Engineering and Cultivation of a Metrics Programme

Jakob H. Iversen[a] & Lars Mathiassen [b]
iversen@cs.auc.dk [a] and larsm@cs.auc.dk [b]
Department of Computer Science
Aalborg University, Denmark

Abstract

Software process improvement is seen by many as a viable strategy for overcoming the software crisis. It is, however, difficult to determine the actual effect of such improvement efforts. This paper reports from an organisation that is implementing a software metrics programme with the expressed purpose of measuring the effects of their improvement project in quantitative terms. The metrics programme is intended to measure key indicators of all completed projects and summarise progress information in a quarterly management report.

The paper describes important events during the implementation of this metrics programme. The process turns out to be long and complex and the software organisation finds itself confronted with dilemmas based on contradictory demands and value conflicts. The paper interprets the implementation process from two complementary perspectives. First, it is viewed as a rational, engineering process in which a metrics programme is constructed and put into use. Second, it is seen as an evolutionary, cultivation process in which basic values of the software organisation are confronted and transformed. This analysis leads to practical advice on how to implement metrics programmes as part of software management practices.

Keywords: Software process improvement, metrics, organisational change

BRT Keywords: DA04, EI0111, DD01, EI0205

Introduction

Software Process Improvement (SPI) is conducted as continuous, evolutionary improvements to the processes in specific software organisations. SPI is not based on a single technique or method, but rather on a collection of different approaches applied as necessary in each company. Common to the many general approaches is a focus on establishing sound project management practices before attempting to implement more advanced and organisation-wide techniques. SPI is today seen as one of the most viable approaches to decisively improve the state of software development (Humphrey 1989;Grady 1997).

A weakness in most SPI efforts is the lack of focus on measuring actual effects in business and other terms that are independent of the maturity models. With little information on the negative or positive impact of improvement initiatives, it is difficult to manage the improvement effort. By measuring relevant phenomena, e.g. software, error

reports, customer satisfaction, or man-hours, and by doing it across projects for the entire company, it becomes possible to paint a picture of the software operation that is different from those emerging from software assessments based on normative maturity models. The literature on this subject is not extensive, but some studies have been published, as referenced in (Emam & Briand 1997).

This paper presents lessons on how to implement a software metrics programme to inform about the improvements of the SPI efforts within a specific company. The argument is based on a single case in which the organisation struggles with the implementation of a software metrics programme. The initiative turns out to be long and complex and the involved actors are confronted with a number of dilemmas based on contradictory demands and value conflicts related to the metrics programme. To make sense of this process we interpret the events from two complementary perspectives.

First, we see the implementation process as a rational, engineering process in which the metrics programme is designed and constructed (Basili & Weiss 1984;Carleton et al. 1992;Fenton & Pfleeger 1997;Dahlbom & Mathiassen 1993;Dahlbom & Mathiassen 1997). The underlying approach is that of instrumental problem solving (Schön 1983). The challenge is to build a device, i.e. the metrics programme, which provides information about the software operation based on a number of indicators. The quality of the effort is related to the usefulness of the provided information in making decisions about the improvement effort. The focus is, in other words, on building a signalling device to support rational decision-making about SPI (Feldman & March 1981).

Second, we view the implementation process as an evolutionary cultivation process in which interests and values within the software organisations are confronted and transformed (Dahlbom & Mathiassen 1993;Dahlbom & Mathiassen 1997;Schein 1985). The underlying approach is in this view that of organisational intervention (Argyris & Schön 1978;Argyris & Schön 1996). The challenge is to transform the software culture from being based on espoused theories, i.e. on what software developers and managers believe and say, to being based on indicators of the theories-in-use, i.e. the deeply rooted assumptions (Schein 1985) that govern the actual software practices. The quality of the effort is, from this viewpoint, related to successful integration of the measurement programme into its use context. The focus is therefore on the symbolic aspects of the metrics programme as an element in changing management practices (Feldman & March 1981).

The next section presents the case organisation, its SPI project, and the underlying research approach. Then follows a mainly descriptive section presenting the sequence of events that took place in implementing the software metrics programme. The next section is then mainly interpretative applying the engineering and cultivation perspectives above to make sense of the described events. Based on these interpretations we propose a number of lessons on how to successfully implement software metrics programmes.

Case: Financial Software Solutions[1]

Financial Software Solutions (FSS) is a subsidiary of Financial Group, a financial institution centred around a bank, Finance Bank. Financial Group provides all aspects of financial services (banking, mortgaging, insurance, etc.). The primary business function

[1] The names of the organisation and all actors have been changed.

of FSS is the development of IT systems and services for Financial Group, but FSS also sells IT systems to other financial institutions across Europe. FSS has expertise in the development of banking, insurance, mortgage and financing applications. FSS has approximately 850 employees located at four geographically dispersed development centres.

Projects in FSS vary greatly in size. Most projects are small and short-term, with 3-5 people on the project team for 6-12 months, but some are major projects with strategic implication for the entire corporation. This includes the Y2K project and a project to integrate the international branch offices in the central computer systems. Such projects are staffed with 10-20 people and run for years rather than months.

FSS is divided into four development divisions, each of which is headed by a Senior Vice President. Each division is composed into departments headed by Vice Presidents. The departments typically have 20-25 people divided into approximately 5 projects. Project managers manage regular projects, whereas a Vice President manages high-profile projects.

FSS mainly develops systems for the large central mainframe installation, which is divided on two operation centres. Systems developed for this platform are based on an advanced event-oriented database principle, which increases the flexibility in processing data. Security and reliability are the two main requirements of the systems as data are mirrored in real-time between the two operation centres. Modern methods for modelling of data, functions, and workflow are used along with the business model, Information FrameWork, which is used intensely to involve stakeholders from the user organisation in the development process.

SPI in Financial Software Solutions

FSS was recently made into an independent subsidiary delivering IT services to Financial Group. In principle then, the mother organisation would be free to choose other providers, if they were dissatisfied with FSS. Faced with this reality, it has become a major thrust for FSS to stay competitive. SPI is seen as one of the strategies for keeping the business with the rest of the corporation. One of the most significant changes that occurred when FSS was established as an independent company was that all development work should be conducted in projects governed by a contract. This has even been extended to internal projects, which are now under contracts with their customers within FSS.

It was recognised from the beginning of the SPI effort that leadership would be very important to the success of the effort. The project was thus given a very high profile in the organisation by letting an experienced Vice President act as project manager, and have other Vice Presidents as team members. The idea was that each of the four divisions would have one representative on the project team. The representatives were appointed, but only a few were actually active on the project. Those who were active were, however, very committed towards the project and well respected in the rest of the organisation. Further adding to the image of a high-profile, organisation-wide project, the SPI project refers directly to a steering committee consisting of all the senior managers (Chief Executive Officer and four Senior Vice Presidents). The SPI project organisation is shown in Figure 1.

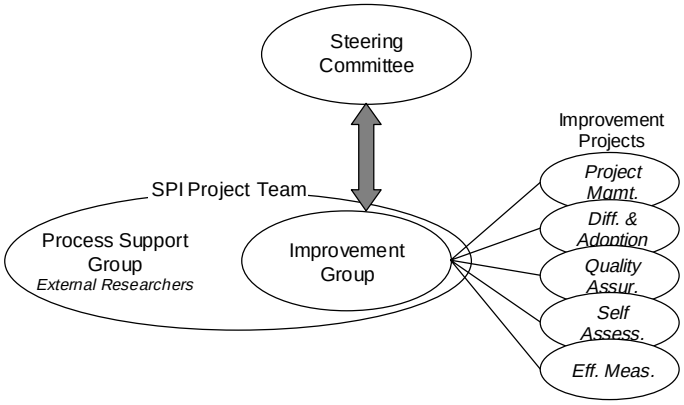

Figure 1: Organisation of the SPI project in Financial Software Solutions.

Since FSS is still at CMM level 1 (Iversen et al. 1998), the improvement group does not function as an actual software engineering process group (Fowler & Rifkin 1990), but acts more as a catalyst for the actual improvement effort. Improvement projects, each of which deals with one improvement area, make the detailed decisions on how and what to improve. The improvement projects are staffed with people that are knowledgeable in the area that they are going to improve and well respected in the organisation. The researchers have also involved themselves in these improvement projects, and are thus able to provide the FSS members of the groups with additional insights and inspiration in return for the added insight and understanding of SPI that being involved in such groups give to the researchers. Currently, the following improvement projects are ongoing: project management; diffusion and adoption of methods and techniques; quality assurance in projects; self assessment; effect measurement.

FSS has hence accepted the challenge of measuring the impact of various improvement strategies in order to provide managers with relevant data upon which they can make informed decisions about the software development process. The focus on measurements is specifically intended to enable the SPI project and senior management to make such informed decisions about the improvement activities as well as to assess the effect and progress of these activities. Apart from giving guidance to the improvement group and the improvement projects, the measurements are also seen as a way of getting some attention from the rest of the organisation on the SPI project.

Research Approach

The presented research is part of a large research project involving four software-developing companies, two universities, and a consultancy company. The researchers and consultants participate actively in the SPI projects of each of the four companies over a three-year period (Johansen & Mathiassen 1998). The SPI project in FSS was initiated along with the research project in January 1997. The research project will end in December 1999, but it is expected that the SPI project will continue beyond that.

At FSS, the researchers and consultants (commonly referred to as 'researchers') are active participants in the improvement group, and the research methodology applied was thus action research (Foster 1972). The main interaction between the researchers and the organisation takes place at the monthly SPI meetings, but also by more informal meetings, working sessions, workshops etc. in which only a single improvement initiative (in this case effect measurement) was discussed. As the researchers became part of the

SPI organisation they were able to obtain real insight into what the issues facing the SPI project were.

Two of the major problems in conducting action research is 1) the limited ability to generalize findings (Mathiassen 1998), and 2) the frequent neglect by many action researchers to collect adequate data to be able to demonstrate convincingly what was learned during the study. The former problem is dealt with in this paper by determining a number of practical lessons that were learned both by the researchers and the internal SPI members. These lessons are conveyed as practical advice to other companies involved in implementing similar programmes. In this research we have attempted to overcome the latter problem by systematically collecting as much data as possible about the organisations. This included all the 'natural traces' of the SPI programme such as project plans, meeting minutes, memos etc. In addition to this, we have tape-recorded the monthly SPI meetings as well as some of the working sessions and workshops. The relevant segments for effect measurement were transcribed.

Key Events in Implementing the Programme

This section presents the process of implementing a metrics programme in FSS. The presentation is structured around eight events that have influenced the implementation process. However, the description of each event in some cases includes what happened immediately prior to the event and what followed immediately after it. Table 1 shows a timeline of the individual events. Events 7 and 8 are currently expected to occur in April 1999. Some of the key stakeholders are listed and described in Table 2.

Table 1: Timeline of key events.

Year	**1997**												**1998**												**1999**											
Month	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D
Event	1		2						3						4 5					6								7 8									

Event #1: Press Conference. January 7, 1997

One of the first things that happened in the overall SPI project was a press conference. At this conference, the CEO of FSS stated that

> "We expect to gain a 10% improvement in efficiency through this project […] that is equal to 30 mill. DKK."

Table 2: Key stakeholders of the effect measurement programme in Financial Software Solutions

Chief Executive Officer (CEO)	Sponsor of the SPI project. Stated that FSS should improve efficiency by 10% through SPI.
Vice Presidents	Responsible for 20-25 people and 3-5 projects, they are what the CMM terms a “first-line software manager”. They are crucial in implementing suggested improvements. Their attitude towards the effect measurement programme is thus important to how the project managers and developers perceive the programme.
Project Managers	Required reporting data on their project in the central time and project registration system (TIRE/POKA), and for providing the effect measurement project team with a few central pieces of information about the project.
Peter	Project manager of the SPI project from March 1997 to February 1998. After a reorganisation, he has become the manager of the central Architecture unit. After June 1, 1998, he has been the acting project manager of the SPI project.
John	Project manager for the SPI project from February 1998 to June 1, 1998.
Ashley	Full time employed on the SPI project. Project manager for the effect measurement project.
Finley	Vice President and member of the SPI improvement group. Was heavily involved in defining the first measurement programme.
Linda	Vice President and member of the SPI improvement group. Has not been directly involved in the effect measurement programme.

This statement has become the focal point of the SPI project in FSS. In the project contract, completed on April 3, 1997, one of two critical success factors of the SPI project were thus

> “That FSS within the project's 3-year time span has achieved an improved efficiency of the system development process of at least 10%.”

Later in the document, the effect measurement programme is introduced:

> “The project's continuous effect measurements throughout the project will create the baseline for specifying the expected efficiency improvement at the project's completion in further detail.”

From the beginning of the project it has thus been very important to be able to show this 10% improvement. However, neither the CEO nor the contract was explicit on what should be measured and how the data should be analysed to show this 10% efficiency improvement. This was left to those members of the SPI team who were made responsible for implementing the effect measurement programme.

Event #2: Decision to Implement Effect Measurements. March 1997

After some input from the researchers, the improvement group decided to implement an effect measurement programme, to measure the 6 factors listed in Table 3. This decision

was discussed at the SPI Project Team meeting on March 25, 1997, and reflected in the minutes from this meeting. At this point, the improvement group had some notions about how to measure each factor, although there were no precise definitions yet. More precise definitions can be found in the May 13 decision memo about the effect measurement programme, as reflected in the 'definition' column in Table 3.

Table 3: Indicators of the metrics programme in FSS.

Factor	*Definition*
Project Productivity	Resources used to develop the system relative to size of project in function points
Quality	Number of error reports both absolute and relative to size of project in function points
Adherence to schedule	Variation from agreed time of delivery both absolute and relative to size of project in function points
Adherence to budget	Variation from estimated use of resources
Customer satisfaction	Satisfaction with the development process and the implemented solution (multiple choice questionnaire)
Employee satisfaction	Satisfaction with the development process (multiple choice questionnaire)

The decision memo also laid down some of the principles that the effect measurements would adhere to:

- Measurements should be relevant in relation to process improvement and quality, but also have general management interest.
- Measurements should as far as possible be made automatic. Information should be interpreted rather than disturb the development organisation.
- Cost of performing the measurements should be minimal.
- Use of questionnaires should be limited as much as possible, as the organisation suffers from 'questionnaire-disgust'. If questionnaires are used, they should be placed at a milestone, and be adapted to the natural system development process.

Data should be collected on projects that were finished (to keep disturbances to a minimum) and results should be published every quarter. The volume of the projects should be calculated using an automatic counting algorithm for function points (IFPUG 1994). Function points are usually calculated by experts with a significant amount of practice in counting function points. Not many organisations have attempted counting function points automatically, and it was therefore a relatively risky endeavour to engage in this work. Therefore, the project concentrated some effort (approximately 1-2 man months) on implementing the system to do this automatic calculation.

Event #3: First Measurement Report. September 1997

The first visible result of the measurement programme was the first measurement report, which was completed in September 1997 with results from 13 of 56 projects that were completed in 3rd quarter 1997. The report had data on 3 of the 6 factors (adherence to budget, time-to-market, and project productivity). The data contained some surprising information especially regarding adherence to budget causing senior management to withhold the report from wide distribution. Parts of the results were instead disseminated to the development organisation through a 'roadshow' conducted by the improvement

group to raise awareness towards the SPI project. The report was also criticised for being too academic. A workshop was held in October 1997 to improve the layout of the report to alleviate this problem.

The problems in gaining acceptance for this first report did not bother the improvement group significantly, as it was, after all, the first report, and was seen as something of an experiment with less than optimal data foundation.

Event #4: Second Measurement Report. March 1998

Data for the second report, covering projects that completed in 4th quarter 1997, were collected in October through December 1997, and the report was completed in March 1998. The results were discussed in an SPI Project Team meeting on February 20, 1998. Data discipline had been greatly improved from the first to the second report as shown in Table 4, although the percentage of complete data sets still ought to be 100.

Table 4: Data discipline in the effect measurement programme.

		Complete data sets	
Period	# projects	#	%
3Q97	56	21	37
4Q97	29	19	65

This improvement in data discipline and thus in data quality was received with much enthusiasm at the meeting. However, the discussion soon centred on the issue of whether the report should be made public or not. At the meeting, there was some disagreement on how to distribute the report and how detailed the information should be:

> John: “It could become publicly available. If we publish the main figures, and then anybody could request a copy. What will happen in the report is that projects will be mentioned with their names. In the first report, they were anonymous. They aren't sissies in the senior management group.”
>
> Linda: “I don't think it should just be a matter of requesting a copy. [The report] should be spread.”
>
> Ashley: “I'm […] a little nervous about including names and so on.”

John and Linda had not been directly involved in the effect measurement programme. Their interest here was to improve the SPI project’s visibility in the organisation; everybody must see what is happening. Ashley, on the other hand, had personally promised many members of the development organisation that they would not be personally blamed for any bad results, so she was naturally not interested in getting personal information spread too widely.

The researchers tried to help by suggesting alternative solutions:

> Lars (researcher): “what is actually unpleasant today is worst-case: there are measurements of people, and they know they are there, but they don't know what the measurements are. […] We can only win by getting these measurements out. There is also a solution that what is published is per department, so that those in department 2 can't see department 3. There are all sorts of intermediate solutions. But they need to get some concrete information back.”
>
> Jan (researcher): “I think it could give an unfortunate effect to make the numbers too

> widely available, because […] someone may try to make their numbers look better than they are. However, if they get an average plus their own project and then are encouraged to [discuss internally] why the numbers look the way they do. […] I think that will give a good effect."

As it can be seen, there was no common understanding of what it would mean to make the numbers public. However, there was general consensus that senior management had accepted the idea that the measurements should be made public:

> Linda: "I think we have [the CEO's] commitment that now he will [make it public], and we should of course make sure he sticks to that, once the report is completed."

The report was finished in March 1998, and had a much clearer layout than the first report. The report did not include productivity data because the automatic calculation of function points was considered faulty. Instead, data on customer and employee satisfaction was included.

Event #5: Decision Not to Disseminate Second Report. March 31, 1998

The report was presented at the steering committee meeting on March 31, 1998. The data was considered insufficiently reliable to warrant a wide distribution of the report. The results of the satisfaction surveys showed very unfavourable results for key business areas for FSS, and the steering committee thus decided to withhold the report. The presentation of the report to the steering committee was discussed at the SPI project team meeting on April 22, 1998:

> Meeting minutes: "The report is not and will not be made public internally in FSS. A strong contributing factor to this is that the credibility of the data validity is insufficient."

> John: "When I presented the report, I did so from a positive angle throughout. Emphasised what was good, and the positive things that had happened since the last [report]. Data discipline has improved. […] Some of the data discipline is due to the information meetings. People have become aware that this is actually used for something. Of course, it can still get better."

> John [On customer satisfaction numbers]: "But this number – here he [the CEO] almost fell off his chair. For what is it that we in FSS should do? We should be business partners with the corporation. We are competent and should show them that. We should provide them with qualified feedback. We can tell them which direction to take the business."

However, it was quickly identified that the customer satisfaction questionnaire was of poor quality and primarily directed towards customers, whereas in most cases users, who had not been involved in negotiating terms and contracts, had answered it.

> Finley: "We send these questionnaires […] to people who were involved in the acceptance test of the system. And what we then ask are managerial, contractual, overall process-related questions on how the project was conducted. Then some random user has to answer if commitments were met. He hasn't seen the contract or anything. It's bound to go wrong, and that's why management can't recognise reality in these numbers."

The issue of management commitment towards effect measurements was discussed as well:

> Finley: "I don't know how much management commitment we have here. This is the second report. We ask for more resources, but nothing much is happening. And now we can't bear any more major discussions about this. We must end up with something that

> gives management a credible picture of reality. Otherwise they will say: this measurement stuff - forget it, I'll be better off trusting my intuition about how the reality actually looks."
>
> Linda: "I get personally disappointed that [the CEO] does not release the report. I can understand that he is afraid of [the bank's central IT co-ordinator], but if we are ever going to get people interested in SPI, then they need to see what the project is doing."

This event is probably the most important in the history of the effect measurement programme. It caused a dramatic increase in the attention given to the programme, and caused establishment of a project to improve the effect measurement programme. This illustrates how difficult implementing such a programme is. A lot of resources had been used on defining each metric, and deciding how to measure them. But some aspects had still been overlooked: the questions in the satisfaction questionnaires had not been carefully formulated, and the customer questionnaire was given to users instead of customers. On top of that, insufficient attention had been given to incentives for the development projects in reporting the necessary data, resulting in poor data discipline.

Event #6: Improvement Project Established. August 1998

After the disappointment that the report was not made public, the discussion in the SPI project team meeting on April 22 focused on actions that could be taken to improve the effect measurement programme enough to enable publication of the next report. The group decided to try and establish a project to improve effect measurements:

> Lars (researcher): "…it's all about planning. If we think of the establishment of the effect measurement programme as a systematic improvement effort, then we need to have a plan for this project. A plan of how to systematically improve the initiative. […] As a first goal we need to give this solidity that the management group dare say that this is public. That should be a baseline for this year. Can you imagine that, John, that we design it as a project?"
>
> John: "I can't see why we shouldn't run it as a project."
>
> Linda: "But of course it's also because it emerges slowly, that it turns out that it has not been so easy to implement. So now we need to do something else, or establish a new initiative."

The project was proposed in June 1998, and the contract was signed in August 1998. The project was established as an improvement project, with the expressed goal to improve the quality of the measurement report so much that it would be impossible for management to deny making it public.

Event #7: Third Measurement Report. Planned for April 1999

The success criteria for the improvement project are that a measurement report is completed in April 1999. This report should contain data on all 6 indicators and from all projects completed in 1Q99. Compared to the second report, this report should have improved the measurement process for all the indicators, and the layout of the report should also be enhanced. Among other things, the data quality of each indicator should be displayed, for instance as a percentage of complete data sets (as in Table 4). However, function points will not be included due to the problems of making an accurate count. The possibility of using other size measures has been examined, and rejected as none of

the measures proposed (lines of code, compiled size, function points, and number of statements) all had severe weaknesses that made it better to not include a size measure and just rely on the rest of the measures to provide information on the effect. Excluding a size measure seriously impedes reaching the original objective of measuring efficiency and productivity, as there then is no measure of the output of the project.

While the improvement project has been running, a bonus system has been introduced, in which one of several factors is the correct reporting of data for the effect measurement programme.

The preliminary results of the improvement project indicate that the data discipline is still considered insufficient, as it has not been possible to persuade projects close to finishing that they should report data in the new format. However, it could be discussed whether this would actually make the data less valid, because one of the factors that is frequently missing, is an estimate of the expected date of system delivery. Reporting this figure when the project is almost complete, would make for complete data, but with very little relevance. For projects that have started more recently, things look a lot more positive. This is also attributable to the establisment of a bonus programme that requires projects to report accurate and correct data.

The satisfaction surveys have been integrated into the quality assurance process, and seems to work well. The projects give the questionnaires to their customers and employees and then the results are discussed in one or two quality meetings towards the end of the project.

Event #8: Report Disseminated Widely. Hoped for April 1999

The fate of the metrics programme will be decided after the first report has been completed. If the report is not made public within FSS, then it is likely that the effect programme will collapse. Not having an effect measurement programme will cause serious problems for the rest of the SPI project, as many of the initiatives are driven by the effect measurements.

Engineering and Cultivation

Interpreting these events from a rational, engineering perspective we see an inefficient effort with rather limited success. The effort got a flying start with explicit goals and maximum management attention (event # 1). A detailed design of the involved metrics was created together with basic principles for the implementation of the programme (event #2). This design was debated with and accepted by management. The first measurement report only included 20% of the projects and only 3 out of 6 factors. In addition, the report was criticised for being too academic (event #3). Data discipline was improved in the next report and the report had a much clearer layout than the first report (event #4). The data were, however, still considered too unreliable by management (event #5) and that led to the establishment of a new, intensified effect measurement project in which the metrics programme was redesigned. As part of this redesign the metrics to provide information on productivity (based on function points) was dropped (event #6). Finally, some improvements were made in gathering information about customer satisfaction (event #7). Still, after 2 years of effort, the organisation did not have an operating metrics and reporting programme in place.

A different picture emerges when the same events are interpreted from an

evolutionary, cultivation point of view. An ambitious statement was announced by top management implying that a transformation of current management practices was needed (event #1). The productivity should improve and a metrics programme should be implemented to evaluate whether this goal was being achieved. The existing management practice, which was mainly based on espoused theories of software developers and managers, was to be complemented with a data-driven intervention approach based on quantitative indicators of the theories-in-use in the software organization. When the first report was presented to management it contained surprising and negative information on current practices causing senior management to withhold the report from wide distribution (event #3). A dilemma emerged between, on the one hand, publication of the report—to increase knowledge, stimulate debate, increase participation in the project, and improve data quality—and, on the other hand, a concern for the validity and negative effects of the data provided. But no shared understanding of possible strategies to overcome this dilemma was reached (event #4). At the same time it turned out that the chosen strategy for measuring function points was far from successful, which confronted the effort with another dilemma between the relevance of the data and the economy of gathering them. As a consequence it was decided to drop the productivity metrics (events #5 and #7). This problematic situation made the involved actors realise that the effort had to be reorganised as a proper improvement initiative with more resources and incentive schemes to make software developers participate more actively in the effort (event #6).

From the first point of view we see a rather unsuccessful engineering project with very slow progress. From the other point of view, we see an initiative that was first conceived as a rather straightforward engineering effort being transformed into a complex transformation process of current management practices. The organisation has, in this second view, gained significant experience in implementing data-driven intervention practices and some data has been collected that has contributed to critical reflections on the beliefs and intuitions on how the organisation operates.

Lessons Learned

The general lesson that can be drawn from this case is not new: Successful use of technology in organisational contexts requires efforts that go beyond instrumental problem solving. Some studies within the field of information and software technology suggest that, in addition to a more traditional engineering approach, we need to focus on the involved formative contexts (Ciborra & Lanzara 1994). Others argue that we must understand and transform the involved communities-of-practice (Brown & Duguid 1991). The reported experiences have, however, in the context of successful implementation of metrics programmes given rise to a number of specific lessons, which may be valuable for other companies attempting to implement similar programmes. These lessons are presented in the following.

Start Simple

The wish to measure six indicators, each of which was of a complex nature with no previous measurement process in place, can only be characterised as extremely ambitious. It is easy to become disappointed when the collected data doesn't have the quality expected, and measuring some indicators must be abandoned. Another approach is to start out by simply measuring one or two indicators, perhaps just collecting data that

are already there and just analysing it. Later, when the measurement team and the development organisation have gained experience in measuring and being measured, other measures could be added to improve the programme. Such a staged introduction of a metrics programme may seem to take longer than the ambitious approach. But the results may be less disappointing.

A Real Project

At first, the metrics programme was considered as an integrated part of the SPI project. In this phase, there was no plan for the work, and the objectives were described in vague terms in the SPI project contract. It was only because of the dedication of a few employees that a metrics programme was being developed at all. Later in the process, when a real project was established, it became far easier for the actors to argue that they should have adequate resources and the focus on the initiative was generally increased. When established as a proper improvement initiative, missing or unreliable data are not only seen as defects. They become baselines for additional initiatives in which improvements in the quality of the metrics can then be tracked.

Establish Incentives

The FSS case illustrates the value of establishing incentives to improve data quality. From the outset, all projects were required to record several numbers in the on-line project and hour registration system, But almost no projects recorded complete and correct information mainly because they saw no immediate use for the data they provided. A marked improvement of data quality was achieved by using a combination of informing the project managers of what data they should report and how to do it, as well as informing about the importance of the data they provided and show some results based on the data. Moreover, when reporting accurate data became part of the bonus system, a very clear incentive scheme was established, and the data quality clearly improved for the projects starting after the bonus system was established.

Publish Widely

Probably the biggest disappointment for the measurement team at FSS has been management’s decisions to withhold the reports from distribution. In order to improve the metrics programme it is vital that the development organisation be given some feedback on measurements. Not making the reports public is a barrier for improvement of the programme. However, it is also important that performance measures of individuals be kept to the individual. Otherwise everybody would do all they can to make their numbers look better, and the entire purpose of the metrics programme to establish indicators of current theories-in-use would be lost.

Facilitate Debate

Implementing an effect measurement programme forms an important shift in culture towards a culture where decisions are based on factual numbers rather than vague intuitions. If the metrics is to reach a sufficiently high level of quality, the organisation must be willing to make less than optimal data available for discussion about the validity

of the data, and the assumptions underlying the measurements. But perhaps more importantly, it should be recognised that even if the validity of data is not as high as could be wished, the information contained in the data might still carry some relevance. Numbers should never be taken as absolute truths and it is vital that the data and their quality can be discussed to continuously improve the foundation of the metrics programme.

At FSS, this discussion has currently been limited to the SPI project team and management. Here, the discussion has been lively and inspiring to those participating. But as long as the discussion is contained to a small number of people, it is difficult to use the measurements as a feedback mechanism to the development organization to improve on the daily practices in the project. Effect measurements are likely to give unpleasant results about the software operation. Being able to cope with such results and use them to improve the organisation rather than just figuring out who is to blame for the bad results is an important part of the cultivation involved in implementing metrics programmes. Facilitating such a culture is no small task as illustrated by the presented case.

Conclusion

Metrics in software development are primarily used to direct the course of a single development project (Basili & Weiss 1984;Carleton et al. 1992;Fenton & Pfleeger 1997;Grady 1992;PSM 1998) and secondarily used for management overview of all projects. However, collecting data after a project is completed is far easier than while it is running. At FSS this has been exploited to develop a metrics programme that uses post-mortem measurements to characterise the overall state of the company's software operation. The information obtained from these measurements can be used as indicators of the effect of ongoing improvement initiatives within the organisation.

The lessons learned from this experiment illustrate that it is far from a simple undertaking to implement such a software metrics programme. There are many challenges involved in engineering useful metrics and ensuring a satisfactory level of reliability of the involved data. On top of this, any effort to implement a metrics programme will necessarily challenge and eventually transform current management practices. It is essential that this cultivation aspect be seen as a key challenge to ensure successful implementation of metrics programmes in software organisations.

Acknowledgement

Financial Software Solutions is thanked for granting access to their SPI project and providing excellent conditions for conducting this research. The Danish National Centre for IT Research has financially supported this research. The following colleagues have provided valuable input to the paper: Ivan Aaen, Jesper Arent, Gro Bjerknes, Karlheinz Kautz, Jacob Nørbjerg, and Rens Scheepers.

References

Argyris, C. & Schön, D. A. 1978, 'Organizational Learning: A theory of Action Perspective',

Addison-Wesley, Reading, Massachussetts.
Argyris, C. & Schön, D. A. 1996, 'Organizational Learning II', Addison-Wesley, Reading, Massachussetts.
Basili, V. R. & Weiss, D. M. 1984, 'A Methodology for Collecting Valid Software Engineering Data', *IEEE Transaction on Software Engineering*, vol. 10, no. 6, pp. 728-738.
Brown, J. S. & Duguid, P. 1991, 'Organizational Learning and Communities-of-practice: Toward a Unified View of Working, Learning, and Innovation', *Organization Science*, vol. 2, no. 1, pp. 40-57.
Carleton, A. D., Park, R. E., Goethert, W. B., Florac, W. A., Bailey, E. K. & Pfleeger, S. L. 1992, *Software Measurement for DoD Systems: Recommendations for Initial Core Measures*, Software Engineering Institute, Pittsburgh, Pennsylvania, SEI-92-TR-19.
Ciborra, C. & Lanzara, G. F. 1994, 'Formative Contexts and Information Technology: Understanding the Dynamics of Innovation in Organizations', *Accounting, Management & Information Technology*, vol. 4, no. 2, pp. 61-86.
Dahlbom, B. & Mathiassen, L. 1993, *Computers in Context - The Philosophy and Practice of System Design*, Blackwell, Cambridge, Massachussetts.
Dahlbom, B. & Mathiassen, L. 1997, 'The Future of Our Profession', *Communications of the ACM*, vol. 40, no. 6.
Emam, K. E. & Briand, L. 1997, *Costs and Benefits of Software Process Improvement*, Fraunhofer - Institute for Experimental Software Engineering, ISERN-97-12.
Feldman, M. S. & March, J. G. 1981, 'Information in Organization as Signals and Symbols', *Administrative Science Quarterly*, vol. 26, pp. 171-186.
Fenton, N. E. & Pfleeger, S. L. 1997, *Software Metrics - A Rigorous and Practical Approach*, 2nd edn, PWS Publishing Company.
Foster, M. 1972, 'An Introduction to the Theory and Practice of Action Research in Work Organizations', *Human Relations*, vol. 25, no. 6, pp. 529-556.
Fowler, P. & Rifkin, S. 1990, *Software Engineering Process Group Guide*, Software Engineering Institute, Pittsburgh, Pennsylvania, USA, CMU/SEI-90-TR-24.
Grady, R. B. 1992, *Practical Software Metrics for Project Managment and Process Improvement*, Prentice Hall, Englewood Cliffs, New Jersey.
Grady, R. B. 1997, *Successful Software Process Improvement*, Prentice Hall PTR, Upper Saddle River, New Jersey.
Humphrey, W. S. 1989, *Managing the Software Process*, 1st edn, Addison-Wesley, Pittsburgh, Pennsylvania.
IFPUG 1994, *Function Point Counting Practices Manual Release 4.0*, The International Function Point Users Group (IFPUG).
Iversen, J., Johansen, J., Nielsen, P. A. & Pries-Heje, J. 1998, 'Combining Quantitative and Qualitative Assessment Methods in Software Process Improvement', in *European Conference on Information Systems 1998*, ed. W. R. J. Baets, Aix-en-Provence, France, pp. 451-466.
Johansen, J. & Mathiassen, L. 1998, 'Lessons Learned in a National SPI Effort', in *EuroSPI'98*, Gothenburg, Sweden.
Mathiassen, L. 1998, *Reflective Systems Development*, Dr. Techn., Aalborg University.
PSM 1998, *Practical Software Measurement: A Foundation for Objective Project Management*.
Schein, E. K. 1985, *Organizational Culture and Leadership*, Jossey-Bass, San Francisco, California.
Schön, D. A. 1983, *The Reflective Practitioner. How Professionals Think in Action*, Basic Books, New York.