

The T_EXspecTool for Computer Aided Software Engineering

Stephen E. Oliver

Whilesell Labs
Pinawa, Manitoba
Canada R0E1A0
olivers@aecl.ca

Introduction

This paper reviews the development of the T_EXspec tool, which assists in the development and documentation of quality assured software in a regulated environment. The tool can assist in the development of a broad range of software, but is targeted the development of software that implements mathematical models. The original application relates to the development of models of a repository for Canada's high level nuclear waste, but is not limited to this use. T_EXspec is particularly strong when documenting models and associated programs which rely on mathematical notations to communicate the intent of the software.

Problem Definition

Atomic Energy of Canada Ltd. (AECL) has developed computer programs to model a deep geologic repository for used nuclear fuel [3] [2]. Regulators require that these programs be of demonstrably high quality to support licence applications.

In 1999, the Canadian Standards Association (CSA) adopted standard N286.7 [6] for the development of nuclear safety related computer programs, a scope that includes the AECL models. While the software development process used to date had been considered robust, it required refinement in order to achieve compliance with this standard.

The T_EXspec project seeks to address the issue of compliance with CSA N286.7. The tool supports a compliant software development procedure, while imposing a minimum of additional overhead. While optimised to meet requirements associated with the modeling nuclear fuel waste, it is hoped that T_EXspec will find more common usage.

Several commercial Computer Aided Software Engineering (CASE) tools will support a robust software development methodology, but none provide support for the mathematical notations that are common in scientific models. T_EXspec provides extensive support for this notation.

Software Development Methodology. Although Object Oriented (OO) analysis and design is appro-

priate for documenting many software applications, there are still applications for procedure/flow based software. In particular, some models which are basically linear in structure are best described using structured (non-OO) methodologies.

Many scientific models have, to date, been described using a modified Yourdon/DeMarco methodology [5, 11]. Although OO methods would perhaps be more appropriate for some models, priority is given to the more common Yourdon/DeMarco analysis methodology. Products associated with this methodology are:

- data flow diagrams (DFDs),
- process descriptions (mini-specs),
- structure charts,
- subprogram design descriptions, and
- data dictionary listings.

DFDs and mini-specs comprise the requirements specification, while structure charts and subprogram design descriptions document the design. Data dictionary listings may be separated into requirements and design, or combined into a single product.

Requirements Specification. Figure 1 illustrates the main concepts of data flow diagrams. Diagram 0 shows the input and output "flows" to/from a single "process". This high level abstraction is intended to allow the reader to identify the functions of the complete system. Process 1 is broken into components in Diagram 1. The diagrams may be thought of as a hierarchy, with higher level diagrams having shorter process numbers (i.e. Diagram 1.2.3 is "higher" than Diagram 1.2.3.4) The numbering convention of the diagrams and the processes allows the decomposition to be clearly seen. Once processes are decomposed to a point where they can be clearly specified in a short textual description, they appear on a DFD with a double circle, as Process 1.4 illustrates. Such "atomic" processes are associated with a mini-spec, rather than a lower level diagram.

Data flows, like processes, can be broken up into constituent parts on lower diagrams. In figure 2, for example, "Implements" on Diagram 0 becomes "Measuring Cup", "Bowl", "oven", and "Pan" on

Diagram 1. These flows are associated with multiple processes on Diagram 1, so they must be shown individually.

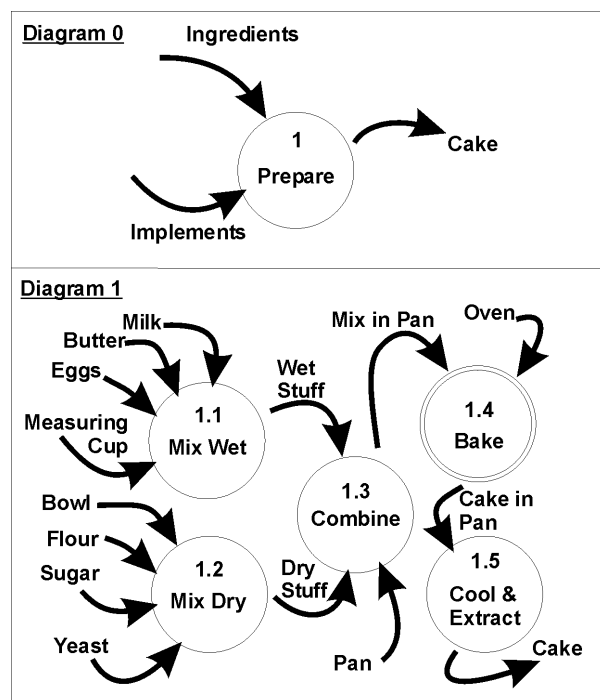


Figure 1: Example Data Flow Diagrams (DFDs)

Mini-specs for each atomic process repeat some of the information on the DFD, and also detail the requirements for a low level process in any manner deemed suitable by the author. For scientific codes, mini-specs often make extensive use of diagrams and mathematical notations.

Design Specification. figure ?? illustrates the main concepts of structure charts. The boxes represent "subprograms" to be composed in a procedural programming language such as FORTRAN. The chart is intended to illustrate the nature of the interface between subprograms. The lines between the subprograms indicate a "calling" relationship, with the subprogram which is closer to the top of the structure chart invoking the lower. The transfer of data at these interfaces is also shown. Data can be passed from one subprogram to another via an argument list (shown along the connecting line), or through common storage that can be accessed from multiple subprograms (shown inside the subprogram box). The interface variables can be input, output, or both, as denoted by arrowheads next to the variable name.

Structure charts do not have a hierarchical organization paralleling DFDs. However, a large struc-

ture chart may span many pages using off-page connectors.

Each subprogram must itself be documented. Subprogram design descriptions repeat some of the information on the structure chart, and document the algorithm and design details of the subprogram. This may include material common with mini-specs, as the design reflects the requirements. In particular, many of the mathematical equations are referenced in both places.

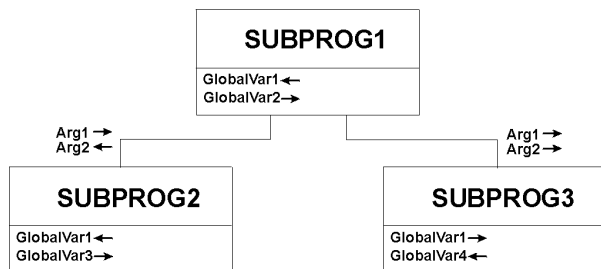


Figure 2: Example Structure Chart

Consistency Between Products. Experience at AECL has shown that lack of consistency between software products has been a major source of software defects [7].

Commercial CASE tools have helped to reduce this inconsistency, but these tools all have difficulty in one or more critical areas:

- Lack of support for scientific and mathematical notations. The nature of scientific software demands that mathematical notations (e.g., $A_i(t) = \int_0^t [F_i^{IN}(\tau)] d\tau$) be permitted in specifications.
- Insufficient accountability. The principle of ownership and accountability for products is not strictly enforced. While a record of who updated products is often kept, the process control is typically inadequate.
- Assembling large products from smaller components is not adequately supported. Many defects originate as transcription errors between products. Mathematical equations are particularly susceptible.
- Insufficient consistency checking between products.

The T_EXspec Solution

T_EXspec takes advantage of the plain text nature of L^AT_EX input to permit processing and tracking of shared components. The main T_EXspec processing is performed by modules which have been implemented in PERL [10], as indicated in figure ??.

A graphical user interface (GUI) captures interactions with the user. Most of this interaction consists of displaying and manipulating ‘component’ files, which form the inputs for the T_EXspec scripts that select components and assemble them into products. These products are primarily L^AT_EX or Noweb [8] input files, which can be post-processed to produce output suitable for viewing, printing, or compiling.

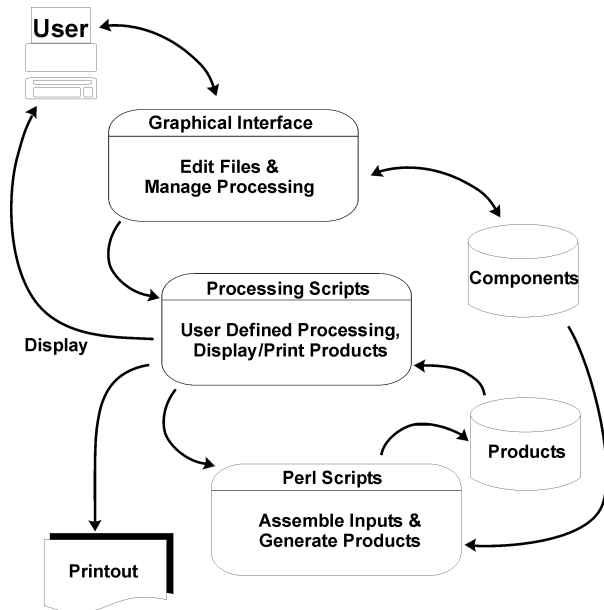


Figure 3: T_EXspec Architecture

While the GUI is a convenient way to construct components and initiate processing, it can be bypassed if required. The components can be generated by any means that can generate a plain text output file, including a text editor. More importantly, the processing can be controlled by any means that can initiate a process, with no requirement for interaction with a GUI. When processing many components, or when a log of processing is required, this ‘batch’ style processing is a useful alternative.

Neither the T_EXspec scripts, nor the GUI can display or print the products. Figure ?? indicates that an intermediate script, which is intended to be edited by the user, initiates T_EXspec to produce the product files, then controls post-processing as appropriate. This flexibility allows the user to integrate T_EXspec into existing procedures. For example, if a static code analyzer such as Floppy [1] is in use, it can be run automatically on code as it is generated. Interaction with a version control system might be desired, or the user may even wish to compile code as it is generated. Alternatively, processing

that is not needed can be removed, such as removing documentation generation (including L^AT_EX processing) until the code is stable.

In order to support sharing of equations and data definitions, while tracking ownership and responsibility for content, T_EXspec supports a fine granularity of components. Each T_EXspec component is tracked independently by placing each in a unique file which is mapped by the file name to the name of the component, and by the file name ‘extension’ (in the tradition of MS-DOS or CP/M) to the type of component.

T_EXspec components, with associated file name extensions, are:

- Requirements Data Dictionary entries (.rdd),
- Design Data Dictionary entries (.ddd),
- Equations (.teq),
- Data Flow Diagrams (.dfd),
- Mini-Specs (.ms),
- Structure Charts (.sc),
- Subprogram Design Descriptions (.ds), and
- Manuals (.tex).

Data Flow Diagrams. DFDs such as figure ??? are produced using the xy-pic package [9] under L^AT_EX.

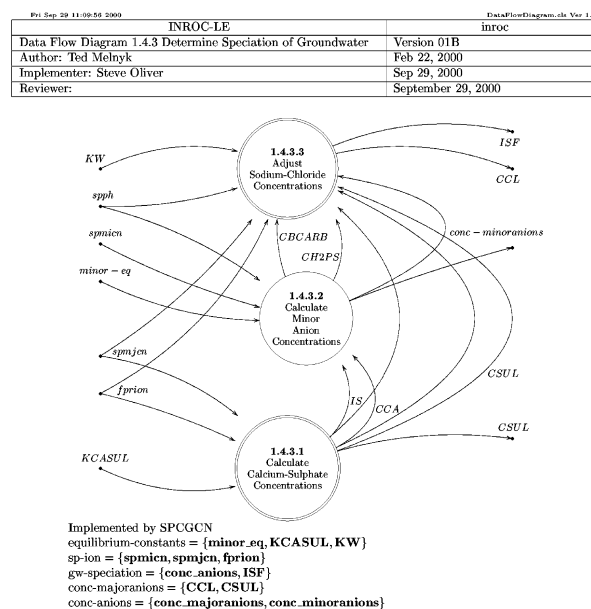


Figure 4: T_EXspec Data Flow Diagram (DFD)

T_EXspec shows the details of composite data flow decomposition explicitly on the DFD. In figure ??, for example, flow “conc-majoranions” (which

would appear on Diagram 1.4.3) is shown as constituent components "CCL" and "CSUL". If a composite flow contains components which do not appear on the current diagram, they are shown in a regular font, while components that do appear on the current diagram are shown in a bold font.

Consistency between DFDs is monitored by T_EXspec. A warning messages is generated for any inconsistency between a DFD and it's parent.

Labels can be shown with mathematical notation, rather than the plain text shown. Switching from plain text to mathematical labels is simple, since flows are taken from the requirements data dictionary (file **name.rdd**), which typically contains both a mathematical plain text label. Although this is an interesting capability, there has been little enthusiasm among users to take advantage of it.

The format of the header in figure 5 is common to all T_EXspec products, detailing the project, the responsible author, implementer, and reviewer, along with an indication of the genealogy of the product (in very small type), which can be used to trace back the source of any defects.

Mini-Specs. Atomic processes are not broken down into lower level DFDs, but are further specified using a mini-spec. This document is intended to be flexible in format, permitting the author freedom to communicate the intent of the process in whatever manner is most effective.

CC4		INROC	
Process: 1.1: Determine Container Failures	Version 01A		
Author: T.E. Molyk	Feb 22, 2000		
Implementer: S.E. Oliver	Apr 10, 2000		
Reviewer:	March 17, 2001		

Determine the number of containers that have failed at the start of the simulation.

Variable	Symbol	Long Name	Units	I/O
NCONS.sec	N_F	Number of containers failed in a sector		O
IFAIL.Q.sec	Q^F	instant container failure quantile		I
IFRACT	P_F	instant failure fraction		I
NCONSC.sec	N_T	containers in a sector		I

The failure probability of any individual container P_F is constant and the same for all containers, so the number of failed containers, N_F , out of N_T total containers is determined from the cumulative binomial distribution:

If $Q^F \leq P(0; N_T, P_F)$

$$N_F = 0 \tag{1}$$

Otherwise determine $N_F \in \{1...N_T\}$ such that

$$P(N_F - 1; N_T, P_F) < Q^F \leq P(N_F; N_T, P_F) \tag{2}$$

Where $P(m; N, p)$ is the cumulative binomial probability distribution. The quantity m is called the number of "successes" (container failures) from N trials (total containers), each having probability p of "success".

$$P(m; N, p) = \sum_{j=0}^m \binom{N}{j} p^j (1-p)^{N-j} \tag{3}$$

This is discussed in the "Container Failures" section of the Inroc Theory Manual.

Figure 5: T_EXspec Mini-Spec (MS)

The standard T_EXspec header is generated for each mini-spec, as shown in figure 5. The author

must explicitly state input and output flows, which are presented in tabular format and verified for consistency with the DFD.

Equations appearing in mini-specs are often referenced in other documents. Authors are encouraged, but not required, to place each equation in a separate file (**name.rdd**), and reference that file from within the mini-spec. The equation can then be reused in subprogram design descriptions or manuals. At AECL, a commercial package is used to create equations that can be saved in L^AT_EX format which includes information encoded as L^AT_EX comments which permits reuse by word processors. By keeping an equation in a single file, available for reuse, transcription errors are reduced.

Structure Charts. T_EXspec structure charts such as figure 6 are produced using the xy-pic package [9] under L^AT_EX. Subprograms can be grouped

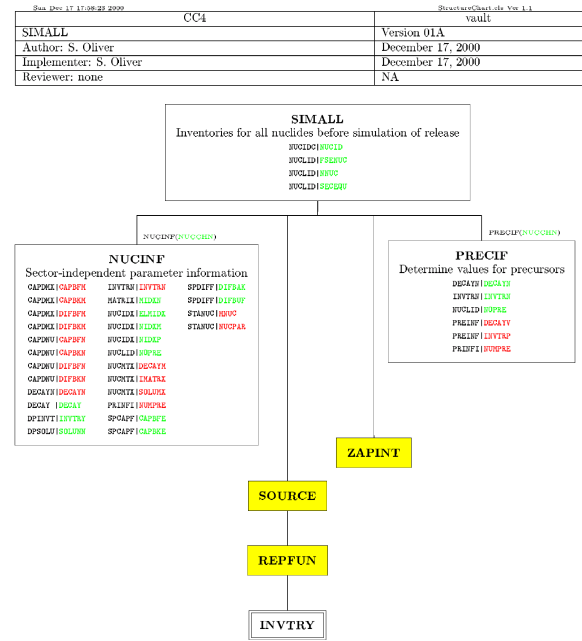


Figure 6: T_EXspec Structure Chart

using colour coded backgrounds. In figure 6, subprograms "SOURCE", "ZAPINT", and "REFFUN" are grouped with a yellow background indicating that they are library routines, not part of the software being documented.

Input and output variables are also colour coded: green for input, red for output, and blue for both. This applies to both the arguments to a subprogram and to the common storage variables.

FORTTRAN groups common storage variable into named blocks, which are indicated to the left of each

and compilable code is guaranteed, since Noweb extracts the code from the subprogram design description. Information for each subprogram on a structure chart is also extracted from the subprogram design description, ensuring that all design documentation is consistent.

Data Dictionaries. T_EXspec distinguishes between dictionary entries for requirements and design specification. Some design information is never applicable to requirements (e.g. a common storage block name).

It is possible to have a close correlation between entries in the requirements and design dictionaries. Design entries may optionally state a requirements dictionary entry which is related. When this is done, fields in the design dictionary acquire default values equivalent to the requirements data dictionary. This is particularly useful to inherit the mathematical symbol and description.

T_EXspec produces a data dictionary listing which can show a cross reference of which products use which dictionary entries.

Graphical User Interface. There is a considerable amount of data contained in many plain text files in a typical T_EXspec documented project. To assist users, a GUI has been developed as a Java application to act as a front end to the process. While there is little new technology embedded in the GUI, it is interesting to note that the GUI, at over 20,000 lines of code, is much larger than the T_EXspec scripts.

An example screen is shown in figure ???. Here, the structure chart "SIMALL" from figure ??? is being edited in the upper window. A subprogram on the chart is being modified in the lower window. The GUI can open many windows, so it is contained within an application desktop, which produces only one icon on the users desktop.

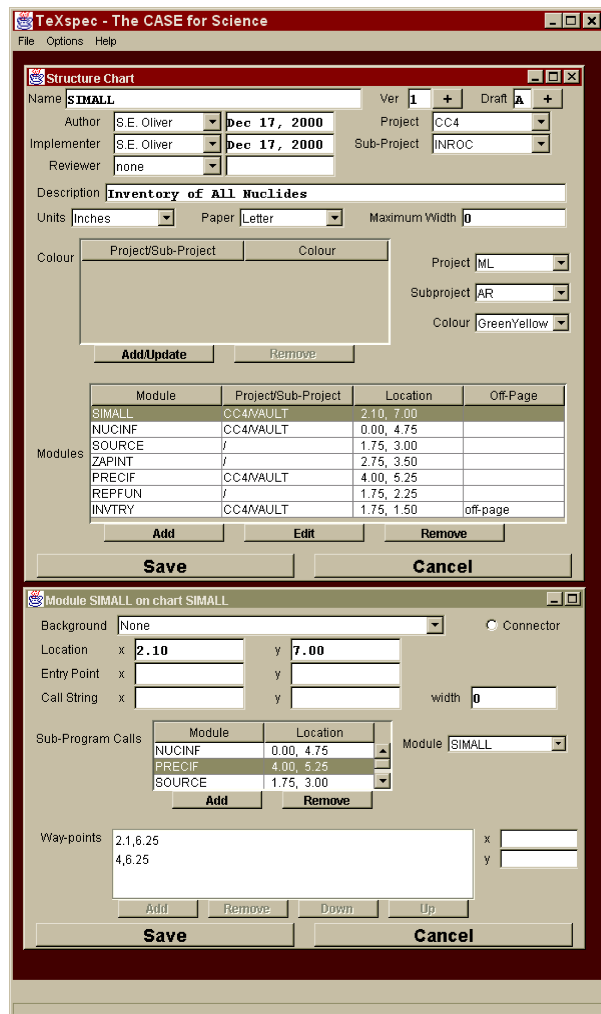


Figure 8: T_EXspec GUI editing a Structure Chart

When editing a subprogram (module) on a structure chart, the user specifies the subprogram name, and sets a position in x,y coordinates. If the user wants to show an entry point on the chart to this subprogram (useful for charts with multiple entry points), then the location of the entry point must be specified. The location of the call string, and the maximum width of that string is also entered. Calls to other subprograms can be added from a drop-down list of all available modules. For each called subprogram, "waypoints" determine the shape of the line connecting the two boxes.

Similar editing capability is provided for the overall structure chart in the upper window. Selecting a subprogram from the scroll list at the bottom of the upper window causes the lower window to appear.

Future Development

The next stage of T_EXspec development will be to add some object oriented programming extensions, and a rudimentary interface between the GUI and an Integrated Development Environment (IDE) to assist in debugging and performance analysis.

The plain text files that store T_EXspec data are formatted to be human readable and editable. This allowed T_EXspec to be used before the GUI was developed. With the advent of a GUI to interface with this data, the file format may be redefined to an XML syntax.

The perl scripts may be reimplemented in Java, to permit a more seamless interface between the GUI and the main application.

The application may be divided into a client and a server. This would improve performance, assist in sharing data between users and projects, and provide more robust auditing and version tracking. The system could allow installation of files into a configuration management system. Dependencies between files would be monitored by the server, and ownership would be enforced.

A number of extensions may be made to the GUI, including preview capability for mathematical notations.

More diagram types and programming languages may be supported. In particular, object oriented diagrams may be added, and the full FORTRAN-9x, Java, or Perl syntax may be added.

The GUI support for the graphical products (Data Flow Diagrams and Structure Charts) could be based on editable graphics, or perhaps provide a 'preview' window. Having to process the file to see the format of the output is not a optimal.

Some allowance for formal tracing between design and requirements could be provided.

Conclusion

T_EXspec provides a workable solution to computer aided software engineering requirements that are peculiar to scientific programs. It is a significant quality assurance device for these programs.

T_EXspec is in use on several projects relating to modeling the disposal of Canada's nuclear waste. As such, it is a working tool, but is still in the early phases of development. Further enhancement will improve the capability of meeting quality assurance requirements imposed by standards such as CSA N286.7.

References

[1] J.J. Bunn. Floppy and flow user manual. 1997.

- [2] B.W. Goodwin, T.H. Andres, D.C. Donahue, W.C. Hajas, S.B. Keeling, C.I. Kitson, D.M. LeNeveu, T.W. Melnyk, S.E. Oliver, J.G. Szekely, A.G. Wikjord, K. Witzke, and L. Wojciechowski. The disposal of canada's nuclear fuel waste: A study of postclosure safety of in-room emplacement of used candu fuel in copper containers in permeable plutonic rock. volume 5: Radiological assessment. Technical Report AECL-11494-5, COG-95-552-5, Atomic Energy of Canada Ltd, 1996.
- [3] B.W. Goodwin, D.B. McConnell, T.H. Andres, W.C. Hajas, D.M. LeNeveu, T.W. Melnyk, G.R. Sherman, M.E. Stephens, J.G. Szekely, P.C. Bera, C.M. Cosgrove, K.D. Dougan, S.B. Keeling, C.I. Kitson, B.C. Kummer, S.E. Oliver, K. Witzke, L. Wojciechowski, and A.G. Wikjord. The disposal of canada's nuclear fuel waste: Postclosure assessment of a reference system. Technical Report AECL-10717, COG-93-7, Atomic Energy of Canada Ltd, 1994.
- [4] D.E. Knuth. *Literate Programming*. Center for the Study of Language and Information, 1992.
- [5] D.M. LeNeveu. Analysis specifications for the cc3 vault model. Technical Report AECL-10970, COG-94-100, Atomic Energy of Canada Ltd, 1994.
- [6] Quality assurance of analytical, scientific, and design computer programs for nuclear power plants. Technical Report N286.7-99, Canadian Standards Association, 178 Rexdale Blvd. Etobicoke, Ontario, Canada M9W 1R3, 1999.
- [7] S. Oliver, K. Dougan, K. Kersch, C. Kitson, G. Sherman, and L. Wojciechowski. Unit testing - a component of verification of scientific modelling software. In T.I. Oren and G.B. Birta, editors, *1995 Summer Computer Simulation Conference*, pages 978-983. The Society for Computer Simulation, 1995.
- [8] N. Ramsey. Literate programming simplified. *IEEE Software*, September 1994.
- [9] K. Rose. Very high level 2-dimensional graphics. In *1997 TeX User Group Conference*. TeX User Group, 1997.
- [10] L. Wall, T. Christiansen, and R. Schwartz. *Programming Perl*. O'Reilly & Associates, 101 Morris Street, Sebastopol, CA 95472, second edition, 1989.
- [11] E. Yourdon. *Modern Structured Analysis*. Yourdon Press/Prentice-Hall, 1989.