

Measure Instrument for Heterogeneous Data Sources Extraction

Watcharee Chitvarodom ¹⁺ and Yachai Limpiyakorn ²

1 Department of Computer Engineering

Chulalongkorn University, Bangkok 10330, Thailand

2 Department of Computer Engineering

Chulalongkorn University, Bangkok 10330, Thailand

Abstract. Measurement and Analysis process is vital for mature organizations. It supports several activities of project management, process management, and quality management. However, data collection in the measurement process is resource consuming and error prone. Moreover, the measurement data may be created in different formats. Extracting the data from heterogeneous sources is thus challenging. This research proposes a method to gather the measurement data from the sources with compatible type under XML. The measurement tags are inserted to locate the position. The extracted data can be exchanged with other applications, imported into a visualization tool for analysis, or exported to the repository. The approach suits for mature organizations that own process assets such as Measurement Specification, measurement plan, and SLCP.

Keywords: Measurement, Data Extraction, Process asset, Process Improvement, XML

1. Introduction

Measurement and Analysis [1] is one of the essential processes in mature organizations. It supports project parameter estimation, project progress monitoring and control, risk management, quality management activities, and so on. For the success program established, the measurement objectives are suggested to be aligned with the organization's business goals, which in turn, determine the information needs be collected from the projects.

The best practices recommended for the Measurement and Analysis process include: Establish measurement objectives, Specify Measures, Specify data collection and storage procedures, Specify analysis procedures, Collect and analyze measurement data, Store data and results, and Communicate results [1]. As a consequence, the organization's measurement repository is established and maintained to store the collected measurement data, the measures or metrics, and the analysis results. In addition, the measurement plan is created before project initiation to specify the measures and the data need be collected from the work products attached to the activities performed during the project life cycle.

This research proposes a method and implements a component for collecting the measurement data from sources in various formats. XML is used as the technique to facilitate data gathering and exchanging. Replacing the manual data collection chore with automation could reduce resource consumption, enhance data consistency, and promote the improvement of organization's processes.

2. Measurement Information Model (Mim)

⁺ Corresponding author. Tel.: (66)-86-965-1812
E-mail address: cwatcharee@gmail.com.

Measurement Information Model [2] presents the hierarchical structure of building information products for responding to information needs. Information products integrate measurement results with decision criteria and present recommendations to project decision makers on alternative courses of action.

MIM can be considered as the mechanism for linking defined information needs to project processes and products, or the entities that can actually be measured. The evolution of information needs to the project measurement plan starts from identifying the measurable concept that addresses the defined information need. The measurable concept will then be formalized as a measurement construct that specifies exactly what will be measured and how the data will be combined to produce results that satisfy information need. Levels of a measurement construct consist of Base measure, Derived measure, and Indicator. Attributes of an entity (e.g. process, product, project) are measured by humans or automated means. Example attributes are size, effort, amount of defects. A Base measure of a single attribute is defined by a specified measurement method. Executing the method produces a value for the base measure, for example, software size, number of defects found/ removed. Data collection involves assigning values to base measures, specifying scales, and/or defining unit of measures. A scale is an ordered set of values, continuous or discrete, or a set of categories to which an attribute is mapped. Types of scale include: 1) Ratio (numeric data ranging from 0- ∞); 2) Interval (numeric data as minimum value is 1); 3) Ordinal (discrete ranking); and 4) Nominal (categorical data). Measurement Specification (Fig.1) is an organizational process asset (OPA) established to guide how to measure the attributes of individual entities. It also informs all the measurement constructs defined for the organization.

Entities and Attributes	
Relevant Entities	Effectiveness of design inspection
Attributes	1. Staff effort in design inspection 2. Defect in design inspection
Base Measure Specification	
Base Measures	1. Preparation time 2. Overview time 3. Inspection time 4. Total defects
Measurement Methods	Calculate all spent effort hours per defect found in this inspection process
Type of Method	Objective
Scale	0 to ∞
Type of Scale	Ratio
Unit of Measurement	Hours, Defects

Entity as a property relevant to information need and compound with attributes

Measures of attributes results from using measurement method

Fig. 1: Example of Measurement Specification [2]

3. Research Methodology

This research has focused on the step of gathering measurement data or base measures from the artifacts (e.g. form, template, checklist) attached to software life cycle (SLC). Referring to IEEE1074 [3], software life cycle is defined as the project-specific sequence of activities that is created by mapping the activities of IEEE/EIA 12207.0 [4] onto a selected software life cycle model (SLCM). Artifacts that help define an organization's software project environment are called organizational process assets (OPAs), for example, policies, standards, procedures, metrics, tools, forms, checklists, and the history of previous projects. Once the SLC has been developed, the available OPAs are applied to the SLC to establish the software life cycle process (SLCP). In this work, it is assumed that Measurement Specification exists. The measurement plan and SLCP are established before project initiation. The project measurement plan contains the list of measurement data needs be collected throughout the SLC. The measurement method of a particular base measure that is contained in Measurement Specification, describes how to gather the data, whether the method is manually or automatically; whether the measure is subjective or objective; type of scale; and unit of measures. The sources of data extraction that are the OPAs attached to the project SLCP are identified by a measurement specialist.

The approach presented in this research requires that the segments in sources containing the measurement data will be formerly tagged by a measurement specialist using the XML (eXtension Markup Language) format as it supports exchanging and integrating data from various document formats. The sources of data extraction are thus limited to those that can be converted into XML files, namely those files

created by Microsoft office, or Open Office etc. Fig.2 shows an example of XML tags that notify the extraction location associated with the details of the measurement method of a base measure as described previously.

Inspection Id: _____ Project id: _____ Project name: _____ Product component/ software item: _____ Material size: _____ (pages/ LOC)	<pre> <ProcessAsset xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" <Landmark isTable="no" startMarker="inspection id" endMarker=""> <MeasureAttribute name="inspection id"> <ScaleType name="Nominal" /> </MeasureAttribute> </Landmark> </pre>
--	---

Fig. 2: Inspection checklist header and underlying XML measurement tags

4. Implementation

The functionalities of the instrument for extracting data from heterogeneous sources are described with UML 2.0 Use Case diagram as shown in Fig. 3

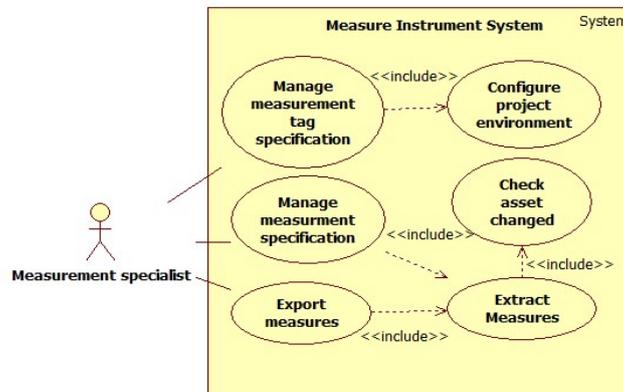


Fig. 3: Use case diagram of data extraction component

A user role as a measurement specialist is responsible for inserting the XML measurement tags into OPAs guided by Measurement Specification and the list of all of the OPAs contained in the Organization's Process Asset Library.

During the project execution, the data sources will be produced. The Extraction module will monitor the SLCP whether any data sources are created or modified. The module will then extract the data located by the XML measurement tags. The extracted details in XML format can be exported for further analysis or usages. For example, they can be imported into MS-Excel for visualization to provide better understanding, or exported to store in the measurement repository.

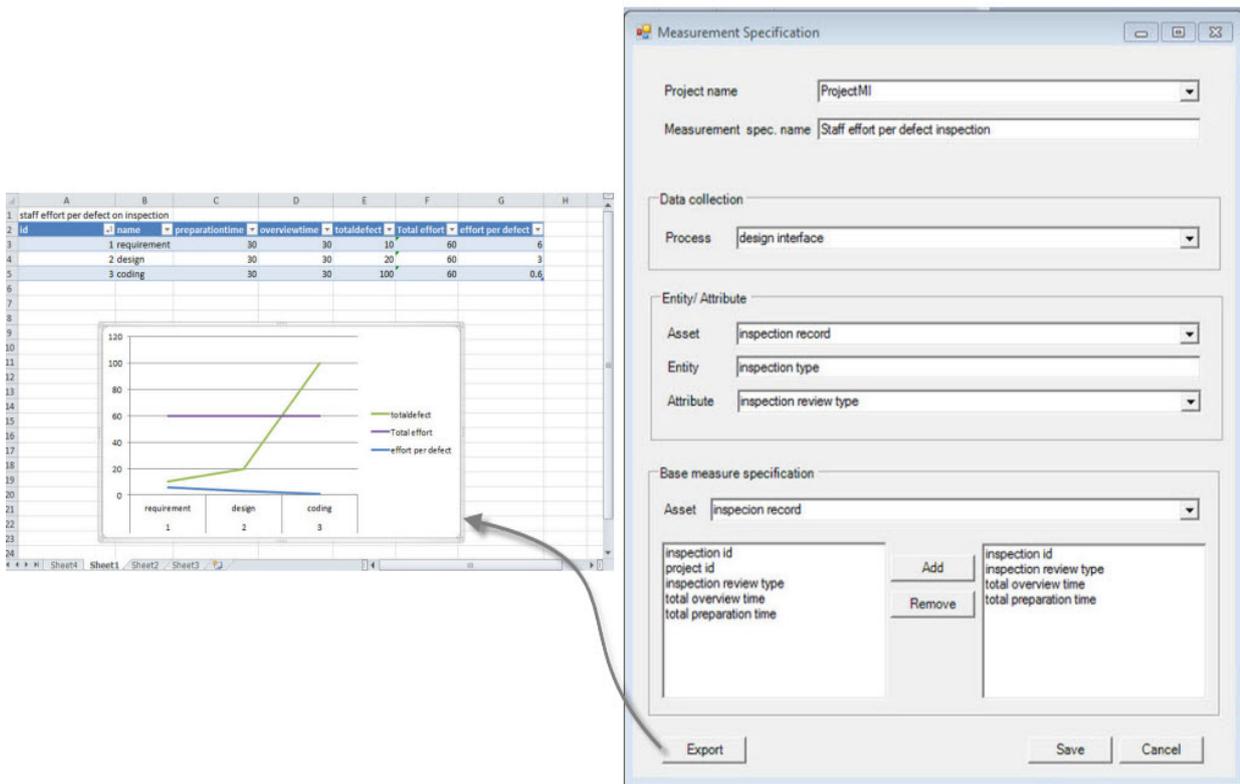


Fig. 4: Screens of project SLCP and measurement tag specification

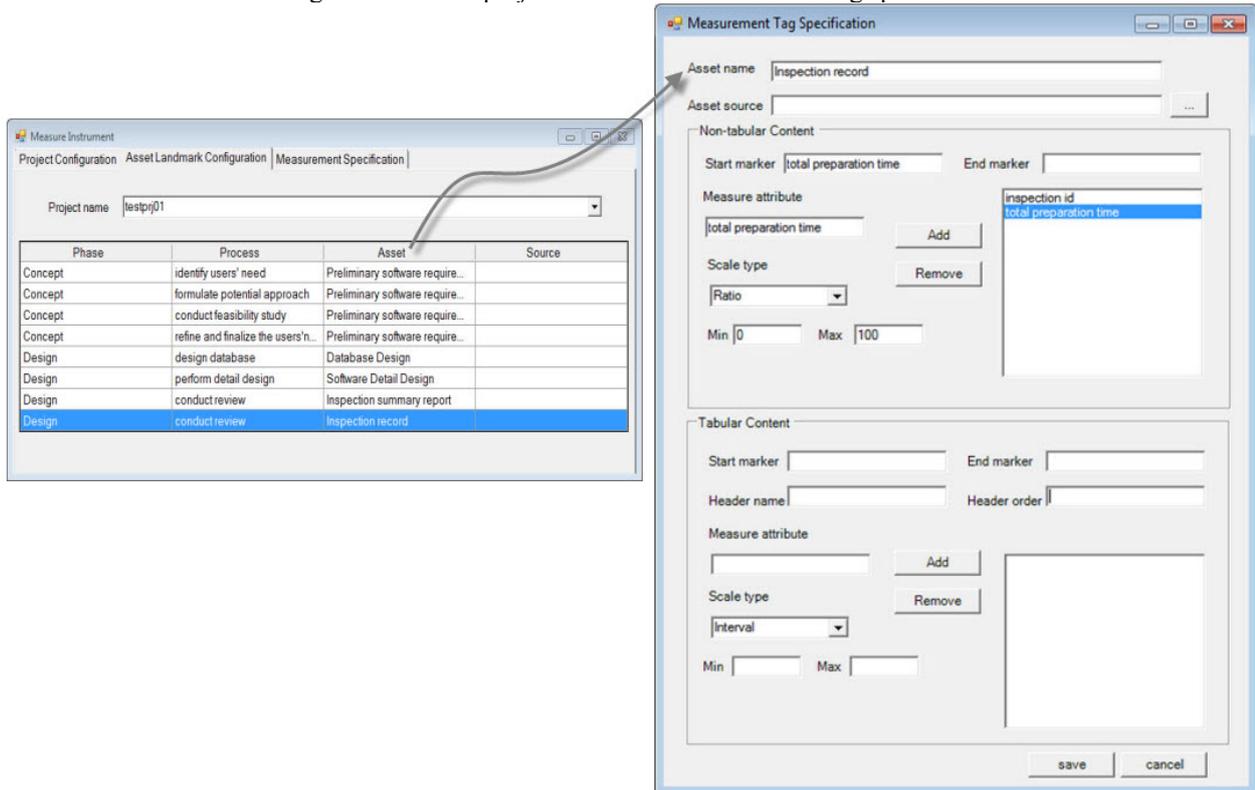


Fig. 5: Screens for exporting data for analysis and the resulting chart

illustrates the user interface for defining the measurement tag specification of the selected data source or OPA listed on the project SLCP. Fig. 5 shows the screen for exporting the collected data to store in the measurement repository or further analysis, such as presenting with a graph.

5. Conclusion

Measurement should begin early if it is to reach its potential fully, and measurement capability must grow over time [5]. Automation could help enhance the capability and productivity of the measurement and analysis process. This paper presents a method and develops a component for extracting measurement data from heterogeneous sources. The XML technology is deployed to help gathering and exchanging data. It is thus required that the data sources can render the XML format. The XML measurement tags are inserted into the sources to locate the position for the extraction module. The approach suits for mature organizations of which the OPAs, such as Measurement Specification, project measurement plan, and SLCP, are established. The collected data can be exported for analysis or stored in the measurement repository.

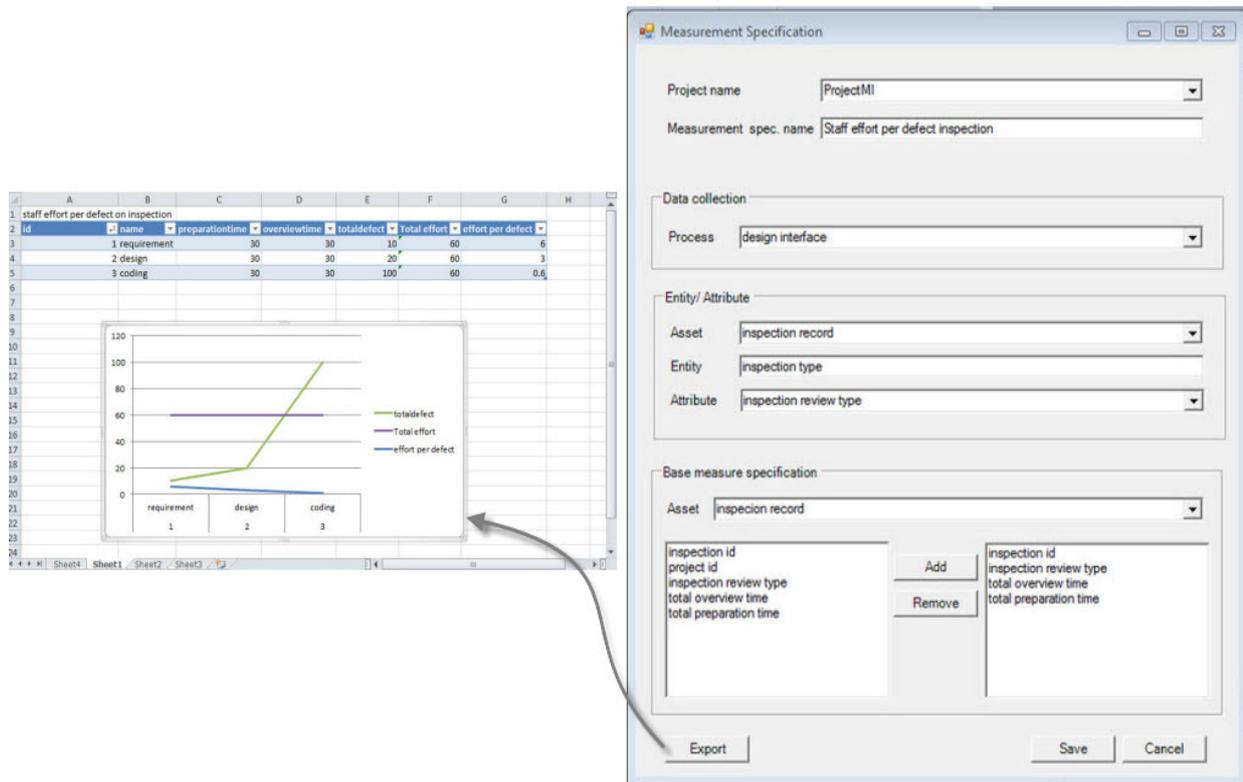


Fig. 4: Screens of project SLCP and measurement tag specification

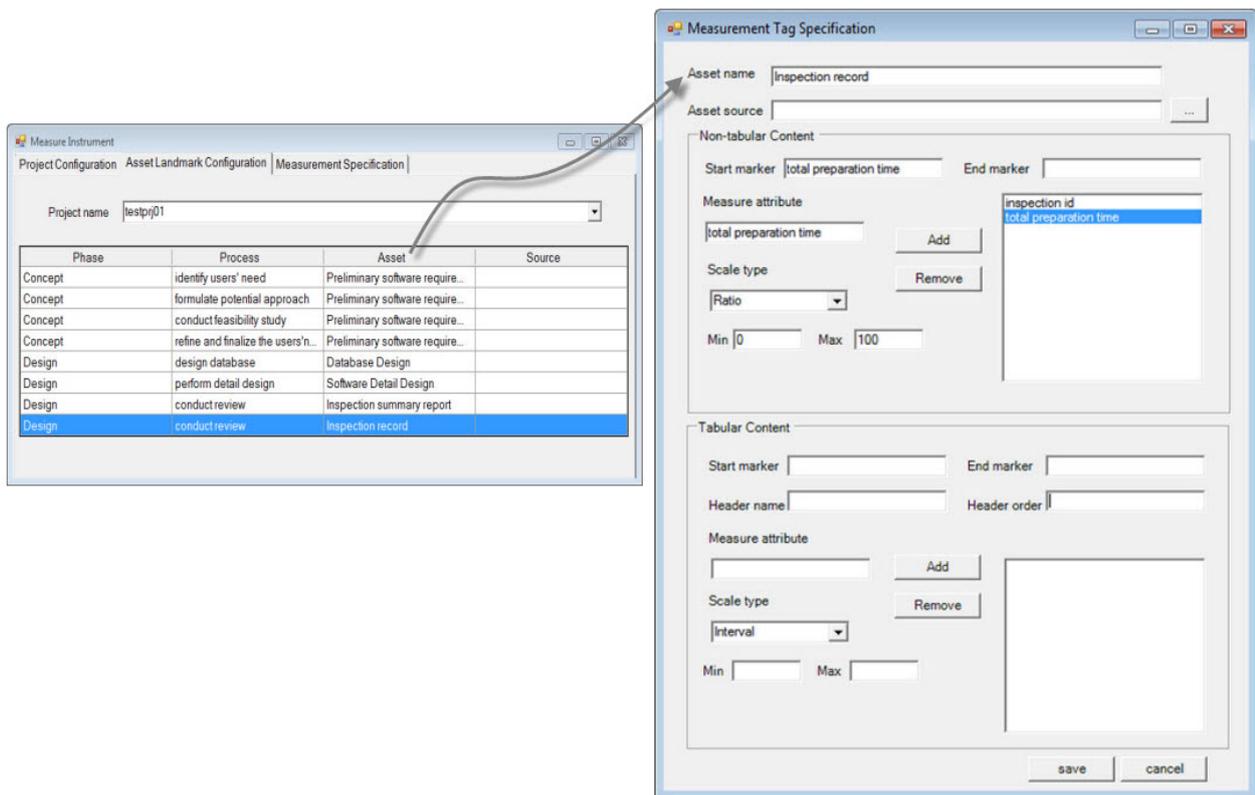


Fig. 5: Screens for exporting data for analysis and the resulting chart

6. References

- [1] Software Engineering Institute, *Capability Maturity Model Integration, CMMI for Development, Version 1.3*. Software Engineering Institute Carnegie Mellon University, Pittsburgh, CMU/SEI-2010-TR-033, 2010.
- [2] J. McGarry, et al. *Practical Software Measurement: Objective Information for Decision Makers*. Addison-Wesley, 2001.
- [3] *IEEE Standard for Developing a Software Project Life Cycle Process*. IEEE 1074:2006.
- [4] *IEEE Standard for Information Technology - Software Life Cycle Processes*. IEEE/EIA 12207.0:1996.
- [5] D.R. Goldenson, J. Jarzombek, and T. Rout. Measurement and Analysis in Capability Maturity Model Integration Models and Software Process Improvement. *CrossTalk: The Journal of Defense Software Engineering*. **16**(2003)7:20-24.