.

\[
\langle \text{dtd: yes/no attribute} \rangle \equiv (\text{yes} | \text{no}) 
\]

#### 2.2.8 Comments

In many places we may want to annotate some of the data being collected about the school with a comment or additional information. A `comment` structuring element is allowed in such places.

\[
\langle \text{dtd: comment declaration} \rangle \equiv \\
\langle \text{!ELEMENT comment (#PCDATA)} \rangle \\
\langle \text{!ATTLIST comment} \rangle \\
\langle \text{!dtd: standard attributes} \rangle \\
\langle \text{!dtd: standard text field attributes} \rangle 
\]

A comment can hold any text, in any language supported by the Unicode standard and XML. It has the standard attributes supported by text elements.

#### 2.2.9 Frequencies

Frequency of occurrences of events are recorded as a set of discrete choices:

\[
\langle \text{dtd: frequency} \rangle \equiv \\
\text{daily} | \text{weekly} | \text{fortnightly} | \text{monthly} | \text{once-in-3-months} | \\
\text{once-in-6-months} | \text{yearly} | \text{once-in-2-years} | \text{other} 
\]
The expected meaning of these symbols should be inferable from their names.

### 2.2.10 Issues

An issue element is used to model a real-life problem or issue that is to be recorded and tracked in the school data.

\[
\langle \text{dtd: issue element} \rangle \equiv (24) \quad \langle \text{dtd: standard attributes} \rangle
\]

An issue comprises of elements that record:

- the description of the issue (description)
- any decision taken on the matter (decision)
- any final result (result)

A description is an element wrapping plain text describing the issue at hand. It has the usual text field attributes.

\[
\langle \text{dtd: issue element} \rangle \equiv (24) \quad \langle \text{dtd: standard attributes} \rangle \langle \text{dtd: standard text field attributes} \rangle
\]

The decision element wraps plain text describing any decision reached by the community on an issue. It has the usual attributes associated with text elements.

\[
\langle \text{dtd: reason element} \rangle \equiv (24) \quad \langle \text{dtd: standard attributes} \rangle \langle \text{dtd: standard text field attributes} \rangle
\]

### 2.2.11 Reasons

In places in the school survey form, “reasons” for observed issues are recorded. The reasons are placed into a reason element, which holds the text describing the “reason”.

\[
\langle \text{dtd: reason element} \rangle \equiv (24) \quad \langle \text{dtd: standard attributes} \rangle \langle \text{dtd: standard text field attributes} \rangle
\]
The reason has the standard attributes present in elements containing text.

### 2.2.12 Results

In a couple of places in the survey form, we keep track of the results of activities, and initiatives undertaken by school committees and other organizations. The results are recorded in a `result` element.

```
<!ELEMENT result (#PCDATA)>
<!ATTLIST result
    issue IDREF #IMPLIED
    status (resolved | unresolved) #IMPLIED>
```

The content of the `result` element is a free-form text description of the result. The attributes of `result` are described in Table 2.3.

A `result` element is linked to an `issue` element via use of the `issue` attribute. If present, the value of this attribute should be the `id` of the reference `issue` element.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>language</td>
<td>one the supported languages (Section 2.2.23)</td>
</tr>
<tr>
<td>encoding</td>
<td>encoding used, if different from UTF-8</td>
</tr>
<tr>
<td>issue</td>
<td>link to an existing <code>issue</code> element elsewhere in this school data file</td>
</tr>
<tr>
<td>status</td>
<td>the status of the referenced <code>issue</code>. One of resolved, unresolved or left unspecified if the status is unknown.</td>
</tr>
</tbody>
</table>

### 2.2.13 Classes

Some elements of the form are associated with specific classes in the school. We provide a standard naming convention for classes: classes are named `std-nn`, where `nn` ranges from 1 to 10. These symbols denote Class 1 to Class 10 respectively.

```
std-1 | std-2 | std-3 | std-4 | std-5 | std-6 | std-7 | std-8 | std-9 | std-10 | std-other
```

Other classes, if any, are classified as `std-other`.

### 2.2.14 Distances

The distances are recorded in the form with a coarse level of granularity. The options are (Table 2.4):

```
m0-50 | m50-100 | m-over-100 | km0-2 | km2-5 | km-over-5 | other
```
Table 2.4 Distances

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>m0-50</td>
<td>within 50 metres</td>
</tr>
<tr>
<td>m50-100</td>
<td>between 50 to 100 metres</td>
</tr>
<tr>
<td>m-over-100</td>
<td>further than 100 metres away</td>
</tr>
<tr>
<td>km0-2</td>
<td>within 2 kilometers</td>
</tr>
<tr>
<td>km2-5</td>
<td>between 2 and 5 kilometers</td>
</tr>
<tr>
<td>km-over-5</td>
<td>further than 5 kilometers away</td>
</tr>
<tr>
<td>other</td>
<td>other distance</td>
</tr>
</tbody>
</table>

2.2.15 Dimensions

In some sections of the survey form we collect information about the sizes of buildings, rooms, and playgrounds. It is convenient to record these in a dimension element, defined as follows:

```xml
<!ELEMENT dimension (#PCDATA)>  
<!ATTLIST dimension unit (mm | cm | m | km | in | ft | yd | mile) #REQUIRED>
```

The content of the dimension element is the numeric value of the dimension. The attribute unit records the unit of measurement. Allowed values for unit are:

Table 2.5 Dimension units

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>mm</td>
<td>millimeters</td>
</tr>
<tr>
<td>cm</td>
<td>centimeters</td>
</tr>
<tr>
<td>m</td>
<td>meters</td>
</tr>
<tr>
<td>km</td>
<td>kilometers</td>
</tr>
<tr>
<td>in</td>
<td>inches</td>
</tr>
<tr>
<td>ft</td>
<td>feet</td>
</tr>
<tr>
<td>yd</td>
<td>yards</td>
</tr>
<tr>
<td>mi</td>
<td>miles</td>
</tr>
</tbody>
</table>

An application processing the XML file needs to verify that the content of the dimension element is a syntactically valid number. It may also need to verify that the numeric value is in “reasonable” bounds. This DTD does not specify what such a reasonable bound would be.

2.2.16 Quality Judgements

In some sections of the form we gather qualitative information about the condition of rooms and books in the school. We use a restricted set of choices in the survey form for gathering information:

```xml
<!ELEMENT qualityJudgement (#PCDATA)>  
<!ATTLIST qualityJudgement good | ok | bad | very-bad #REQUIRED>
```

The expected usage of these symbols is:
2.2. BASIC DTD ELEMENTS

Table 2.6 Quality judgements

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>good</td>
<td>no scope for improvement or no repairs needed</td>
</tr>
<tr>
<td>ok</td>
<td>useable but minor repairs or improvements are needed</td>
</tr>
<tr>
<td>bad</td>
<td>just useable and major repairs are needed</td>
</tr>
<tr>
<td>very-bad</td>
<td>unusable or dangerous to use</td>
</tr>
</tbody>
</table>

2.2.17 Contractors and Constructors

The survey form keeps track of the people and organizations that constructed the rooms, walls and buildings in the school. The currently allows list of choices is:

\[
\text{⟨.dtd: constructed by} \equiv (62d 65b 66a) \text{⟩}
\]

land-army | zp-contractor | tp-contractor | pvt-contractor | volunteer | other

The meanings of these symbols is as in Table 2.7.

Table 2.7 Infrastructure constructed by

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>land-army</td>
<td>constructed by the government land army</td>
</tr>
<tr>
<td>zp-contractor</td>
<td>constructed by a contractor of the Zilla Parishad</td>
</tr>
<tr>
<td>tp-contractor</td>
<td>constructed by a contractor of the Taluk Parishad</td>
</tr>
<tr>
<td>volunteer</td>
<td>constructed by a volunteer group</td>
</tr>
<tr>
<td>ngo</td>
<td>constructed by an NGO</td>
</tr>
<tr>
<td>community</td>
<td>constructed by the local community</td>
</tr>
<tr>
<td>other</td>
<td>constructed by others</td>
</tr>
</tbody>
</table>

XXX: This table has new kinds of constructed-by values, not in the form.

2.2.18 Recording Sexes

The sex of a person is recorded using a \textit{sex} attribute to the relevant elements of the school data file.

Allowed values for the \textit{sex} attribute are the tokens \text{male} and \text{female}.

\[
\text{⟨.dtd: sex} \equiv (48c 55c 92b) \text{⟩}
\]

\text{male} | \text{female}

2.2.19 Transportation options

When describing schools we sometimes need to record the available transportation options for students and teachers.

As of the moment, we are allowing free-form descriptions for transportation options. We use a \texttt{CDATA} attribute to hold the description of a transportation option.

\[
\text{⟨.dtd: transportation options} \equiv (36b 48c) \text{⟩}
\]

\texttt{CDATA}
2.2.20 Representing dates

Dates can be represented in many forms. For example, “1-1-1991”, “1st Jan 1991” and “January 01, 1991” are all valid examples of dates, but “234-2003” is not. XML DTDs do not have a date type as a primitive type, nor do we have a way to specify restrictions on the format of strings using a DTD. Consequently, where-ever a date is expected, we can only inform the XML parser to expect a string value.

The responsibility for checking that the given string can actually represent a date lies with the application processing the school data file.

2.2.21 Medium of Instruction

The medium of instruction in a school is sometimes used as an attribute to qualify elements. The currently understood values for these attributes are (Table 2.8):

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>english</td>
<td>instruction is predominantly in English</td>
</tr>
<tr>
<td>kannada</td>
<td>instruction is predominantly in Kannada</td>
</tr>
<tr>
<td>other</td>
<td>instruction is in some other language</td>
</tr>
</tbody>
</table>

Table 2.8 Medium of instruction

2.2.22 Numbers

XML does not have a mechanism to mark that an element or attribute is to be restricted to be a numeric quantity. Hence we allow strings (CDATA) where-ever we expect a number.

Applications have to check that the string value recorded is a valid representation of a numeric quantity.

2.2.23 Supported Languages

We want to give the freedom to the users to enter local language using roughly phonetic (transliterated) ASCII.

On the encoding front, the UNICODE standard allows us to encode data in almost every language in the world. Processing tools today have not caught with the standard.

We thus need a way to annotate text items with both the language of the textual item and the encoding used.
2.2. BASIC DTD ELEMENTS

We will initially allow text items to be entered in the few South Indian languages shown above.

Every element whose content can be text is given an attribute language that can be one of the supported languages (Table 2.9)

```xml
<dt:standard text field attributes>
  language (dt: supported languages) #IMPLIED
  encoding CDATA #IMPLIED
</dt:standard text field attributes>
```

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>language</td>
<td>one of the supported languages</td>
</tr>
<tr>
<td>encoding</td>
<td>encoding used, if different from UTF-8</td>
</tr>
</tbody>
</table>

The optional language attribute (part of the standard text field attributes) can be used to specify the language of a comment. We differentiate between the language of the comment and the encoding (Unicode) used, since we can have comments in a local language transcribed approximately into ASCII (Roman) text.

XXX: to be modified to use the standardized xml:lang notation.

2.2.24 Academic Qualifications

We record the academic qualifications of people in a few places. These are modelled by an attribute that records free-form text. In subsequent versions of the DTD, this attribute could be restricted to one of a list of choices.

```xml
<dt:qualification>
  CDATA
</dt:qualification>
```

2.2.25 Designations

Individuals designations are also recorded as free-form text, as an attribute value.

```xml
<dt:designation>
  CDATA
</dt:designation>
```

2.2.26 Occupations

Occupations of people are recorded as free-form text, as an attribute value.

```xml
<dt:occupation>
  CDATA
</dt:occupation>
```

1 Restricting the allowed values helps to “tighten” the DTD, allowing for stricter validation of the school data file by an XML parser. It also allows easier searching for information in the overall database of school information.
CHAPTER 2. OVERVIEW

2.2.27 Ownership of Premises

Schools can be run in many locations, for example, in rented houses, in rent-free accommodation, or in the open. The kinds of premises used by the school are recorded in the value of an appropriately named attribute. The following kinds of ownership of premises is recognized:

\[ (\text{dtd: premise ownership} \equiv \text{own | rented | rent-free | no-premises | other} \]  

The meaning of these symbols is as in Table 2.10.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>own</td>
<td>the school or anganwadi owns the building and the rooms</td>
</tr>
<tr>
<td>rented</td>
<td>the school or anganwadi rents the premises from someone</td>
</tr>
<tr>
<td>rent-free</td>
<td>community owned and rent free</td>
</tr>
<tr>
<td>no-premises</td>
<td>the school or anganwadi has no premises</td>
</tr>
<tr>
<td>other</td>
<td>any other kind of ownership of premises</td>
</tr>
</tbody>
</table>

Table 2.10 Premises ownership

2.2.28 Yes/No answers

In some places in the DTD, we need to keep track of “yes”/“no” answers. For such cases we follow the convention of using an attribute that can take as values one of yes and no. The absence of a value means that the answer was unknown.

\[ (\text{dtd: yes or no} \equiv \text{yes | no} \]  

2.2.29 Roles

Sometimes we would like to keep track of the role played by a person in the running of a school. Elements that represent persons will have an additional role attribute to record this information.

\[ (\text{dtd: role attribute} \equiv \text{role (staff | volunteer | partner | government-representative | other) #REQUIRED} \]  

When used, the role attribute will be mandatory. The list of values for the role attribute are limited to the values in Table 2.11.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>staff</td>
<td>a person paid by CIEE</td>
</tr>
<tr>
<td>volunteer</td>
<td>an unpaid person working for CIEE</td>
</tr>
<tr>
<td>partner</td>
<td>an employee of a CIEE partner organization</td>
</tr>
<tr>
<td>government-representative</td>
<td>a government representative working with CIEE</td>
</tr>
<tr>
<td>other</td>
<td>any other person</td>
</tr>
</tbody>
</table>
2.3 Assembling the overall DTD

An XML parser expects the DTD to be specified using a defined formal syntax laid down in the XML specification. We need to assemble our DTD to match this syntax.

The structure of the CIEE WEB DATABASE DTD file is as follows:

```
<cieedb.dtd>
<!--
CIEE Web Database DTD.
.dtd version number
(a warning about generated files)
-->}
<dtd: notation declarations>
(dtd: miscellaneous declarations)
(dtd: form information element)
(dtd: community information element)
(dtd: school general information)
(dtd: student information)
(dtd: staff information)
(dtd: admin information)
(dtd: infrastructure)
(dtd: other information)
(dtd: neighbour information)
(dtd: school pictures element)
(dtd: improvement opinions)
(dtd: school information element)

This DTD has a version associated with it. The number is currently at 0.3.

```}

**Warning about generated files**

In the header comment, we have a warning for people reading the DTD file. Since the DTD file is generated from the NOWEB source for this document, any changes made to the generated file will be lost the next time the file is regenerated.

```
(a warning about generated files)
-->}
```
CHAPTER 2. OVERVIEW

Following this warning we reference definitions for the various elements that make up the document type specification. XML DTD syntax requires that elements are defined before they are referenced and this requirement dictates the order of the chunk reference above.

2.3.2 Miscellaneous Declarations

We collect all the miscellaneous elements into one chunk in the DTD file.

```xml
<dtd: miscellaneous declarations [4] ≡ (23a)
  <dtd: comment declaration [5c]
  <dtd: phone declaration [12a]
  <dtd: address declaration [12c]
  <dtd: name declaration [15a]
  <dtd: reason element [16d]
  <dtd: issue element [16a]
  <dtd: result element [17a]
  <dtd: dimension element [18a]
```
Chapter 3

Form Information

School surveys are conducted by CIEE staff and volunteers visiting schools and recording their observations in a printed survey form. This survey form then gets sent to the CIEE office where the collected data is added to the data file for the school.

A form element in the school data file records information about such updates to the recorded content about the school. The element records:

- the dates during which the school data was gathered
- the people who did the actual school visit and survey
- the version numbers of the printed form used
- the serial numbers (filing numbers) of the form used
- the person who entered the data into the system

In the school data file, one form element corresponds to one printed form used in the survey. New survey results about a school would result in new form elements being added to the school data file.

The form elements associated with a school data file are grouped inside a form-list wrapper element.

3.1 DTD

Formally, a form-list element comprises one or more form sub-elements. At least one form element has to be present.

```
<![dtd: form information element]>
<![dtd: form sub elements]>
<!ELEMENT form-list (form+)>
<!ATTLIST form-list
  ![dtd: standard attributes]>
<!ELEMENT form (form-filled-by+, form-entered-by)>
<!ATTLIST form
  ![dtd: standard attributes]>
  version (v1.4 | v1.5 | v1.6 | other) #REQUIRED
  filing-number CDATA #REQUIRED
  start-date (![dtd: date type]) #REQUIRED
```
CHAPTER 3. FORM INFORMATION

Table 3.1 Attributes of element form

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>optional XML ID for this element (Section 2.2.1)</td>
</tr>
<tr>
<td>version</td>
<td>version number of the printed survey form used</td>
</tr>
<tr>
<td>filing-number</td>
<td>CIEE’s filing number for this survey form</td>
</tr>
<tr>
<td>start-date</td>
<td>start date for this school’s survey</td>
</tr>
<tr>
<td>entry-date</td>
<td>processing date for this form</td>
</tr>
<tr>
<td>end-date</td>
<td>end date for this school’s survey</td>
</tr>
</tbody>
</table>

The attributes of form are summarized in Table 3.1.

The form element has an attribute version that records the version number of the printed form used in the survey.

Every incoming form processed at the CIEE office has a filing number used to locate the document in CIEE’s filing system. We record this filing number using the filing-number attribute. The content of the attribute can be any text, and not necessarily a “number”.

The attribute start-date indicates when the form started being filled in the field, while the end-date attribute indicates when the form was completed. Of these two, only the start-date attribute is mandatory. If end-date is not specified, it is assumed to be the same as start-date.

The entry-date attribute records the date when the data from the survey form was entered into the system.

The child elements of form, namely, form-filled-by and form-entered-by, track the persons who conducted the survey and who entered the data into system.

3.1.1 Element: form-filled-by

A form-filled-by element records the name of a person who conducted the school survey. The content of this element holds text (#PCDATA). There can be more than one form-filled-by elements in a form, if a team of people conducted the survey.

```
<!ELEMENT form-filled-by (#PCDATA)>
<!ATTLIST form-filled-by
  role (volunteer|partner|staff) #REQUIRED

The form-filled-by element is qualified by the role attribute which tracks whether the person was a volunteer, partner or a member of CIEE staff.

3.1.2 Element: form-entered-by

The form-entered-by element records the name of the person who entered the data into the system from the printed form.

```

The form-entered-by element has a role attribute (Section 2.2.29) that records whether the person was a volunteer, staff member or partner.

3.2 Sample Data

In the sample data below, we show form elements for a school that was surveyed twice. The data from the first survey was from a survey form with filing number “CIEEDB-2001/32” and the second was from a form with filing number “CIEEDB-2002/64”. The first survey used revision 1.4 of the printed survey form, and the second used revision 1.5.

```
<sample:form-info>
<forms>
  <form version='v1.4' filing-number="CIEEDB-2001/32"
   start-date="1-1-2001" end-date="3-1-2001"
   entry-date="21-1-2001">  
    <form-filled-by role="volunteer">P. C. Thangappa</form-filled-by>
    <form-entered-by role="staff">P. C. Munjappa</form-entered-by>
  </form>
  <form version='v1.5' filing-number="CIEEDB-2002/64"
  start-date="1-1-2002" entry-date="22-2-2002">  
    <form-filled-by role="volunteer">P. C. Doddappa</form-filled-by>
    <form-entered-by role="staff">P. C. Munjappa</form-entered-by>
  </form>
</forms>
</sample:form-info>
```

Note that we do not track (minor) edits to the school data itself. This functionality would be best done by a revision control system like RCS.
Chapter 4

Community Information

The community information section in the sample file records demographic information about the locality and social environment in which the school is operating.

4.1 Community Information

A sample XML file containing some fictitious community information is presented for discussion.

```xml
<community-information>
  ⟨sample:population table⟩
  ⟨sample:communities in village⟩
  ⟨sample:anganwadis in village⟩
  ⟨sample:local institutions in village⟩
  ⟨sample:special conditions in village⟩
  ⟨sample:educational institutions in village⟩
  ⟨sample:other educational activities⟩
  ⟨sample:other facilities in village⟩
</community-information>
```
There are many subsections to this element. However, not all these subsections are mandatory. Only the population table and community information is required to be present. All other sections are optional.

**DTD**

The DTD models this structure as follows:

```xml
<!ELEMENT community-information (population-table+, communities+, anganwadi-list*, local-institutions*, special-conditions*, educational-institutions*, other-educational-activities*, other-facilities*)>
```

At least one of the population-table and communities sub-elements are required. All others are optional. We can have multiple sub-elements, one for each academic year.

We now look at each of these sub sections in detail.

### 4.2 Population Table

The population table maps the age distribution of the student population of the village in terms of social classes. Each social class is represented by a population-segment element in the population table.\(^1\)

In the following example, we have two population classes, sc-m and sc-f denoting “schedule class” males and females respectively. The number of children in the 0–5 and 6–13 age group have been specified in these elements.

```xml
<sample: population-table academic-year='1999-2000'>
  <population-segment segment='sc-m' n-0-5='35' n-6-13='30'/>
  <population-segment segment='sc-f' n-0-5='15' n-6-13='10'/>
</sample: population-table>
```

\(^1\)The two dimensional table in the paper form has been organized as a collection of columns.
4.2. POPULATION TABLE

4.2.1 DTD

We first consider the definition of the population-segment element.

\[
\langle \text{.dtd: population table} \rangle \equiv
\langle \text{!ELEMENT population-segment EMPTY} \rangle
\langle \text{!ATTLIST population-segment (\text{dtd: standard attributes})} \rangle
\langle \text{segment (\text{dtd: population segments}) #REQUIRED} \rangle
\langle n-0-5 (\text{dtd: number type}) \#IMPLIED \rangle
\langle n-6-13 (\text{dtd: number type}) \#IMPLIED \rangle
\langle n-14-17 (\text{dtd: number type}) \#IMPLIED \rangle
\langle n-18-above (\text{dtd: number type}) \#IMPLIED \rangle
\langle n-drop-out (\text{dtd: number type}) \#IMPLIED \rangle
\langle n-working (\text{dtd: number type}) \#IMPLIED \rangle
\langle n-disabled (\text{dtd: number type}) \#IMPLIED \rangle
\langle n-orphan-homeless (\text{dtd: number type}) \#IMPLIED \rangle
\langle n-never-enrolled (\text{dtd: number type}) \#IMPLIED \rangle
\langle \text{population-table (\text{population-segment}+)} \rangle
\langle \text{population-table (\text{dtd: mandatory academic year attribute})} \rangle
\]

The only mandatory attribute is the segment attribute that names class of population that the element refers to. The set of population classes recognized is:

\[
\langle \text{dtd: population segments} \rangle \equiv
\langle \text{sc-m, sc-f, st-m, st-f, group1-m, group1-f, muslim-m, muslim-f, others-m, others-f, migrant-m, migrant-f} \rangle
\]

The meaning of these symbols is explained in Table 4.1.

XXX: What about migrants who are SCs or STs or Muslims?

Table 4.1 Population Segment Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>sc-m, sc-f</td>
<td>Scheduled Caste children, males and females</td>
</tr>
<tr>
<td>st-m, st-f</td>
<td>Scheduled Tribe children, males and females</td>
</tr>
<tr>
<td>group1-m, group1-f</td>
<td>Group 1 (forward caste) children</td>
</tr>
<tr>
<td>muslim-m, muslim-f</td>
<td>Muslim children</td>
</tr>
<tr>
<td>others-m, others-f</td>
<td>Children belonging to all other minorities</td>
</tr>
<tr>
<td>migrant-m, migrant-f</td>
<td>Migrant children</td>
</tr>
</tbody>
</table>

The population-segment element can have a number of optional attributes (see Table 4.2) which represent the number of students in each age group. If any of these attributes is present, it is expected to have as its value the number of children in the grouping specified. If the attribute is left unspecified, the meaning is that the number of children is not known (as opposed to zero).

XXX: How are we to handle overlaps? Some children can be in 5–13 years of age, but also disabled and also orphaned and also never enrolled! Where do such children get counted? This needs to be specified.

The population-table element is a collection of population-segment elements. It has an academic-year attribute that is mandatory.
Table 4.2 Population Segment Attributes

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>n-0–5</td>
<td>Age group 0–5 inclusive</td>
</tr>
<tr>
<td>n-5–13</td>
<td>Age group 5–13 inclusive</td>
</tr>
<tr>
<td>n-14–17</td>
<td>Age group 14–17 inclusive</td>
</tr>
<tr>
<td>n-18-above</td>
<td>Age group 18 and older</td>
</tr>
<tr>
<td>n-drop-out</td>
<td>Children who have dropped out of school</td>
</tr>
<tr>
<td>n-working</td>
<td>Children who are working</td>
</tr>
<tr>
<td>n-disabled</td>
<td>Disabled children</td>
</tr>
<tr>
<td>n-orphan-homeless</td>
<td>Orphaned and homeless children</td>
</tr>
<tr>
<td>n-never-enrolled</td>
<td>Never enrolled in a school</td>
</tr>
</tbody>
</table>

4.3 Communities

The kinds of communities in the village are recorded in the Communities section. Some sample data are presented:

```xml
<communities academic-year='1999-2000'>
  <community name="SC">
    <number-of-households>2</number-of-households>
    <total-population>20</total-population>
    <occupation>grave digger</occupation>
    <language-spoken>esperanto</language-spoken>
    <yearly-average-income>31416</yearly-average-income>
  </community>
</communities>
```
Each community is modelled by a `community` element. This element in turn has as sub-elements, `number-of-households` denoting the number of households belonging to the community, `total-population` representing the total strength of the community, a list of occupations represented by one or more `occupation` elements, a list of languages spoken represented by one or more `language-spoken` elements, and a yearly average income in rupees represented by `yearly-average-income`.

The `number-of-households`, `total-population` and `yearly-average-income` elements are to hold numbers. Since we have no way of restraining element content in XML, we have specified these are `#PCDATA`, meaning any character data would be valid. Applications processing these fields have to ensure that their contents are valid numbers.

Allowable community names are:

- `SC` | `ST` | `I` | `muslim` | `migrant` | `other`

The meanings of these symbols should be self-explanatory.

- **XXX**: How do we deal with migrants who are also SCs or STs?

The `communities` element comprises of one or more `community` elements. This element has a mandatory academic-year element that records when the data was collected.
4.4 Anganwadis

The anganwadi-list element records the anganwadis in the village. Here is some sample data.

```xml
<anganwadi-list>
  <anganwadi date-started="1-1-1991" n-children="20" n-teachers='1'>
    <name>M. C. Ponnappa Bala Vikasa Kendra</name>
  </anganwadi>
</anganwadi-list>
```

The section comprises one or more anganwadi elements.

4.4.1 DTD

```xml
<!ELEMENT anganwadi (name)>  
<!ATTLIST anganwadi 
  date-started  #IMPLIED 
  n-children  #IMPLIED 
  n-teachers  #IMPLIED 
  n-helpers  #IMPLIED 
  room-ownership ((<!ATTLIST premise-ownership>)) #IMPLIED>
```

XXX: Are we really tracking if the teacher/helpers were absent?

4.5 Local Institutions

The local institutions section records if there are any womens groups, youth groups and other organizations active in the village.

```xml
<local-institutions>
  <local-institution is-active="no">
    <name>J. C. Ponnappa Abimaaniga Sangha</name>
    <reason>Admirers of J.C. Ponnappa</reason>
    <comment>Inactive because all members have migrated to Timbuktu.</comment>
  </local-institution>
</local-institutions>
```
4.6 SPECIAL CONDITIONS

4.5.1 DTD

Each local institution is represented by a local-institution element. The is-active attribute records if the institution is active. This attribute is optional; if not present, it means that the activity level of the institution is unknown. Structurally, the local-institution element has a mandatory name element that contains the name of the institution. Following this, we can have an optional reason-for-formation element that records the reason for formation of the organization. A comment element records extra comments, if any.

The overall local-institutions element serves as a wrapper for individual local-institution elements.

4.6 Special Conditions

The Special Conditions section records any unusual conditions in the village that could have an effect on schooling. This is a plain text field, used to hold any text.

Examples of special conditions could be: a history of communal tension in the village, prevalent alcoholism, widespread practice of child labour etc.

4.6.1 DTD

This element is simple, holding only plain text. Accordingly, its formal model has only #PCDATA as content.
The element has one mandatory attribute, academic-year that records the year of collection of data.

### 4.7 Educational Institutions

The *Educational Institutions* sub section in the collected data records nearby educational institutions (other than the school being surveyed) and records the distances of these from the village.

```xml
<sample: educational institutions in village≡
<educational-institutions>
  <educational-institution institution="govt-lps" distance='km0-2'/>
</educational-institutions>
```

### 4.7.1 DTD

Formally, an educational-institutions element is a list of educational-institution elements.

```xml
<!ELEMENT educational-institution EMPTY>
<!ATTLIST educational-institution
  transportation #IMPLIED
  distance (km|km0-2|km2-5|km>5) #IMPLIED
  institution (govt-lps | govt-hps | govt-hs | pvt-lps | pvt-hps | pvt-hs | puc-college | other) #REQUIRED>

<!ELEMENT educational-institutions (educational-institution+)>
<!ATTLIST educational-institutions
  transportation #IMPLIED
  distance (km|km0-2|km2-5|km>5) #IMPLIED
  institution (govt-lps | govt-hps | govt-hs | pvt-lps | pvt-hps | pvt-hs | puc-college | other) #REQUIRED>
```

Each educational-institution element records as attributes the distance of the institutions from the village (distance, see 2.2.14), the kind of the institution (institution), and transportation options available to reach the institution (transportation).

The list of recordable educational institution types is:

```xml
<sample: educational institutions≡
govt-lps | govt-hps | govt-hs | pvt-lps | pvt-hps | pvt-hs | puc-college | other
```

The meaning of these symbols is described in Table 4.3.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>govt-lps</td>
<td>Government Lower Primary School</td>
</tr>
<tr>
<td>govt-hps</td>
<td>Government Higher Primary School</td>
</tr>
<tr>
<td>govt-hs</td>
<td>Government Higher Secondary School</td>
</tr>
<tr>
<td>pvt-lps</td>
<td>Private Lower Primary School</td>
</tr>
<tr>
<td>pvt-hps</td>
<td>Private Higher Primary School</td>
</tr>
<tr>
<td>pvt-hs</td>
<td>Private Higher Secondary School</td>
</tr>
<tr>
<td>puc-college</td>
<td>Pre-University College</td>
</tr>
<tr>
<td>other</td>
<td>any other educational institution</td>
</tr>
</tbody>
</table>
4.8 Other Educational Activities

In Other Educational Activities, we record any other education related activities that are going on in the community.

In Other Educational Activities, we record any other education related activities that are going on in the community.

4.8.1 DTD

Each other-educational-activity element has a mandatory activity attribute that identifies the kind of activity that it represents. The optional is-functioning attribute records if the activity is still in progress. This attribute if not present means that the state of the activity is unknown.

The list of educational activities recognized is:

govt-nfe | evening-classes | adult-literacy | ngo-nfe | other

Explanations for these symbols are given in Table 4.4.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>govt-nfe</td>
<td>Government run non-formal education</td>
</tr>
<tr>
<td>evening-classes</td>
<td>Evening classes run by anyone</td>
</tr>
<tr>
<td>adult-literacy</td>
<td>An official adult literacy programme</td>
</tr>
<tr>
<td>ngo-nfe</td>
<td>An NGO run non-formal education programme</td>
</tr>
<tr>
<td>other</td>
<td>Any other educational activity</td>
</tr>
</tbody>
</table>

Table 4.4 Educational Activities
CHAPTER 4. COMMUNITY INFORMATION

The other-educational-activities element is a list of one or more other-educational-activity elements.

4.9 Other Facilities

In this section we record the presence or absence of other facilities like libraries, playgrounds and banks in the community.

4.9.1 DTD

The other-facilities element is a collection of other-facility elements.

The meaning of these symbols should be evident from their names.
Chapter 5

School General Information

In the section School General Information we store information about the school that can be considered reasonably static; for example, the full name of the school, its number, circle, ward etc.

5.1 Sample Data

```xml
<sample:school-general-info>
  <school-general-information>
    <school-name>Biligiri Government Primary School</school-name>
    <school-number>#123</school-number>
    <ward>#123</ward>
    <year-started>1942</year-started>
    <school-address date-from="1982" is-current="yes">
      <address>
        <additional-text>42 Murugesh Palya</additional-text>
        <village-or-place>Chikka Biligirihalli</village-or-place>
        <gram-panchayat>I. G. Gram Panchayat</gram-panchayat>
        <taluk>Negamandala Taluk</taluk>
        <district>Bangalore</district>
        <pin>5000011</pin>
      </address>
      <phone-list>
        <phone phone-number="0801112222">in HM office</phone>
        <phone phone-number="0801113333" />
      </phone-list>
      <location-comments is-access-easy="no" does-location-affect-attendance="yes" neighbourhood-kind="low-income">too far away for SC/ST community children</location-comments>
    </school-address>
    <working-days-list>
      <working-days working-days="156" academic-year="1999-2000" />
    </working-days-list>
  </school-general-information>
</sample:school-general-info>
```
CHAPTER 5. SCHOOL GENERAL INFORMATION

5.2 DTD

5.2.1 School Name and Number

We model the schools name and number as plain text, using the school-name and school-number elements.

\[
\text{school-name} \quad \equiv \\
!ELEMENT school-name (#PCDATA) > \\
!ATTLIST school-name \\
\text{standard attributes} \\
\text{standard text field attributes} > \\
\text{school-number} \\
\equiv \\
!ELEMENT school-number (#PCDATA) > \\
!ATTLIST school-number \\
\text{standard attributes} > \\
\]

5.2.2 Circle, Cluster or Ward

Only one of circle, cluster and ward can be specified for one school.

XXX: is this true?

Each of these elements is a free form text field.

\[
\text{circle} \quad \equiv \\
!ELEMENT circle (#PCDATA) > \\
!ATTLIST circle \\
\text{standard attributes} > \\
\text{cluster} \\
\equiv \\
!ELEMENT cluster (#PCDATA) > \\
!ATTLIST cluster \\
\text{standard attributes} > \\
\]

5.2.3 Year Started

The year the school started is recorded in the year-started element, as free form text.

\[
\text{year-started} \quad \equiv \\
!ELEMENT year-started (#PCDATA) > \\
!ATTLIST year-started \\
\text{standard attributes} > \\
\]
5.2. DTD

5.2.4 School Address

A school address element holds a premise postal address along with some commentary on the location of the school.

An address element holds the schools current premise address. The location-comments element records whether the address allows easy access to the children or whether it has any drawbacks.

The comment element holds any extra comments if any.

The school-address has a mandatory attribute is-current. When this is yes, the school address element records the current address of the school. Prior addresses of the school would have a value no for this attribute. The date-from attribute records the date from which the school started functioning in the given premises. The date-till attribute records the date when the school stopped functioning in the premises.

5.2.5 Working Days

The element working-days-list records the working days of the school, per academic year. It is a wrapper around individual working-days elements.

Each working-days element holds the number of actual working days in one academic year. It is an empty element, holding the number of working days in its working-days attribute. The academic year is stored in the mandatory academic-year attribute.

5.2.6 Range of Classes

The range of classes in the school is recorded in a range-of-classes element.

```
<!ELEMENT range-of-classes EMPTY>
<!ATTLIST range-of-classes
  academic-year (dd:standard-attributes) #REQUIRED
  juniormost-class (dd:classes) #REQUIRED
  seniormost-class (dd:classes) #REQUIRED>
```

This element has three mandatory attributes: the academic-year attribute records the academic year when this data was collected, the juniormost-class attribute records the lowest (juniormost) class present in the school and the seniormost-class attribute records the highest (seniormost class) that is present in the school.

5.2.7 School General Information

```
<!ELEMENT school-general-information (school-name, school-number, (circle | cluster | ward),
  year-started, school-address+, working-days-list, range-of-classes+)>
<!ATTLIST school-general-information
  (dd:standard-attributes)>
```
Chapter 6

Student Information

The Student Information section of the form collects academic information about the children attending school. The number of children on the rolls, and the school results in prior examinations are recorded.

Since the school information file holds data for more than one year, we model the data in this section as a collection of one or more student-survey-info elements wrapped in a top-level student-information element.

\[
\langle \text{sample:student-info} \rangle \equiv \langle \text{sample:student-info} \rangle \end{array} \]

\[
\langle \text{sample:student-info} \rangle \equiv \langle \text{student-information} \rangle \end{array}
\]

\[
\langle \text{student-information} \rangle \equiv \langle \text{student-survey-info} \rangle \end{array} \]

\[
\langle \text{student-information} \rangle \equiv \langle \text{student-results} \rangle \end{array}
\]

\[
\langle \text{student-information} \rangle \equiv \langle \text{student-survey-info} \rangle \end{array} \]

\[
\langle \text{student-information} \rangle \equiv \langle \text{student-results} \rangle \end{array}
\]

6.1 DTD

In the DTD, we specify that the student-information section has two major kinds of sub-elements: the student-survey-info element contains the information about student attendance in one survey form. The student-results element contains the prior academic performance of the school.

\[
\langle \text{dtd: student information} \rangle \equiv \langle \text{dtd: student survey info} \rangle \end{array} \]

\[
\langle \text{dtd: student information} \rangle \equiv \langle \text{dtd: student results} \rangle \end{array}
\]

\[
\langle \text{dtd: student information} \rangle \equiv \langle \text{dtd: student results} \rangle \end{array}
\]
6.1.1 Information collected per form

The student-survey-info element is repeated, once per survey form. It comprises, a students-on-roll element that tracks the children on the school rolls, and an additional comment element that records if anything unusual was going on when the school data was being collected.

\[
\langle\text{student-survey-info}\rangle \equiv \langle\text{students-on-roll}, \text{comment}\rangle
\]

\[
\langle\text{students-on-roll}\rangle \equiv \langle\text{students-on-roll}\rangle^* \langle\text{student-attendance}\rangle^+
\]

\[
\langle\text{student-attendance}\rangle \equiv \langle\text{class}\rangle^* \langle\text{n-sc-m}\rangle^* \langle\text{n-sc-f}\rangle^* \langle\text{n-st-m}\rangle^* \langle\text{n-st-f}\rangle^* \langle\text{n-others-m}\rangle^* \langle\text{n-others-f}\rangle^* \langle\text{n-attendance-m}\rangle^* \langle\text{n-attendance-f}\rangle^*
\]

If something unusual (for example, a festival) was ongoing at the time of the survey, then the data collected on attendance may not be very reliable. The data-reliability attribute is set to high or low to note this fact. The collection-date attribute records the date and time when the data was collected.

Student Attendance

The students-on-roll element records the number of children attending each class and their communities. It is modelled as containing one or more student-attendance elements. This is to be one students-on-roll element per medium of instruction. We allow more than one students-on-roll elements since some schools can have classes in more than one medium of instruction.

Each student-attendance element records the break-up of children in the community.
6.1. DTD

6.1.2 Student Results

We use one student-result element to model the results from one academic year. Each student-result element contains as content one or more class-academic-result elements.

```
<student-results>
  <class-academic-result/>
</student-results>
```

The element contains a mandatory academic-year attribute which records the academic year for which the data was collected.

Each class-academic-result element holds the results declared for each class. This element has no further content: its attributes record the relevant data.

```
<class-academic-result>
  <class> </class>
  <n-boys-passed>0</n-boys-passed>
  <n-boys-failed>0</n-boys-failed>
  <n-boys-total>0</n-boys-total>
  <n-girls-passed>0</n-girls-passed>
  <n-girls-failed>0</n-girls-failed>
  <n-girls-total>0</n-girls-total>
</class-academic-result>
```

Of the attributes, the class attribute is mandatory and the rest are optional. The rest of the attributes record the number of each category of results (boys and girls, passing or failing their examinations).
Chapter 7

Staff Information

The Staff Information section of the school survey form captures information on the staffing of the school. It comprises a list of staff members, along with their qualifications and other personal details, and other information pertaining to the extra work that teachers have to do.

A sample section of the data file would be as follows:

```
<staff-information>
  <staff-member qualification="BA" sex="male" date-of-birth="1-1-1970">
    <name>T. C. Doddappa</name>
    <staff-role role="headmaster"/>
    <staff-role role="PE"/>
    <staff-role role="teacher"/>
    <class-list>
      <class class="std-6"/>
    </class-list>
    <subject-list>
      <subject subject="hindi"/>
    </subject-list>
  </staff-member>
  ...
</staff-information>
```

The teachers’ opinion on community participation would go here.

```
<incentive-program-info>
  <incentive-program scheme="midday-meal" frequency="daily" n-teachers="2"
  hours-per-session="1"/>
</incentive-program-info>
```

```
<admin-work>
  <admin-work-element admin-work="hm-meeting" frequency="weekly" n-teachers="4"
  n-days="1"/>
</admin-work>
```
CHAPTER 7. STAFF INFORMATION

7.1 DTD

The content of this section of the school information form is recorded in a staff-information element. This element comprises in turn of staff-member-info, staff-attendance, teachers-opinion, incentive-program-info and admin-work sections, matching the structure of the school information survey form.

\[
\langle \text{dtd: staff information} \rangle \equiv (23a)
\]

\[
\langle \text{dtd: staff member info} \rangle \langle \text{dtd: staff attendance} \rangle \langle \text{dtd: teachers opinion} \rangle \langle \text{dtd: incentive program info} \rangle \langle \text{dtd: admin work} \rangle
\]

\[
<!ELEMENT staff-information (staff-member-info?, staff-attendance?, teachers-opinion?, incentive-program-info?, admin-work?)>
\]

\[
<!ATTLIST staff-information
\langle \text{dtd: standard attributes} \rangle
collection-date \langle \text{dtd: date type} \rangle #IMPLIED>
\]

The collection-date attribute of this element records when the data was collected.

7.1.1 Staff General Information

This section records information on the staff employed at this school. This information is modelled as a list of staff-member records.

\[
\langle \text{dtd: staff member info} \rangle \equiv (48b)
\]

\[
\langle \text{dtd: staff member} \rangle
\]

\[
<!ELEMENT staff-member-info (staff-member+)>
\]

\[
<!ATTLIST staff-member-info
\langle \text{dtd: standard attributes} \rangle
qualification \langle \text{dtd: qualification} \rangle #IMPLIED
sex \langle \text{dtd: sex} \rangle #IMPLIED
date-of-birth \langle \text{dtd: date type} \rangle #IMPLIED
distance \langle \text{dtd: distance options} \rangle #IMPLIED
transportation \langle \text{dtd: transportation options} \rangle #IMPLIED>
\]

Each staff-member element records the staff members name, one or more roles played by the member in school and if the staff member is a teacher, the subjects and classes taken by him or her.
In the staff-member element, the sex attribute records the sex of the staff member, if available. The date-of-birth element records the member's date of birth. The distance attribute records how far away from school the staff members home is. The transportation attribute records the transportation method used by the staff member to reach school from his or her home. The qualification of the staff member is recorded in the qualification attribute.

A given staff member can be shouldering many responsibilities in school. For example, a subject teacher could also be handing craft or music teaching. This data is captured by allowing a staff member record to contain one or more staff-role sub-elements.

```
<!ELEMENT staff-role EMPTY>
<!ATTLIST staff-role role (headmaster | teacher | PE | craft | music | office-assistant | peon | other) #REQUIRED>
```

Each staff-role element has a mandatory attribute role that lists which role this staff member is playing in school.

The allowed list of staff roles are

```
headmaster | teacher | PE | craft | music | office-assistant | peon | other
```

A class-list records the list of classes this staff member is involved with. It is represented as a list of class elements. Each class element has an attribute class that records the class.

```
<!ELEMENT class EMPTY>
<!ATTLIST class class (class+) #REQUIRED>
<!ELEMENT class-list (class*)>
<!ATTLIST class-list (dd: standard attributes)>
```

The subject-list element is similarly organized as a list of subject elements. Each subject element records the name of the subject.

```
<!ELEMENT subject EMPTY>
<!ATTLIST subject subject (subject+) #REQUIRED>
<!ELEMENT subject-list (subject*)>
<!ATTLIST subject-list (dd: standard attributes)>
```

```
<!ELEMENT school-subject-list EMPTY>
<!ATTLIST school-subject-list (dd: standard attributes)>
```

CDATA
The list of subjects is current marked as CDATA, allowing free-form input but this will be replaced with a choice list later.

7.1.2 Staff Attendance

Staff attendance at the time of the survey is recorded in a staff-attendance element. We are only interested in the number of people absent for a given reason. This information is modelled by having the staff-attendance element in turn be comprised of one or more staff-absent-reason elements.

\[
\text{staff-attendance} \equiv (\text{staff-absent-reason}+) \quad (48a)
\]

Each staff-absent-reason element records only the reason for being absent (in attribute reason-absent) and the number of staff absent for this reason (in attribute n-absent).

The list of recognized reasons for being absent are:

\[
\text{absenteeism reasons} \equiv \text{personal} | \text{in-training} | \text{administrative-work} | \text{field-visit} | \text{other}
\]

7.1.3 Teachers Opinion

The teachers-opinion element records the teachers suggestions on the areas where the community can participate in the operation of the school.

\[
\text{teachers-opinion} \equiv (\#PCDATA) \quad (48a)
\]

7.1.4 Incentive Schemes

This section records the government incentive schemes that the teachers in this school are required to participate in.

Incentive schemes are recorded in the incentive-program-info element. This element comprises a list of incentive-program elements.

\[
\text{incentive-program-info} \equiv (\text{incentive-program}+) \quad (48a)
\]
Each incentive-program element pertains to one government scheme (recorded in its scheme attribute). The frequency of the scheme is recorded in the frequency attribute. The number of teachers that participated in the scheme is recorded in the n-teachers attribute and the number of hours per session spent is recorded in the hours-per-session attribute.

### 7.1.5 Administrative Work

This section of the DTD records the administrative work done by teachers at the school.

The section is modelled by a `admin-work` element, which is in turn composed of one or more `admin-work-element` elements.

```xml
<!ELEMENT admin-work-element EMPTY>
<!ATTLIST admin-work-element
    n-teachers (dtd: number type) #REQUIRED
    n-days (dtd: number type) #REQUIRED
    frequency ((dtd: frequency) #REQUIRED
    admin-work ((dtd: admin work kind) #REQUIRED

<!ELEMENT admin-work (admin-work-element+)>
<!ATTLIST admin-work
    (dtd: standard attributes)
```

In the `admin-work-element` element, the attribute `admin-work` identifies the kind of work being done. The `frequency` attribute records the frequency of the activity. The `n-teachers` and `n-days` attributes record the number of teachers who participated and the average number of days per session.

The kinds of administrative work recognized are:

```xml
<!ATTLIST admin-work kind
    in-service-training | hm-meeting | BEO-meeting | cluster-meeting | on-deputation | other
```

```xml
<!ELEMENT admin-work kind (dtd: admin work kind) #REQUIRED>
```
Chapter 8

School Administration

The School Administration section tracks how the school is being looked after by the government administration and by the local community.

There are three major sections in the survey form: the first collects information on the frequency of visits to the school by various officials. The second records information about the school betterment committees and village education committees monitoring the school. The third section records information about special meetings termed (Samudayadatta Shale).

We keep the name of the block educational officer as part of the overall administration information about the school. We also record the kinds of official visitors to the school and their frequency of visits.

```xml
<school-administration>
  <BEO date-of-joining-office='1999-2000'>
    <name>Sri Sri Sri Chikka Dodda Swami</name>
  </BEO>
  <visitor-frequency academic-year='1999-2000'>
    <visitor visitor-kind='BEO' frequency='weekly' />
  </visitor-frequency>
</school-administration>
```
Here is a sample of how SBC/VEC information would be represented.

```
⟨sample: sbc-vec-info⟩≡
⟨sbc-vec-info⟩
  ⟨sbc-vec-member-table⟩
    ⟨sbc-vec-member sex='male' qualification='Std-10' occupation='farmer'
      n-children-in-govt-school='3' n-children-in-pvt-school='2'>
      ⟨name⟩P.C. Doddachinnapa⟩⟨/name⟩
    ⟨/sbc-vec-member⟩
  ⟨/sbc-vec-member-table⟩
  ⟨sbc-vec-activity-table academic-year='1999-2000'>
    ⟨sbc-vec-activity activity-name='enhancing-enrollment' n-members='2'⟩
    ⟨/sbc-vec-activity⟩
  ⟨/sbc-vec-activity-table⟩
  ⟨sbc-vec-meeting-table academic-year='1999-2000'>
    ⟨sbc-vec-meeting n-members='23' date='1-1-2000'>
      ⟨sbc-vec-meeting-agenda⟩Discussion whether monkeys can eat java beans.⟩⟨/sbc-vec-meeting-agenda⟩
    ⟨/sbc-vec-meeting⟩
  ⟨/sbc-vec-meeting-table⟩
⟨/sbc-vec-info⟩
```

8.1 DTD

```
⟨dtd: admin information⟩≡
  ⟨dtd: visitor frequency⟩
  ⟨dtd: sbc vec info⟩
  ⟨dtd: samudayadatta shale info⟩
  <!ELEMENT BEO (name)>  
  <!ATTLIST BEO date-of-joining-office #IMPLIED>
  <!ELEMENT school-administration (BEO*, visitor-frequency*, sbc-vec-info*, samudayadatta-shale-table?)>
  <!ATTLIST school-administration date-of-joining-office #IMPLIED>

8.1.1 Visitors

```
⟨dtd: visitor frequency⟩≡
  <!ELEMENT visitor EMPTY>  
  <!ATTLIST visitor visitor-kind #REQUIRED
    frequency #REQUIRED>
  <!ELEMENT visitor-frequency (visitor+)>
  <!ATTLIST visitor-frequency date-of-joining-office #IMPLIED>
```
The kinds of visitors tracked are:

\[
\text{⟨dtd: visitor kind⟩} \equiv \text{BEO} \mid \text{educational-coordinator} \mid \text{cluster-resource-person} \mid \text{block-resource-person} \mid \text{GP-secretary} \mid \text{taluk-EO} \mid \text{ZP-CEO} \mid \text{corporator} \mid \text{MLA} \mid \text{MP} \mid \text{GP-member} \mid \text{ZP-member} \mid \text{TP-member} \mid \text{other-educational-department-official} \mid \text{other-government-official} \mid \text{other}
\]

Surveyors are expected to choose the closest match from the list of frequencies (2.2.9) when filling up the form.

### 8.1.2 Data on SBC/VECs

\[
\text{⟨dtd: sbc vec info⟩} \equiv \text{⟨dtd: sbc vec member table⟩} \text{⟨dtd: sbc vec activity table⟩} \text{⟨dtd: sbc vec meetings table⟩}
\]

\[
\text{<!ELEMENT sbc-vec-info (sbc-vec-member-table, sbc-vec-activity-table*, sbc-vec-meeting-table*)>}
\]

\[
\text{<!ATTLIST sbc-vec-info ⟨dtd: standard attributes⟩>}
\]

\[
\text{⟨dtd: sbc vec member table⟩} \equiv \text{⟨dtd: sbc vec member⟩}
\]

\[
\text{<!ELEMENT sbc-vec-member (name)>}
\]

\[
\text{<!ATTLIST sbc-vec-member ⟨dtd: standard attributes⟩>}
\]

\[
\text{sex } \langle \text{dtd: sex} ⟩ \text{#IMPLIED}
\]

\[
\text{qualification } \langle \text{dtd: qualification} ⟩ \text{#IMPLIED}
\]

\[
\text{occupation } \langle \text{dtd: occupation} ⟩ \text{#IMPLIED}
\]

\[
\text{designation } \langle \text{dtd: designation} ⟩ \text{#IMPLIED}
\]

\[
\text{n-children-in-govt-school } \langle \text{dtd: number type} ⟩ \text{#IMPLIED}
\]

\[
\text{n-children-in-pvt-school } \langle \text{dtd: number type} ⟩ \text{#IMPLIED}
\]

\[
\text{<!ELEMENT sbc-vec-member-table (sbc-vec-member+)>}
\]

\[
\text{<!ATTLIST sbc-vec-member-table ⟨dtd: standard attributes⟩>}
\]

\[
\text{⟨dtd: sbc vec activity⟩ empty}
\]

\[
\text{<!ELEMENT sbc-vec-activity (name) #IMPLIED}
\]

\[
\text{activity-name } \langle \text{dtd: sbc vec activities} ⟩ \text{#REQUIRED}
\]

\[
\text{n-members } \langle \text{dtd: number type} ⟩ \text{#IMPLIED}
\]

\[
\text{<!ELEMENT sbc-vec-activity-table (sbc-vec-activity*)>}
\]

\[
\text{<!ATTLIST sbc-vec-activity-table ⟨dtd: standard attributes⟩>}
\]

\[
\text{⟨dtd: mandatory academic year attribute⟩} \
\]

The element \textit{sbc-vec-activity} tracks an activity undertaken by an SBC/VEC. Each activity has a mandatory \textit{activity-name} attribute that records the kind of activity. An \textit{n-members} field records the number of members from the SBC/VEC that participated in this activity.

All activities undertaken in one academic year are grouped in a \textit{sbc-vec-activity-table} element.
The `sbc-vec-activity-table` element has a mandatory `academic-year` attribute that records the academic year for which data was collected.

The list activities tracked is:

```xml
<enabling-parent-teacher-discussions | other
```

The meaning of these symbols is explained in Table 8.1.

### Table 8.1 SBC/VEC Activities

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>enhancing-enrollment</td>
<td>Enhancing school enrollment</td>
</tr>
<tr>
<td>improving-infrastructure</td>
<td>Improving school infrastructure</td>
</tr>
<tr>
<td>learning-level-enhancement</td>
<td>Learning level improvement</td>
</tr>
<tr>
<td>organizing-activities</td>
<td>Organizing activities like summer camps, sports competitions</td>
</tr>
<tr>
<td>participating-in-school-functions</td>
<td>Participation in school functions</td>
</tr>
<tr>
<td>raising-funds</td>
<td>Fund raising</td>
</tr>
<tr>
<td>mobilising-resources</td>
<td>Mobilising resources for building construction etc.</td>
</tr>
<tr>
<td>enabling-parent-teacher-discussions</td>
<td>Facilitation parent teacher meetings</td>
</tr>
<tr>
<td>other</td>
<td>Any other activities</td>
</tr>
</tbody>
</table>

XXX: what does learning-level-enhancement mean?

Meetings conducted by the SBC/VEC are tracked in a `sbc-vec-meetings-table` element. There is one `sbc-vec-meetings-table` element per academic year. Each such element comprises one or more `sbc-vec-meeting` elements.
Each SBC/VEC meeting has an agenda, recorded in the free form text field element `sbc-vec-meeting-agenda`. The date when the meeting was conducted is recorded in attribute `date` of element `sbc-vec-meeting` and a count of the number of SBC/VEC members who attended it, kept in attribute `n-members` of element `sbc-vec-meeting`.

### 8.1.3 Samudayadatta Shale

A summary of each Samudayadatta Shale is recorded in a `samudayadatta-shale` element (Table 8.2).

The content of the element comprises zero or more `issue` elements, with each `issue` element recording one issue discussed in the meeting.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>date</td>
<td>the date on which the meeting was held</td>
</tr>
</tbody>
</table>

The `date` attribute of `samudayadatta-shale` records the date on which the meeting was held.

All `samudayadatta-shale` elements are grouped in a `samudayadatta-shale-table` element.

Sample Data

The `samudayadatta-shale` meetings that take place are recorded as follows.

```xml
<samudayadatta-shale-table>
  <samudayadatta-shale date='10-10-2001'>
    <issue>
      <description>Village monkeys now demanding java beans.</description>
      <decision>Compromise: provide peanuts.</decision>
      <result status='unresolved'>awaiting supply of peanuts.</result>
    </issue>
  </samudayadatta-shale>
</samudayadatta-shale-table>
```
Chapter 9

Infrastructure

The *Infrastructure* chapter in the survey form is one of the longest and most detailed. It records in detail the kinds of facilities provided in the school, the current condition of the school, and the people responsible for the upkeep of each part.

A sample section in the school information file would have the structure matching the major sections of the survey form.

```
<s:infrastructure>
<s:infrastructure building ownership / site>
<s:infrastructure has verandah or corridor>
<s:infrastructure rooms>
<s:infrastructure water>
<s:infrastructure toilets>
<s:infrastructure compound wall>
<s:infrastructure flooring>
<s:infrastructure has electricity>
<s:infrastructure furniture list>
<s:infrastructure teaching material>
<s:infrastructure playground>
<s:infrastructure sports equipment>
<s:infrastructure library>
<s:infrastructure medical>
<s:infrastructure location of classes>
<s:infrastructure maintenance>
<s:infrastructure environment>
<s:infrastructure development>
<s:infrastructure previous complaints>
</infrastructure>
```
In formal terms, the structure of the school data file is described by the infrastructure element. This element has child elements which correspond to each aspect of “Infrastructure”.

```xml
<infrastructure>
  <building-ownership/>
  <building-site-details/>
  <has-verandah-or-corridor/>
  <rooms/>
  <water/>
  <toilet-facilities/>
  <compound-wall/>
  <flooring/>
  <has-electricity/>
  <furniture-list/>
  <teaching-material-list/>
  <playground/>
  <sports-equipment-list/>
  <library/>
  <medical/>
  <location-of-classes/>
  <maintenance/>
  <environment/>
  <development/>
  <previous-complaints/>
</infrastructure>
```

```xml
<!ATTLIST infrastructure
  (building-ownership)>
```

```xml
<!ELEMENT infrastructure
```

```xml
<!ATTLIST infrastructure
  (dd: standard attributes)>
```
9.1 Building Ownership

This section records the specifics of ownership of the school premises and school buildings.

The land that the school is situated on could be of different kinds of legal stature. This is recorded by a building-site-details element.

XXX: What about “open-air” schools? What should we fill for these?

The type attribute records the kind of the land on which the school is situated.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>urban-development-authority</td>
<td>land given by an urban development authority</td>
</tr>
<tr>
<td>corporation</td>
<td>corporation land</td>
</tr>
<tr>
<td>municipality</td>
<td>land given by the municipality</td>
</tr>
<tr>
<td>gram-panchayat</td>
<td>land given by a Gram Panchayat</td>
</tr>
<tr>
<td>private-land</td>
<td>privately owned land</td>
</tr>
<tr>
<td>other</td>
<td>any other form of land ownership</td>
</tr>
</tbody>
</table>

The ownership of the school building itself can similarly be of different kinds. This is recorded in a building-ownership element.

The ownership attribute of this element records the kind of ownership of the school buildings.

A sample of this section of the school data file is as follows:

```xml
<building-ownership ownership="own" />
<building-site-details type="urban-development-authority" />
```
9.2  Verandah / Corridor

The survey form asks whether the school has a verandah or corridor for use by people in the school.

XXX: any reason why verandahs and corridors are significant?

We model the presence or absence of a verandah or corridor in the school building by an empty element, has-verandah-corridor.

\[
\langle \text{dtd: has verandah or corridor} \rangle \equiv (60)
\]

\[
\langle !\text{ELEMENT has-verandah-or-corridor} \text{ EMPTY} >
\langle !\text{ATTLIST has-verandah-or-corridor} \text{ dtd: standard attributes} >
\]

This presence of this attribute as a child element of the parent infrastructure element indicates that the school has at least one verandah or corridor.

\[
\langle \text{sample: infrastructure has verandah or corridor} \rangle \equiv (62)
\]

\[
\langle \text{has-verandah-or-corridor} / \rangle
\]

9.3  Rooms

The survey form has space to record information about the rooms in the school building. In the school data file, this information is kept inside a rooms element.

The rooms element records whether the school has a staff-room, a room for the headmaster using attributes has-staff-room and has-hm-room respectively. These attributes can have value yes, indicating that the room is present, or no indicating that the room is not present. If the survey form does not have any clear indication of whether a staff room or headmaster’s room is present, then this attribute should be left unspecified.

\[
\langle \text{dtd: rooms} \rangle \equiv (62a)
\]

\[
\langle \text{dtd: room info} \rangle
\]

\[
\langle !\text{ELEMENT rooms} (\text{room-info}+) >
\langle !\text{ATTLIST rooms} \text{ dtd: standard attributes} >
\quad \text{has-staff-room} (\langle \text{dtd: yes or no} \rangle) \ #IMPLIED
\quad \text{has-hm-room} (\langle \text{dtd: yes or no} \rangle) \ #IMPLIED>
\]

9.3.1  Room information

Information collected for each room in the school is recorded in a room-info element, one element per room.

\[
\langle \text{dtd: room info} \rangle \equiv (62b)
\]

\[
\langle !\text{ELEMENT room-info} (\text{dimension}*) >
\langle !\text{ATTLIST room-info} \text{ dtd: standard attributes} >
\quad \text{roofing-type} (\langle \text{thatched | tiled | asbestos | RCC | other} \rangle) \ #IMPLIED
\quad \text{year-built} (\langle \text{dtd: date type} \rangle) \ #IMPLIED
\quad \text{constructed-by} (\langle \text{dtd: constructed by} \rangle) \ #IMPLIED
\quad \text{condition} (\langle \text{dtd: quality judgement} \rangle) \ #IMPLIED>
\]
9.4 Water

Water facilities available in the school are recorded in a `water` element in the school data file. This element in turn is comprised of zero or more sub-elements `water-source`.

```
<rooms has-staff-room="yes" has-hm-room="yes">
  <room-info roofing-type="RCC" condition="ok" constructed-by="land-army"
    year-built="1989">
    <dimension unit="ft">10</dimension>
    <dimension unit="ft">15</dimension>
  </room-info>
</rooms>
```

The attributes of the `room-info` element are given in Table 9.2.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>condition</td>
<td>the current condition of the room</td>
</tr>
<tr>
<td>constructed-by</td>
<td>who built the room</td>
</tr>
<tr>
<td>year-built</td>
<td>the year in which the room was built</td>
</tr>
<tr>
<td>roofing-type</td>
<td>the type of roofing for the room</td>
</tr>
</tbody>
</table>

If any of these attributes are not known, they should be left unspecified in the school data file.

Sample data in the school data file would look like:

```
<room has-staff-room="yes" has-hm-room="yes">
  <room-info roofing-type="RCC" condition="ok" constructed-by="land-army"
    year-built="1989">
    <dimension unit="ft">10</dimension>
    <dimension unit="ft">15</dimension>
  </room-info>
</room>
```

9.4 Water

Water facilities available in the school are recorded in a `water` element in the school data file. This element in turn is comprised of zero or more sub-elements `water-source`.

```
<!ELEMENT water-source EMPTY>
<!ATTLIST water-source
  capacity (idd: number type)  #IMPLIED
  is-potable ((idd: yes or no))  #REQUIRED
  source ((idd: water source))  #IMPLIED
  storage ((idd: water storage))  #IMPLIED
  who-fills ((idd: helpers))  #IMPLIED>
```

A school can have zero or more sources for water. Each water source is modelled by a `water-source` element.

The attributes of the `water-source` element are described in Table 9.3.

Sources for water could be:

```
municipal-supply | open-well | river-or-tank | bore-well | other
```
### Table 9.3 Attributes of element water-source

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>capacity</td>
<td>storage capacity in litres</td>
</tr>
<tr>
<td>is-potable</td>
<td>if yes, then water is drinkable</td>
</tr>
<tr>
<td>source</td>
<td>source of this water</td>
</tr>
<tr>
<td>storage</td>
<td>where the water is stored</td>
</tr>
<tr>
<td>who-fills</td>
<td>role of the person in-charge of providing this water</td>
</tr>
</tbody>
</table>

### Table 9.4 Water sources

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>bore-well</td>
<td>from a bore well in the school</td>
</tr>
<tr>
<td>municipal-supply</td>
<td>supplied by the municipality or gram panchayat</td>
</tr>
<tr>
<td>open-well</td>
<td>from an open well in the school</td>
</tr>
<tr>
<td>river-or-tank</td>
<td>from a river or nearby tank</td>
</tr>
<tr>
<td>other</td>
<td>any other source</td>
</tr>
</tbody>
</table>

The meanings of these symbols are given in table 9.4.

We distinguish between the following kinds of water storage facilities:

\[
\text{tap | steel-vessel | drum | pot | tank | other}
\]

The usage of these symbols is described in table 9.5.

### Table 9.5 Kinds of water storage

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>tap</td>
<td>no storage, but tap-water is readily available</td>
</tr>
<tr>
<td>steel-vessel</td>
<td>water is stored in steel (kitchen) vessels</td>
</tr>
<tr>
<td>drum</td>
<td>water is stored in drums</td>
</tr>
<tr>
<td>pot</td>
<td>water is stored in earthen pots</td>
</tr>
<tr>
<td>tank</td>
<td>water is stored in a tank</td>
</tr>
<tr>
<td>other</td>
<td>some other kind of storage</td>
</tr>
</tbody>
</table>

Sample data in a school data file would look like:

\[
\text{tap} | steel-vessel | drum | pot | tank | other
\]
9.5 Toilet

Toilet facilities available in the school premises are recorded using the `toilet-facilities` element. This element comprises of one or more `toilet` elements, with one `toilet` element holding information for one toilet in the school. If the school has no toilet facilities whatsoever, then the `toilet-facilities` element will not be present in the school data file.

```xml
<!ELEMENT toilet-facilities (toilet+)>
<!ATTLIST toilet-facilities
  ⟨dtd: standard attributes⟩
  user (teachers | boys | girls | others | common) "common" #IMPLIED
  year-built ⟨dtd: date type⟩ #IMPLIED
  constructed-by ⟨dtd: constructed by⟩ #IMPLIED
  has-water ⟨dtd: yes or no⟩ #IMPLIED
  condition ⟨dtd: quality judgement⟩ #IMPLIED
  cleaner ⟨dtd: helpers⟩ #IMPLIED>
```

Each toilet is represented by an element with no children, with the information collected in the survey form being recorded as values of attributes.

```xml
<!ELEMENT toilet EMPTY>
<!ATTLIST toilet
  ⟨dtd: standard attributes⟩
  user (teachers | boys | girls | others | common) "common"
  year-built ⟨dtd: date type⟩ #IMPLIED
  constructed-by ⟨dtd: constructed by⟩ #IMPLIED
  has-water ⟨dtd: yes or no⟩ #IMPLIED
  condition ⟨dtd: quality judgement⟩ #IMPLIED
  cleaner ⟨dtd: helpers⟩ #IMPLIED>
```

The attributes defined for this element are described in Table 9.6.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>constructed-by</td>
<td>who built the toilet</td>
</tr>
<tr>
<td>user</td>
<td>who uses this toilet</td>
</tr>
<tr>
<td>cleaner</td>
<td>who cleans the toilet</td>
</tr>
<tr>
<td>condition</td>
<td>the current condition of the toilet</td>
</tr>
<tr>
<td>year-built</td>
<td>the year during which the toilet was built</td>
</tr>
<tr>
<td>has-water</td>
<td>whether adequate water is available for use</td>
</tr>
</tbody>
</table>

Sample data for toilets would look like:

```xml
<toilet-facilities>
  <toilet user="teachers" year-built="1979" constructed-by="zp-contractor" has-water="no" condition="very-bad" cleaner="paid-help" />
</toilet-facilities>
```
9.6 Compound Wall

The presence of a compound wall separating the school premises from the rest of community is modeled by a compound-wall element. This element does not have any further content; its attribute record information about the wall.

\[
\text{<compound-wall} \quad (66a) \equiv (66b)
\]

\[
\text{<!ELEMENT compound-wall EMPTY>}
\]

\[
\text{<!ATTLIST compound-wall}
\]

\[
\text{wall-type (66c) \quad #IMPLIED}
\]

\[
\text{year-built (66d) \quad #IMPLIED}
\]

\[
\text{constructed-by (66e) \quad #IMPLIED}
\]

\[
\text{condition (66f) \quad #IMPLIED>}
\]

The attributes for this element are described in Table 9.7.

Table 9.7 Attributes of element compound-wall

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>wall-type</td>
<td>kinds of walls described in Table 9.8</td>
</tr>
<tr>
<td>year-built</td>
<td>the year in which the wall was constructed</td>
</tr>
<tr>
<td>constructed-by</td>
<td>who constructed the wall</td>
</tr>
<tr>
<td>condition</td>
<td>the condition of the wall</td>
</tr>
</tbody>
</table>

The kinds of walls we recognize are (see Table 9.8):

\[
\text{<wall-types} \quad (66g) \equiv (66h)
\]

\[
\text{fence | stone-wall | brick-wall | other}
\]

Table 9.8 Kinds of compound walls

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>fence</td>
<td>a fence of some kind</td>
</tr>
<tr>
<td>stone-wall</td>
<td>a wall made of stone</td>
</tr>
<tr>
<td>brick-wall</td>
<td>a wall made of brick</td>
</tr>
<tr>
<td>other</td>
<td>any other kind of wall</td>
</tr>
</tbody>
</table>

Sample data in the school information file would look like:

\[
\text{<compound-wall} \quad (66i) \equiv (66j)
\]

\[
\text{wall-type="fence" year-built="1991"}
\]

\[
\text{condition="good"} />
\]

9.7 Flooring

Data on the kinds of flooring present in the school is recorded in a flooring element. This element has no content; information collected in the survey form is recorded in its attributes (Table 9.9).

\[
\text{<flooring} \quad (66k) \equiv (66l)
\]

\[
\text{<!ELEMENT flooring EMPTY>}
\]

\[
\text{<!ATTLIST flooring}
\]

\[
\text{flooring (66m) \quad (66n) \quad #IMPLIED>}
\]

\[
\text{<!ATTLIST flooring}
\]

\[
\text{flooring (66o) \quad (66p) \quad #IMPLIED>}
\]

\[
\text{<!ATTLIST flooring}
\]

\[
\text{flooring (66q) \quad (66r) \quad #IMPLIED>}
\]
9.8. **ELECTRICITY**

<table>
<thead>
<tr>
<th>Table 9.9 Attributes of element flooring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribute</td>
</tr>
<tr>
<td>flooring</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(dtd: flooring choices</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>mud</td>
<td>cement</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 9.10 Kinds of flooring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbol</td>
</tr>
<tr>
<td>mud</td>
</tr>
<tr>
<td>cement</td>
</tr>
<tr>
<td>red-oxide</td>
</tr>
<tr>
<td>tile</td>
</tr>
<tr>
<td>stone</td>
</tr>
<tr>
<td>other</td>
</tr>
</tbody>
</table>

Sample data in the school data file could look like:

```
<flooring flooring="mud" />
```

9.8 **Electricity**

We record whether the school has electricity using a *has-electricity* element. The presence of this element signals that the school has been electrified. The attributes of this element are described in Table 9.11.

<table>
<thead>
<tr>
<th>(dtd: has electricity</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;!ELEMENT has-electricity EMPTY&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;!ATTLIST has-electricity source (self-generated</td>
<td>state-electricity-board</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 9.11 Attributes of element has-electricity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribute</td>
</tr>
<tr>
<td>source</td>
</tr>
<tr>
<td>quality</td>
</tr>
</tbody>
</table>

Sample school data could look like:

```
<has-electricity source="private-producer" quality="ok" />```
CHAPTER 9. INFRASTRUCTURE

9.9 Furniture

The furniture in use by the school is recorded in a furniture-list element. This element comprises of one or more furniture elements.

Each item of furniture in the school is recorded using an empty furniture element. The attributes of this element (Table 9.12) store the information in the survey form. There is to be one furniture element in the school data file for each kind of furniture present.

The kinds of furniture include:

Sample data could look like:

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>kind</td>
<td>specifies the kind of furniture item, see Table 9.13</td>
</tr>
<tr>
<td>sufficient-quantity</td>
<td>whether there are enough items of this kind of furniture for all users</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>table</td>
<td>tables</td>
</tr>
<tr>
<td>chair</td>
<td>chairs</td>
</tr>
<tr>
<td>lockable-storage</td>
<td>lockable storage space</td>
</tr>
<tr>
<td>low-bench</td>
<td>low benches used by children</td>
</tr>
<tr>
<td>bench-desk</td>
<td>bench and desk combination for students</td>
</tr>
<tr>
<td>mats-carpets</td>
<td>mats and carpets for children to sit on</td>
</tr>
<tr>
<td>other</td>
<td>other kinds of furniture</td>
</tr>
</tbody>
</table>

Sample data could look like:

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>kind</td>
<td>specifies the kind of furniture item, see Table 9.13</td>
</tr>
<tr>
<td>sufficient-quantity</td>
<td>whether there are enough items of this kind of furniture for all users</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>table</td>
<td>tables</td>
</tr>
<tr>
<td>chair</td>
<td>chairs</td>
</tr>
<tr>
<td>lockable-storage</td>
<td>lockable storage space</td>
</tr>
<tr>
<td>low-bench</td>
<td>low benches used by children</td>
</tr>
<tr>
<td>bench-desk</td>
<td>bench and desk combination for students</td>
</tr>
<tr>
<td>mats-carpets</td>
<td>mats and carpets for children to sit on</td>
</tr>
<tr>
<td>other</td>
<td>other kinds of furniture</td>
</tr>
</tbody>
</table>

Sample data could look like:

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>kind</td>
<td>specifies the kind of furniture item, see Table 9.13</td>
</tr>
<tr>
<td>sufficient-quantity</td>
<td>whether there are enough items of this kind of furniture for all users</td>
</tr>
</tbody>
</table>
9.10 Teaching Material

The kinds of teaching material present in school is recorded in a teaching-material-list element. This element comprises of one or more teaching-material elements.

```xml
<!ELEMENT teaching-material-list (teaching-material+)>
<!ATTLIST teaching-material-list
  ⟨dtd: standard attributes⟩>
```

The school data file will have one teaching-material element for each kind of teaching material present. teaching-material elements have a single optional `reason` which records the reason (if any) for the material not being in use. Other attributes of this element (Table 9.14) contain the rest of the information collected in the survey form.

```xml
<!ELEMENT teaching-material (reason?)>
<!ATTLIST teaching-material
  ⟨dtd: standard attributes⟩
    kind ((⟨dtd: teaching material choices⟩)) #REQUIRED
    n-total (⟨dtd: number type⟩) #IMPLIED
    n-useable (⟨dtd: number type⟩) #IMPLIED
    n-in-use (⟨dtd: number type⟩) #IMPLIED
    have-storage ((⟨dtd: yes or no⟩)) #IMPLIED
    language CDATA #IMPLIED>
```

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>kind</td>
<td>kind of teaching material as defined in Table 9.15</td>
</tr>
<tr>
<td>n-total</td>
<td>total number of items of this kind</td>
</tr>
<tr>
<td>n-useable</td>
<td>number of items in useable condition</td>
</tr>
<tr>
<td>n-in-use</td>
<td>number of items of this kind actually in use</td>
</tr>
<tr>
<td>have-storage</td>
<td>the items have storage space</td>
</tr>
<tr>
<td>language</td>
<td>language of the material if applicable</td>
</tr>
</tbody>
</table>

Kinds of teaching material recognized are (Table 9.15):

```xml
<teaching-material-list>
<teaching-material kind="computer" n-total="1" n-in-use="0" have-storage="no">
  <reason>No qualified teacher to handle this equipment</reason>
</teaching-material>
<teaching-material kind="blackboard" n-total="3" n-useable="2" n-in-use="2" />
</teaching-material-list>
```
Table 9.15 Kinds of teaching materials

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>blackboard</td>
<td>blackboards or equivalent</td>
</tr>
<tr>
<td>charts</td>
<td>subject specific charts</td>
</tr>
<tr>
<td>globe</td>
<td>a globe</td>
</tr>
<tr>
<td>maps</td>
<td>maps of the world, country or local regions</td>
</tr>
<tr>
<td>science-kit</td>
<td>a kit for doing science experiments</td>
</tr>
<tr>
<td>math-kit</td>
<td>material for learning mathematics</td>
</tr>
<tr>
<td>vcr-tv</td>
<td>a VCR or TV</td>
</tr>
<tr>
<td>tape-recorder</td>
<td>a tape recorder</td>
</tr>
<tr>
<td>laboratory</td>
<td>any kind of science laboratory</td>
</tr>
<tr>
<td>tools-equipment</td>
<td>other tools like carpentry tools</td>
</tr>
<tr>
<td>tailoring-machine</td>
<td>tailoring machines</td>
</tr>
<tr>
<td>computer</td>
<td>a computer of any kind for use by the children</td>
</tr>
<tr>
<td>other</td>
<td>other teaching material</td>
</tr>
</tbody>
</table>

9.11 Playground

We record whether the school has a playground using a playground element. The presence of this element indicates that the school has a playground. The dimensions of the playground if known are marked up using dimension sub-elements as children.

```xml
<dtd: playground>
  <!ELEMENT playground (dimension*)>
  <!ATTLIST playground (dtd: standard attributes)>
</dtd: playground>
```

Sample data could look like:

```xml
<playground>
  <dimension unit="ft">30</dimension>
  <dimension unit="ft">40</dimension>
</playground>
```

9.12 Sports Equipment

The sports equipment present in the school are recorded in a sports-equipment-list element. This element comprises of one or more sports-equipment elements.

```xml
<dtd: sports equipment>
  <!ELEMENT sports-equipment-list (sports-equipment*)>
  <!ATTLIST sports-equipment-list (dtd: standard attributes)>
</dtd: sports equipment>
```
9.12. SPORTS EQUIPMENT

We track the different kinds of sports equipments that are present using a `sports-equipment` element.

```xml
<!ELEMENT sports-equipment (reason?)>
<!ATTLIST sports-equipment
  kind ((dtd: sports equipment kinds) #REQUIRED)
  n-total (dtd: number type #IMPLIED)
  n-usable (dtd: number type #IMPLIED)
  n-in-use (dtd: number type #IMPLIED)
  has-storage ((dtd: yes or no) #IMPLIED)>
```

The attributes of this element are described in Table 9.16.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>kind</td>
<td>the kind of equipment, described in Table 9.17.</td>
</tr>
<tr>
<td>n-total</td>
<td>total number of items of this kind</td>
</tr>
<tr>
<td>n-usable</td>
<td>number of items in useable condition</td>
</tr>
<tr>
<td>n-in-use</td>
<td>number of items of this kind actually in use</td>
</tr>
<tr>
<td>have-storage</td>
<td>the items have storage space</td>
</tr>
</tbody>
</table>

The child `reason` element records the reason for the equipment not being in use.

The kinds of sports equipment that we track are:

- cricket-kit
- football
- volleyball
- lajim
- ring
- dumb-bells
- skipping-rope
- band-set
- musical-instruments
- other

These are described in Table 9.17.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>cricket-kit</td>
<td>balls, bats, wickets and other cricket items</td>
</tr>
<tr>
<td>football</td>
<td>football equipment</td>
</tr>
<tr>
<td>volleyball</td>
<td>balls and nets for volleyball</td>
</tr>
<tr>
<td>lajim</td>
<td>equipment for the game of lajim</td>
</tr>
<tr>
<td>ring</td>
<td>a tennicoit ring</td>
</tr>
<tr>
<td>dumb-bells</td>
<td>dumb-bells for weight lifting</td>
</tr>
<tr>
<td>skipping-rope</td>
<td>skipping ropes</td>
</tr>
<tr>
<td>band-set</td>
<td>a band set (drums, musical instruments)</td>
</tr>
<tr>
<td>musical-instruments</td>
<td>any musical instruments</td>
</tr>
<tr>
<td>other</td>
<td>any other kinds of sports equipment</td>
</tr>
</tbody>
</table>

XXX: Should musical instruments be here? These are not considered “sports”.

Sample XML data could look like:

```xml
<sports-equipment-list>
  <sports-equipment kind="cricket-kit" n-total="1" n-usable="1" has-storage="no" />
</sports-equipment-list>
```
9.13 Library

The library facilities available in the school are recorded in a library element.

We record the person or persons in-charge of the library using an in-charge attribute. Other than this, the library element comprises of one or more library-item elements, with one library-item element recording information on one kind of item in the library.

```
<library in-charge="teachers">
  <library-item kind="books" n-total="3" n-usable="0">
    <reason>lost key to cupboard</reason>
  </library-item>
</library>
```

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>in-charge</td>
<td>the person(s) in charge of the library (see 2.2.5)</td>
</tr>
</tbody>
</table>

Each library item is recorded using a library-item element (Table 9.19).

```
<library-item kind="books" n-total="3" n-usable="0">
  <reason>lost key to cupboard</reason>
</library-item>
```

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>kind</td>
<td>the kind of item, see Table 9.20</td>
</tr>
<tr>
<td>n-total</td>
<td>total number of items of this kind</td>
</tr>
<tr>
<td>n-usable</td>
<td>number of items in useable condition</td>
</tr>
<tr>
<td>n-in-use</td>
<td>number of items of this kind actually in use</td>
</tr>
<tr>
<td>where-stored</td>
<td>the location where the item is stored if known</td>
</tr>
</tbody>
</table>

XXX: where-stored should be a list of choices? Free form text is difficult to handle.

The kinds of library items tracked are (Table 9.20):

```
books | magazines | newspapers | other
```

Sample data on a library in a school could look like:

```
<library in-charge="teachers">
  <library-item kind="books" n-total="3" n-usable="0">
    <reason>lost key to cupboard</reason>
  </library-item>
</library>
```
9.14. MEDICAL FACILITIES

Table 9.20 Kinds of library items

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>books</td>
<td>printed and bound books</td>
</tr>
<tr>
<td>magazines</td>
<td>all kinds of periodicals</td>
</tr>
<tr>
<td>newspapers</td>
<td>newspapers in any language</td>
</tr>
<tr>
<td>other</td>
<td>anything else in the library</td>
</tr>
</tbody>
</table>

9.14 Medical Facilities

Medical facilities available at the school are tracked using a medical element (Table 9.21).

The medical element contains zero or more health-checkup elements that track when health checkups were conducted for the students. Each health-checkup element in-turn comprises of one or more class elements that record the classes that participated in the health checkup.

The list of organizations that conducted health checkups that we track in the form are (Table 9.23):

PHC | NGO | local-medical-college | clubs | health-dept | others
### Table 9.23 List of health checking organizations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHC</td>
<td>the primary health centre in the area</td>
</tr>
<tr>
<td>NGO</td>
<td>a non-government organization</td>
</tr>
<tr>
<td>local-medical-college</td>
<td>any local medical college</td>
</tr>
<tr>
<td>clubs</td>
<td>clubs like the Rotary or Lions clubs</td>
</tr>
<tr>
<td>health-dept</td>
<td>the state health department</td>
</tr>
<tr>
<td>others</td>
<td>anyone else</td>
</tr>
</tbody>
</table>

Sample data in the school info file could look like:

```xml
<medical has-first-aid-kit="no" checkup-frequency="monthly">
  <health-checkup date="1-19-1970" conducted-by="PHC">
    <class class="std-1" />
    <class class="std-2" />
    <class class="std-3" />
  </health-checkup>
</medical>
```

### 9.15 Location of classes

In some schools, classes may be held in open-air due to a shortage of classrooms. We record the location of classes in a `location-of-classes` element.

```xml
<!ELEMENT location-of-classes (class-location+)>
<!ATTLIST location-of-classes (did: standard attributes)>
```

Each `location-of-classes` element comprises one or more `class-location` subelements, one per class in the school. These are elements with no content, with attributes (Table 9.24) holding the information in the survey form.

```xml
<!ELEMENT class-location EMPTY>
<!ATTLIST class-location (did: standard attributes)>
```

#### Table 9.24 Attributes of element class-location

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>class</td>
<td>the class name, see 2.2.13</td>
</tr>
<tr>
<td>location</td>
<td>the location where class is usually held, see Table 9.25</td>
</tr>
<tr>
<td>teaching-style</td>
<td>the style of teaching employed, see Table 9.26</td>
</tr>
</tbody>
</table>

Tracked locations where classes are held include:

```xml
<did: location of classes>
  <class-location>
    under-a-tree | verandah | in-the-open | classroom | other
  </class-location>
</did: location of classes>
```
9.16 MAINTENANCE

The meanings of these symbols are given in Table 9.25.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>under-a-tree</td>
<td>under a tree</td>
</tr>
<tr>
<td>verandah</td>
<td>in the verandah of the school</td>
</tr>
<tr>
<td>in-the-open</td>
<td>anywhere else outside a building</td>
</tr>
<tr>
<td>classroom</td>
<td>in a designated classroom</td>
</tr>
<tr>
<td>other</td>
<td>any other place</td>
</tr>
</tbody>
</table>

Teaching styles tracked are recorded in Table 9.26.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>single-grade</td>
<td>only grade in this class</td>
</tr>
<tr>
<td>multi-grade</td>
<td>a combined class with children of different grades</td>
</tr>
</tbody>
</table>

Sample data in a school data file would look like:

```xml
<location-of-classes>
  <class-location class="std-1" teaching-style="single-grade"
    location="under-a-tree" />
</location-of-classes>
```

9.16 Maintenance

The degree of maintenance of the school premises is tracked in an maintenance element. This element contains attributes (Table 9.27) that track individual aspects of the schools maintenance.

Sample data in the school data file could look like:

```xml
<maintenance has-garden="no" has-trees="no"
  who-maintains-classrooms="paid-help" present-condition="bad" />"
CHAPTER 9. INFRASTRUCTURE

Table 9.27 Attributes of element maintenance

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>who-maintains-classrooms</td>
<td>maintainer of classrooms (see 2.2.5)</td>
</tr>
<tr>
<td>last-painted-date</td>
<td>date the school was last painted</td>
</tr>
<tr>
<td>present-condition</td>
<td>qualitative judgement of the schools current condition (see 2.2.16)</td>
</tr>
<tr>
<td>has-garden</td>
<td>whether the school has a garden</td>
</tr>
<tr>
<td>has-trees</td>
<td>whether the school has trees</td>
</tr>
<tr>
<td>who-maintains-garden-trees-Compound</td>
<td>maintainer of garden if any (see 2.2.5)</td>
</tr>
</tbody>
</table>

9.17 Environment

The environment in and around the school is recorded in an environment element. Subelements track specific kinds of pollution and health hazards in the vicinity of the school.

A hazard element tracks hazards in the vicinity of the school. Attributes of this element (Table 9.28) describe the kind of hazard. The optional comment element holds any additional comments about the hazard.

Table 9.28 Attributes of element hazard

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>proximity</td>
<td>proximity of the hazard to the school, see 2.2.14</td>
</tr>
<tr>
<td>kind</td>
<td>kind of hazard, see Table 9.29</td>
</tr>
<tr>
<td>cause</td>
<td>cause of hazard, see Table 9.30</td>
</tr>
</tbody>
</table>

The kinds of hazards tracked are:

industrial-pollution | noise-pollution | alcohol-shop | safety | health | other

Some causes of hazards that are tracked are:

textile-mill | highway | bus-stand | motor-repair-shop | market | place-of-worship | community-hall | theatre | hospital | other
### Table 9.29 Kinds of hazards

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>industrial-pollution</td>
<td>a source of industrial pollution</td>
</tr>
<tr>
<td>noise-pollution</td>
<td>a source of noise pollution</td>
</tr>
<tr>
<td>alchohol-shop</td>
<td>a shop selling consumable alchohol</td>
</tr>
<tr>
<td>safety</td>
<td>a safety hazard of any kind</td>
</tr>
<tr>
<td>health</td>
<td>any kind of health hazard (drains, garbage pits etc)</td>
</tr>
<tr>
<td>other</td>
<td>any other hazard</td>
</tr>
</tbody>
</table>

### Table 9.30 Causes of hazards

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>textile-mill</td>
<td>a textile mill</td>
</tr>
<tr>
<td>highway</td>
<td>highway or main road</td>
</tr>
<tr>
<td>bus-stand</td>
<td>a bus stand</td>
</tr>
<tr>
<td>motor-repair-shop</td>
<td>a motor mechanics shop</td>
</tr>
<tr>
<td>market</td>
<td>a market</td>
</tr>
<tr>
<td>place-of-worship</td>
<td>a temple, mosque or other place of worship</td>
</tr>
<tr>
<td>community-hall</td>
<td>any hall used by the community</td>
</tr>
<tr>
<td>theatre</td>
<td>a cinema hall or theatre</td>
</tr>
<tr>
<td>hospital</td>
<td>a hospital</td>
</tr>
<tr>
<td>other</td>
<td>any other cause of a hazard</td>
</tr>
</tbody>
</table>

XXX: is this really going to be tracked?

Sample data in the school data file could look like:

```xml
<environment>
  <hazard kind="industrial-pollution" proximity="m0-50">
    <comment>Sri Balaji Tannery</comment>
  </hazard>
  <hazard kind="noise-pollution" proximity="m0-50" cause="motor-repair-shop">
    <comment>Very noisy in the mornings</comment>
  </hazard>
  <hazard kind="safety">
    <comment>open drain without a proper bridge which children have to use</comment>
  </hazard>
  <hazard kind="alchohol-shop" proximity="m0-50" />
</environment>
```

### 9.18 Development

Additional development needs of the school are recorded in a development element. This element comprises of one or more development-item elements.
Each development-item element tracks one kind of development need. The attribute `kind` (Table 9.31) tracks the kinds of development items recorded.

```
<dtd: development> + ≡
  <!ELEMENT development-item (comment?, additional-classrooms?)>
  <!ATTLIST development-item
    (dtd: standard attributes)
    kind (dtd: development item kinds) #REQUIRED
    estimate (dtd: number type) #IMPLIED>
```

Table 9.31 Attributes of element `development-item`

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>kind</td>
<td>kinds of development items, see Table 9.32</td>
</tr>
<tr>
<td>estimate</td>
<td>the estimate of the cost of the development work in Rupees</td>
</tr>
</tbody>
</table>

The following kinds of development needs are tracked:

```
<dtd: development item kinds> ≡
  compound-wall | additional-classrooms | other-repairs
```

Table 9.32 Kinds of development items

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>compound-wall</td>
<td>the school needs a compound wall</td>
</tr>
<tr>
<td>additional-classrooms</td>
<td>the school needs more classrooms</td>
</tr>
<tr>
<td>other-repairs</td>
<td>any other repairs</td>
</tr>
</tbody>
</table>

This element can contain an optional comment, recorded in a `comment` child element which records further details about the development item. In case the development item is the need for additional classrooms, the `additional-classrooms` element records further details about the need for additional classrooms.

```
<dtd: development> + ≡
  <!ELEMENT additional-classrooms (sanctioning-scheme?)>
  <!ATTLIST additional-classrooms
    (dtd: standard attributes)
    n-required (dtd: number type) #REQUIRED
    n-under-construction (dtd: number type) #IMPLIED
    is-site-available (dtd: yes or no) #IMPLIED>
```
Table 9.33 Attributes of element additional-classrooms

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>n-required</td>
<td>how many classrooms are needed</td>
</tr>
<tr>
<td>n-under-construction</td>
<td>how many classrooms are currently under construction</td>
</tr>
<tr>
<td>is-site-available</td>
<td>if a site is available for new classrooms</td>
</tr>
</tbody>
</table>

If construction of classrooms has already commenced and if the money for this is being sanctioned by the government, the scheme under which the money is being disbursed is recorded in a sanctioning-scheme element.

Sample data in the school data file could look like:

```xml
<development>
    <development-item kind="compound-wall" estimate="25000">
        <comment>goats are getting into the garden</comment>
    </development-item>
    <development-item kind="additional-classrooms" estimate="200000">
        <comment>for the higher classes</comment>
        <additional-classrooms n-required="3" n-under-construction="1" is-site-available="yes">
            <sanctioning-scheme>name of the scheme here</sanctioning-scheme>
        </additional-classrooms>
    </development-item>
</development>
```

9.19 Previous Complaints

The history of prior complaints given to the authorities about this school are recorded in previous-complaints element. This element comprises of one or more complaint elements.

```
<complaint (comment, authority, result)>
    date (dd: date type) #REQUIRED>
```

Each complaint comprises a comment element that records the complaint, a authority element that records the authority to whom the complaint was made, and a result element (2.2.12) that records status of the complaint. Additional attributes (Table 9.34) record other information about the complaint.
Table 9.34 Attributes of element complaint

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>date</td>
<td>the date when the complaint was made</td>
</tr>
</tbody>
</table>

The authority element records the name of the authority to whom the complaint was made.

```xml
<!ELEMENT authority (#PCDATA)>
```

XXX: authority should be a list of fixed choices

Results and complaints can be linked using the id attribute of complaint and the issue attribute of element result.

Sample data in the school data file could look like:

```xml
<previous-complaints>
  <complaint date="1-1-1997" id="complaint123">
    <comment>school roof was stolen</comment>
    <authority>the grand panjandrum</authority>
    <result issue="complaint123" status="unresolved">grand panjandrum is never in office</result>
  </complaint>
</previous-complaints>
```
Chapter 10

Other Information

In this section, we keep track of other educational support activities like,

- whether local groups like the Lions Club or the Rotarians have adopted the school
- which government incentive schemes the school has been the beneficiary of
- other extra curricular activities held

Each of these is examined below.

Local group activities include those run by volunteer groups like the Lions Club and the Rotary club. Examples of such activities include text book distribution, providing snacks, holding supplementary classes and so on. Each such activity is tracked by a group-activity element in the XML file.

The government runs various programmes that target its schools. These include midday meal schemes, scholarships, distribution of free uniforms and the like. Each such scheme is tracked using a government-programme element.
Other extra-curricular activities are also tracked using `extra-curricular-activity` elements.

```xml
<extra-curricular-activities>
  <extra-curricular-activity activity='essay-competition' n-students='21'
    organized-by='school' academic-year='1999-2000'>
    <comment>Remarks on this extracurricular activity</comment>
  </extra-curricular-activity>
</extra-curricular-activities>
```

Educational support activities for the academically weak and disadvantaged classes are also recorded.

```xml
<educational-support-activities>
  <educational-support-activity n-students='21' academic-year='1999-2000'
    activity='supplementary-classes'>
    <comment>Some remarks</comment>
  </educational-support-activity>
</educational-support-activities>
```

### 10.1 DTD

Formally, the `other-information` element comprises of zero or more subelements `local-group-activities`, `government-programmes`, `extra-curricular-activities`, `educational-support-activities`, with each sub element recording one of the major subsections of the data collected.

```xml
<!ELEMENT other-information (local-group-activities*, government-programmes*,
  extra-curricular-activities*, educational-support-activities*)>
```

#### 10.1.1 Local Group Activities

A local group activity comprises one or more `group-activity` elements.

```xml
<!ELEMENT local-group-activity (name,comment?)>
```

```xml
<!ATTLIST local-group-activity
  area-aspect (area-aspect-type)
    #REQUIRED>
```
The `local-group-activity` element has two mandatory attributes. `academic-year` tracks the year when the activity happened, and `area-aspect` records what the activity targeted.

For now, we allow a free format field for the `area-aspect` attribute. This may change later as we gain experience with the form in the field.

10.1.2 Government Programmes

Each government programme is represented by a `government-programme` element.

The list of government schemes includes

```xml
<!ELEMENT government-schemes (midday-meal | uniform | textbooks | school-bags | notebooks | scholarship-sc | scholarship-st | rashtriya-gramena-school | merit | girls-attendance | other)>```
The meaning of these symbols is explained in Table 10.1.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>midday-meal</td>
<td>Midday meal schemes for children</td>
</tr>
<tr>
<td>uniform</td>
<td>Distribution of free uniforms</td>
</tr>
<tr>
<td>textbooks</td>
<td>Distribution of free or subsidized textbooks</td>
</tr>
<tr>
<td>school-bags</td>
<td>Distribution of free school bags</td>
</tr>
<tr>
<td>notebooks</td>
<td>Distribution of free or subsidized notebooks</td>
</tr>
<tr>
<td>scholarship-sc</td>
<td>Scholarship programmes for Scheduled Caste children</td>
</tr>
<tr>
<td>scholarship-st</td>
<td>Scholarship programmes for Scheduled Tribe children</td>
</tr>
<tr>
<td>rashtriya-grameena-school</td>
<td>Scholarship under the Rashtriya Grameena School programme</td>
</tr>
<tr>
<td>merit</td>
<td>Other merit scholarship</td>
</tr>
<tr>
<td>girls-attendance</td>
<td>Girls attendance scholarship</td>
</tr>
<tr>
<td>other</td>
<td>Any other incentive scheme</td>
</tr>
</tbody>
</table>

The \( n \)-girls and \( n \)-boys attributes count the number of girls and boys who benefited from the scheme. These attributes are optional, and if not present mean that the number of beneficiaries were unknown.

The element comment records a remark about the functioning of the scheme, if any.

The government-programmes element is a wrapper containing one or more government-programme elements.

### 10.1.3 Extra Curricular Activities

Extra curricular activities are tracked using extra-curricular-activity elements.

The mandatory attributes for this element are the academic-year attribute that records the academic year in which the activity happened and the kind of activity recorded in attribute activity. Who the organizers were (attribute organized-by), and the number of students who participated (attribute \( n \)-students) are optional attributes.

The list of extracurricular activities includes

- essay-competition
- sports-day
- music-competition
- craft-work
- dramatics
- annual-day
- field-visit
- picnic
- other
The list of organizers includes
\[\text{parents-and-community} | \text{sbc-vec} | \text{government-beo} | \text{school} | \text{other}\]
The explanations for the symbols are given in Table 10.2.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>parents-and-community</td>
<td>Parents and the community</td>
</tr>
<tr>
<td>sbc-vec</td>
<td>School Betterment committee or Village Education Committee</td>
</tr>
<tr>
<td>government-beo</td>
<td>Government / Block Education Officer</td>
</tr>
<tr>
<td>school</td>
<td>School</td>
</tr>
<tr>
<td>other</td>
<td>Any other person or group</td>
</tr>
</tbody>
</table>

The extra-curricular-activities element is a wrapper containing one or more extra-curricular-activity elements.

**10.1.4 Education Support Activities**

Each educational support activity is recorded in an educational-support-activity element in the school data file.

This element has two mandatory attributes, academic-year which records the academic year this element is for, and name which is the activity name.

Recognized kinds of educational support activities include:

- supplementary-classes | other

The educational-support-activities element is a wrapper containing one or more educational-support-activity elements.
Chapter 11

Neighbours

A major section of the CIEE survey concerned itself with the neighbouring institutions in the village.

```
<sample:neighbours>
  <neighbour-information>
    <neighbour>
      <name>Sri Sri Sri Sri Mookambikeshwara Agribusiness Institute</name>
      <address>
        <additional-text>Near Govt Ration Shop, Minanahalli</additional-text>
      </address>
      <reasons-for-preference>
        <preference-reason reason="better-sports-facilities"/>
      </reasons-for-preference>
    </neighbour>
  </neighbour-information>
</sample:neighbours>
```

The Neighbour Information section comprises of zero or more neighbour elements.

### 11.1 DTD

Each neighbour element records the name of the institution, the address and the reasons why this institution would be preferred over the government school being surveyed.

There can be more than one reason to prefer a neighbour over the current school, so in general we need a list of options.

```
<!ELEMENT preference-reason (preference-reason)>
<!ATTLIST preference-reason reason (academic-excellence | better-cocurricular-activities | better-sports-facilities | better-infrastructure | better-motivated-teachers | better-teaching | better-english-teaching | cheaper-overall | other) #REQUIRED>
```

87
A single reason of preference is represented by the empty element `preference-reason`. This element has a mandatory attribute `reason` that has to take a value that is one of the choice list given (the choice list should be self explanatory).

All the reasons for preferring an institution are collected in one `reasons-for-preference` element.

```
88a ⟨dtd: neighbour declaration⟩+≡(88b)
  <!ELEMENT reasons-for-preference (preference-reason+)>
  <!ATTLIST reasons-for-preference ⟨dtd: standard attributes⟩>
```

Each `neighbour` element has to have a name (`name`), an address (`address`) and a list of why it is preferred to the school being surveyed (`reasons-for-preference`).

```
88b ⟨dtd: neighbour declaration⟩+≡(88c)
  <!ELEMENT neighbour (name, address, reasons-for-preference)>
  <!ATTLIST neighbour ⟨dtd: standard attributes⟩>
```

Finally, the top-level `neighbour-information` element is defined as a list of one or more `neighbour` elements.

```
88c ⟨dtd: neighbour information⟩≡(88d)
  ⟨dtd: neighbour declaration⟩
  ⟨!ELEMENT neighbour-information (neighbour+)⟩
  ⟨!ATTLIST neighbour-information ⟨dtd: standard attributes⟩⟩
```
Chapter 12

Improvement Opinions

In the Improvement Opinions section, data is collected on what could be done to improve the school, from the perspective of the teachers, children, community members and so on.

This section has three major sections:

- an Improvement Table where people rank various improvement options in order of importance
- a list of past initiatives that were undertaken
- a childrens section where the children voice their opinion.

A sample section in the school information data file looks like the following:

```xml
<school-improvement-opinions>
  <improvement-opinion-table academic-year='1999-2000'>
    <improvement-opinion name='improve-water' person='teacher' rank='5' />
    <improvement-opinion name='improve-compound-wall' person='group-representative' rank='5' />
  </improvement-opinion-table>
  <past-initiatives>
    <past-initiative date='Jan 2, 2000' id='PI.1.2.2000'>
      <issue-description>Yamadooths haunting the classroom.</issue-description>
      <activities-undertaken>Asked them to help with bureaucracy in Zilla Parishad office.</activities-undertaken>
      <result status='unresolved' issue='PI.1.2.2000'>Lord Yama now worried about missing yamadooths.</result>
    </past-initiative>
  </past-initiatives>
  <childrens-input academic-year='1999-2000'>
    <childrens-improvement-opinion item='no-physical-punishment' rank='3' />
    <child class='std-1' sex='male'><name>Basavanna</name></child>
  </childrens-input>
</school-improvement-opinions>
```
CHAPTER 12. IMPROVEMENT OPINIONS

12.1 DTD

The School Improvement Opinions section is modeled by a school-improvement-opinions element. This element comprises of zero or more improvement-opinions-table elements one for each academic year, a possible list of past initiatives recorded in past-initiatives, and zero or more childrens-input elements that hold the childrens inputs on school improvement, one per academic year.

\[ \langle \text{dtd: school-improvement-opinions} \rangle \equiv \langle \text{dtd: improvement-opinion-table-declaration} \rangle \langle \text{dtd: past-initiatives-declaration} \rangle \langle \text{dtd: childrens-input-declaration} \rangle \]

\[ \langle \text{dtd: improvement-opinion-table-declaration} \rangle \equiv \langle \text{dtd: improvement-opinion} \rangle \]

\[ \langle \text{dtd: improvement-opinion} \rangle \equiv \langle \text{dtd: standard-attributes} \rangle \]

\[ \langle \text{dtd: standard-attributes} \rangle \equiv \]

\[ \langle \text{improve-water} \mid \text{improve-rooms} \mid \text{improve-compound-wall} \mid \text{improve-toilets} \mid \text{improve-playground} \mid \text{improve-teaching-learning-material} \mid \text{improve-timely-distribution-of-incentives} \mid \text{improve-teacher-attendance} \mid \text{improve-teacher-training} \mid \text{improve-english-teaching} \mid \text{improve-activity-based-teaching} \mid \text{improve-curriculum} \mid \text{improve-extra-curricular} \mid \text{improve-community-support} \mid \text{improve-childrens-attendance} \mid \text{improve-administration} \mid \text{other} \rangle \]

The meanings of the symbols should be self-explanatory.

The list of people who are being asked for their opinions include:

\[ \langle \text{dtd: improvement-opinion-people} \rangle \equiv \langle \text{gram-panchayat-member} \mid \text{community-leader} \mid \text{group-representative} \mid \text{vec-sbc-representative} \mid \text{teacher} \mid \text{other} \rangle \]

The meanings of these symbols too, should be self-explanatory.

Ranks of items are restricted to be in the range 1 \cdots 5. “1” means the most important and “5” means the least important.

\[ \langle \text{dtd: improvement-ranks} \rangle \equiv \langle 1 \mid 2 \mid 3 \mid 4 \mid 5 \rangle \]
The improvement-opinion-table element is a list of improvement-opinions.

\[
\langle \text{dtd: improvement opinion table declaration} \rangle + \equiv \langle \text{improvement-opinion} \rangle
\]

\[
\langle !\text{ELEMENT improvement-opinion-table (improvement-opinion+)} \rangle
\]

\[
\langle !\text{ATTLIST improvement-opinion-table (dtd: standard attributes) (dtd: mandatory academic year attribute)} \rangle
\]

Each improvement-opinion-table has a mandatory academic-year attribute which records when the survey was undertaken.

### 12.1.2 Past Initiatives

All past initiatives taken to improve the school are recorded in the element past-initiatives. This element is a collection of one or more past-initiative elements.

\[
\langle \text{dtd: past initiatives declaration} \rangle + \equiv \langle \text{past-initiative} \rangle
\]

\[
\langle !\text{ELEMENT past-initiative (issue-description?, activities-undertaken?, result?)} \rangle
\]

\[
\langle !\text{ATTLIST past-initiative (date) (#REQUIRED)} \rangle
\]

The past-initiative element has a mandatory attribute date, that tracks the date when the initiative was initiated. Three sub-elements record the description of the issue (issue-description), activities undertaken to resolve the issue (activities-undertaken) and the result (result).

\[
\langle \text{dtd: past initiative elements} \rangle + \equiv \langle \text{issue-description} (#PCDATA) \rangle
\]

\[
\langle !\text{ATTLIST issue-description (dtd: standard attributes)} \rangle
\]

\[
\langle !\text{ELEMENT activities-undertaken (#PCDATA)} \rangle
\]

\[
\langle !\text{ATTLIST activities-undertaken (dtd: standard attributes)} \rangle
\]

We allow arbitrary text in the issue-description and activities-undertaken elements—their content models are hence #PCDATA.

All past-initiative elements are enclosed in a past-initiatives wrapper.
12.1.3 Childrens Input

The survey also collects input from the school children using the school on what could be improved about the school. This input is modelled by childrens-input elements. There is one childrens-input per academic year and the element itself comprises one or more childrens-improvement-opinion elements followed by one or more child elements.

Each child is modelled as follows. A name element is mandatory and records the child’s name. Attribute sex tracks the sex of the child—it should be male, female or left unspecified if not known. The class attribute tracks the class the child is attending. The date-of-birth attribute records the child’s date of birth.

The childrens-improvement-opinion elements are EMPTY: all the information they record are in the form of attributes.

The list of improvement options present to children for them to choose from are:

- improve-water
- improve-infrastructure
- improve-playground
- improve-toilets
- no-physical-punishment
- improve-activity-based-teaching
- improve-teachers-attendance
- improve-teaching-learning-material
- improve-curriculum
- improve-teacher-training
- improve-timely-distribution-of-incentives
- improve-extra-curricular
- other
Chapter 13

School Pictures

Part of the survey involves taking actual photographs of the school premises. We would like to make this data also available for processing and presentation.

Photographic data associated with a given school are named in the school data file. It is the processing applications responsibility to ensure that the named files actually exist and are useable.

In the school data file the primary data about a photographic image recorded are:

- the system name (filename) associated with the picture
- the format of the photographic data
- the view of the photograph
- and the date of the photograph

13.1 DTD

Formally, a school picture section comprises zero or more school-picture elements wrapped in a school-pictures wrapper.

```
<!ELEMENT school-pictures (school-picture*)>
<!ATTLIST school-pictures
    ⟨dtd: standard attributes⟩>
```

The top-level school-picture element has no special attributes of its own.

Each picture is described in a school-picture element (Table 13.1).

```
<!ELEMENT school-picture (comment?, photographer?)>
<!ATTLIST school-picture
    ⟨dtd: standard attributes⟩
    view (front | side | classroom | other) #REQUIRED
    systemname CDATA #REQUIRED
    systemtype NOTATION ⟨(dtd:picture-notations)⟩ #REQUIRED
    date ⟨dtd: date type⟩ #IMPLIED>
```
The element has two optional child elements, comment and photographer, that record additional information about the picture.

Element comment holds any additional comment about the picture. Element photographer (see 13.1.2), if present, records information about the photographer of this picture.

Table 13.1 Attributes of element school-picture

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>view</td>
<td>one of the list in Table 13.2</td>
</tr>
<tr>
<td>systemname</td>
<td>a string specifying the location of the photograph on the system</td>
</tr>
<tr>
<td>systemtype</td>
<td>one of the recognized processing image types as described in Table 13.3</td>
</tr>
<tr>
<td>data</td>
<td>the date and time at which the photograph was taken</td>
</tr>
</tbody>
</table>

Pictures are classified according to their view of the school, as in Table 13.2.

Table 13.2 Views of the school

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>front</td>
<td>a view facing the entrance of the school</td>
</tr>
<tr>
<td>side</td>
<td>a side view</td>
</tr>
<tr>
<td>back</td>
<td>a view of the rear of the school</td>
</tr>
<tr>
<td>classroom</td>
<td>a photograph of the interior of a classroom</td>
</tr>
<tr>
<td>interior</td>
<td>a photograph taken inside the school building</td>
</tr>
<tr>
<td>playground</td>
<td>a view of the playground</td>
</tr>
<tr>
<td>other</td>
<td>any other view of the school</td>
</tr>
</tbody>
</table>

The location of the photograph on the system is given in attribute systemname. This is stored as an unstructured string (CDATA, in XML terminology). The interpretation of the contents of this string is not specified by this DTD.

The date on which the picture was taken is recorded in the date attribute. The date attribute is optional. The content of the element is a textual field that can hold any additional comment about the picture.

13.1.1 Known Picture formats

Digitally, pictures can be stored in a number of formats, but our processing software may only be able to handle a few of the possibilities. We can specify restrictions on the use of picture formats in the DTD by restricting the kinds of pictures to one of a limited set.

\[
\langle \texttt{dtd:picture-notations} \rangle \equiv \texttt{png} \mid \texttt{gif} \mid \texttt{jpeg} \]
By declaring the `systemtype` attribute of the `school-picture` element to have type `NOTATION`, the XML parser can ensure that valid school data files only refer to recognized file formats.

The recognized formats are specified in Table 13.3.

We also need to declare these notations to the XML parser. Consequently, the overall DTD file has notation declarations for each of the allowed formats.

13.1.2 Photographer

In addition to the details of the picture, if the identity of the photographer is known, we record the name and role of the person who took the photograph in a `photographer` element.

The `role` attribute of the `photographer` element tracks whether the person was a volunteer, partner or staff member (Table 2.11).

13.2 Sample

Sample data in the school data file could look like:
Chapter 14

Copyright and License

This DTD, and its accompanying documentation are being distributed under a BSD-style license. The copyright and license are described below:

Copyright 2001 A. Joseph Koshy. All rights reserved.

Redistribution and use in source and derived forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
2. Redistributions in derived forms must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
3. Neither the name of A. Joseph Koshy nor the names of other contributors may be used to endorse or promote products derived from this software without specific prior written permission.

DISCLAIMER:

THIS SOFTWARE IS PROVIDED BY A JOSEPH KOSHY AND CONTRIBUTORS ‘‘AS IS’’ AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE A JOSEPH KOSHY OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.
Part II

Appendices
Appendix A

Glossary

A glossary of terms used in this document.

**DTD Document Type Definition.** A precisely defined specification of the form and contents of a data file.
Appendix B

An XML Primer

B.1 What is XML?

XML stands for eXtensible Markup Language.

B.2 What is a DTD?

A DTD is a Document Type Definition, a formal way of describing the structure of a document. By “formal” we mean that the structure of a document is described in language with precisely defined structure and semantics, in a computer oriented language in other words. This structure definition language allows you to clearly specify the structure of a document (e.g. what elements have to be present in the document, in what order, allowed values of elements etc.).

This precise definition allows us to easily write computer programs to:

- check if any given document conforms to the structure laid down in the DTD
- derive user interface programs that can take input from users conformant to the structure specified in the DTD
- write programs to transform the data in the document into other forms (for example, upload the data in the document into a database).

The formal language used to specify DTD is however complex. In this document we will try to explain the rationale behind the DTD specification side by side with the actual formal constructs that make up the specification.
Appendix C

Rationale

Data collected in the CIEE project is expected to have a useful life of a few decades at the least. Only a handful of software companies have lived this long. Few software packages have lasted out even a single decade.

We need to keep school data in a form that can survive changes in:

**storage media**. Twenty years ago, magnetic tapes were the norm. Will hard-disks and CD-Rs be around twenty years from now?

**operating systems** Operating systems popular twenty years ago are today being treated as historical relics.

**data formats** Along with changes in hardware and operating systems, the storage formats used by the older computers are now inaccessible.

**operational modes** Today most computers work in client-server mode. A decade earlier, large, centralized mainframes ruled computing.

In the past, humans were the primary consumers of processed data. Tomorrow, machines will be increasingly consuming, pre-processing and condensing data.

Next, we need to be prepared for the data being collected on Karnataka’s schools to be used in unexpected ways in the future. The easier we can make the process of extracting data from the survey results the better.

In order to ensure that the CIEE project is viable even after a decade, we need to take at least the following steps:

- we need to completely avoid the use of proprietary tools and data formats to represent our data.
  
  This is necessary because we should not put ourselves in the position of having a lifetimes data jeopardized because a company decided to discontinue support for its product.

- our data needs to be stored in a format that is close to being self-describing.
  
  A formally described, well understood data format can be adapted to technology trends.

- we need to conform to published ISO and other international standards on data representation and management to the extent possible.
Applications conforming to published standards are likely to survive technology and platform changes. Further, in the long run we are more likely to find people (volunteers, developers, reviewers, whoever) who are familiar with standards compliant environments.

C.1 Data description using SGML

Standard Generalized Markup Language (SGML) standard [ISO 8879:1986] addresses many of these issues. SGML provides a way of describing the structure of text as a formal model. Doing so allows use to hand off to the computer many tasks like

- verifying if a given file of data is complete and consistent
- transforming a given file from one format to another
- extracting, inserting or editing select elements in the file

SGML works by adding markup to the data content being represented in the file. In particular the markup added is generic markup, in that it specifies the logical structure of the data in the file. SGML allows us to completely abstract out many implementation dependent details.

SGML has been around for a long time, over twenty years. It has been widely deployed in industries where there is a need for long-lived data. It is well understood and there are a number of non-proprietary tools that can process data represented in SGML format.

We will be representing data collected as part of the CIEE project in the form of XML files. XML is a simplified variant of SGML, designed to be light-weight and easily to write processing software for.

The choice of XML offers us the following benefits:

- adherence to a well-understood, popular and simple to understand standard
- hardware and operating system independence
- a rich selection of free software tools to choose from
- availability of developer talent
- a high degree of future-proofing for our data

The downsides of using XML are:

- is that it is still new compared to SGML, less than a decade old when this document was created
- it is not, at this time, a formal, internationally sanctioned standard like SGML

Nevertheless, the simplicity of XML and its near universal acceptance today, coupled with its SGML roots make it an attractive method of representing CIEE school data.
C.2 Design Decisions

This section explains the design decisions taken at the system architecture level. The major design decisions are:

- one file per school, containing the data for multiple years
- school data will be kept as Unicode encoded text, initially in UTF-8 format
- file structure defined formally using an XML DTD
- automated tool to enter or edit school data

In the rest of the sections, we explain these design decisions.

C.2.1 Design: One File Per School

Filled-in survey forms are structured as one form per school. Updates to this data, whenever they happen, will also done on a per-school basis.

Keeping all the information about a school in one logical unit, that is edited and updated as the years go by seemed the easiest way to manage. Further, with a single logical unit (i.e. one file) per school, we can support “standalone” data entry for the survey data.

C.2.2 Design: Unicode encoding

The reasons for deciding on Unicode as the character encoding were:

- transparently handles nearly every living language in the world, including Kannada which we are initially interested in
- makes the project reusable in other contexts
- already supported by XML

The alternative was to mandate the use of the ASCII character encoding only with adhoc transliteration of local language items to Roman script. This approach would suffer from lack of precision (non-English text could be transliterated in many ways by the authors, leading to the useability of the stored data being reduced). Transliteration to Roman script also makes further processing of the data more complex.

The downside: full fledged Unicode support is complex to implement and the technology available in the open-source world to process Unicode is still in its infancy.

C.2.3 Design: Formal file structure expressed in XML

Any program that processes school data will need to have its input and output formats clearly defined. With an ad-hoc data format, we will need to write programs that read and write to these formats. Writing programs is an error-prone process. By specifying the data format in the formal language used to define an XML DTD, we gain in many ways:

- parsers to parse the data file are available with extra effort. We can concentrate on what we want to do with the data.
- there are many open-source libraries and utilities available that can parse XML
C.2.4 Design: Data Entry Tool

An automated entry/editing tool is planned so that data entry is easy for unskilled operators. While XML can be edited as plain text, this adds an extra cognitive load on the operators. An automated tool also allows “standalone” data entry.

C.3 Alternative approaches

Various alternate approaches were considered for representing school data. These are briefly discussed below.

C.3.1 Alternate: Use a database

Instead of keeping data in text files, keep it directly in a database, perhaps a relational (or object-relational) database.

While good relational database software is available as open-source, I felt that there were various problems with this approach that would reduce its effectiveness:

- Data models are fragile. The data being collected from each school has changing in each revision of the survey form. With each such change, we would need to change the data model inside the database. While such changes are feasible, but are tricky to carry out.

- Data gets distributed inside the database (depending on the way the database data model is organized). Entering and updating data for a given school will involve querying and updating a number of database tables—thus tools which manipulate the database become complex.

- Vendor independence is lost, since we are bound to the workings of the database system.

- Standalone data entry is difficult, since we need a connection to the database. In order to make standalone data entry work we need a formally defined text file format that could later be fed into the database.

- Backups strategies become more complex than archiving text files.

Placing data in a relational database helps in efficiently searching and querying for data. Using the database as the primary data store is what I felt was a bad idea.

C.3.2 Alternate: Use plain text files

Another alternate considered was to use plain text files to hold data avoiding the use of XML or other such structured formats. This approach while it looks simpler, doesn’t scale.

- Some kind of structure is needed even for these “plain text” files, so we will end up writing parsers and other file manipulation tools for our ad-hoc format. This is extra work that is automatically done for us when using XML.

1 Someone suggested a comma separated value format!
C.3. ALTERNATIVE APPROACHES

• Handling changes to the structure of the “plain text” file becomes complex—for example, when a newer generation of survey form adds (or deletes) collected data.

• searching and otherwise processing this data is difficult due to its lack of formal structure.

Using XML structured text files gives us a good balance between simplicity and functionality.
Index

C.4 Chunk Index

(a warning about generated files 23d)
(ciedb.dtd 23d)
(copyright and license 7)
(dd: absenteeism reasons 60d)
(dd: address declaration 22d)
(dd: admin information 34d)
(dd: admin work 51d)
(dd: admin work kind 51h)
(dd: anganwadi 34c)
(dd: anganwadi list element 34b)
(dd: area aspect type 83a)
(dd: building ownership / site 61a)
(dd: checkup conducted by 73c)
(dd: child declaration 92b)
(dd: childrens improvement items 92d)
(dd: childrens improvement opinion 92c)
(dd: childrens input declaration 92a)
(dd: class academic result 45b)
(dd: class list 49c)
(dd: class location 74d)
(dd: classes 17b)
(dd: comment declaration 54d)
(dd: communities element 33a)
(dd: community information element 30a)
(dd: community names 33b)
(dd: compound wall 66d)
(dd: constructed by 62a)
(dd: date type 15d)
(dd: designation 21d)
(dd: development 77d)
(dd: development item kinds 78a)
(dd: dimension element 83d)
(dd: distance options 87d)
(dd: educational activities 37c)
(dd: educational institutions 56d)
(dd: educational institutions element 56a)
(dd: educational support activities declaration 55d)
(dd: educational support activity 55d)
(dd: environment 76d)
(dd: extra curricular activities declaration 54b)
(dd: extra curricular activity 54d)
(dd: facility kind 56c)
(dd: flooring 66d)
(dd: flooring choices 67a)
(dd: form information element 25)
(dd: form sub elements 56d)
(dd: frequency 55d)
(dd: furniture kinds 66d)
(dd: furniture list 28d)
(dd: government programmes declaration 53d)
(dd: government schemes 53d)
(dd: has electricity 57d)
(dd: has verandah or corridor 52d)
(dd: hazard causes 56d)
(dd: hazard kinds 56c)
(dd: helpers 14b)
(dd: improvement item names 50d)
(dd: improvement opinion people 50d)
(dd: improvement opinion table declaration 50b)
(dd: improvement opinions 50b)
(dd: improvement ranks 50d)
(dd: incentive program info 50d)
(dd: infrastructure 54)
(dd: issue element 56a)
(dd: kinds of library items 22d)
(dd: library 22d)
C.5. IDENTIFIER INDEX

- activities-undertaken: 89, 71b, 91c
- additional-classrooms: 78a, 78b, 78c, 79b
- additional-text: 12c, 13, 14a, 39, 87a
- address: 12c, 14a, 39, 41b, 42b, 87a, 88b
- admin-work: 77, 88c, 88d
- admin-work-element: 77, 77b
- anganwadi: 50b, 51b, 51c, 51d
- anganwadi-list: 50b, 51b, 51c
- authority: 61a, 61c, 79d, 80a
- BEO: 51b, 53, 53c, 53d
- building-ownership: 50, 51a, 51c
- building-site-details: 50, 51a, 51c
- child: 89, 92a, 92c
- childrens-improvement-opinion: 89, 92a, 92c
- childrens-input: 89, 92a, 92c
- circle: 12b, 74a
- class: 59, 42a, 45b, 53a, 53d, 58a, 59, 79, 92c
- class-academic-result: 43a, 45a, 55b
- class-list: 47, 45c, 55e
- class-location: 47a, 47b, 55b
- cluster: 42b, 55b
- comment: 12a, 14a, 15c, 34d, 35a, 39, 41a, 43a, 44a, 76b, 77a, 78a, 79b, 79d, 80b, 81b, 81c, 82a, 82b, 82d, 83c, 84b, 85c, 93b, 95c
- communities: 50a, 52, 55c
- community: 11a, 14b, 29, 50a, 52
- community-information: 11a, 23, 42a
- complaint: 79c, 79d, 80a
- compound-wall: 80, 66a, 66c, 78a, 79b, 80a
- decision: 16a, 16b, 57d
- description: 16a, 16b, 57c, 57d, 89

C.5 Identifier Index
C.5. IDENTIFIER INDEX

reasons-for-preference: 87a, 88a
result: 16a, 17a, 43a, 55a, 77c
room-info: 62a, 62b, 63a
rooms: 61, 62a, 62b, 63a
samudayadatta-shale: 54b, 57c
samudayadatta-shale-table: 54b, 57c
sanctioning-scheme: 78a, 79a
sbc-vec-activity: 54a, 55a, 55d
sbc-vec-activity-table: 54a, 55a, 55d
sbc-vec-info: 54a, 54b, 55a
sbc-vec-meeting: 54a, 55a, 56a
sbc-vec-meeting-agenda: 54a, 55a
sbc-vec-meeting-table: 54a, 55a, 56a
sbc-vec-member: 54a, 55a, 55d
sbc-vec-member-table: 54a, 55a, 55d
school-address: 39, 41b, 42b
school-administration: 11a, 53
school-general-information: 11a, 39, 42b
school-improvement-opinions: 11a, 89, 90a
school-information: 9b, 10, 11a
school-name: 39, 40a, 42b
school-number: 39, 40a, 42b
school-picture: 93a, 93b
school-pictures: 11a, 93a, 95c
special-conditions: 50a, 55a
sports-equipment: 60, 70a, 71a
sports-equipment-list: 60, 70a
staff-absent-reason: 47, 50a
staff-attendance: 47, 48b, 50a
staff-information: 11a, 47, 50a
staff-member: 47, 50a, 50b
staff-member-info: 77, 50b, 50a
staff-role: 47, 48b, 50b
student-attendance: 43a, 44a
student-information: 11a, 11b
student-results: 43a, 43b, 44a
students-on-roll: 43a, 44a, 44b
student-survey-info: 43a, 43b
subject: 17, 48a, 50b
subject-list: 47, 48a, 50b
taluk: 12c, 13, 14a, 39, 55a
teachers-opinion: 47, 50a, 50b
teaching-material: 60, 60a, 60b
subject-list: 47, 48a, 50b
teaching-material-list: 60, 60a, 60b
toilet: 60, 60c, 65a
toilet-facilities: 60, 60a, 60b
total-population: 32, 33a
village-or-place: 12c, 13, 14a, 39
visitor: 53, 54b, 55b
visitor-frequency: 9a, 54b, 55b
ward: 12c, 13, 14a, 39
water: 60, 63b, 64a, 65b, 65a, 65b, 90c, 92d
water-source: 63b, 64a
working-days: 39, 41c, 41d
working-days-list: 39, 41c, 41d
yearly-average-income: 32, 33a
year-started: 39, 41c, 42b
Colophon

About the Author

Joseph Koshy <jkoshy@freebsd.org>, is an open-source enthusiast, and a volunteer developer with the FreeBSD Project.

Typography

This document has been typeset using the \LaTeX typesetting system, using pdfTeX to generate a PDF file from \LaTeX sources. The \textit{Times} family of fonts were selected over the default \textit{Computer Modern} that \LaTeX uses for better compatibility with PDF.

Literate Programming

\textit{Literate programming} techniques were used in the process of creating this document. Literate programming evolved as a way of documenting computer programs, allowing the design of the programs to explained side by side with the code. The best way to explain the workings of a program to another programmer is generally not the way a compiler requires the program to be structured. Literate programming allows a program to be described to the reader in a reader-friendly way, while keeping the ability to extract and rearrange the code from the description to fit a compilers requirements.

The literate programming tool used was NOWEB, a simple and freely available literate programming tool. In the NOWEB system, the program and the description of the program are written together as a single document. Two separate transformations are used to extract either the explanation or the code from this source. These transformations are called \textit{tangling} and \textit{weaving} respectively. The NOWEB home page at http://www.eecs.harvard.edu/~nr/noweb/ has more information about the NOWEB system.

With the advent of data structuring languages like XML, we can now bring the same level of formality and precision enjoyed by computer programs to the descriptions of data too. However, the formal description is also a computer processable language and consequently has rigid constraints on its internal structure. Thus, to explain the description, literate programming techniques are useful.