IQ PRINCIPLES IN SOFTWARE DEVELOPMENT: IQ-2005 PROCEEDINGS

(Best Practice)

Michael Mielke
Die Bahn, DZB, TQM Team, Germany
Michael.Mielke@bahn.de

Abstract: This is an ongoing project, which is scheduled until 2006. This paper covers the so far reached project status / sharable public knowledge that is allowed to present here. The main objective is to implement DQ Principles in software development and to assure continuously measurement of Data Quality. This paper shows how we developed a 7-Steps-Method to determine, measure and improve DQ in our team.
Based on an actual business process description, we worked out the critical process points. Then we interviewed all Team Members to get a basic idea of how our Information Customers rate Data Quality at all and within the sub-processes looking at the critical points. After we collect all that data we could set up a system that allows us to measure DQ in a Customers focus. This System is the basis for all our further work. Currently we are somewhere between Step 4 and 5, this paper will show consolidated and anonymous data till Step 3.
We plan to continue publishing our results when further steps are taken.

Key Words: Data Quality, Information Quality, TDQM, Information Product, Software Development

FOREWORD

In the world today no one will complain that our production systems have a new dimension: Information. But do we already manage Information as a product? A few company do. We were faced with DQ while developing a new software for our consulting department. The old system was not able anymore to handle all projects and skills. After I took a look at the existing software and had a lot of discussions with several different users, I got the impression that the Data Quality was the core problem. Most of the users did not maintain their data on a timely basis. On the other hand they did not trust the Data very much. Both problems are very much related to each other. We then focused to the Data Entry Problem. To answer the question why users do not maintain their data properly was not easy. Most of the users told us that they had no access, or time to maintain their data. They all had in common very less profit of maintaining the data, while only a few of them where using the system to plan and control all consulting activities.
Our team decide to focus on two points, first to provide benefits for all kind of users and to determine and to measure DQ Prospective of our users continuously.
We are looking at our system users as customers first, even if they are collectors, custodians or consumers. Our customers define what quality is, so we want to know how they determine DQ according to the 16 dimensions that Richard Y. Wang describes in many of his published articles.

Introduction

To give our management and ourselves an idea about the level of DQ we have achieved as well as the level our customers require we need a certain mechanism in place that measures DQ continuously. The approach to manage information as a product, introduced by Richard Y. Wang is considering that quality in general depends on customer requirements and how the product fit into those requirements, took our group to the following idea.

- Each individual has an individual impression about DQ,
- Each individual has individual requirements about the level of DQ which is necessary for their work,
- The individual impression and/or requirements vary in between different business processes,
- The individual impression and/or requirements vary based on the organisational level they are looking at DQ,
- The individual impression and/or requirements are influenced in some case by the voice of the crowd.

The level of complexity is very high, while we are dealing with n-requirements, n-organisational levels, n-individual impressions and n-business-processes.

To get a working solution in place we developed the following 7 steps method:

1. Questionnaire your customers to get an idea which DQ Dimensions are the most important for them.
2. Analyse the business processes looking at the data streams that are involved.
3. Determine together with the involved users the critical points in their business processes
4. Questionnaire the involved users again, focusing on the process, sub-process and critical control points about their DQ requirements.
5. Determine metrics to measure if you fit the user requirements based on processes and critical control points
6. Measure and analyse
7. Questionnaire your customers, redesign your metrics if necessary and start with step 1 again.

1. Problem Description

We are an internal consulting group of 80 team members. Our organisational structure looks like a matrix. Beside our boss we have a sales-team, a production-team, a service and IT-team and 5 different customer focus consulting groups. Actually 65 consultants are working in more than 100 different projects within our holding. The team had started five years ago with three members and was growing rapidly during the last years. All tools they used to organise their work were developed internally by their own. There are two main databases, which are used to manage the team, one focuses on projects the other is focusing on the skills of our staff. Both databases are maintained basically looking more on data collection than on DQ.

Most of the users are concerned about systems accessibility, reliability of data, usability, understandability and so on. On the other hand the users do not keep the system uptodate timely, which is the most important for reports. Over all the system does not meet actual needs in case of reports, simulation and process control. Our team got the advantage to develop a new software system for the consulting department. This paper describes how far we have got today, our project is still ongoing.

2. *Problem Description*

To avoid very high complexity we based our metrics on the 16 dimensions Richard Y. Wang published. We also determine the user groups of our organisation and the basic business processes. All members of our department are members of one or more of the following user groups:

a) Head of Department,
b) Board of Managers,
c) Marketing and Sales,
d) Human Resources,
e) Production and Customer Focus Teams,
f) IT and Services,

The service group contains: Secretary, Accounting, Quality Management, DQ-Management, and IT-Management. In our team philosophy IT and Services is not a department. We look at it as independent services providers addressing all department members. Each service unit is lead by a Service-Unit-Manager who assists, consults and coaches also the Head of Department. The major processes are:

a) Marketing and Sales,
b) Product development (Training and Consulting),
c) Production,
d) Project management and Multi Project Management,
e) Business Consulting,
f) Management Coaching and Training,
g) IM Consulting and Training,
h) Medium sized Software Tools

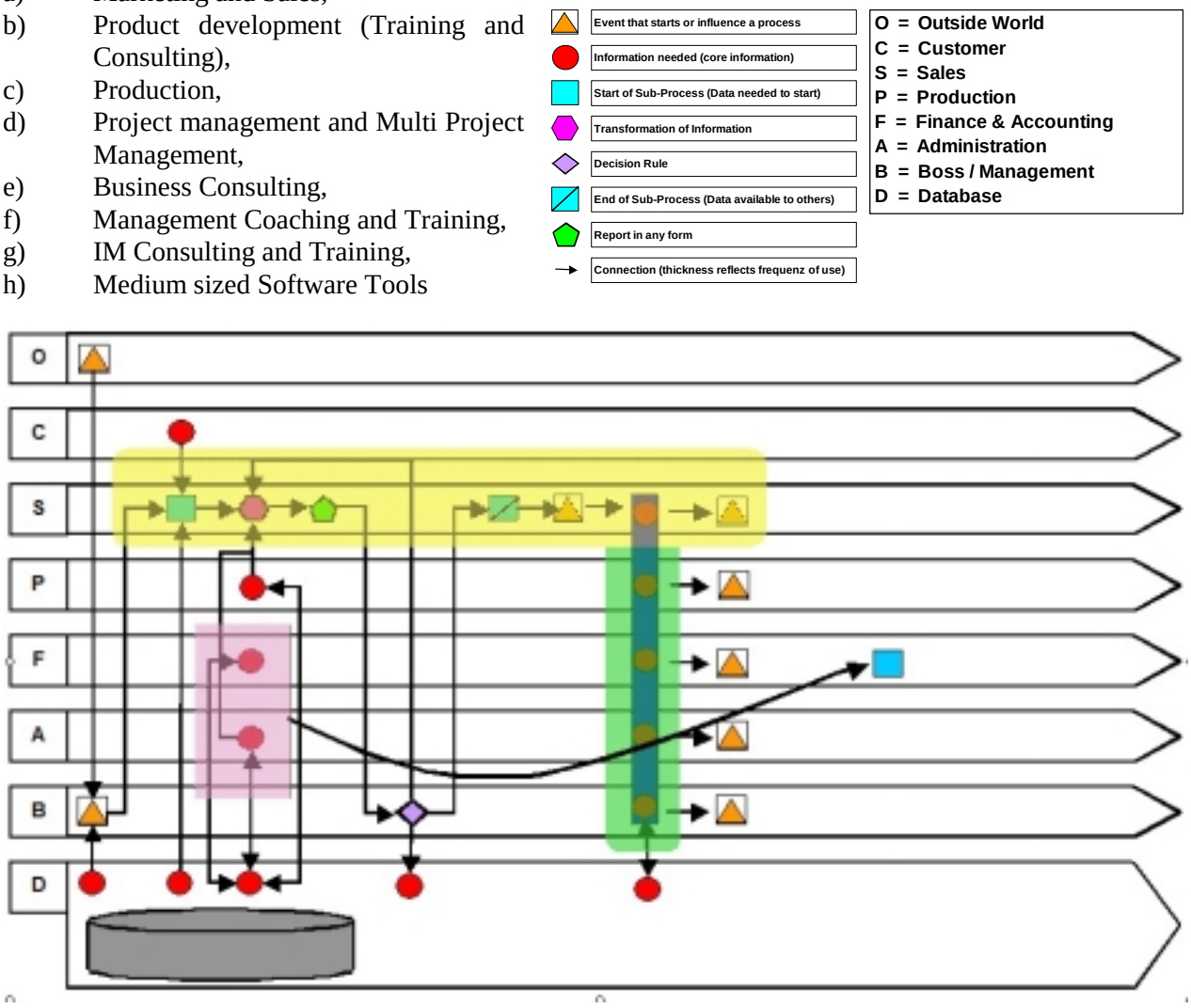

Figure 1 Abstract Process View (Example)

The abstract business process view illustrates the way we describe our processes and their relationship

basically using the concept of Data Flow Diagrams mentioned by Rich Wang, in conjunction with business layers. Our group prefers a sequence approach model so we could easily put time layers on the model to see when a specific information, product or duty is needed as an input for subsequent processes. This technique allows us to read the complex business process model from left to right without getting lost in complexity. This common business process modelling approach is often used by our team and allows us to review the business process separately in first place. Each process owner and his members can describe their core business activities and the relation to other mayor process owners. After we have got the descriptions we could combine them while focussing on the conjunctions.

The conjunctions were defined together with the involved process leaders. Taking a look on the quality requirements for information we always used the fully joint data quality-requirements of the involved groups.

3. Measure Customer Focus

When we where asked to develop a new management tool for consulting projects and consultant skills, we had to answer the following questions our managers asked:

a) How good is our DQ throughout the department and is there a continuous improvement?
b) How good is DQ in my Sub department / Team,
c) What are our requirements within a business process and between them?
d) Do we achieve the requirements?
e) What are the expectations and how are our customers think we achieve them?

To answer all that questions we set up a framework and designed a calculation model following our 7-Steps-Method.

1. Questionnaire your customers to get an idea which DQ Dimensions are the most important for them.
2. Analyse the business processes looking at the data streams that are involved.
3. Determine together with the involved users the critical points in their business processes
4. Questionnaire the involved users again, focusing on the process, sub-process and critical control points about their DQ requirements.
5. Determine metrics to measure if you fit the user requirements based on processes and critical control points
6. Measure and analyse
7. Questionnaire your customers, redesign your metrics if necessary and start with step 1 again.

The basic concept is to get an idea about the dimensions our “customers” within their business process are most interested in. We applied a basic Quality Function Deployment Technology known as the relationship matrix to prioritise the 16 DQ Dimension invented by Rich Wang and others. We decide for our work to focus on the first five dimensions, that’s the dimensions with the highest-ranking values. The following figure shows the framework we are using as well as some sample results to illustrate the ranking results.

Row is (0,1,2) related to column where: 0 means not important, 1 means less important and 2 means more important. Example: Accuracy is less important than Objectivity put in Cell C2 1	Accuracy	Objectivity	Believability	Reputation	Accessibility	Access security	Relevancy	Value-Added	Timeliness	Completeness	Amount of Information	Interpretability	Ease of understanding	Ease of manipulation	Consise representation	Consistent represantation	Cross-added value	Standardization on Base of 10 10*Cross-added-value divided by max Cross-added-value
Accuracy		1	1	1	1	1	1	1	1	1	2	1	1	2	1	1	**17**	6,07
Objectivity	2		1	2	2	2	2	2	0	0	0	1	1	1	1	1	**16**	5,71
Believability	2	2		2	2	2	1	2	1	1	1	1	1	1	1	1	**19**	6,79
Reputation	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1	**14**	5,00
Accessibility	1	1	1	2		1	1	0	0	0	0	0	0	0	0	0	**6**	2,14
Access security	1	1	1	1	1		1	1	1	1	1	1	1	1	1	1	**14**	5,00
Relevancy	2	1	1	2	2	2		2	2	1	1	1	1	1	1	1	**19**	6,79
Value-Added	1	1	2	2	2	2	1		0	0	2	1	1	2	2	2	**20**	7,14
Timeliness	2	2	2	2	2	2	2	2		2	2	2	2	2	2	2	**28**	10,00
Completeness	2	2	2	2	2	2	2	2	2		2	2	2	2	2	2	**28**	10,00
Amount of information	1	1	1	0	1	1	2	2	2	2		2	2	2	1	1	**20**	7,14
Interpretability	0	2	1	1	1	0	2	1	1	0	0		2	1	1	0	**13**	4,64
Ease of understanding	2	2	1	1	2	2	2	2	2	2	2	2		2	2	2	**26**	9,29
Ease of manipulation	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	**14**	5,00
Consise representation	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	**0**	0,00
Consistent represantation	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		**0**	0,00

Figure 2 Ranking Framework (Example)

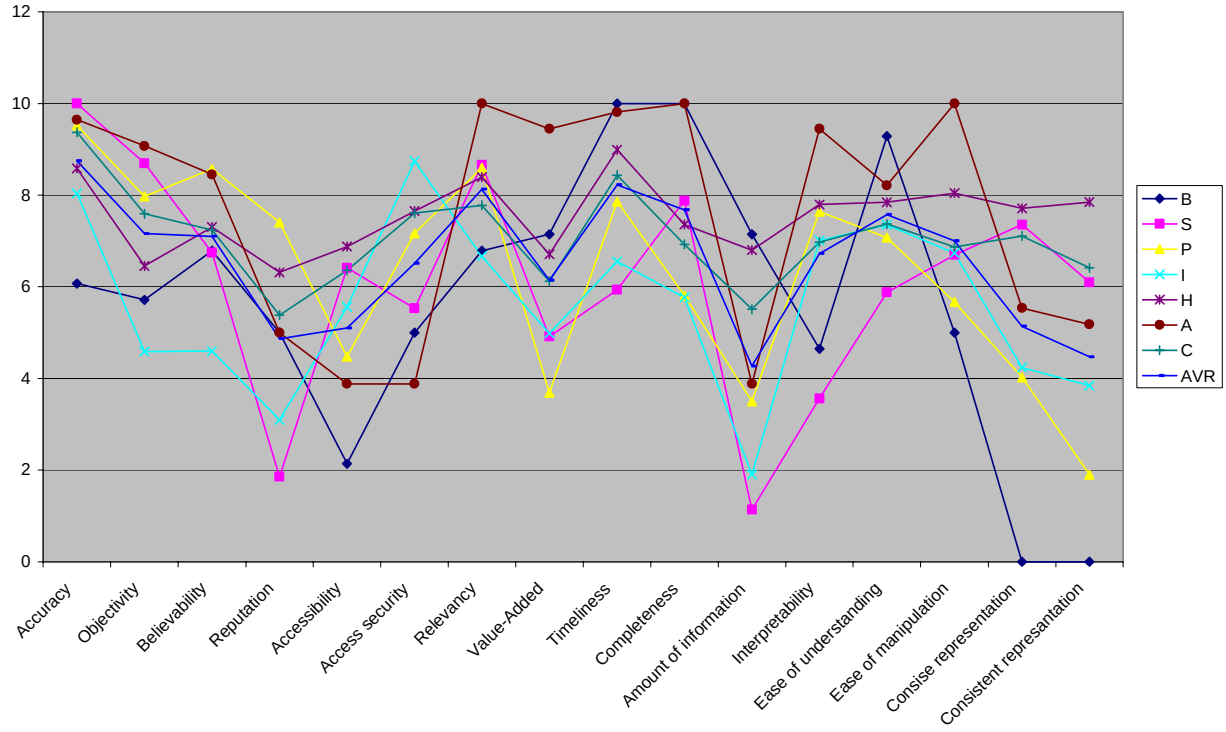

Figure 3 Cross Organizational IQ Dimension Rating

B = Boss, S = Sales, P = Production, I = Information Systems, H = Human Resources, A = Accounting, C = Consultant, AVR = Average (not weighted)

The different ranking of our “customers” was far away from what we have expected. First we thought that we would not be able to focus on only a few dimensions instead of all sixteen, so we first put a full joined table of the first five dimensions together. As a result we got 12 dimensions, then we put weights to each value as follows Rank 1 = 5 Points, Rank 2 =4 Points, Rank 3 = 3 Points, Rank 4 = 2 Points and Rank 5 = 1 Point.

	B	S	P	I	H	A	C	AVR
Accuracy	6,07	10	9,52	8,04	8,58	9,65	9,4	8,75
Objectivity	5,71	8,7	7,98	4,58	6,45	9,08	7,6	7,16
Believability	6,79	6,75	8,57	4,6	7,3	8,45	7,2	7,1
Reputation	5	1,86	7,4	3,1	6,31	5	5,4	4,86
Accessibility	2,14	6,41	4,48	5,56	6,87	3,88	6,4	5,1
Access security	5	5,53	7,17	8,74	7,65	3,88	7,6	6,51
Relevancy	6,79	8,66	8,6	6,67	8,39	10	7,8	8,13
Value-Added	7,14	4,92	3,69	4,98	6,71	9,45	6,1	6,14
Timeliness	10	5,93	7,86	6,55	8,99	9,82	8,4	8,23
Completeness	10	7,88	5,81	5,77	7,36	10	6,9	7,68
Amount of information	7,14	1,14	3,5	1,9	6,8	3,88	5,5	4,27
Interpretability	4,64	3,56	7,64	7,01	7,79	9,45	7	6,72
Ease of understanding	9,29	5,88	7,07	7,35	7,85	8,21	7,4	7,57
Ease of manipulation	5	6,7	5,67	6,74	8,04	10	6,9	7
Consise representation	0	7,36	4,02	4,24	7,71	5,54	7,1	5,14
Consistent represantation	0	6,1	1,9	3,84	7,85	5,18	6,4	4,47

Figure 4 Oriented Rating of IQ-Dimension

In our next step we accumulate this information about our “clients” requirements on a control card and provided it to developers and business analyst so they where able to check their designs against these requirements. We summed up the weights for each Dimension to get a better idea about the importance for the organizations view itself. This also gave us the possibility to rank the dimensions and make sure that the first tree dimensions were tracked even if they did not accrued in the individual view of the business process members. We still were locking overall at 12 Dimensions, but could focus as an organisation on tree (Accuracy, Relevancy and Timeliness) while each “customer” was focussing on 5 up to 7 dimensions.

Dimension	B	S	P	I	HR	A	C	Sum	General Rank
Access Security				5			1	6	5
Accuracy	1	5	5	4	4	1	5	25	1
Amount of Information	4							4	6
Believability	2		4					6	5
Completeness		2				4		6	5
Concise Representation		1						1	8
Ease of manipulation				1	2	3		6	5
Ease of understanding	5			3	1			9	4
Interpretability			1	2				3	7
Objectivity		4	3				2	9	4
Relevancy	3	3	X	X	3	5	3	17	2
Timeliness	X	X	2	X	5	2	4	13	3

Figure 5 Main focus IQ Dimension Control Card

4. *Information Flows in Business Processes*

Looking on the information flow in business processes is the starting position for our ways of developing ip-maps and determine which information products we have within our future system. This kind of view is based on the following determinations:

a) Business process description is well documented, describing the production, administration and connecting processes that in total map the organisation completely.
b) Data and Information has to fit user requirements based on the 16 Dimensions
c) Information in a business process has a value related to the risk of missing those data/information or it's tribute to the outcome
d) Information Pre-Products occur in more than one specific business process or sub-process and are a part of a complex information product.
e) Information Products are primarily used for decision purpose and contains out of more than one pre-product and transformation or presentation rules.
f) Transformation rules contain of business knowledge that adds value to data / information which was not in before the transformation.

The basic idea behind this concept is that information flows in however repeatable preproducts through our business processes and have a certain impact of the revenues generated in the organisation. Actually we cannot estimate this impact precisely but we can estimate the risk of bad data instead of perfect data or the additional work involved for the subsequent operation based on their experience.

On the following pages we will show a sample taken out of our business process model and how we applied our methodology. The project is ongoing due to this situation not all information products have been defined and valuated yet. We will publish additional results in the future.

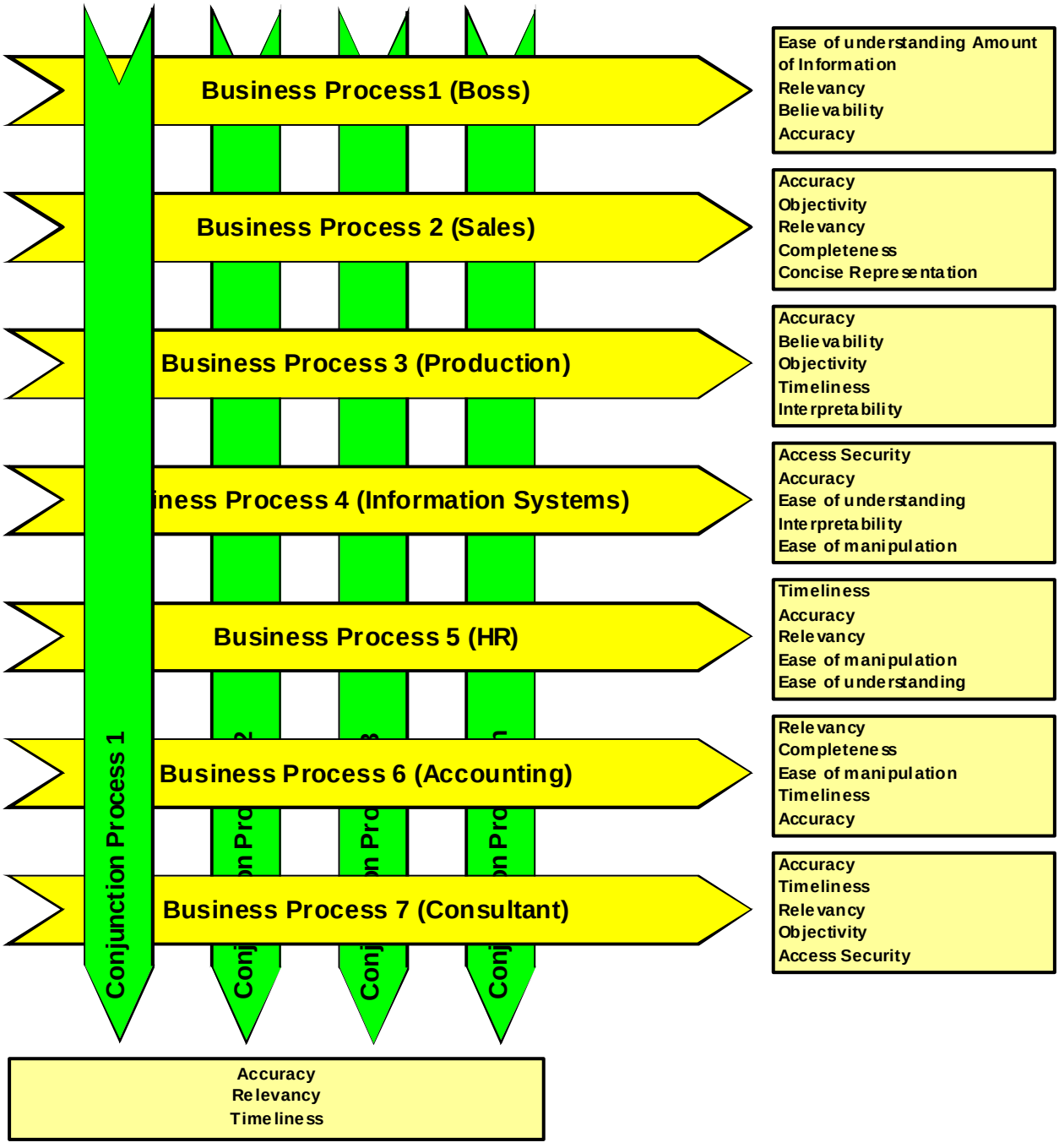

Figure 6 Business Process Meta Model (Example)

On the following pages we will show a sample taken out of our business process model and how we applied our methodology. The project is ongoing due to this situation not all information products have been defined and valuated yet. We will publish additional results in the future.

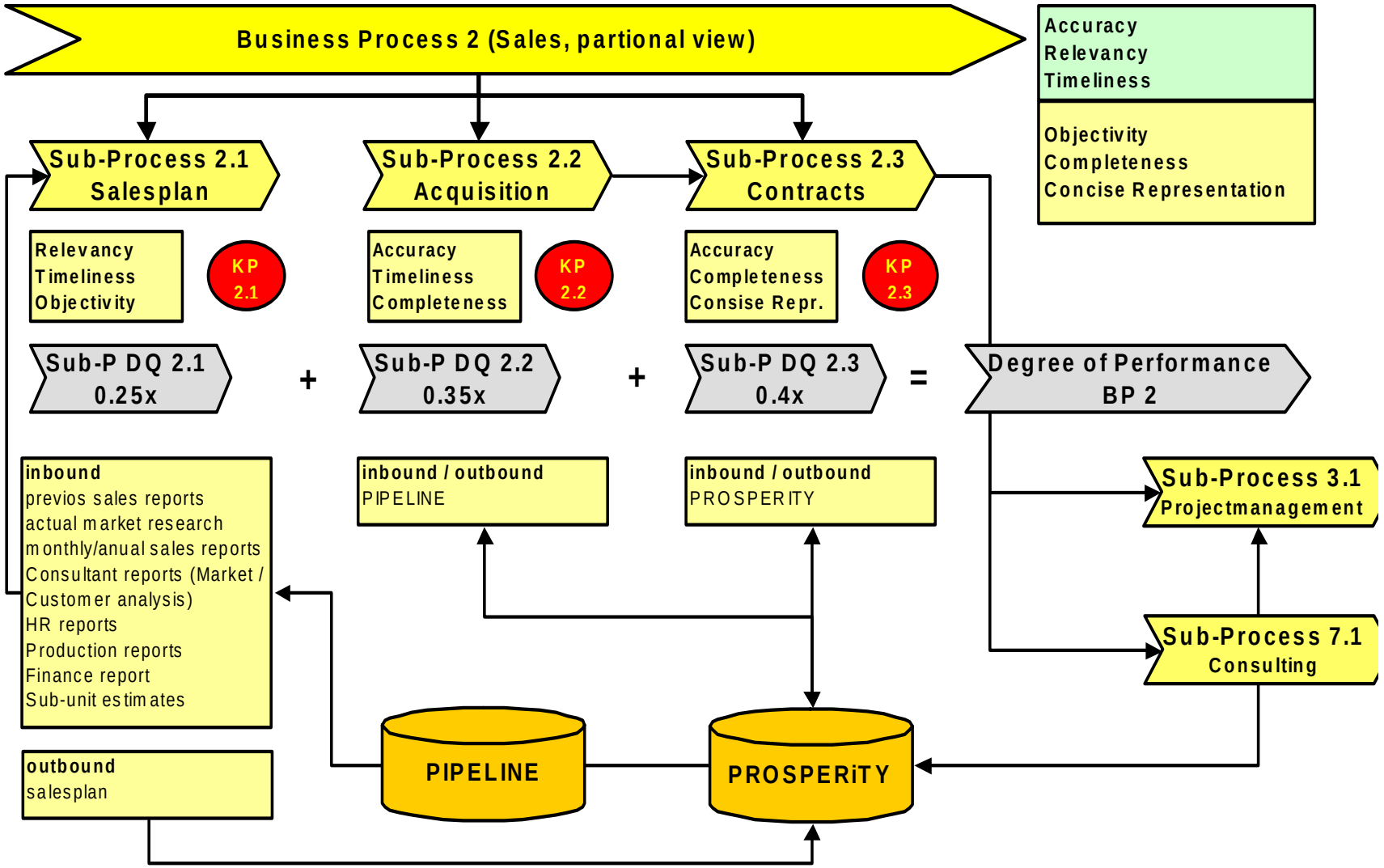

Figure 7 **Business Process (Example))**

The partial business process view as shown in Figure 7 displace our approach. This is a step-by-step model where we focus on the main sub-processes and what information is needed or used there the most. The sub-process 2.1 for example is about the sales plan development. According to our sales manager and his department they need the following information for analysis and preparation of a sales plan:

a) previous sales reports
b) actual market research
c) monthly / annual sales reports
d) Consultant reports about markets and clients
e) HR Reports
f) Production reports
g) Financial Reports
h) Sub unit estimates

It was very easy to get this information from the sales department while we were not looking at "database tables" so our colleagues from sales did not get lost in IT-Language. Taking the next step we collect examples of the "information products" the already use like monthly and annual sales reports. We then discussed with the sales department how these information products fits the sales iq-requirements like Relevancy, Timeliness and Objectivity. We developed a form we are using for all further questionnaires which is shown on the next page. The basic concept for our approach is, starting on a meta view and going in more deep step by step to determine what exactly the requirements on data are, were the data is stored or provided, how to measure and what values your metrics should have.

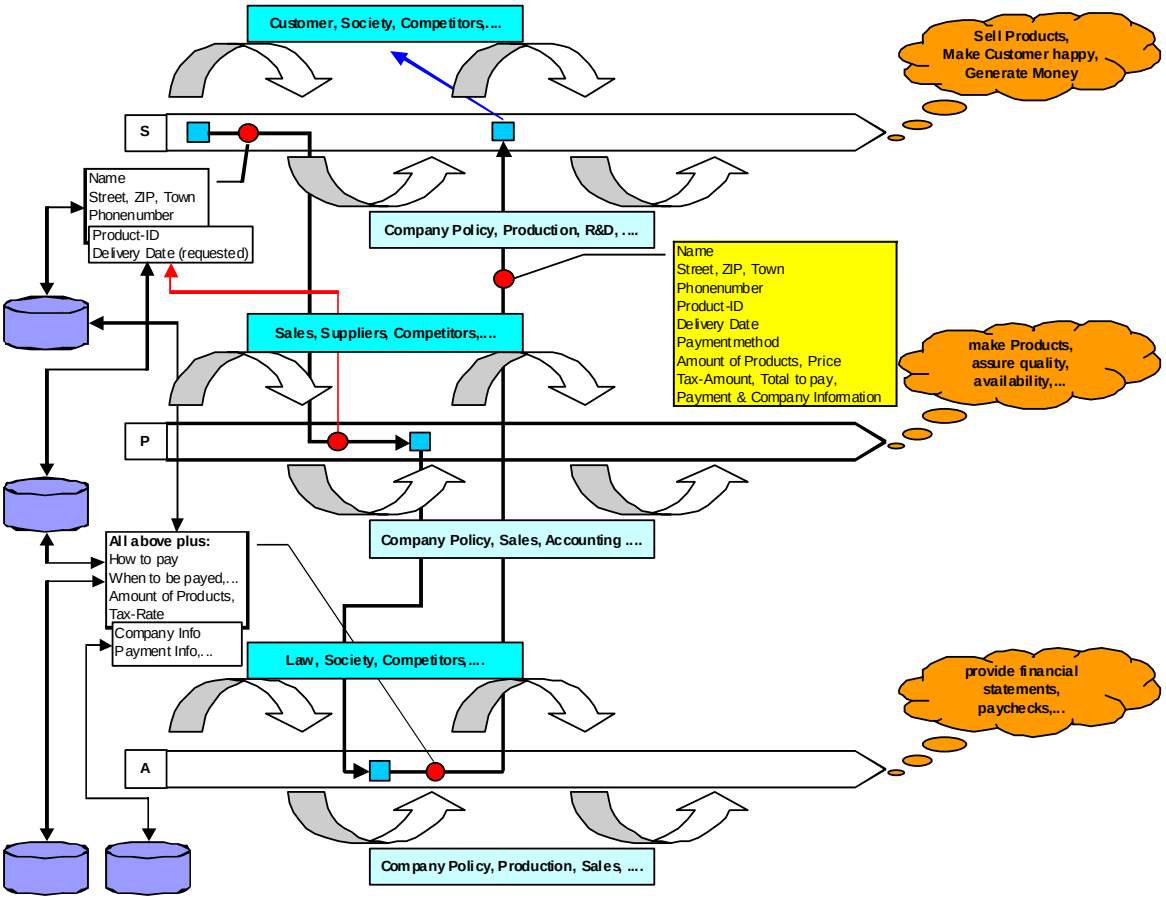

Figure 8 Example Sub Process (Sales Information Data Elements)

A monthly sales report for example (including the data you are looking for) has to be in the sales office not later than 5 days after computed at the end of the previous month (Timeliness). These sales reports have to include all sales sorted by main clients and products displaying the difference to the contract-data-base (Relevancy and Objectivity).
We expect the monthly-sales-report to fit all these requirements. To compute the key-performance-indicator 2.1 the following calculation method is used:
1/3 * P(Timeliness) + 1/3 * P(Objectivity) + 1/3 * P(Relevancy)
An sales report which is 2 days later provided than required, 4 main clients instead of 5 and with a 20% overall difference between sales-data-base and contract-data-base comes up to:
1/3 * (5/7) + 1/3 * (4/5) + 1/3 * (100 – 20) = 0,2380 + 0,2667 + 0,2664 = 0,7713

5. *DQ-Metric System*

To get an idea about the data quality of the information collected, transformed and provided by the new Software System we need to measure it throughout the sub-processes as well as in conjunction with other processes. We adopt a methodology we already have in place for quality management system metrics and allocation. The concept is to use the degree of performance instead of the current values, which allows us to allocate different key performance indicators independent from its origin dimension (miles versus pounds).
Allocation of various KPI´s has two perspectives, horizontally within a business process and vertically in conjunction with other business processes. Both views are based on weights, which we estimate with the involved process owners. Our group in total agrees that the value we are generating as is primary the result of a team than on individuals. Therefore we weighted the conjunction in a special way by multiplication. In the conjunction would fail (value of 0) the whole calculated value comes up to Zero.

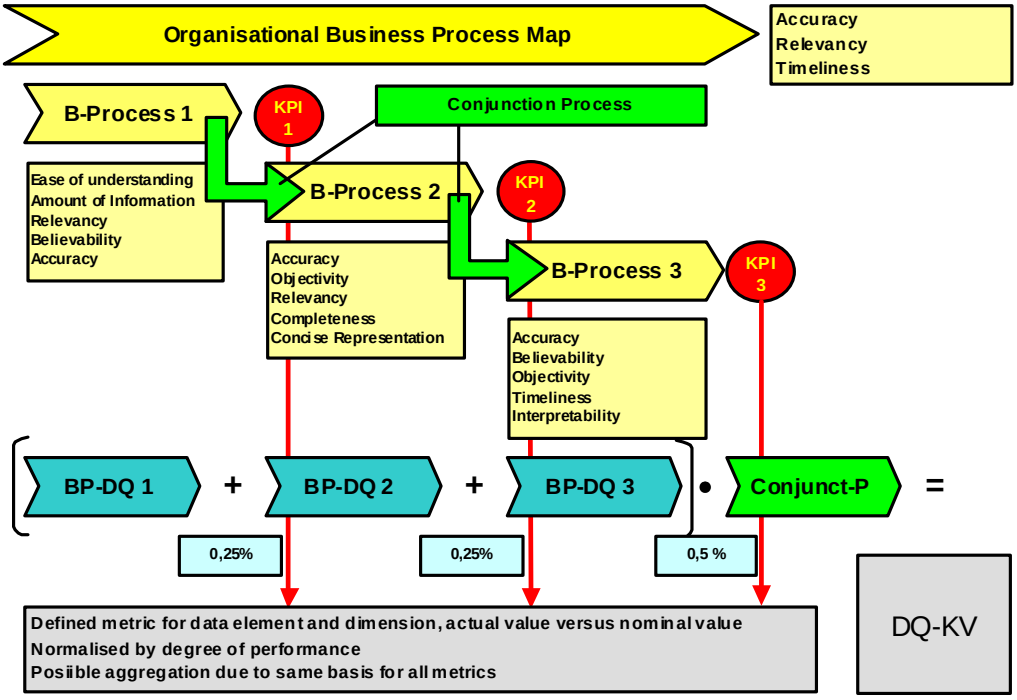

Figure 9 DQ-Metric Meta Model

6. *Lessons learned*

Getting "customers" data quality requirements into new designed software systems is not easy. It took and still takes us more time to define the measurable requirements than we expected. On the other hand this step-by-step approach already helps us to clarify our information needs and has result in a better cross organisational understanding and less complaints.

When we started the project I was the only one who has a basic idea about data quality, today our whole team knows that data quality is important or even the key for success. The software developers are working more customer-information flow-information quality oriented than before.

Implementing metrics in advance (timestamps, user-id, focus-id..) is a lot easier than trying to measure data quality requirements without appropriate values. It is also less cost intensive while we don not need to change or enhance a system. Last but not least it helps us to focus on the information (Tables, Elements) that are really needed, in most cases we where able to reduce the amount of data elements.

REFERENCES

[1] Huang, K., Y. Lee and R. Wang, Quality Information and Knowledge. Prentice Hall, Upper Saddle River: N.J., 1999.

[2] Juran, J. M., Juran on Quality by Design: The New Steps for Planning Quality into Goods and Services. Free Press, New York, 1992.

[3] Madnick, S. and R. Y. Wang, Introduction to Total Data Quality Management (TDQM) Research Program, TDQM-92-01. Total Data Quality Management Program, MIT Sloan School of Management, 1992.

[4] Mielke, Michael, Quality Function Deployment, Weka 1999

[5] Richard Y. Wang, Mostapha Ziad, Yang W. Lee, „Data Quality",Kluwer Academic Publishers, Nov 2000

[6] Yang, Y. and Y. R. Wang, Data Quality Calculus: A data-consumer-based approach to delivering quality data, CISL-91-08. Composite Information Systems Laboratory, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, 02139, 1991