Conversion of SEED format to XML representation for a new standard of seismic waveform exchange

Bristy Chakraborty

May 19, 2022

Abstract

Your abstract.

1 Introduction

The Standard for the Exchange of Earthquake Data (SEED) has been designed as an international standard format for the exchange of digital seismological data [SEED Reference Manual, 1993]. It is now widely used among the community which maintains the broadband seismograph network and recognized as a standard format for data exchange. SEED volume consists of headers and data records and blockettes are stored in headers. The format for data records is called mini-SEED and it is closely related to the format recorded in data loggers. Since SEED blockettes are defined as a collection of named fields with fixed length data, this introduces difficulties of extension of data structures. However, because there already exists huge amount of waveform data saved in mini-SEED format, it is a formidable task to fully revise the current SEED format to allow future flexible extensions. Although it has been recognized that the revision of SEED format is necessary, there has been no attempt for revision since its latest release of Ver. 2.3 in February, 1993 because of this difficulty. Here we propose an XML(eXtended Mark-up Language) representation of SEED header structure and show that flexible design and robust validation in data models will be realized at the same time. Technical difficulty for constructing network-based system will also be reduced by introducing XML to SEED data description. We also mention about the extension of XML-SEED format to synthetic seismogram database.

2 SEED format

SEED was adopted as a standard format for international digital seismic data exchange in 1987 by the Federation of Digital Seismographic Network (FDSN), which was formed under the International Association for Seismology and Physics of the Earth's Interior (IASPEI). Before the SEED format was adopted, digital seismic data exchange was complicated by different data logger formats. SEED was designed so that it accommodates differences in data format originated from the type of data logger comprehensively. The SEED format consists of one logical volume, which contains two format objects: (1) control headers and (2) time series. The former object is formatted in ASCII and contains auxiliary information about the volume. The latter object contains raw binary data, which is digital seismogram. Control headers are categorized as (1) volume index control headers, (2) abbreviation dictionary control headers, (3) station control headers and (4) time span control headers. These headers are used to provide such as abbreviations used in the control headers, operating characteristics for a station and its channels, and time span of data. Because of these comprehensive descriptions of the SEED volume in control headers, SEED format cbe used to provide digital seismograms recorded by almost any kinds of data loggers. Each control header consists of a series of blockettes, which contain sequence of data filed specific to that blockette type. Because blockettes are defined as a collection of named fields with fixed length data, this introduces difficulties of extension of data structures. On the other hand, header structure is designed to be modular, which is similar to XML. This suggested us to try to represent SEED format structure in XML. Data structures of XML document are very flexible, because lengths of any fields in its format are not fixed. In order to define a new field and blockettes,

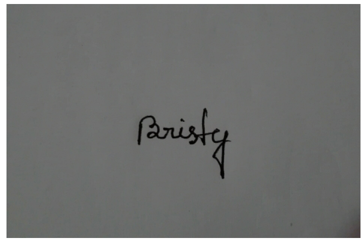

Figure 1: an image of my sign

item1	item2
Widgets	gadgets
Gadgets	widgets

Table 1: a new table created

just give a new tag name and its hierarchy to make a room for new blockettes. To describe types of data, XML has its schema language, which is called as XML-Schema. This schema language is also used for validation of XML document. By introducing XML into SEED, it is apparent that flexible design and robust validation in data models will be realized.

2.1 adding table

header1	header2
un¡http¿www. ¡un¿	¡drive id¿my drive ¡drive id¿
data server on network	storage media

2.2 adding a Mathematical equation

$$sum = \frac{X_1 + X_2}{n}$$

2.3 Data

When we represent SEED header structure in XML, we do not modify anything regarded with the format of time series data. To include binary mini-SEED format digital seismograms in XMLrepresented SEED volume, we consider two scenarios. The first scenario is separated header file and data (Fig. 1). Data location could be other data files or data servers connected via networks. One can get a standalone header file to know about an event, properties of stations and data locations. This is the same concept as dataless-SEED volume. However, when we create these separated header files only volumes, to get seismic wave data, one accesses data server or look for data files according to data location described in this header file. The second scenario is the same as the current Full-SEED volume so that XML-SEED volume includes both SEED header represented in XML and the binary seismic wave data (Fig. 2). Header specifies the location of data that is also stored in the same file. This composition is basically possible in following way. For example, the first line describes the length of header, followed by a blank line. Header XML document starts at the third line. The format of header part is plain-ASCII, and is not based on logical records. Data part starts at the position specified at the first line

2.4 Programs

Programs to convert from current full-SEED volume to XMLSEED volume and to read XML-SEED volume to extract seismograms are available. Currently we provide digital broadband seismograms from Ocean Hemisphere Project geophysical network by XML-SEED format through IFREE data center (http://www.jamstec.go.jp/xmlninja/).

2.5 XML-SEED for synthetic database

Recently we have demonstrated that we can calculate global theoretical seismograms for realistic 3D Earth models based upon the combination of a precise numerical technique (the spectral-element method) and a sufficiently fast supercomputer (the Earth Simulator) [Tsuboi et al., 2003]. It has now become possible to routinely calculate synthetic seismograms for earthquakes greater than a certain magnitude. Starting in 2003, we select earthquakes with magnitudes greater than 6.5 from the Harvard CMT catalog and calculate theoretical seismograms for the Stations in the Global Seismographic Network. To distribute this synthetic seismogram database to the seismological community, we modify the XML-SEED to include metadata entries which are characteristic to the synthetic seismogram database, such as numerical technique we used to generate synthetic seismograms [Tsuboi et al., 2004]. We plan to distribute these theoretical seismograms through mirrored IFREE/JAMSTEC and Caltech web interfaces soon.

2.6 Summary

7. Summary Here we have shown that current SEED Format can be directly translated to XML representation without introducing any modifications to current format. The advantage of using XML representation of SEED format is obvious. It will be straightforward to add any necessary information afterwards. What we should do is to define tag name and include this into schema. Although we have not modified current SEED control headers, there should be various ways to extend SEED format by taking full advantage of XML. One of the examples should be the status report of data logger. If the data logger reports its status or parameter settings in XML format with its digital seismograms, this information can be directly incorporated into database directory in the data center. This should greatly simplify data quality checks done at the data center. Another example may be the data distribution through the web service. Because the protocol of data exchange for the web service is done by XML, if the SEED data itself is expressed by XML, we may use the control header content described in XML format for data exchange and distribution. We have developed network data center system based on Java RMI [Takeuchi et al., 2002]. We may distribute our XML-SEED formatted digital seismograms through our network data center system to fully utilize the XML represented header structure

References

Standard for the Exchange of Earthquake Data, Reference Manual, SEED Format Version 2.3, Incorporated Research Institution for Seismology, 1993. Takeuchi, N., S. Watada, S. Tsuboi, Y. Fukao, M. Kobayashi, Y. Matsuzaki and T. Nakashima, Application of distributed object technology to seismic waveform distribution, Seismological Research Letters, Vol. 73, No. 2, 166-172, 2002. Tsuboi, S., Komatitsch, D., Ji, C. and Tromp, J., Broadband modeling of the 2002 Denali Fault earthquake on the Earth Simulator, Physics of the Earth and Planetary Interior, vol. 139, 305-312, 2003. Tsuboi, S., Tromp, J., and Komatitisch, D., An XML-SEED Format for the Exchange of Synthetic Seismograms, EOS Transactions of American Geophysical Union, suppl., SF31B-03, 2004.