STREAM: a method to facilitate efficient data exchange and archiving

P.E.L. Schaminée
Deltares, Delft, The Netherlands

A.A. Klapwijk
NCIM, Leidschendam, The Netherlands

D.J. White
COFS UWA, Perth, Australia

ABSTRACT: The results of physical experiments are complex and valuable. To facilitate efficient exchange and archiving of test results STREAM has been developed. STREAM means Standardized Test Results Exchange and Archiving Method. A formal distinction is made between the description of the information and the data itself. Standardizing the format of the description file for a given type of test - the so-called Testdefinition - instead of the data files themselves, permits the utilization of generic software tools and data formats for documentation, reporting and data storage. The formal breakdown of a test into five phases – the so-called SMARF phases - has several benefits: (i) testing activities are structured, (ii) it simplifies the process of documentation, and (iii) it facilitates a structured data handling process. These complete, consistent, fully-checked and quality-assured files may be easily utilized by personnel who were not directly involved in the conduct of the tests.

1 INTRODUCTION

Everywhere are needs for an efficient way to exchange and to archive information. In this paper the focus is on information obtained by means of experimental testing. In general a test is performed in a test set up designed by engineers. They may be involved in the current program of tests, or they may have long since moved on, with the test activity becoming standardized and routine. During testing, the actual values (the instrument readings, or observations) are obtained by operators. These values are analyzed and reported by experimental specialists (often someone different to the operator). Then, it is usual for the reported values to be transferred to a client, who will use a selection of the results for calculations, and so on. It is important to observe that at different stages different people are involved, who all use the results for their specific purposes. It is a small step to recognize that they all want a different set of information to be archived for later use. This is in contrast with the idea that one single standard data file for all types of test fits all purposes.

In this paper a method is described for the exchange and storage of data from single tests. The method considers the initial set-up of the experiment through the gathering of raw data until the factual report of the test results (which may involve some level of interpretation and collation of the raw measurements).

The reasons that the method has been developed are (i) to improve the quality of the reported and stored test results, (ii) to minimize the effort required to apply the method and (iii) to maximize the benefit of using the method. The method was developed with many types of geotechnical tests in mind, spanning from routine laboratory tests (e.g. a moisture content determination) through to complex centrifuge model tests. It is equally applicable to experimental activity that is conducted in the field.

The method described in this paper is centred around two new elements:

1 A chronology that captures the activity involved in any form of experimental activity: SMARF (Set-up, Measurement, Analysis, Reporting, Filing – see Section 2);
2 A method that structures information generated throughout the SMARF process: STREAM (Standardized Test Results Exchange and Archive Method – see Section 3).

STREAM uses a standard format for defining types of tests and accepts several formats for storing the corresponding data. The purpose is to allow the results of a single test to be collated as a complete set of information, which can be accessed by anyone who has the corresponding standard definition. There are two different kinds of information involved:

1 the values of the quantities to be exchanged (i.e. the readings of one test). These are stored in a separate data file for each test.

2 The precise description of these quantities (i.e. an explanation of what the readings are of, in a standard way, comprehensible by any user). These descriptions of the quantities for a type of test are stored in a so-called Testdefinition, which is a document that is formatted in a standard way.

In STREAM, conversion between various formats of data files is accepted and various formats of data file are already implemented. In addition a program - the so-called Testdefiner® - is available to easily create a Testdefinition and a Matlab® library has been created to make optimum use of STREAM.

2 THE CONCEPT OF SMARF

It is useful first to note some difficulties that commonly arise with data handling during the process of obtaining experimental results. Firstly, different personnel are involved at different stages of the activity. Also, if multiple tests of the same kind are conducted at different times, different users may undertake them. The resulting data may be stored in different ways and it may also be measured in different ways (which may not be documented).

To resolve these differences may require double-handling of information before it is finally reported (for example, transferring information from spreadsheets of different formats, or between some combination of multiple lab notebooks and multiple spreadsheets). In other cases, if the procedures are ill-defined and not logged then the resulting data may be inconsistent – as a simple example, if different oven temperatures were used but not noted during moisture content determinations.

The SMARF concept identifies five chronologically-ordered phases from the design of a test to its factual reporting. These phases lead to the SMARF acronym – set-up, measurement, analysis, reporting and filing.

- *Set-up phase* in which the equipment and samples are prepared;
- *Measurement phase* in which the actual experiment takes place, i.e. the gathering of sensor readings;
- *Analysis phase* in which the recordings are analyzed and new derived quantities are calculated or key values at particular times are extracted;
- *Reporting phase* in which both the measured and calculated results are presented attractively by means of figures, tables, etc;
- *Filing phase* in which the results are prepared for long term storage and future accessing.

The SMARF concept provided moments to transfer to a next phase by means of a well defined phase report. At this moment the current set of data for a particular test can be stored in a data file and 'frozen'. The format and operation of STREAM is specifically developed for transfer of experimental test results. The target is to facilitate well-defined storage and exchange of properly-described data during these processes.

3 THE CONCEPT OF STREAM

The STREAM method, linked to the SMARF chronology, is meant to be useful for both simple standardized common tests (i.e. a moisture content measurement) and very complex custom tests (i.e. a geotechnical centrifuge test). In the case of performing a geotechnical centrifuge investigation, the test series consists only of a limited number of tests which use the same Testdefinition. Even then the method must be efficient and beneficial if it is to prove desirable for users to apply.

Documentation of exchanged or archived results is an integral part of the process. Tests of the same type share the same procedures and therefore share the same documentation with respect to the description of quantities and equipment. It was concluded that the method can be optimized by using the shared information for tests of the same type. Therefore, only one Testdefinition is created for all tests of the same type. All values (i.e. readings) from each single test are stored in a separate data file.

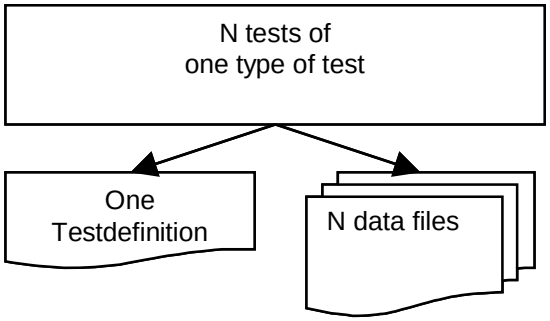

Figure 1. Relation between a Testdefinition and the corresponding data files

In STREAM the five SMARF-phases are defined separately. In the Testdefinition it is clear in which of the five SMARF phases an item is created. An item is referred to as a test result or analysis result to be exchanged. Using the Testdefinition each data file can be checked before entering the next SMARF phase for completeness and correctness.

The SMARF process is set out in Figure 2, indicating broadly the different steps that are conducted at each stage. The data file is updated throughout the test, as the experiment moves through the SMARF phases.

A side effect is that this allows reuse of (parts of) common measurement or analysis definitions. The concept of the five SMARF-phases has proved to be efficient (i) with respect to process of defining a test, because the kind of information to be exchanged and

archived can be optimized for each phase and (ii) with respect to the completeness after each phase, because it is well documented which items are required in later phases.

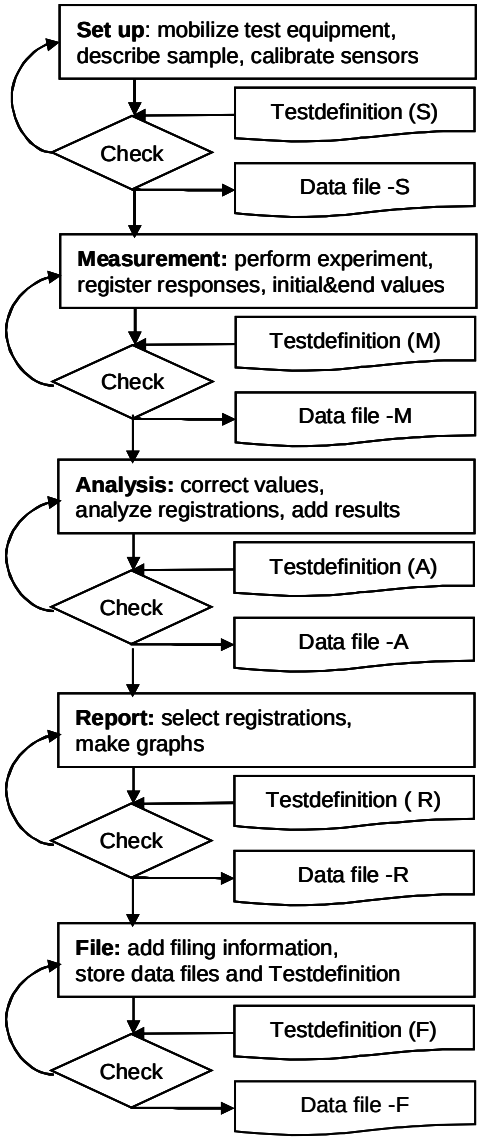

Figure 2. Flow diagram of one test.

When the one of the phases changes, which might be the case when new insights are gained, the Testdefinition also has to be changed. However it becomes very easy to redo processing a test from a well described point. For example, the best way to analyze and report a test may only be established, after the last test of a short series.

All persons involved of defining one or more phases will use the Testdefiner to edit the single Testdefinition which corresponds to that test. When the type of test is completely new, an empty template has to be used. However, when a Testdefinition for a similar type of tests already exists, sets of items can be copied from previous Testdefinitions. Even the client can be involved in the process of defining a test, because he can easily be informed about the items and their definition using the information concerning the report-phase in the Testdefinition.

4 USING STREAM: AN EXAMPLE

In this section STREAM is illustrated by applying it to an example test. Imagine a test to determine the moisture content of a sand sample as a function of time and suppose the test procedure consists of weighing a moist sample in a bin placed in an oven until it is dry. The use of STREAM starts by describing all items to be obtained in the five SMARF phases.

In the set up phase the mass of the empty bin has to be determined and the sand has to be classified. In Figure 3 a screen shot from the Testdefiner interface is shown.

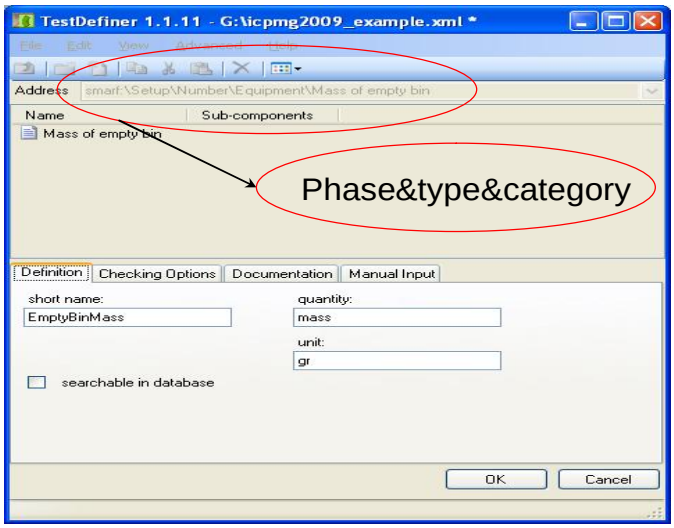

Figure 3. Graphical user interface in Testdefiner® for item.

Figure 3 shows that the quantity “mass of the empty bin” requires a unique reference in the Testdefinition, here chosen as ‘EmptyBinMass’ – the so called *shortname*. No other quantity is allowed with the same shortname in this Testdefinition. In the address field at the top are the *phase* (set up), the *type* of data (a number) and the *category* (‘equipment’) are visible. The available phases and types are set by the Testdefiner and the relevant ones for each quantity are selected by the user. The category name is defined by the user. The purpose of the category name is to permit the user to group sets of items within a phase. Clicking an other tab will open new pages to enter information about the item EmptyBinMass, such as how it should be measured, the valid range (i.e. minimum and maximum values). This information is used to check the data file.

The next step is defining the measurement phase. During this phase the mass of the bin together with the mass of the moist sand has to be determined as a function of time. So, time and mass are to be measured, these are quantities with more than one observation: these are called *columns*. Figure 4 shows a Testdefiner input screen for a column.

That sensor values in columns often require more information than a single number –for example, they

should be accompanied by the location of the transducer in the experiment and the calibration information of the sensor.

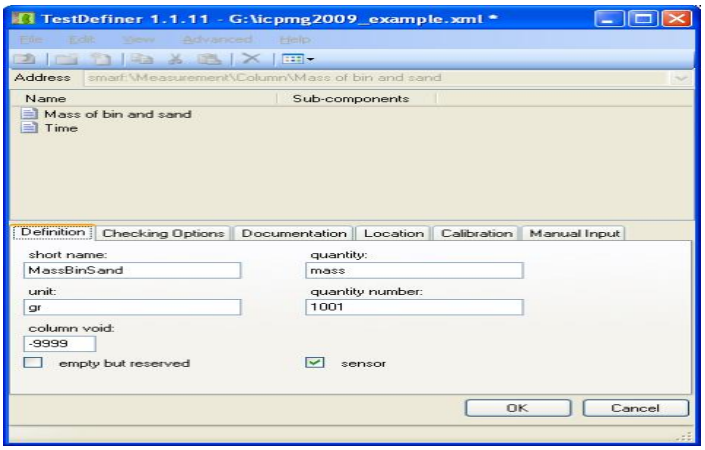

Figure 4. GUI in Testdefiner for a column quantity

In the analysis phase some calculations have to be performed to calculate the moisture content as a function of time, from the measured weight. The result of this analysis will be stored in a new *column* item 'MoistureContent'. Figure 5 is a screen view of its documentation tab in the Testdefiner.

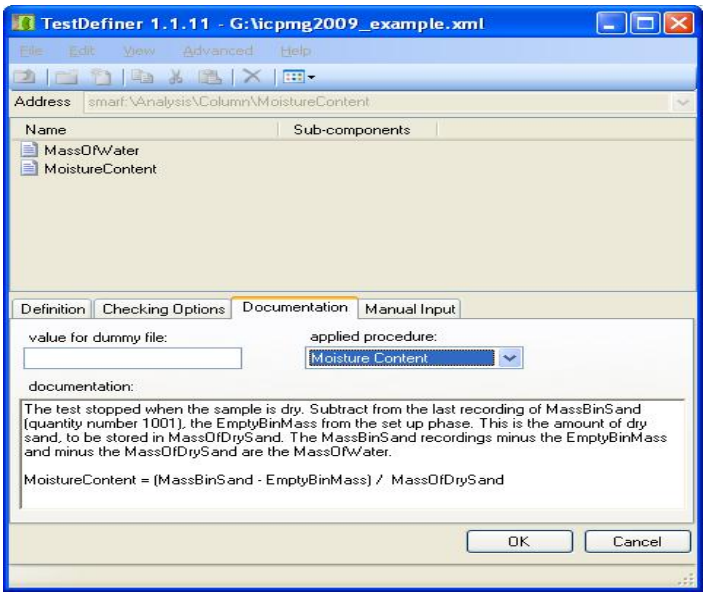

Figure 5. GUI in Testdefiner for documentation

The procedure – that is, the formula to be applied to calculate the moisture content – is described in the documentation field, and refers to other items in the Testdefinition by their shortnames. It becomes clear that for the analysis the EmptyBinMass is essential, so this item must be stored in the data file after the set up phase.

When preferably all phases, but at least the set up and measurement phases are defined, the data acquisition system can be set up to store the information directly in the correct format. In most cases a simple conversion program will be required to convert the data file produced by the data acquisition (DAQ) system into a STREAM data file that conforms with the Testdefinition. Suitable conversion routines have been written for the DAQ systems at Deltares and COFS/UWA.

When the measurement phase has ended and the data is stored in a STREAM data file, it is checked for completeness and correctness. For example, the existence of all defined *column* items can be checked and the *column* data can be compared with the instrument range limits. When the file passes these checks, it can be 'frozen', to prevent it from being changed, and transferred to the analysis phase.

The analysis procedure is mostly known beforehand, the Testdefinition can also be complete and presumably also most of the analysis software is in place. The STREAM library allows efficient importing of information from or export of information to a STREAM data file. The main characteristic of the library is that all references to the information in the data file are made by using the shortnames defined in the Testdefinition. The implementation is by reading and writing the data file through the corresponding Testdefinition. All of the items that have been defined in the Testdefinition can be accessed. This guarantees that all information in a data file is described in the Testdefinition all the time. When all analysis is ready, it can be checked for completeness and correctness and, when it passes these checks, can be frozen as an analysis data file. Based on the frozen file a selection will be made for the report to the client. This selection can be presented and sent to the customer and the data file can be archived.

However, when analyzing the first test of a new type of test, it might become apparent that new items have to be created and defined within the Testdefinition. Describing these new items instantly in an update of the Testdefinition, allows all standard STREAM tools to be used continuously. It is recognized that the Testdefiner should function in a user friendly way. The next tests of the same type might lead to modification in the items set in the Testdefinition. In this case the stringent 'borders' between the phases, especially between the measurement and analysis phase, permit new analysis to be performed based on previously frozen measurement files. For this reason the STREAM library facilitates batch processing.

After the last test in a series, all of the tests can be easily processed using the same software because all items are well described in the Testdefinition and all values are available in the (frozen) data files.

5 ABOUT THE TESTDEFINITION

All STREAM Testdefinitions have the same standardized structure and are written as an XML

file. The Testdefinition does not store the data itself, but contains a structured list of the data elements, and additional information that helps the user to correctly perform the test, input the resulting data in a datafile, to describe the applied formulae and procedures and to check corresponding data files for correctness (Figure 6).

Each Testdefinition consists of two classes of information: general and user-defined. In the following paragraphs both are described in detail.

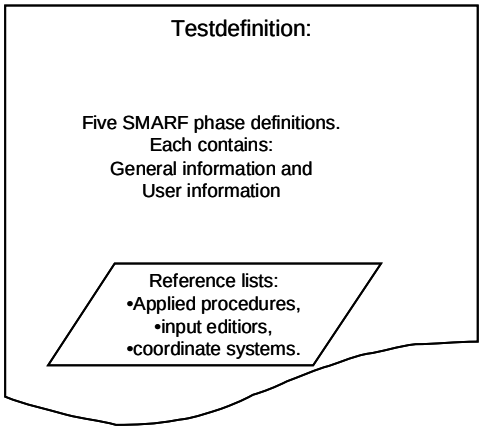

Figure 6. Types of information in a Testdefinition

5.1 *General information*

General information is unconditionally required for all data files and is therefore predefined. These general information items are well defined. They can be grouped in three categories: (i) file tracing, i.e. information to identify for example the organization where the test was performed, the person responsible for the test (ii) test description, i.e. the reference to the corresponding Testdefinition, (iii) data format description, i.e. information on the characters used as column and record separators in the data file.

5.2 *User-defined information*

The users have to define each item they want to exchange and store in the data file separately. STREAM provides the user three types of information (i) numbers, which can be used for information such as the duration of the test, the temperature of the oven and the target centrifuge g-level, (ii) strings, which can be used for names of standards, equipment, procedures, people and text labels and (iii) columns, which are a set of numbers belonging to one quantity: in most cases a series of measurements of a quantity varying through time.

The information to describe each item depends on the type of information. For example, each number or column requires a unit in contrast to a string which does not. To facilitate the user to describe each item, the program Testdefiner is available. It guides the user and stores the information in the correct XML-based Testdefinition format.

6 ABOUT THE DATA FILE

During the search for a standard data file format it became clear that only the specific application of the data determines the most suitable format. For example, the DIGGS data format is tailored for road maintenance (with its strong links to GIS databases), the XML language is very useful for structured data exchange (where many single data values, each with a different definition, are collated) and worksheets in a tabular program (such as Excel) are very useful for intuitive interpretation, and for storing time histories of transducer data. Similar tests will be performed for different applications; therefore the STREAM method for defining and structuring data has to be application-independent.

So, it was concluded that the exchange method has to incorporate conversion of data files, instead of one standard format fitting all uses. Once the Testdefinition is set up, correct conversion tools can be easy developed Currently STREAM is implemented with a GEF formatted data file and an MS Excel data file.

Each STREAM data file contains all of the results from a single test and is regarded as an independent entity. The method does not provide the means to describe relations between tests, other than belonging to the same type of test. For example, the method does not relate tests from the same location as a GIS-database application would do.

7 STREAM AND STAND ALONE TOOLS

For the Testdefinition two stand alone tools have been developed, the Testdefiner® and a report tool. As described above, a wide range of information is required to define each item in a Testdefinition completely. The Testdefiner supports the user to enter all required information by a structured graphical user interface. Input accelerators are the capabilities to copy parts of existing Testdefinitions, to copy information from a single quantity using automated counters to maintain required uniqueness. The Testdefiner is programmed in C#. The resulting Testdefinition is a structured XML file that conforms to the STREAM Testdefinition format. In a Testdefinition all information describing the data file is available. However, being an XML file it is not directly suitable for a summary of this description in a factual report. The Testdefinition report tool constructs excerpts from a Testdefinition that can be included into Word or Excel documents.

Specific stand alone tools are also available for presenting and interpretation data files in the GEF format, which has been developed over many years since the GEF format was first created in 1999. There are two tools to reduce the number of readings (or instrument scans) of a GEF file, one by only

keeping the n-th row of data and one by graphically selecting a continuous part of the data. The GEFviewer®, is a powerful plotting tool.

8 FUTURE DEVELOPMENTS FOR TC2

The STREAM method was developed (and is continuing to develop) by means of Deltares funding. During this development process both the use at Deltares and the potential use in the international geotechnical community (as defined by the working group for data exchange and archiving of TC2 of ICCSGME) were taken into account. Since no additional funding is expected, no developments solely for working group purposes can be honored. However, it has been attempted to provide a platform which is as flexible as possible. A user software exchange library can be easily arranged based on the basic functions, library and tools provided. Some of the expected developments for the various aspects of STREAM are described below.

- *The concept.* The concept of a definition file and corresponding data files is an essential part of STREAM and will not be changed.
- *The Testdefinition*. All non-user item information in a Testdefinition has to be agreed upon. It will be brought in line with international data models for geological testing, especially with the accuracy of identifying parameters such as the location. These modifications have to be made due to integration of the concept into a national governmental program to make a Dutch national database for hydrological and geotechnical parameters.
- *The Testdefiner*. The Testdefiner itself will be adapted for the modifications described above. The interface may evolve through user feedback.
- *The collection of Testdefinitions*. Each Testdefinition is defined by the people involved with a type of test. If, for example, TC2 agrees a standard type of testing for a routine event, such as a model CPT or T-bar penetrometer test, a standard Testdefinition can be agreed upon. In this case it might be appropriate to make this Testdefinition available by means of a TC2 or ISSGME web site. It might also be useful to make only parts of definitions standard. For example, what particular properties of the centrifuge being used have to be stored. This can be defined by means of the concepts of Testdefinitions as well.
- *The Matlab® STREAM system library*. The basic functions make use of Deltares software. It is expected that the library of STREAM functions will be made freely available. The underlying software will be delivered in a shielded form.
- *The Matlab® STREAM user library*. This is a new development: a public site to exchange generic functions. However, when there are plenty of tools and functions available, the benefit of using STREAM increases and the possibility to exchange test results is enhanced.
- *The GEF library*. This library is now widely used in the Netherlands and supports several field testing exchange formats, so it will remain available. The increase of tools available for GEF will partly depend on the growth of the STREAM libraries described above.

9 CONCLUSIONS

The concept of a standard description document – a Testdefinition – for each type of geotechnical test provides significant benefits to experimentalists. It provides a method that can

1 streamline experimental activity;
2 improve its quality;
3 and lead to well-documented tests;
4 that are archived in a time-proof manner.

The Testdefinition provides a structure for logging and storing in a structured form all of the information that arises from any geotechnical test – ranging from a simple moisture content determination to a complex centrifuge model.

The structure of the Testdefinition is divided into a chronology that reflects the different stages of experimental activity, given by the acronym SMARF – Set-up; Measurement; Analysis; Reporting and Filing. By dividing the experimental activity into this chronology, there are clear gateways at which the experiment advances to the next stage. At each gateway the completeness and quality of the data can be automatically checked according to criteria within the Testdefinition.

Software tools for generating the documentation for defining and inputting test data have been developed and are described in this paper. A basic Matlab library to import, export and convert data based on the Testdefinition is available.

This approach also facilitates the process of defining specific minimum standards for the information to be stored, particularly for cases where no standard currently exists such as centrifuge model tests. It is regarded as an advantage that these standards can develop gradually in time, leading to easier exchange and more efficient interpretation of test results.