metaware.

the complete IT modernization solution

REVERSE ENGINEERING
Accuracy: The Initial Requirement

A Metaware White Paper

EXECUTIVE SUMMARY

While the success rate of IT projects is still below expectations, we have to recognize that modernization projects do not escape this rule if we do not meet a number of requirements that more than twenty years of practice have taught us collectively.

This document is the first of a series and aims to define and provide details of the critical success factors for modernization projects. With Metaware "industrialization" approach of modernization projects, we refer to the structured process that describes the performance of tasks, the organization of resources, the skills and the automation resources needed to meet these requirements.

The purpose of this document is to deal with the first requirement to be met in order to be within reach of achieving a modernization project: automation and its corollary, accuracy, have already been mentioned.

1 ACCURACY

Productivity in any modernizations project depends on the accuracy with which code transformations are carried out before entering the test phase. We measure the accuracy of the transformations by means of the defect injection rate, which calculates the number of anomalies generated by the transformation tool – before entering the test phase – against the total number of lines of code (LOC) processed.

The table below shows the number of anomalies generated by applying a tool that has an injection rate of 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5} between 100,000 and 10,000,000 LOC.

	100,000	**1,000,000**	**10,000,000**
10^{-2}	1,000	10,000	100,000
10^{-3}	100	1,000	10,000
10^{-4}	10	100	1,000
10^{-5}	1	10	100

In practice, processing 1,000 anomalies represents the upper limit above which any project is doomed to fail. The marginal cost of processing an anomaly is not constant but on the contrary, increases incrementally with the total number of anomalies due to the:

- Multiplication of code releases to be produced and retested
- Multiplication of test iterations
- Identification of 1 anomaly out of 10 per test plan is much more difficult and therefore more expensive than identifying one anomaly per test plan

In this way, if this figure is selected as a limit that is not to be exceeded, the minimum accuracy required of a transformation tool (and processes associated with it) is as follows on the basis of the volume of code to be processed:

	100,000	**1,000,000**	**10,000,000**
Accuracy rate	10^{-2}	10^{-3}	10^{-4}

Controlling a modernization project depends directly on controlling the accuracy of the transformation tool and of the implemented processes. Note that the tool implementation processes (configuration, feeding in, processing error logs) may degrade this tool's intrinsic accuracy significantly.

Note: Creating a transformation driver requires an accuracy rate that is much less than the rate required to be able to process the entire legacy asset. It is therefore quite tempting to go into the production phase (generalization to the entire legacy) using the rate obtained at the end of the prototyping feasibility studies. Many people have made this error.

The volumetric analysis of a legacy defines an accuracy threshold beyond which the project cannot be carried out within a reasonable budget and time. The candidate transformation tool MUST guarantee – under pain of failure – a minimum accuracy rate based on the number of LOC to be processed (e.g. 10^{-3} per MLOC).

2 CAPACITY

The technology used to develop the tool should be POWERFUL enough to allow the developers of the tools to achieve the accuracy rate required for a cost and within a time frame that are compatible with the project.

The sophistication of the technology increases virtually exponentially with the degree of accuracy aimed at. Therefore, for simple transformations such as changes in dialect, you can get by with text manipulation technologies (Perl, awk, Java, COBOL) to achieve a rate of 10^{-2}. For more complex transformations such as language translation or field extension, increasingly sophisticated technologies that come from compilation must be used, technologies such as:

- parsing: i.e. production of more or less detailed syntactic trees
- transformation of syntactic trees
- modeling data flow and control flow

The last 30 years have seen the development of a great number of language manipulation techniques in research laboratories, which have allowed IT to have, for example, more and more powerful languages and compilers.

To be able to achieve the degree of accuracy required for a modernization project, the sophistication of the transformation tool technology must be proportional with the volumetric analysis to be processed and the complexity of transformations.

Complexity	Transformation examples
Simple	Change of COBOL dialect, Y2K adaptation, Euro, cross references
Medium	Script language translation (JCL), translation of screens
Complex	language translation, field extension, CODASYL to SGBDR conversion

Note: there are very few experiences and expertise available on the market to define and validate if a given technology is powerful enough to meet the accuracy requirement of a given project.

In this respect, the best guarantee is provided by the jobs passed to this technology.

A number of projects have been undertaken using a "language" technology that is not adapted to the accuracy requirement of their projects. The tools developed using this technology therefore require very great development efforts from the tool team, development that was prematurely abandoned in favor of a manual process that led to the catastrophe.

3 INDUSTRIALIZATION

To achieve the required rate of accuracy, you must foresee an industrialization phase during which the transformation tool (and the processes associated with it) must in order be:

- developed to achieve the accuracy requirement on the basis of the selected technology by design
- run-in and optimized to prove that the accuracy requirement was achieved in practice (the phase referred to as Pre-Production)

During the Pre-Production phase, you must measure and improve transformation tool accuracy rate until you achieve the required rate before you apply the tool to the entire legacy (Production phase).

Allowing for exceptions, all anomalies detected during this phase must be corrected **not by intervening at translated code level but by doing so at development tool level**.

Only when the required accuracy has been observed (after complete tests) **in a completely repeatable way**, can the production phase begin.

Note: The notion of industrialization phase is too often skirted around in favor of a prototyping phase that is limited to establishing technical feasibility without really trying to prove the feasibility of the required accuracy rate whereas this is a key success factor in the project.

4 REPEATABILITY

This notion – used in manufacturing engineering – qualifies the capacity of a system, a procedure to reproduce a series of actions. The higher the repeatability of a unit, the more the characteristics of the action are stable.

Repeatability is an absolute requirement for:

- The load and the time frame of the Pre-production phase to be minimized in the course of the iterations. This is because only an iterative process is a guarantee of the convergence of the tool's accuracy with the threshold required on the basis of a Pre-production batch
- The characteristics observed in Pre-production phase to be able to be reproduced in a stable way over the entire legacy
- In the event of improvement or correction of anomalies in Production phase, to be able to intervene at tool level and not be constrained to "do everything manually"

Repeatability is another requirement that is essential to guarantee that the accuracy requirement of a modernization project is achieved and maintained. In the event that iterations are abandoned, it becomes impossible to improve tool accuracy and consequently, the project becomes uncontrollable.

Note: There is no point in mentioning projects that ended up in mid or end project with the need to change a transformation rule because of an anomaly or insufficient performance and to choose between:

- Changing tool but with the obligation to reintegrate reliably all manual corrections to date,
- Processing all new changes by hand, i.e. forcing yourself that much more to finish the project manually.

5 AUTOMATION

The automation requirement is a natural consequence of the requirements for accuracy and repeatability.

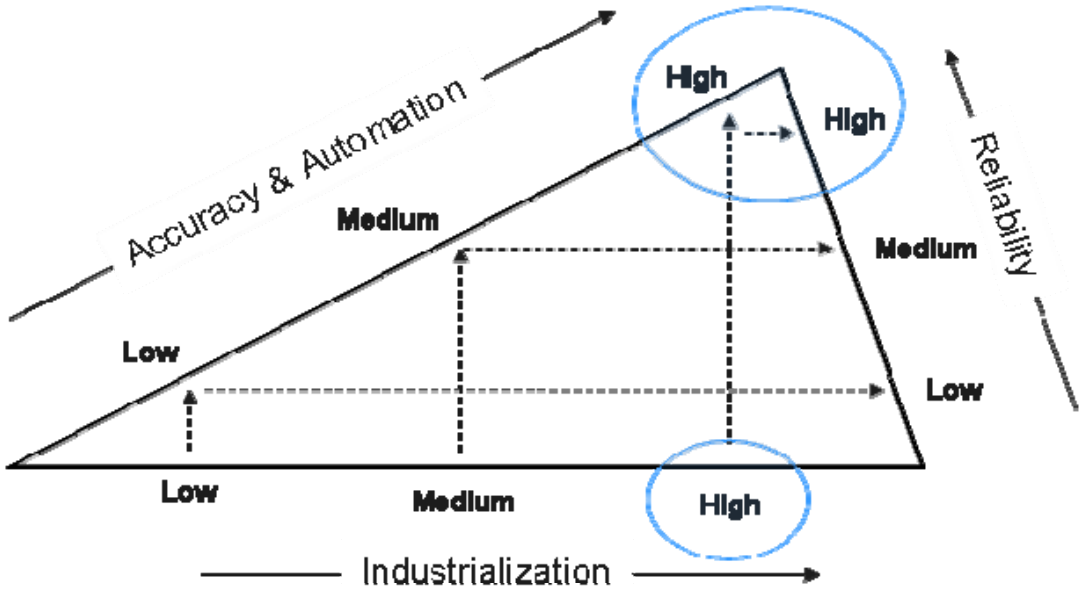
By way of corollary, to reduce the cost of each iteration, the automation rate must be better than the accuracy rate by a factor of at least 10.

CONCLUSION

In conclusion, we have described a number of requirements for a modernization project to succeed. By the specific character of these projects (with code analysis and transformation activities affecting high volumetric analysis and implementing language technologies), very few IT professionals are familiar with these notions.

The table below provides a summary of the requirements imposed on all modernization projects.

	100,000	1,000,000	10,000,000
Accuracy rate	10^{-2}	10^{-3}	10^{-4}
Automation rate	10^{-3}	10^{-4}	10^{-5}
Repeatability	Yes	Yes	Yes



metaware.
the complete IT modernization solution

www.metaware.fr
Tel. +33 1 30 15 60 00
Fax. +33 1 30 15 06 71

Global Headquarter:
1. Parc des Grillons
60, route de Sartrouville
78230 Le Pecq Cedex – FRANCE