The Working Group III on data exchange

Bernard Dost (*)

The present Working Group III (WG-III) on data exchange started in 1992 as a follow-up on the former WG-III on data collection and exchange formats (Romanowicz, 1991). WG-III was chaired by Jim Scheimer (1986-1989) and later by Ray Buland (1989-1992) and focused on the definition of the first Standard for the Exchange of Earthquake Data (SEED) format. This working group was very successful and its work resulted in a general acceptance of SEED by the entire broadband community. After SEED was defined and accepted it was decided to enlarge the scope of WG-III. A working document was prepared to define new issues and to set priorities. During the 1992 FDSN meeting in Seattle a lively discussion started and concentrated on data request managers. In this paper I report on the discussions during the Wellington 1994 IASPEI meeting, where WG-III met twice. Apart from a continuing discussion on SEED issues, that resulted in minor upgrades, key issues were: FDSN CD-ROM, Data Request Managers, quality control and the station book.

SEED issues

SEED evolved as a very flexible, and therefore complex archival format. Its use is widespread, partly because software is made available to read and write data in SEED. The SEED format is described in detail in a manual, that was published in two issues. The latest version is 2.3 (IRIS, 1993). In the manual the changes between the different versions are explained.

One of the important features of SEED is the definition of a Data Description Language (DDL). In the DDL any digital data format should be coded in a set of keys. SEED reading software is then able to reconstruct the original data values automatically. This ensures that data users are not bothered by adapting their SEED reader software each time a new station comes in. Although the DDL was regarded as important it was not used in practice. Apart from the programming effort, the main reason was a concern on loss of speed in data decoding. Tim Ahern reported that Allen Nance at the IRIS Data Management Center (DMC) developed DDL decode software and in the process uncovered problems. These problems will be verified by Ray Buland and if necessary, the SEED DDL definition will be improved. It is foreseen that the current DDL will be extended in the future, since a number of new data compression schemes are being used.

The SEED header consists of a sub-set of a total of 47 blockettes. A blockette is defined as «a self identifying sequence of data fields that assists computer readability, storage efficiency and flexibility» (IRIS, 1993). A minimum set of blockettes is required. Joe Steim proposed the addition of a blockette that includes manufacturer specific information. His proposal was easily adopted, since the blockette is optional. In SEED version 2.3 network codes were assigned to the data producers. A discussion started on who will be responsible for a specific data set. It was agreed that the Network Data Centers (NDCs) should be responsible.

(*) ORFEUS Data Center, P.O. Box 210, 3730 AE De Bilt, the Netherlands.

FDSN CD-ROM: event criteria and production schedule

The FDSN maintains a station inventory of digital broadband stations (see report of WG-I). A selection of these stations is made to represent a Federation network. It is the intention to produce data from this FDSN network on CD-ROM as SEED event volumes. The NEIC, as FDSN archive for event data, offered to produce six CD-ROM per year of data.

These CD-ROMs will then be distributed. Data from the FDSN network forms a subset of the total available digital, broadband data set. The FDSN network data set will be used for global studies and as a first selection of data by users who want to focus on a specific region. To service both needs and to take into account the limited number of CD-ROMs that can be produced, a magnitude threshold is required. The determination of this threshold was discussed over electronic mail. A proposed 100 station network formed the basis of this discussion. In 1993 a consensus was reached at a magnitude threshold of 5.7. This threshold was based on the inclusion of broadband and long period data and an average of 40 events per two months exceeding this magnitude. Very Long Period (VLP) data (sample rate at .1 Hz) are not included, but will be available on-line from the FDSN continuous archive (IRIS DMC). The duration of the recordings will continue to be governed by the FDSN criteria that were already adopted in the past. In practice SEED event volumes are requested by the NEIC from the FDSN continuous archive at the IRIS DMC, where raw data is assembled and SEED volumes created. The NEIC puts the data on CD-ROM and produces copies. It is important to secure data collection and a time planning was agreed upon (table I).

Table I

Data period	Date of submission
01.01.1990-06.30.1990	06.30.1994
07.01.1990-12.31.1990	09.30.1994
01.01.1991-06.30.1991	12.31.1994
07.01.1991-12.31.1991	03.31.1995
01.01.1992-06.30.1992	06.30.1995
07.01.1992-12.31.1992	09.30.1995
01.01.1993-06.30.1993	12.31.1995
07.01.1993-12.31.1993	03.31.1996
01.01.1994-06.30.1994	06.30.1996
07.01.1994-12.31.1994	09.30.1996
01.01.1995-06.30.1995	12.31.1996
07.01.1995-12.31.1995	03.31.1997
01.01.1996-06.30.1996	06.30.1997
07.01.1996-12.31.1996	09.30.1997
01.01.1997-06.30.1997	12.31.1997

Data Request Managers

The main subject of discussion in Wellington was related to automatic Data Request Managers (autoDRMs). DRMs have long been available in an interactive version (*e.g.* Hanka, 1991; SPYDER, see Dost this issue, page 1071). Data users can access a data center or seismograph station directly. Automated versions accept requests by electronic mail and either return the requested data by electronic mail or send a message that the data can be retrieved from disk by anonymous ftp. A semi automated version is in use at the IRIS DMC and a fully automated version (Kradolfer, 1993) has been developed for the Group of Scientific Experts (GSE; Ringdahl, 1991). Many FDSN members are installing or are about to install autoDRMs and there was a feeling to discuss a minimum standardization between them. In Canada, the autoDRM is replacing their earlier on-line DRM. The prototype autoDRM developed by Kradolfer forms the basis of our discussion.

The only format currently supported by the autoDRM is the GSE format. The GSE format is designed to be simple and is only available as an ASCII format. Data compression is allowed for. FDSN members would also like to see support for data in SEED, but SEED is a binary data format. An ASCII format is needed to distribute data via electronic mail. Alternatives, such as using uuencode of SEED volumes, were discussed and may be offered as an option. It was decided to adopt the GSE format as the FDSN ASCII format, but only if the GSE format would accept some changes or extensions. In a separate meeting WG-III members discussed the necessary changes and a final document was prepared by Ray Buland. This document was presented at a GSE meeting in February 1994. A GSE working group was formed to discuss the final form of the GSE format, taking into account the input from the FDSN. Recently the autoDRM was adapted to allow file transfer using ftp. This opened the way to include SEED volumes. In order to increase the efficiency in transfer of SEED volumes, it was proposed to send only SEED headers for a specific station if there is a change in the header information. It is the intention of the ORFEUS Data Center to develop an autoDRM interface to SPYDER.

In conclusion, the FDSN endorsed the following statement: «the FDSN proposes to adopt the autoDRM, as developed by Urs Kradolfer in Zurich (Kradolfer, 1993), with the possibility to extend its functionality. All new request formats should be backward compatible. The FDSN members are encouraged to install the autoDRM and evaluate its functionality».

Quality control

Raw data coming directly from seismograph stations will have to be quality controlled at the Data Collection Center. Up to now there has been no formal structure of what the minimum requirements for quality control are and how to communicate data problems to the end user. As a starting point it was suggested to make an inventory of how networks and data collection centers carry out their quality control. This inventory will form the basis of a set of common guidelines. Data problems should be reported in a common (electronic) form. Preferably in data less SEED volumes (only header information). A plan was proposed to develop portable data collection center software. The IRIS DMC offered to develop this software with as much input as possible from FDSN members.

Station book

At the IRIS DMC a major effort was made to produce a station book, where all FDSN stations will be described in detail. Apart from details on the instrumentation and location, noise characteristics are also included. In order to produce an objective noise measurement of each station, a pro-

cedure was designed and all processing done at the IRIS DMC. A preliminary version of the station book was presented. Although the book was very well received, it was too large. It will be investigated how to produce the volume on CD-ROM.

REFERENCES

HANKA, W. (1991): The German Regional Broadband Seismic Network (GRN) Project, in *Workshop on «MEDNET; the Broadband Seismic Network for the Mediterranean» 1990*, edited by E. BOSCHI, D. GIARDINI and A. MORELLI (Il Cigno Galileo Galilei Edizioni di Arte e Scienza, Roma), 83-95.

IRIS (1993): SEED manual, version 2.3.

KRADOLFER, U. (1993): Automating the exchange of earthquake information, *EOS Trans. Am. Geophys. Union,* **74**, 442-445.

RINGDAHL, F. (1991): A global seismic network for monitoring earthquakes and underground explosions: A review of work conducted by the Geneva Group of Scientific Experts, in *Workshop on «MEDNET; the Broadband Seismic Network for the Mediterranean» 1990*, edited by E. BOSCHI, D. GIARDINI and A. MORELLI (Il Cigno Galileo Galilei Edizioni di Arte e Scienza, Roma), 276-286.

ROMANOWICZ, B. (1991): The Federation of digital broadband seismic networks, in *Workshop on «MEDNET; the Broadband Seismic Network for the Mediterranean» 1990*, edited by E. BOSCHI, D. GIARDINI and A. MORELLI (Il Cigno Galileo Galilei Edizioni di Arte e Scienza, Roma), 287-300.