COSMIC Measurement Manual for ISO 19761

Part 3d: Standardized Requirements with Big Data Cleaning Examples

Version 5.0

May 2022

Foreword.

The COSMIC Measurement Manual for ISO/IEC 19761:2011 consists of the Parts:

Part 1: Principles, definitions & rules* (17 pages)
Part 2: Guidelines* (18 pages)
Part 3: Examples of COSMIC concepts and measurements, consisting of:

- Part 3a Standard Measurement Strategy Examples (13 pages)
- Part 3b Real-time Examples (32 pages)
- Part 3c MIS Examples. (58 pages)
- Part 3d Standardized Requirements with Big Data Cleaning Examples (12 pages)

* Parts 1 and 2 describe the entire material necessary for certification.

A public domain version of the COSMIC Measurement Manual and other technical reports, including translations into other languages, can be found at the Knowledge Base of www.cosmic-sizing.org.

Purpose of this document.

The purpose of this document is to show COSMIC can be used to define structured requirements (called 'standardized FUR'), to specify Big Data Cleaning software examples with these requirements and that size follows, i.e. that a separate measurement of this standardized FUR is unneeded. See the Introduction to learn that standardized FUR is also useful when there is no need for size.

The COSMIC Group recommends that users study and master the COSMIC method, before applying the specification of standardized FUR.

May 2022 editing.

Section 5.4 Data movements for keys added. Section 5.7 Purpose mentions to merge 'relevant parts', however full tables are merged, specification of relevant columns added. Section 5.8 Requirement was added to display quartile values and enable verification of the results.

Editors.

Alain Abran, Ecole de technologie supérieure – University of Quebec (Canada),
Arlan Lestherhuis (The Netherlands).

Other members of COSMIC Measurement Practices Committee.

Jean-Marc Desharnais, Ecole de technologie supérieure – University of Quebec (Canada),
Peter Fagg, Pentad (UK),
Dylan Ren, Measures Technology LLC (China),
Bruce Reynolds, Tecolote Research (USA),
Hassan Soubra, German University in Cairo (Egypt),
Sylvie Trudel, Université du Québec à Montréal - UQAM (Canada),
Frank Vogelezang, IDC Metri (The Netherlands).

Table of Contents

1 INTRODUCTION.

The approach to a COSMIC measurement is to use functional user requirements (FUR) to determine functional size. To do so, the functional processes with their data movements must be identified. Specification of FUR and measurement are separate activities. However, if the FUR *themselves* are structured according to both COSMIC concepts, the size of such 'standardized FUR' is a consequence of that structure, so a separate measurement is unneeded.

Standardized FUR has several other interesting benefits:

- Applying the structure of the COSMIC concepts results in uniformly structured, 'standardized' specifications of FUR. If size isn't needed it can be ignored, without losing the other benefits.
- Standardized FUR can make user stories concrete. Assigning functional processes to user stories enables an early COSMIC estimate. This estimate can be refined when the data movements have been identified.
- Standardized FUR provides clearly structured specifications, facilitating verification of their completeness, correctness and consistency.
- As standardized FUR and its measurement coincide a measurement strategy is unneeded. It is replaced by a Requirements Agreement, a 'contract' between stakeholders and specifier. The stakeholders know what they get and whether the sizes are comparable to existing sizes. The specifier knows exactly what and how to specify.
- Standardized FUR's explicit structure of functional processes and data movements facilitates adaptation to new requirements and thereby supports 'living documentation'. It also provides an obvious input format for tools, including tools that generate testcases, test scripts and source code.

2 THE REQUIREMENTS AGREEMENT.

This section describes the key parameters that must be considered before the actual specification of requirements for a piece of software.

Capture

- the purpose of the specification and measurement exercises;
- the functional users of the software to be realized or changed, and the types of data movements that the software may handle;
- if applicable the layered architecture and components separately to be specified;
- the artefacts which will be needed for the specifications exercise.

Note. It is helpful to draw a context diagram of the software being specified.

3 APPLYING THE COSMIC CONCEPTS STRUCTURE.

3.1 Specification of a functional process

Specify

- triggering event (omit if the triggering event obviously follows from the name of the functional process)
- name of the functional process
- functional user(s)

3.2 Specification of a data movement

Note. Requirements that in a measurement are accounted for by the data movement functionality are indicated by '(*)'.

3.2.1 Specification of Entry functionality

Specify

- the data group of the triggering Entry and possibly other input data groups
- the functional user(s)
- (*) the required formatting and presentation manipulations along with all associated validations of the entered data attributes, to the extent that these data manipulations do not involve another type of data movement

If one of more Reads are required as part of the validation process, these are specified as separate Read data movements

3.2.2 Specification of Exit functionality

Specify

- the data group(s) moved to the functional user(s)
- (*) the required data formatting and presentation manipulations, including processing required to send the data attributes to the functional user, to the extent that these manipulations do not involve another type of data movement.

3.2.3 Specification of Read functionality

Specify

- the data group(s) to be retrieved from persistent storage
- (*) the logical processing and/or mathematical computation needed to read the data, to the extent that these manipulations do not involve another type of data movement

3.2.4 Specification of Write functionality

Specify

- the data group(s) to be stored to persistent storage

- the data group(s) to be deleted or validity to be ended
- (*) the logical processing and/or mathematical computation to create the data attributes to be written, to the extent that these manipulations do not involve another type of data movement.

3.3 Specification of functional changes

For details about measurement of changes, see Part 2 of the Measurement Manual, section ‘Measurement of the size of changes to software’.

4 EXAMPLE SPECIFICATION

(Abbreviations: TEv = triggering event, E = Entry, X = Exit, R = Read, W = Write, FU = functional user).

Note. Pay particular attention to updating the standardized FUR after realization or change of the software.

4.1 Example MIS specification

(from the C-Reg Case study)

FP: Add a Professor.			Size: 4 CFP
Specification	*Data group*		
Registrar (FU) enters details for the Professor	Professor details	E	1
C-Reg checks if the data describe a Professor who already exists and if so, displays an error message, else validates the entered data	Professor details	R	1
C-Reg creates a new Professor	Professor details	W	1
Display error message	Error Message	X	1

4.2 Example Real-time specification

(from the Rice cooker Case study)

FP: Start cooking (start button pressed)			Size: 3 CFP
Specification	*Data group*		
When the start button (FU) is pressed it sends a start signal to the software	Start signal	E	1
The software sends a Turn ON command to the Heater (FU)	Turn ON command to Heater	X	1
The software sends a Turn ON command to the Cooking Lamp (FU)	Turn ON command to Cooking Lamp	X	1

4.3 Example change of Real-time specification

(see the Rice cooker Case study)

To the first prototype a hardware ‘convenience timer’ has been added: a start signal can also be provided by this new timer. Cooking can start at a convenient point of time set by the human use (function implemented by the hardware). To prevent malfunction the software must ensure that any new start signal is ignored when the cooker is processing.

FP: Start cooking (TEv: button or convenience timer start) Size of change: 2 CFP			
Specification	*Data group*		
When the start button (FU) is pressed it sends a start signal to the software.	Start signal	E	-
At the set time the convenience timer (FU) sends a start signal to the software (data movement added).	Start signal	E	1
Ignore any start signal if the cooker has been started (data movement added).	Cooking mode	R	1
The software sends a Turn ON command to the Heater (FU)	Turn ON command to Heater	X	-
The software sends a Turn ON command to the Cooking Lamp (FU)	Turn ON command to Cooking Lamp	X	-

5 BIG DATA CLEANING EXAMPLE SPECIFICATION

5.1 The Requirements Agreement.

The Requirements Agreement for the Big Data cleaning software could be:

Purpose of the specifications exercise. Purpose of the specifications is twofold. The specifications describe the functionality of the software to be realized. In addition, the integrated functional size is input for estimating effort of its realization.

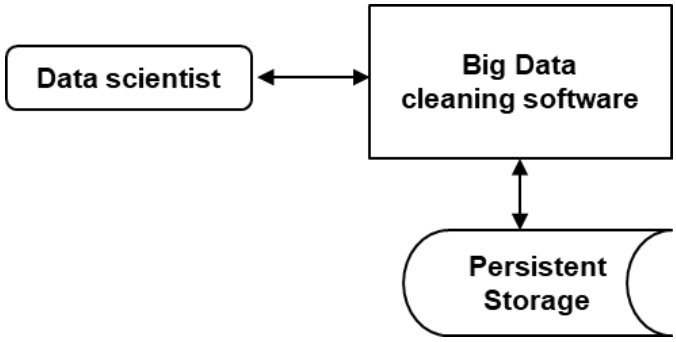
Functional users of the software. In the examples the data scientist is the only functional user.

Types of data movements that the software may handle. The types of data groups pertain to the data, their location (i.e. data set IDs and/or column IDs), defects, constraints and phenomena that the data scientist needs to cope with.

Layered architecture and components. Not applicable.

Artefacts needed for the specifications exercise. Not applicable. Usually the specifications exercise will be based on interview reports of data managers.

The context diagram of the software being specified.



5.2 Record cleaning operation details.

Purpose: Keep record of every cleaning operation (in plain text).

FP: Create cleaning operation.			**Size: 3 CFP**
Specification	*Data group*		
Data scientist (FU) enters details of a cleaning operation	Cleaning operation details	E	1
Store cleaning operation details	Cleaning operation details	W	1
Display error message	Error Message	X	1

FP: Display list of cleaning operations.			**Size: 4 CFP**
Specification	*Data group*		
Data scientist (FU) enters cleaning operation search term	Cleaning operation search term	E	1
Read cleaning operation details	Cleaning operation details	R	1
Display cleaning operation details	Cleaning operation details	X	1
Display error message	Error Message	X	1

FP: Display cleaning operation details.			**Size: 4 CFP**
Specification	*Data group*		
Data scientist (FU) enters cleaning operation ID	Cleaning operation details to be changed	E	1
Read cleaning operation details	Cleaning operation details	R	1
Display cleaning operation details	Cleaning operation details	X	1
Display error message	Error Message	X	1

FP: Update cleaning operation.			**Size: 3 CFP**
Specification	*Data group*		
Data scientist (FU) enters changed details of the cleaning operation	Cleaning operation changed details	E	1
Store changed cleaning operation details	Cleaning operation changed details	W	1
Display error message	Error Message	X	1

5.3 Test column data quality.

Note: in the following ‘constraint’ includes ‘text pattern’

FP: Check Column data quality			**Size: 6 CFP**
Specification	*Data group*		
Data scientist (FU) enters data set ID	Data set ID	E	1
Data scientist (FU) enters column ID	Column ID	E	1
Data scientist (FU) enters constraint to compare with	Constraint	E	1
Read value	Value	R	1
Check constraints and display violation	Constraint violation	X	1
Display error message	Error Message	X	1

5.4 Test relationship foreign key

FP: Test relationship foreign key			**Size: 8 CFP**
Specification	*Data group*		
Data scientist (FU) enters foreign data set ID	Foreign data set ID	E	1
Data scientist (FU) enters foreign key ID	Foreign key ID	E	1
Data scientist (FU) enters primary data set ID	Primary data set ID	E	1
Data scientist (FU) enters primary key to compare with	Key ID	E	1
Read foreign key value	Foreign key value	R	1
Read primary key value	Primary key value	R	1
Display key violation	Both data set IDs and foreign key	X	1
Display error message	Error Message	X	1

5.5 Test business rule.

Note: in the following prime customers (marked somehow) get a given percentage discount on an order.

FP: Test business rule			**Size: 6 CFP**
Specification	*Data group*		
Data scientist (FU) enters data set ID of prime customers	Data set ID of prime customers	E	1
Data scientist (FU) enters selection of prime customer	Customer type to be tested	E	1
Read customer data	Customer data	R	1
Read associated bill data	Customer bill data	R	1
Display business rule violation	Customer ID, bill ID, actual percentage	X	1
Display error message	Error Message	X	1

5.6 Test data violations in data set.

Purpose: to test a data set before processing it.

FP: Test data violations in data set			**Size: 6 CFP**
Specification	*Data group*		
Data scientist (FU) enters data set ID	Data set ID	E	1
Data scientist (FU) enters column ID	Column ID	E	1
Data scientist (FU) enters constraint to be tested	Constraint	E	1
Read value and test constraint	Value to be tested	R	1
Display violation	Value, violated constraint ID	X	1
Display error message	Error Message	X	1

5.7 Merging datasets.

Purpose: combining relevant parts of datasets to create a new file.

FP: Merging datasets			**Size: 10 CFP**
Specification	*Data group*		
Data scientist (FU) enters first data set ID to be merged	First data set ID	E	1
Data scientist (FU) enters linking column ID of first data set to be merged	Linking column ID of first data set	E	1
Data scientist (FU) enters column ID of first data set to be merged	Column ID	E	1
Data scientist (FU) enters second data set ID to be merged	Second data set ID	E	1

Data scientist (FU) enters linking column ID of second data set to be merged	Linking column ID of second data set	E	1
Data scientist (FU) enters column ID of second data set to be merged	Column ID	E	1
Read first data set	Record of first data set	R	1
Read second data set	Matching record of second data set	R	1
Merge records to new data set	Merged record	W	1
Display error message	Error Message	X	1

5.8 Detecting Outliers.

FP: Detect Outliers.			**Size: 6 CFP**
Specification	*Data group*		
Data scientist (FU) enters data set ID	Data set ID	E	1
Data scientist (FU) enters ID of column to be tested	Column ID	E	1
Read value	Value	R	1
Compute and display the quartiles Q1 and Q3 and the interquartile range	Quartile info	X	1
Read value	Movement of data accounted for by previous Read	-	-
Compare value with interquartile range and display it if outside interquartile range	Data group including outlier	X	1
Display error message	Error Message	X	1

5.9 Test for duplicates.

FP: Test for duplicates.			**Size: 5 CFP**
Specification	*Data group*		
Data scientist (FU) enters data set ID	Data set ID	E	1
Data scientist (FU) enters column ID(s) of combination of column(s) that should be different	Column ID	E	1
Read data set record	Data set record	R	1
Display duplicate if combined values equal	Data group with possible duplicate	X	1
Display error message	Error Message	X	1

5.10 Data transformation.

Transform data and store the result into an added column

FP: Data transformation.			**Size: 6 CFP**
Specification	*Data group*		
Data scientist (FU) enters data set ID	Data set ID	E	1
Data scientist (FU) enters the column ID of the data to be transformed	Column ID	E	1
Data scientist (FU) enters transformation rule	Transformation rule	E	1
Read value	Value	R	1
Transform value and store it	Transformed value	W	1
Display error message	Error Message	X	1

5.11 Test for missing value.

FP: Test for missing value.			**Size: 5 CFP**
Specification	*Data group*		
Data scientist (FU) enters data set ID to test for missing value	Data set ID	E	1
Data scientist (FU) enters the column ID to test for missing value	Column ID	E	1
Read data record	Value	R	1
Notify if value is missing	Data record ID	X	1
Display error message	Error Message	X	1