

Creating a Metrics Program

Step 5: Define Data Collection Procedures

Step 5 is to decide where to find the raw data identified in Step 4, and how to collect it.

Defining a set of generic data collection procedures that would be effective for every company is certainly an impossible task. For this reason, this discussion will be limited to a set of guidelines for collecting data. For convenience, the data identified in Step 4 has been partitioned into the following groups:

Figure 3.21 :
Data types

Data group	Description
estimation data	initial cost and schedule estimation data
effort data	data measuring person hours worked
staffing data	project staffing level data
project management data	activity, unit, and task status information
financial data	data concerning project financials
requirements data	project requirements and specifications
implementation data	data generated during implementation
testing data	data generated during testing
defect data	data concerning software defects
performance data	measured software performance data

Use the guidelines that follow to decide which procedures you need for collecting (and recording) the data. Document the procedures in the metrics program documentation.

Here are some points to keep in mind as you devise the collection procedures:

1. *How is the data collected?*

Describe where the data comes from and how the measurement team will obtain it. Ensure that the units of measurement are clear and conform to the standards laid out within the metrics program.

If the data is the output of a tool or an automated procedure, indicate the following:

- where are the tools found?
- who is responsible for keeping the tools in good working order?
- how is the data acquired from the tools?
- are the procedures for using the tools documented? where?

If the data is collected as part of a planning or scheduling task, or introduced in a project document, it is necessary to indicate the following:

- where is the information documented?
- when is the information considered "final"?
- how will the measurement team be informed when data changes?

If the data is collected during the course of development (e.g., defect data), indicate the following:

- who is responsible for providing the data?
- how to ensure that all relevant data has been collected?

2. *When is the data collected?*

Make sure the procedures indicate at exactly what point each piece of data is to be collected.

3. *Who is responsible for collecting and recording the data?*

Describe who is responsible for collecting the data, who is responsible for entering it in the metrics database, and who is responsible for transforming the raw data into any complex metrics.

4. *Where is the collected data stored?*

Describe any transitional means of data storage such as forms, documents, or email, as well as the name and location of the metrics database.

5. *How do we ensure that the data is correct?*

Describe any consistency checks that can be performed to verify that the data is reasonable. Describe the procedure for dealing with blatantly erroneous data.

Detailed Data Collection Procedures

When devising collection procedures, keep in mind that the simpler the procedure, the more likely it will be followed. Try not to inconvenience the project staff who will be doing the collecting.

Estimation Data

Software project estimation data is calculated during the early stages of the software project. In most cases, an estimate of the project cost is required to obtain the appropriate approval and funding for the project to proceed. The estimation data is usually presented in a project proposal or project planning document.

To collect the estimation data, the measurement team simply has to obtain a copy of the appropriate document.

Initial project estimation data is required for project cost, effort, staff, and schedule. This includes the following:

- estimated size of new code (SLOC)
- estimated size of reused code (SLOC)
- estimated total size (SLOC)
- estimated effort to complete each activity
- initial cost estimate in dollars for each activity
- labor rate for each activity

- estimated start date for each activity
- estimated completion date of each activity
- estimated project completion date
- estimated number of staff required for each activity

An estimate of the complexity of the software is often used to determine the development effort, as complex projects often require more effort per unit than less complex projects. If software complexity is being estimated, you should include this value with the other estimation data.

A brief discussion of software estimation is provided in Appendix C.

Effort Data

Effort data is a count of the actual person hours spent on a project. Tabulate this amount from information provided by individual employees. In most companies, employees provide this information on time sheets used for salary purposes. You can collect the effort data from the timesheets, but this is not always the most accurate method: employees often work hours that are not included on their timesheets. To get an accurate count of the actual hours worked, the employees may have to provide the data in another way, such as on their status reports.

Collect the following information:

- actual number of person hours to complete each activity
- number of person hours worked on management tasks in each activity
- number of person hours worked on support tasks in each activity
- number of overtime hours worked
- number of overtime hours worked in each activity

The metrics coordinator (or the database) should calculate the above, from the following base information provided by every employee:

- number of hours worked on each task
- number of overtime hours worked on each task

Classify tasks as being management, support, or development and link them with a particular activity.

Staffing Data

The project manager, or whoever is responsible for project staffing, can provide staffing data. The following information is required:

- number of staff members at each experience level
- number of development staff for each activity
- number of managers for each activity
- number of staff people for each activity
- total number of development staff

Project Management Data

Project management data is dynamic, changing throughout the life of the project. It typically includes the status of activities, units, and tasks. The project management team must ensure that they receive this information from the development team members, at agreeable intervals.

Some status information must be collected periodically:

- actual start date for each activity
- number of units completed in each activity
- number of units required to complete each activity
- actual completion date for each activity
- number of process exception reports

Some status information will be collected only once:

- project start date

Financial Data

There are two kinds of financial data to collect. First, collect the data concerning how the budget (the money available for development) is allocated:

- amount of the budget allocated to each activity
- total budget in dollars
- total dollars allocated to support staff
- total dollars allocated to tools

The second kind of financial data describes how much money is actually spent on the project:

- total cost in dollars
- total dollars spent on development tasks
- total dollars spent on management tasks
- total dollars spent on support tasks
- total dollars spent to date

The accounting or financial planning department should be able to provide this information to the measurement team.

Requirements Data

Collect the following requirements data over the life of the project:

- initial number of requirements
- number of requirements added
- number of requirements changed
- number of requirements deleted

The measurement team can collect this information by counting the number of requirements in the requirements documents. As well, a section could be added to the requirements documentation to aid in the tracking of added, changed, or deleted requirements.

It is also useful if the person who is adding or changing a requirement prepares a short *Requirements Change Impact Notice*. This notice would outline changes to the schedule, work tasks, completed work, estimated effort, etc.

Implementation Data

The development staff should provide the necessary implementation data on a regular basis. To do so, each team member can run programs to count the number of lines of source code and documentation produced, and put the information in a status report. Implementation data includes the following:

- total SLOC produced
- total SLOC produced (new)
- total SLOC produced (reused)
- total LOD
- number of units inspected

During development, units of software that have been coded and unit-tested are typically added to the configuration management (CM) system. Use reports from this system to obtain:

- number of units coded (new)

These systems may also provide information on the SLOC added or changed for a particular project (or set of files).

Testing Data

Testers are typically responsible for reporting their test status at regular intervals. The following information should be available to the measurement team:

- total number of tests
- number of tests executed to date
- number of tests executed successfully to date

Defect Tracking Data

Most companies use some form of database for tracking problems and defects during development. The measurement team should be able to obtain reports containing the following information from the database:

- total number of defects corrected in each activity
- total number of defects detected in each activity
- average duration between defect detection and defect correction
- average effort to correct a defect
- total number of defects remaining at delivery

Software Performance Data

Software performance data is usually generated during system testing, once the software has been integrated and functional testing is complete. Tools that are specific to the environment for which the software is being developed provide this data:

- average CPU utilization
- average memory utilization
- measured I/O transactions rate

The measurement team could include a performance data sheet as part of any testing results documentation as an effective means of collecting this data.

Actions Required for Step 5

1. Decide which data collection procedures are applicable to the data identified in Step 4.
 2. Create any forms that are required for the collection.
 3. Assign responsibility for metrics collection and ensure that the responsibility is agreed upon and documented.
 4. Update any design or project documentation templates to include sections for data that must be collected.
 5. Update the metrics documentation to include details on metrics collection procedures for the metrics coordinator(s).
 6. Optional: update any development process documentation to include relevant data collection procedures. Indicate which tasks and activities are affected.
-