iCART Data-Files

Dr. Stefan Messmer, IWM - Institut für Werkstoff-Fragen und Materialprüfungen <s.messmer@iwm.ch>

Table of Contents

Introduction 1
File-System Structure 2
The Stream File 3
The Notes file 4
The Analysis File 4
Data Safety 5
Fault Tolerance 5
Bibliography 6

iCART is IWM's company own wire rope testing software. To enable customers, competitors and other software manufacturers to access our data files we have developed public data files that will be used by iCART version 1.1 and higher. The aim of this article is to provide an introduction into the data file structure and the corresponding documentation.

Introduction

Several discussions with our customers have shown the strong wish to have full access to IWM's internal wire rope testing related documents. But not only customer demand has driven us to develop a modern file format for iCART. The requirements related to document safety, exchangeability and long term readability of our files initiated a development procedure for a new file format. Beginning with iCART version 1.1 we will use a state-of-the-art XML file format [1] for storing all program data.

As mentioned before we finally have selected XML as standard file format for iCART. This decision has several drastic consequences for the development of iCART. Both applications, iCART Acquisition Module and iCART Analysis Module have been re-written from the base to make full use of the advantages of the xml data file structure. The following reasons (beside others) have convinced us to choose the XML format as native file format for iCART:

- Flexible file format, completely human readable and inspectable;
- Well structured;
- Easy access from LabVIEW;
- Accessible from Web-Browser (Firefox, Safari or Opera, Internet Explorer with limitations) using XSLT technology [2]. No other software required;
- Acceptance in the Computer community. XML-files are widely used (Open Office, XHTML, several modern Apple-Programs use XML as native file formats, Microsoft plans to migrate their Office file formats, etc.)

We have designed our XML files close to National Instruments Lab-VIEW XML-implementation. This structure allows us to generate a very flexible XML file structure that can be well accessed with LabVIEW.

File-System Structure

iCART generates many files that need to be organized and kept together. We deliver all files concerning a rope test on a compact disc (CD) or digital versatile disc (DVD) with the following structure:

- **Root level**
  - A file named `Open Me.html`
  - A folder named `Resources`
  - A folder with the exact name of the Installation

  The Folder `Resources` contains resource files to open iCART files with a web browser. The installation folder contains all test data.
- **Installation level**
  - A file named `Index.xml`
  - One or more folders corresponding to the mounted ropes of the installation.

  The `Index.xml` file serves as helper file to navigate through the database. The rope folders contain all test data.
- **Rope level**
  - A file named `Index.xml`
  - One or more folders corresponding to the tests of the rope (named `*Test????.iCART` [1]).

  The `Index.xml` file serves as helper file to navigate through the database. The test folders contain all test data.
- **Test level**
  - A file named `Index.xml`
  - A file named `Analysis.xml` (may not be present).
  - One or more files named `Notes????.xml`
  - One or more files named `Stream????.xml`
  - Files named `Pict????.jpg` may be present.
  - Files named `Report????.pdf` may be present.
  - For each of the above mentioned files a corresponding digital signature `*.sig`.

  The function of the above mentioned files corresponds to their names: Analysis is a analysis file created by the iCART Analysis Module, Streams are data streams of the testing instrument, Notes are notes and measurement values collected and entered by the tester, Picts are pictures imported from a digital camera and Reports are final test reports.

[1] * denotes any character allowed in filenames and ? denotes any hexadecimal number written in capital letters.

iCART stores all information in the three files Analysis, Stream and Comment. Raw data is stored only in the files Stream (testing instrument) and Notes (collection of raw values acquired by the tester).

The Stream File

The stream file is iCART's main data storage. This XML file consists of two key elements: The `<Header>` element stores all relevant informations about the test, including installation, rope, instrument and settings. The `<Stream>` element contains an alternating sequence of `<Data>` and `<Preview>` elements.

The `<Header>` Element does not require further explanation, as all information is stored as human readable text.

The `<Stream>` element needs further explanation. The information stored here is binary. The binary data must be encoded to follow the rules of XML. The `<Preview>` element contains a base 64 encoded [7] portable network graphic (PNG) picture [3].

The `<Data>` contains a base 64 encoded [7] and ZIP-compressed [6] 2 byte word Integer (LabVIEW I16) array of size (number of tracks * size). The values are stored in the sequence first value / first track, second value / first track, ..., first value / second track, ..., ...

The last two tracks are high- and low word of a 4 byte unsigned (LabVIEW U32) integer array.

The physical meaning of the tracks are[2]:

- **Track 1.** Measurement result of the inner coil sensor (coil 1)
- **Track 2.** Measurement result of the outer coil sensor (coil 2)
- **Track 3.** Measurement values of the second tacho track. Needed to distinguish direction of movement of the testing instrument.
- **Track 4 and 5.** Windows milliseconds timer or number of counts of the system clock during one tacho step. The presence (or absence) of the `<DAQClock>` element in the header defines the meaning of these tracks. If the `<DAQClock>` element is missing, the values correspond to the Windows milliseconds timer, else the values correspond to the counts of the system clock.

Scaling the raw data is rather simple: Tracks 1 to 3 are converted to physical values with following formula:

(physical value) = (raw value) * (scaling factor) + (offset)

If tracks 4/5 contain counts of the DAQ clock, this value can be converted to instrument speed by:

(instrument speed) = (Tacho Increment) * (DAQ Clock) / (number of counts)

Reconstructing the instrument speed on the basis of the milliseconds timer is possible, but not reliable.

Detailed instructions to reconstruct the physical data are included in the file itself (see the comment at the beginning of the file) or at the end of each stream file when accessed with a web browser.

The stream file may be accessed with Apple's Safari, Opera or with Mozilla Firefox. Microsoft's Internet Explorer doesn't load the preview pictures because it lacks the support of so called data-links in <img>-tags [5], [4]. We will not provide support for Internet Explorer until Microsoft implements RFC 2397. Before rendering the document, Safari, Firefox or Internet Explorer transform the XML file to a HTML file with the help of an extensible style sheet (XSLT). The style sheet is provided in the `Resources` folder on the top level of each CD or DVD (`iCART-Stream.xsl`).

[2] iCART currently stores 5 tracks, but this may be subject of change.

The file format is defined by a XML schema. This schema is included in the `Resources` folder on the top level of each CD or DVD (`iCART-Stream.xsd`). The complete technical documentation of the iCART-Stream format in HTML format is also included within the `Resources` folder (`iCART-Stream.html`).

The Notes file

The notes file is iCART's storage for notes and measurement values acquired by the tester. This XML file consists of two key elements: The `<Header>` element that stores all relevant informations about the test, including installation and rope. The `<Notes>` element contains an a sequence of `<Entry>` elements.

The file is almost completely human readable, except a `Preview` element, which stores a base 64 encoded JPG preview of imported digital photos.

The notes file may be accessed with Apple's Safari, Opera or with Mozilla Firefox. Microsoft's Internet Explorer doesn't load the preview pictures because it lacks the support of so called data-links in <img>-tags [5], [4]. We will not provide support for Internet Explorer until Microsoft implements RFC 2397. Before rendering the document, Safari, Firefox or Internet Explorer transform the XML file to a HTML file with the help of an extensible style sheet (XSLT). The style sheet is provided in the `Resources` folder on the top level of each CD or DVD (`iCART-Notes.xsl`).

The file format is defined by a XML schema. This schema is included in the `Resources` folder on the top level of each CD or DVD (`iCART-Notes.xsd`). The complete technical documentation of the iCART-Stream format in HTML format is also included within the `Resources` folder (`iCART-Notes.html`).

The Analysis File

iCART's analysis file documents the final analysis of a rope test. This XML file consists of several elements that we will not discuss in detail. Most of the elements are designed for internal use by the iCART Analysis Module. The following Elements are of interest for other users:

- **`Header`.** This element stores all informations about the test, including installation, rope, instrument and settings. iCART allows these parameters override the corresponding parameters defined in the stream and notes file.
- **`Diagram`.** This element contains a sequence of `Preview`-elements. Each of these elements contains a base 64 encoded PNG picture of 4 m test data. The preview is speed compensated and, if necessary flipped, to ensure all wire break indications point to positive direction. Signals are low pass filtered and offset corrected to ensure maximal readability. The filter parameters are carefully chosen to guarantee undistorted signals. The data is organized by position, double or multiple passes were cut off.

  If you prefer unfiltered data please consult the corresponding stream file.
- **`Indications`.** This is a complete list of all indications. Each `Entry`-element contains the informations of one indication. The `Entry`-element contains human readable elements, with one exception. The `Preview`-element contains a base 64 encoded JPG preview of imported photographs.

Other elements in the Analysis file are either human readable or private binary elements for internal use by the iCART Analysis module.

The analysis file may be accessed with Apple's Safari, Opera or with Mozilla Firefox. Microsoft's Internet Explorer doesn't load the preview pictures because it lacks the support of so called data-links in <img>-tags [5], [4]. We will not provide support for Internet Explorer until Microsoft implements RFC 2397. Before rendering the document, Safari, Firefox or Internet Explorer transform the XML

file to a HTML file with the help of an extensible style sheet (XSLT). The style sheet is provided in the `Resources` folder on the top level of each CD or DVD (`iCART-Analysis.xsl`).

The file format is defined by a XML schema. This schema is included in the `Resources` folder on the top level of each CD or DVD (`iCART-Analysis.xsd`). The complete technical documentation of the iCART-Stream format in HTML format is also included within the `Resources` folder (`iCART-Analysis.html`).

Data Safety

The public and human readable file format of iCART influences data safety in several aspects. There are desired effects as the excellent long time readability of file - a text file can be read and inspected on any computer system, completely independent of the original software used to create the file. But there is one critical point that must be discussed here. In principle everybody capable in using a text editor may alter one of iCART's raw data files. To detect data manipulation the following techniques are implemented in the iCART software:

- **MD5 checksums.** iCART calculates MD5 checksums even before the data is stored to disk. This prevents data manipulation during program execution. If the checksums are stored in a safe place, data manipulation may be detected at any time. Checksums stored in an unsafe place (as your hard drive for example) may be manipulated by an intruder. The MD5 checksum itself is currently considered as "strong"

- **Redundant data storage.** iCART stores raw data in more than one place. Raw data is stored in previews and in form of binary data. A possible intruder must manipulate previews as well as binary data. Both storages are encoded. A detailed programming knowledge is necessary to do such a task consistantly.

- **Distribution of iCART data.** We will deliver all wire rope testing data to our customers. This procedure will practically prevent any undetectable modification of the raw data.

- **PGP (Pretty Good Privacy) signatures.** iCART calculates PGP signatures for all critical files. The de-facto standard for public cryptography is currently considered as "safe". Any manipulation of the data files can be detected.

We have chosen a different approach to store wire rope testing data than other program manufacturers. Instead of protecting the file by cryptographic encryption methods, iCART writes human readable, standardized XML files. This enables our customers and other testing institutes to access iCART files. Such files are exposed to manipulations and must be protected with state-of-the-art techniques. The above discussed methods do not prevent modifications of the files, but it enables iCART, or any person to detect any modification of critical iCART files. The approach does not have any leaks - even the programmers have no option to modify files without leaving detectable traces. Protecting files by encryption is less safe: The programmer must know method and pass phrase and can, therefore, modify files without leaving traces.

Fault Tolerance

Another reason for choosing XML file as iCART's native file format is fault tolerance. XML files are pure text files and, therefore, can be read under any circumstances, even when parts of it have been altered. XML specifications force XML editors to abort reading at the position of the first error. Therefore, it will be very easy to detect the position of the errors in the file and, because of the well structured format of the file, it will be an easy task for experienced user to repair the file.

The `<Data>` and `<Preview>` blocks itself are not fault tolerant. If defects occur within one ore more of these two elements, neither the preview nor the raw data can be reconstructed. The PNG format as well as the ZIP encoding format include methods to discover defects of the binary data and these methods will prevent processing invalid data.

One of the main advantages of the XML approach is that iCART files are easy repairable with very simple methods and the affected region is limited to the direct environment of the defects. As iCART stores raw data in two layers (`<Data>` and `<Preview>`) there is even a greater chance to reconstruct the full data (perhaps with some accuracy restrictions) in the case of a file defect.

Bibliography

[1] W3C - World Wide Web Consortium. *Extensible Markup Language (XML)*. http://www.w3.org/TR/2006/REC-xml-20060816/ . 2006. Fourth Edition.

[2] W3C - World Wide Web Consortium. *XSL Transformations (XSLT)*. http://www.w3.org/TR/xslt . 1999. Version 1.0.

[3] W3C - World Wide Web Consortium. *Portable Network Graphics (PNG) Specification*. http://www.w3.org/TR/PNG/ . 2003. Second Edition.

[4] Joyce K. Reynolds and Sandy Ginoza. *Internet Official Protocol Standards*. STD 1 [http://rfc.net/std1.html] . 2004.

[5] Larry Masinter. *The "data" URL scheme*. RFC 2397 [http://rfc.net/rfc2397.html] . 1998.

[6] Peter Deutsch. *DEFLATE Compressed Data Format Specification*. RFC 1951 [http://rfc.net/rfc1951.html] . 1996.

[7] Simon Josefsson. *The Base16, Base32, and Base64 Data Encodings*. RFC 4648 [http://rfc.net/rfc4648.html] . 2006.