



SOEN 6611 – Software Measurement

PROJECT STEP 4

Source: SEI implementing Goal-Driven Measurement course material (adapted).

Submitted to - Prof. Dr. Olga Ormandjieva

Team – 7

Rutwikkumar Sunilkumar Patel – (40160646)

Charit Pareshbhai Patel – (40160658)

Deep Pareshkumar Patel – (40185585)

Bhoomi Shah – (40169655)

NOTE: the scope of the Steps 3, 4, and 5 is reduced to Validity, Vincularity and Veracity.

Project Step 4 /S22 (5 points, due before midnight on July 30th)

Objective: Planning of the measurement process

Summary of Step 4.

The objective of this step 4 is to identify and plan the activities that must be accomplished in order to collect, store, process, and report the measurements necessary to build your 3V's indicators.

To help you with this portion of the job, here are some guidelines (the order may differ from the listed below):

- a) Review the action checklist in section 1;
- b) Analyze the tasks in the checklist to see if they are sufficient to collect, store, analyze, etc. the required measures (data elements) for your indicators.

Specific tasks should be defined for:

- Prepare [specific data collection]
 - Collect [defined data]
 - Analyze [the results]
 - Report [the results]
- c) Document your tasks using the template provided below. Label each measurement task as MTXX (XX is the sequential number of the task). Trace it to the corresponding DAXX / INXX / MGXX. [DAXX is the label of the corresponding Data Element, INXX is the label of the corresponding Indicator, MGXX is the label of the corresponding measurement goal).

You must remain consistent with all of the base and derived measures defined in the previous step 3.

2. Checklist to complete:

#	Checklist	
a.	List and label as DAXX the data elements (base measures) (XX is the sequential number of the data element).	<input checked="" type="checkbox"/>
b.	Define the frequency of collection and the points in the measurement process where the measurements will be made.	<input checked="" type="checkbox"/>
c.	Identify the supporting tools that must be developed or acquired to help you automate and administer the measurement process.	<input checked="" type="checkbox"/>
d.	Prepare a short process guide for collecting the data, how the data are to be stored and how the data will be accessed, and how the data will be analyzed and reported.	<input checked="" type="checkbox"/>

A: Labels

Measurement Goal	Labels
Increasing veracity of the big data	MG04
Enhancing the validity of the big data	MG05
Improving vincularity of the big data	MG06

Indicator	Labels
Mver	IN01
Mval	IN02
Mvin	IN03

Measure	Labels
Length of Big Data (LBD)	DA01
Rec_no_null (MDS)	DA02
Rec_cc_age (MDS)	DA03
N_succ_req (MDS)	DA04
N_req (MDS)	DA05
Number of Distinct Data Elements (Ndde)	DA06
Nrec_comp	DA07
Nds_cr	DA08
Nds	DA09
Length of the record in the dataset (Ldst)	DA10
Rec_Trace	DA11

B: Frequency of Data Collection

The data is collected in an incremental way, meaning, new data which is added to the raw dataset will be processed and analyzed in a timely manner. We are dividing the dataset into three sub-parts or timeframes, namely T1, T2, and T3, having equal intervals.

C: Identifying supporting tools

Data Visualization and collaboration tools like Google Collab, Tableau, Power BI, Pandas library in Python, NumPy and similar Scripting languages.

3. Plan tasks / activities

#	Task / activity (what / how)	Trace to DAXX / INXX / MGXX	Responsible (who)	Participants (with whom)	Estimated duration (in days)	Estimate d effort (in person- hours)	Schedule (when)	Tool (with what)	Rationale
MT01	Preparing the programming scripts or analytical tools for collecting the base measures and calculating the derived measures	DA01-DA11	Data Scientist	Technical team (Project manager, Developers, etc.)	3	15	Before the start of the big data pipeline cycle	Data Visualization and collaboration tools like Google Collab, Tableau, Power BI, Pandas library in Python, NumPy and similar Scripting languages	To calculate the measures more efficiently and faster rather than doing it manually
MT02	Collect the number of datasets (Nds) in the big data	DA09	Data Scientist	Technical team (Project manager, Developers, etc)	0.2	1	After every phase of the big data pipeline and for each time frame	Datasets are collected manually from an online source. https://www.kaggle.com/	Provides the total datasets count
MT03	Collect the number of distinct data elements	DA06	Data Scientist	Technical team (Project manager,	0.2	0.5	After every phase of the big data pipeline and	Data Visualization and collaboration tools like Google Collab, Tableau Power BI, Pandas,	Provides the count of distinct elements across multiple datasets

#	Task / activity (what / how)	Trace to DAXX / INXX / MGXX	Responsible (who)	Participants (with whom)	Estimated duration (in days)	Estimate d effort (in person- hours)	Schedule (when)	Tool (with what)	Rationale
	(Ndde) in the big data			Developers, etc)			for each time frame	NumPy library in Python and similar Scripting languages	
MT04	Collect the length (LBD) of the big data	DA01	Data Scientist	Technical team (Project manager, Developers, etc)	0.2	0.5	After every phase of the big data pipeline and for each time frame	Data Visualization and collaboration tools like Google Collab, Tableau Power BI, Pandas, NumPy library in Python and similar Scripting languages	Provides the total existing records count across multiple datasets
MT05	Record the collected measurement data	DA01- DA11	Data Scientist	Technical team (Project manager, Developers, etc)	1	4	After every phase of the big data pipeline and for each time frame	Microsoft Excel, Tableau	Storing base measures to calculate the derived measures
MT06	Check the quality of the measurement data collected	DA01- DA11	Data Scientist	Technical team (Project manager, Developers, etc)	0.5	2	After every phase of the big data pipeline and for each time frame	Box Plot	Validating the collected data

#	Task / activity (what / how)	Trace to DAXX / INXX / MGXX	Responsible (who)	Participants (with whom)	Estimated duration (in days)	Estimate d effort (in person- hours)	Schedule (when)	Tool (with what)	Rationale
MT07	Calculate Mver (includes finding out the corresponding base measures and evaluating its derived measures)	IN01, DA01, DA02, DA03, DA04, DA05, DA06	Data Scientist	Technical team (Analyst, Project manager, Developers, etc)	1.5	2	After every phase of the big data pipeline and for each time frame	Data Visualization and collaboration tools like Google Collab, Tableau Power BI, Pandas, NumPy library in Python and similar Scripting languages	Returns indicator for the veracity amongst multiple datasets
MT08	Calculate Mval (includes finding out the corresponding base measures and evaluating its derived measures)	IN02, DA07, DA08, DA09	Data Scientist	Technical team (Analyst, Project manager, Developers, etc)	1.5	2	After every phase of the big data pipeline and for each time frame	Data Visualization and collaboration tools like Google Collab, Tableau Power BI, Pandas, NumPy library in Python and similar Scripting languages	Returns indicator for the validity amongst multiple datasets
MT09	Calculate Mvin (includes finding out the corresponding base measures and evaluating	IN03, DA09, DA10, DA11	Data Scientist	Technical team (Analyst, Project manager, Developers, etc)	1.5	2	After every phase of the big data pipeline and for each time frame	Data Visualization and collaboration tools like Google Collab, Tableau Power BI, Pandas, NumPy library in Python and similar Scripting languages	Returns indicator for the vincularity amongst multiple datasets

#	Task / activity (what / how)	Trace to DAXX / INXX / MGXX	Responsible (who)	Participants (with whom)	Estimated duration (in days)	Estimate d effort (in person- hours)	Schedule (when)	Tool (with what)	Rationale
	its derived measures)								
MT10	Report for Veracity	IN01, MG04	Data Scientist	Technical team (Analyst, Project manager, Developers, etc)	0.2	0.8	After every phase of the big data pipeline and for each time frame	Data Visualization and collaboration tools like Google Collab, Tableau Power BI, Pandas, Numpy library in Python and similar Scripting languages	Insights on how relevant the collected data is across the multiple datasets in a time frame during different phases
MT11	Report for Validity	IN02, MG05	Data Scientist	Technical team (Analyst, Project manager, Developers, etc)	0.2	0.8	After every phase of the big data pipeline and for each time frame	Data Visualization and collaboration tools like Google Collab, Tableau Power BI, Pandas, Numpy library in Python and similar Scripting languages	Quality of the data if it is valid or not across the multiple datasets in a time frame during different phases
MT12	Report for Vincularity	IN03, MG06	Data Scientist	Technical team	0.3	1	After every phase of the big data	Data Visualization and collaboration tools like Google Collab, Tableau	Summarizing the connections and links

#	Task / activity (what / how)	Trace to DAXX / INXX / MGXX	Responsible (who)	Participants (with whom)	Estimated duration (in days)	Estimate d effort (in person- hours)	Schedule (when)	Tool (with what)	Rationale
				(Project manager, Developers, etc)			pipeline and for each time frame	Power BI, Pandas, NumPy library in Python and similar Scripting languages	between the data across the multiple datasets in a time frame during different phases
	Total :				10.3	31.6			

4. Measurement process guide

Write a measurement data collection guide, how the data are to be stored and how the data will be accessed, how the data will be analyzed and reported. to make it easier for the different people involved to collect/analyze/report measurement data/results. This guide can be organized by time of data collection/analysis/reporting (daily, specific days of the week, start or end of an iteration, etc.). This short guide should be used as a reminder and should fit on one page.

Data Collection Guide

⇒ Data source (where the measurement data will be extracted from)

IBM Analytics Dataset: <https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset>

- This all phases will be executed during **every** time frame.
- All the phases will run automatically in big data pipeline.

#	Phase (when)	Task (what)	Data/Tools (using/with what)	Responsible (who)
1	Before the start of the big data pipeline cycle	Collect the data from an online source such as Kaggle. Finalizing the tools or programming language scripts to be used. Finalize the tools or the process for storing, retrieving, and analyzing the measurement data. Determine the tools for report generation.	Data visualization tools like Matplotlib, Google Collab, Tableau, Power BI, Pandas library in Python and similar Scripting languages	Project Manager
2	Post-data extraction phase	Executing the programming scripts or tools that were finalized in the previous phases.	Base and derived measures / Data visualization tools like Matplotlib, Google Collab, Tableau, Power BI, Pandas library	Data Scientist

		Load the raw data in big data pipeline using data visualization and collaboration tools such as google collab, tableau for pre-processing.	in Python and similar Scripting languages	
3	After the data pre-processing phase	<p>Executing the programming scripts or tools that were finalized in the previous phases for data collection.</p> <p>All the unnecessary rows and values will be removed. In addition to this, all unnecessary columns and duplicate values will be removed. Also, all null values will be handled here.</p> <p>Record and store the collected measurement data.</p>	Base and derived measures / Data visualization tools like Matplotlib, Google Collab, Tableau, Power BI, Pandas library in Python, and similar Scripting languages	Data Scientist
4	After data processing	<p>Executing the programming scripts or tools that were finalized in the previous phases for data collection.</p> <p>After the data processing phase, all the null values will be replaced with valid values and the data will become computer readable.</p> <p>Record and store the collected measurement data.</p>	Base and derived measures / Data visualization tools like Matplotlib, Google Collab, Tableau, Power BI, Pandas library in Python, and similar Scripting language	Data Scientist
5	Post completion of the big data pipeline cycle	<p>Perform the measurement.</p> <p>Plot the results and/or generate reports accordingly</p>	Indicator graphs or reports	Data Scientist or Analyst