Data Management in Analytical Customer Relationship Management

Dirk Arndt, Wendy Gersten

DaimlerChrysler AG, Research & Technology, Data Mining Solutions, FT3/AD,
PO BOX 2360 89013 Ulm, Germany
{dirk.arndt, wendy.gersten}@daimlerchrysler.com

Abstract. Customer Relationship Management (CRM) is a strategy to acquire new customers, to retain them and to recover them if they defected. The corresponding CRM goals can only be achieved if the right data sources are combined. This paper discusses what external and internal data are available along the CRM process and how they support the achievement of the specific CRM goals. Starting with defining a CRM process model and the belonging goals within the CRM programs - acquisition, loyalty and recovery - we explain the internal data situation. From here, we derive the need for external data and how one can merge and manage the information along the CRM process. The aim is to provide a rough guideline for the selection and combination of data sources among the CRM programs and to give hints how to overcome possible problems.

1 Introduction

Uniform products, along with individualization of customers has brought pressure for change in marketing practices. In the automotive industry, that implies generating additional product benefits by means of communication and services that are designed and delivered to match the individual needs of customers [3]. This is one of the main goals of Customer Relationship Management (CRM) [7].

Although CRM is an advanced concept, its implementation still requires the development of feasible approaches. From a practical point of view, a crucial issue concerns the systematical collection, storage, usage, and continuous improvement of customer data. The only companies able to construct lasting relationships with their customers are those that properly process and maintain an adequate volume of customer information [8].

The overall aim of this paper is to discuss the different data situations along the CRM process and to give hints on how to overcome data problems. We try to explain how one has to consider the whole process when handling data within a certain stage. These suggestions are valid for many CRM-oriented companies. But, for the ease of comprehension and to cover the specific issues of automotive companies, we restrict ourselves to this industry. (For more information about characteristics in car industry especially in acquisition campaign see [12]). In addition, we focus solely on private customers.

In section two of this article we aim at generating a homogeneous and complete understanding of the CRM process, since there is no generally accepted approach concerning this topic. Afterwards, we describe what data sources can be used in the car industry and how their importance varies along the CRM process. In section four we discuss options for handling these sources in order to fulfill the activities of operational CRM (oCRM). Here we talk about both, single activities and long-term strategy. We conclude by summarizing the main points and outlining open issues for further research.

2 The CRM Process Model

According to one of the earliest definitions Relationship Marketing (RM) is a "... strategy to attract, retain and enhance customer relationships" [6]. The term CRM is a later version of RM, having similar meaning, but used differently in literature. Within this article we use these terms synonymously.

Operational CRM includes all activities concerning the *direct customer contact*, such as campaigns, hotlines or customer clubs. Every oCRM activity is generally implemented in one of the three enterprise processes: sales, marketing or service, since these are the processes concerned with direct customer contact [10].

Analytical CRM (aCRM) provides all components to *analyze customer characteristics* (behaviors) in order to accomplish oCRM activities, with respect to the customers' needs and expectations [18]. There, the idealistic goal is to provide all information necessary to create a tailored cross-channel dialogue with each single customer on the basis of his or her actual reactions.

To reach this goal and aiming to show only one company image to the customer, it is necessary to look at CRM (oCRM and aCRM) as a cross enterprise process. Marketing, sales and service departments have to coordinate their responsibilities, activities, information systems and data. Fig. 1 illustrates this demand.

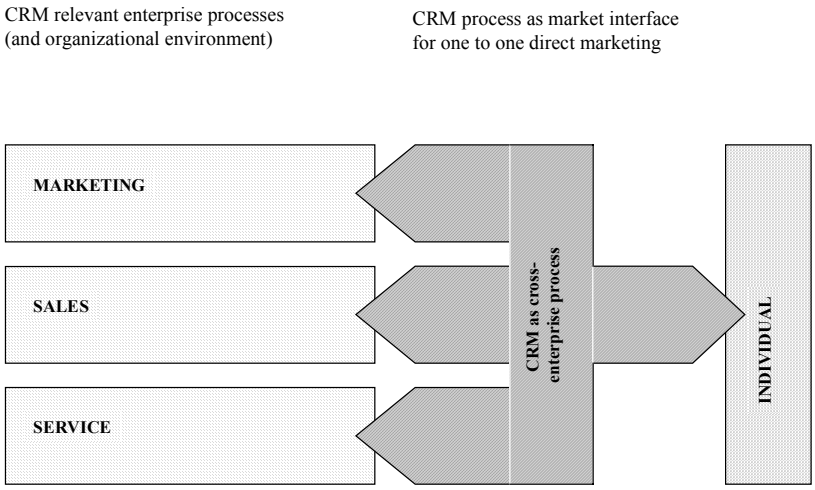

Fig. 1. CRM as cross functional process

There are many theoretical ways to organize the cross-functional CRM process within an enterprise. Generally, we divide the approaches into two broad clusters: *organization-oriented* approaches and *cycle-oriented* approaches. Fig. 2 gives an overview.

For our purpose, to discuss the issue in general, we base the CRM process on the Customer Life Cycle (CLC). Since CRM centralizes the satisfaction of the individual customers and their needs, it seems logical to do so. Furthermore, CLC is the most time stable approach and provides the ability to arrange oCRM activities according to each single life stage of the targeted subjects.

Nevertheless, in real life there may be good reasons to choose another approach, at least temporarily. If, for example, an enterprise enters a new market segment with a new product for the first time, it could be reasonable to organize the CRM process according to the product life cycle. Examples of such events in the automotive industry are the introduction of DaimlerChrysler's Smart in Europe and Volkswagen's New Beetle in North America.

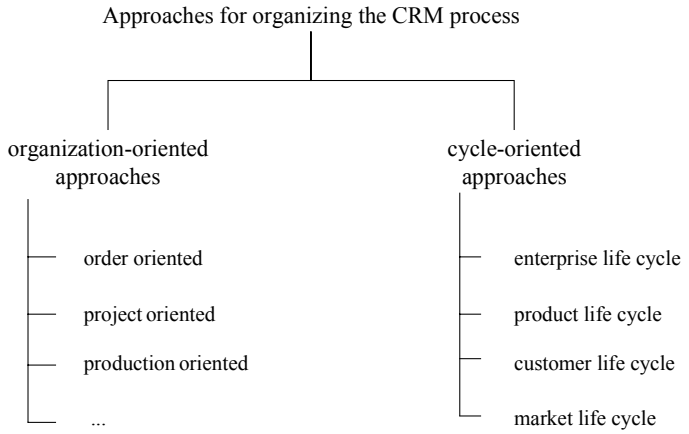

Fig. 2. General possibilities of the CRM process

Before enterprises can develop marketing or CRM strategies, they must understand how consumers make their purchase decisions. This decision process is called Customer Buying Cycle (CBC) [13]. We assume that the chain of all CBCs a single customer runs through, is his or her Customer Life Cycle. The process ends with the final stop of consumption. Fig. 3 illustrates the overall system and is described below.

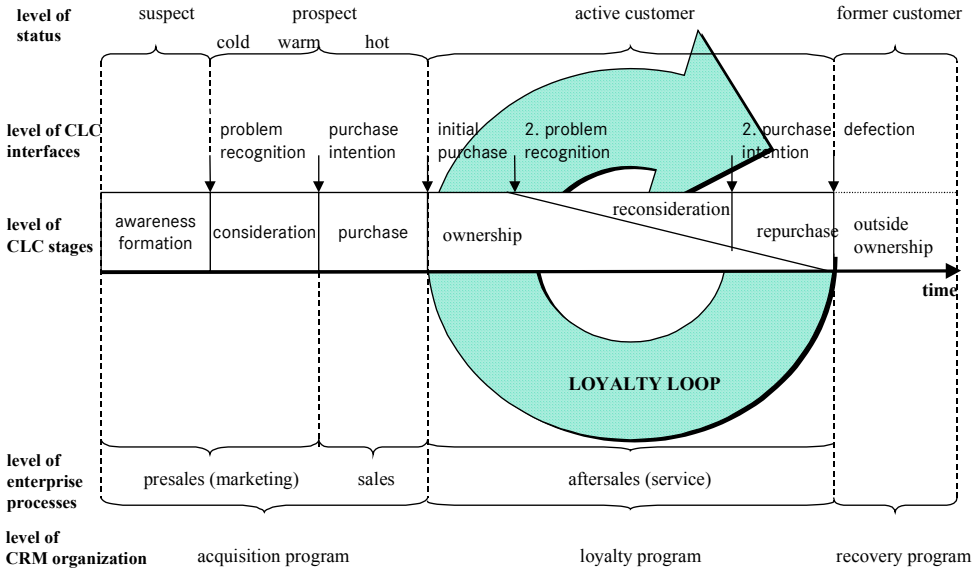

Fig. 3. The CRM-process based on Customer Life Cycle

We start with explaining the level of CLC stages. As mentioned before, the CLC refers to the various stages of the relationship between businesses and a (potential) customer. Now we take a closer look at these stages, assuming that the product in question is a car.

Each relationship starts theoretically long before the first "physical" contact. During the awareness formation stage a person is called "suspect" and gets in touch with lots of information on available brands in a determined market. From there the suspect builds a set of product brands known to him and suitable to satisfy his needs in general. This set is called the awareness set.

The first stage ends when the suspect recognizes a problem or an unsatisfied need or desire [17] and starts to seriously consider several alternatives. Based on the individual criteria and their importance, he evaluates and ranks the brands. Therefore, he typically gets in touch with the respective companies and turns into a "prospect." The result of this stage is the prospects evoked set [9].

Forming a concrete purchase intention, the person becomes a "hot prospect" (sometimes called hot lead) and enters the purchase phase. Here, he decides what kind of car or brand to buy, chooses where to buy the product, and negotiates the terms of sale (e.g. price, delivery, credit arrangement). This stage is closed by the initial purchase.

Now the prospect has become an active customer and starts to use the car. He evaluates it in order to determine whether it is performing as expected. The outcome of this evaluation lies in a range between satisfaction and dissatisfaction, which influences subsequent CBCs.

After a certain time, the customer begins to consider replacement. If he is satisfied with his car experiences, he is most likely to remain an active customer. With the loss of the actual car (selling or stop of usage) the ownership stage of the first CBC ends and the person either stays with the company (loyalty loop) or ceases being a customer. For the latter, he might come back during any subsequent CBC.

As mentioned above, the CRM relevant enterprise processes are cross-functional to the CLC based CRM process. But they can be differentiated depending on their target groups and depending on the events "direct customer contact" and "purchase" [15]. This enables us to relate them to our process model, at least for one CBC. The first direct customer contact might be in the awareness, consideration or purchase stages. To simplify the model, we assume that the first direct customer contact is identical to the purchase intention point. Therefore, we can construct the relationship shown on the level of enterprise processes of Fig. 3.

In order to organize the oCRM and aCRM activities along the CRM process, they are implemented as separate programs (as illustrated in Fig. 3 on the level of CRM organization) with clear interfaces, special goals and corresponding direct marketing activities, like acquisition campaigns or road shows [7]. Nevertheless, the programs have to be coordinated closely, as we will discuss later. Fig. 4 illustrates how the program-related CRM goals fit into the general goal pyramid of the enterprise [5].

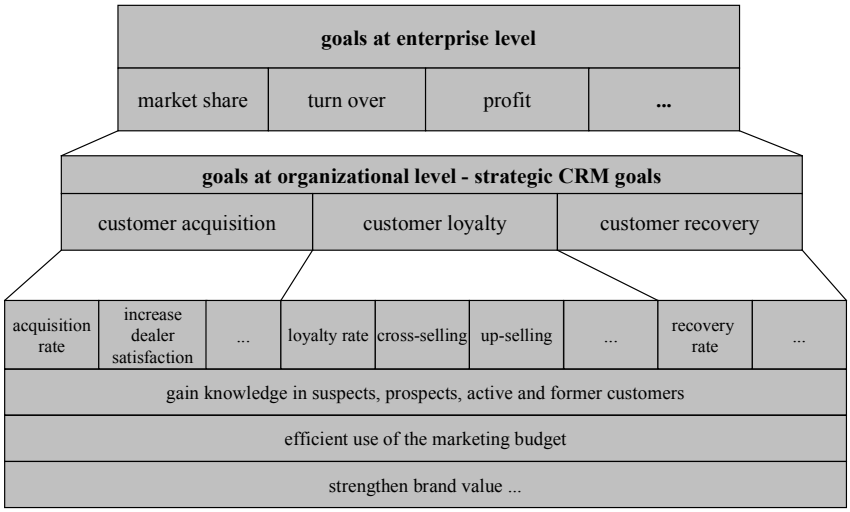

Fig. 4. Systematic of CRM goals

Within the operational CRM goals we distinguish between goals related to single programs, e.g. increase of usage or cross selling for the loyalty program, and overlapping goals, such as efficient use of the budget or creating customer insights. Both require strong interactions among the CRM programs, since program related goals partly need data gathered at some time in the process. For example, customers with a high probability to defect can be better detected in the loyalty program, if patterns of already defected customers and reasons for defection are known from the recovery program. Summing up, we may say that the synergetic effects of program interactions should not be ignored. We will refer to this point in section 4 again.

3 Data Sources for Analytical CRM

Analytical CRM supports the oCRM activities, as discussed before, through systematical collection, storage and evaluation of data. Data management is, therefore, the starting point for effective CRM and will continue to grow in scale and importance. We have to consider three aspects: the quantity, the quality and , the actuality of data available. Based on the continuous cycle of data management, we learn about people's behavior and needs. We want to use this knowledge to optimize our oCRM activities through targeting the right people with the right information and offers at the right time.

In order to provide high quality information for the CRM programs, we must discover first rate customer related data and evaluate them with suitable data mining techniques. Even if it seems that there is a wide variety of data, it is hard to find and merge the "right" data, even within the own company [14]. Generally, we can find data sources within and outside the enterprise. Typically, internal data is considered more valuable data, because it reveals true insights belonging to our company and products. Additionally, we aim to generate a competitive internal data situation in the long run. Since competitors have the same access to external databases as we do, it seems to be the only way to generate real advantages.

	Internal data sources				**External data sources**							
	Status of subjects				Important data sources for automotive industry							
	Suspect	Prospect	Active customer	Former customer	Regi-stration	NCBS	Micro-geogr.	Census	Life-style	Lists	Panel	NCE
Identification data												
name, postal address	👎	👍	👍	👍	👉	👍	👍	👍	👍	👍	👎	👍
telephone, email-address	👎	👉	👍	👉	👎	👎	👉	👉	👉	👉	👎	👉
Descriptive data												
socio/demographical	👎	👉	👉	👉	👎	👎	👍	👍	👍	👎	👍	👉
psychological	👎	👉	👉	👉	👎	👎	👍	👍	👍	👎	👍	👉
behavioral	👎	👉	👍	👉	👎	👎	👉	👎	👍	👎	👍	👉
geographical	👎	👉	👍	👉	👎	👎	👎	👍	👍	👎	👎	👎
financial	👎	👉	👍	👉	👎	👍	👉	👎	👎	👎	👎	👉
usage	👎	👎	👍	👉	👉	👍	👎	👎	👉	👎	👎	👉
purchase	👎	👎	👍	👉	👎	👎	👎	👎	👉	👎	👎	👉
product	👎	👎	👍	👉	👍	👍	👉	👎	👉	👎	👎	👎
Communication data												
channel	👎	👍	👍	👍	👎	👎	👎	👎	👎	👎	👎	👎
first contact date	👎	👍	👍	👍	👎	👎	👎	👎	👎	👎	👎	👎
first contact content	👎	👎	👍	👍	👍	👎	👎	👎	👎	👎	👎	👎

NCBS - New Car Buyer Study, NCE - Non-competitive Enterprises

Fig. 5. Overview of general data sources

In the first column of Fig. 5 we list the main information categories concerning customer characteristics [16]. The most important examples for external sources, as used in automotive industry, are shown on the right hand side of the figure. For internal data we show only the availability during the CLC, as they can be stored at different databases in any company. The hands

indicate for the internal data, when the data categories are most likely to be available. For external data they give evidence as to which data sources these categories can be found in. Because it is very difficult to measure information quality [11], they do not indicate the excellence of information.

After the short overview, we now want to relate the potential information sources to the CRM process. In order to do so, we start by describing the data situation for the CLC of a *single person* (illustrated by the gray area in Fig. 6). After that, we change the perspective to the *enterprise*'s point of view. This is necessary because we want to look at the data situation of each CRM program. Consequently, the CRM programs must be discussed independently of the fact of whether or not an individual passes through the whole process. The relationship between these two perspectives is that programs are static in their nature, but are passed through by a large number of evolving subjects.

Examining the data situation from a single customer's perspective, we lack internal information at the beginning of the CLC (see internal data in Fig. 5). There are two general ways to overcome this gap: we either buy data from external providers or collect it by means of surveys, lotteries, etc.. With the evolvement from a suspect to a hot prospect, the person releases more and more information. This improves the data situation within our company and the proportion of internal and external data changes slightly. At the point of the initial purchase there is a sudden jump. Buying the product, the person becomes an active customer and we receive personal information, e.g. about his favorite car (including equipment) and his financial situation. Over time, the proportion of internal and external data of an active customer changes more rapidly than before. Now we constantly gather information like usage behavior or service interest. If the customer stays active for several buying cycles we are theoretically able to develop the internal data up to a point where no additional external data is needed. If the customer defects to another enterprise, the situation changes again. Shortly after the defection, we still have lots of internal data. If we cannot win back the subject, these data age or get lost. Additionally, the subject creates new data without revealing them to us.

Note that the absolute amount of data is not considered. We also have to keep in mind that this is only a rough guideline for the automotive industry, and generally depends on industry, previous internal knowledge, CRM goals and CRM programs.

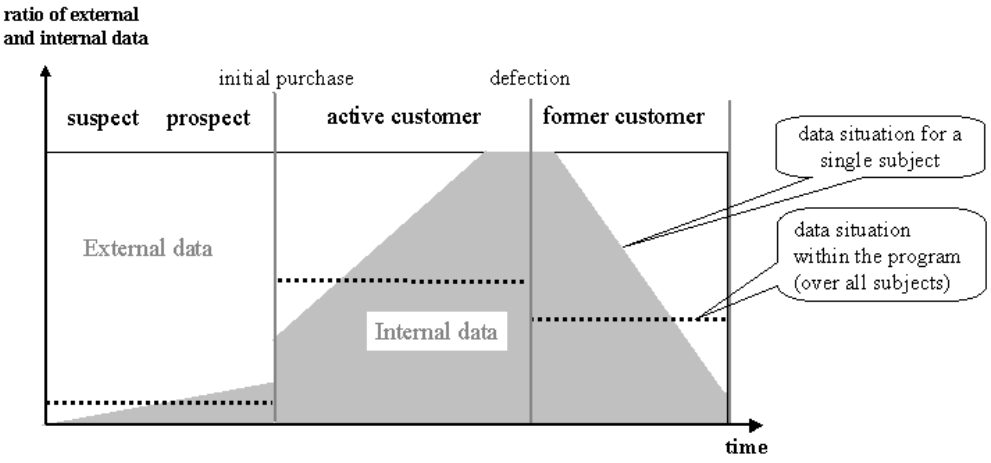

Fig. 6. The general data situation along the CRM process

If we now want to connect the data situation of a subject with the data situation of CRM programs, we just have to understand that each program focuses generally on another status level of the CLC (see Fig. 3).

This means that we always have a certain number of subjects with an identical status level and with a similar data situation within a specific CRM program. Consequently, the data situation of each program is almost static over time. If we take all subjects belonging to a certain CRM program for a certain spot in time and calculate the data average (in terms of quality, quantity and actuality), the results are the dotted lines in Fig. 6. As it becomes clear, we have different information levels of internal data amongst the programs. Therefore, we need more or less external data and for that reason the ration of external and internal data varies also.

4 Data Integration in Analytical CRM

In this section, we explain how the data described before can properly support the CRM process. To do so, we discuss the single programs as well as overlapping issues.

Sub-section 4.1 clarifies basic requirements and assumptions. From section 4.2 to section 4.4 we look at each CRM program. Before we outline special questions related to the corresponding program, we first restate the specific goals and then point out the internal and external data situation. Section 4.5 deals with the program overlapping goals and the resulting interactions between CRM programs.

4.1 Basic Remarks

The most likely situation is that the CRM programs have to be developed for companies that are already existing. In this case, the company must first capture the status quo of the internal data. Depending on their quality in terms of accuracy, actuality, completeness, etc., external data have to be purchased. Two or even all three CRM programs could be started simultaneously. Then, benefits of interactions between the programs are possible from the very beginning. But in companies that are already existing, we face difficulties with existing (non-optimal) conditions, e.g. in terms of internal data quality or given information systems.

If a company starts with CRM for the first time, in theory it could start from scratch. This would mean, that there are no existing customers and, therefore, no customer data. Naturally, the company in question would first develop an acquisition program, immediately buy external data, and form the relationship with the new customers along their CLC. Although this situation is possible (e.g. when smart AG entered the car market for the first time in the 90's), it is rare, especially in established industries. For the purpose of our article we will discuss both situations if there are any significant differences.

4.2 Acquisition

Acquisition mainly aims at establishing a dialogue with the suspects belonging to pre-defined target groups and at gradually converting them from suspects into customers. When a suspect goes through the acquisition program, the major challenge is to become our brand part of the awareness and later the evoked set of the subject. If the subject does not leave the acquisition program unplanned (e.g. because of no interest or unrecognized change of address), the program ends with the initial purchase (see Fig. 3).

As we have shown in Fig. 6, external data sources play an important role, especially in the acquisition program. Therefore, we would want to take a closer look at their suitability. In order to rank data sources in terms of their usefulness we have to find ranking criteria. In [2] we explain that the measurement of information quality is a complex task with various facets. Aiming to make a first rating of the general suitability of an external data source for acquisition, we try to restrict ourselves to four main aspects here. These are the number and quality of descriptive variables, the market coverage, the costs per contact and the expected response rate for suspects.

In order to run *continuous acquisition programs*, we need many new addresses every time and hence, the *market coverage* is a fact to take into account. Because we *lack internal data* about suspects as well as prospects we have to buy this information. For that, the *number and quality of descriptive variables* is important in acquisition. Especially at the beginning of a CRM project we are often obliged to proof its *usefulness and efficiency* within the company. Thus, we must consider the *expected response rate* and of course the overall *costs per contact*.

Considering the market coverage for passenger cars, registration data (for legal aspects), NCBS data (New Car Buyer Survey), panel data and data from non-competitive enterprises (NCE) can only be supplements here. Census data is not only highly aggregated, but also mostly available within micro-geographical data sources. If we focus on the remaining data sources (cf. Fig. 5), we receive the portfolio illustrated in Fig. 7.

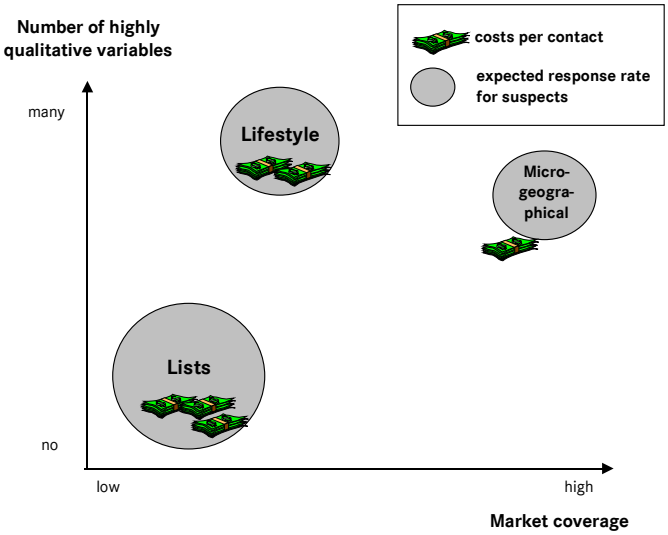

Fig. 7. Ranking portfolio for most important external data sources

Now we want to see how these data sources fall in place during acquisition. An ongoing aim of acquisition programs is the generation of new suspects. As we have learned in section 3 (Fig. 6), usually we have only internal information about them. As a result, in acquisition one aims to buy external data continuously.

If a company starts acquisition with direct marketing for the first time or is new in business, normally it lacks data about responders or customers, and thus it misses a proper target variable for the use of data mining techniques like classification or predictive modeling. Consequently, the company does not need any descriptive variables for prediction. Furthermore, because it is just starting acquisition campaigns and therefore did not use lots of publicly available addresses in the past, market coverage is not very relevant. Considering these two aspects as well as the highly expected response rates for properly selected lists, we recommend using them. This will help to achieve quick results in terms of generated prospects (responders) and to strengthen the position of CRM within the enterprise.

Nevertheless, there are two aspects to bear in mind about lists. First, addresses generated from lists are contacted frequently by many enterprises and therefore show a certain saturation. And second, suited lists are limited and certain social classes are neglected.

In the case of running acquisition programs the potential value of lists is less. For reason of their small market coverage, appropriate lists will at times be nearly exhausted. We can still use lists as frequent sources for the acquisition program, because there will appear new needs for the subjects already on the lists, new subjects in existing lists or completely new lists, but we also have to look for alternatives.

After running the first campaigns, the internal data situation is improved. The reactions to our mailings and the identification data (name, address, etc.) of the contacted suspects from the lists are the core of our internal database. Additional data can be added, if we send questionnaires with the mailings. They inquire e.g. about the next replacement need, the product, and the communication preferences of the prospect and give hints how to correspond with him properly. Since during acquisition we aim mainly to distinguish between potential customers and non-customers, we focus on data mining techniques like classification and prediction. Because of the now improved data situation we have positive responders and therefore a valid target variable. This can be used to predict the response behavior of subjects from other data sources. In order to do so, we have to find data that refer to both our responders and the targeted subjects. For this reason, it is advisable to refer to external data sources again.

If we look at the dimensions of our portfolio from Fig. 7, we now face new requirements. For generating predictive models we clearly need descriptive variables, although the specific

requirements vary for different data mining approaches and techniques. Additionally, we call for high market coverage. This is essential for maintaining ongoing acquisition campaigns. Micro-geographical and lifestyle data mostly fulfill the two requests. Because of the higher expected response rates, it makes sense to start using lifestyle data. If the data is exhausted, micro-geographical data can take their place. The more knowledge we possess about the responders (and responders that become customers) the finer we can select new subjects to contact. Thereby, the usually lower response rates of micro-geographical data compared to lists and lifestyle data, can be partly compensated by better targeting (and lower address costs).

4.3 Loyalty

Loyalty clearly focuses on establishing a permanent dialogue with active customers. The aim is to build a mutually profitable long-term relationship. Thereby, we pursue operational goals like increasing usage, up-selling and repurchase rates.

A person becomes a customer and therefore a potential target for the loyalty program, if he signs a contract (initial purchase). He stays within the program as long as he remains an active customer. With defection to another company (voluntary or forced) he will be handed over to the recovery program (Fig. 3).

From a data point of view, loyalty is the key program of CRM. Neither during acquisition nor recovery are there so many and such close contacts to customers. As a result we can gather lots of internal data with supposedly high quality. Therefore, the ratio between internal and external data will change over time, in favor of the internal data (cf. section 3). If many customers stay with our company and pass through several CBCs, theoretically there is a point where no additional external data is needed.

Taking a closer look at the quantitative and qualitative growth of internal data, we discover three general ways for improvement. First, we can collect information (like sociological or financial data) freely given by the customer. This data has very high reliability, if the customer has to prove the accurateness (e.g. for leasing or credit purposes). The data is less reliable if it is based on volunteer statements (e.g. from questionnaires). The second way for obtaining data is observing customer behavior. Here we get hard facts like mileage, product characteristics or use of services. As the third category there is data derived from the internal database. For deriving the data, we use basic statistical or data mining techniques like regression or neural networks (e.g. for building averages or estimating and prediction of variables).

Because of the growing internal data, the need for external data decreases the longer the loyalty program runs and is clearly lower than in the acquisition program. Nevertheless, there may always be a good reason to buy external data. As we will explain, there are two basic applications for external data in the loyalty program.

First, we can buy external data on a regular basis and thus extend the customer database with valuable information, *permanently*. For example, we could add socio-demographical data from micro-geographical sources in order to generate selection criteria for direct marketing activities. For applications like these, we still need data that refers to most of the customers, and consequently, we need high market coverage as found by micro-geographical or lifestyle data sources. Lists do not represent an additional value here, because they lack the descriptive variables and market coverage.

Second, we can use external data on a *non-permanent* base, e.g. for special tasks in market research. For these tasks we work mostly with samples. Consequently, we do not need high market coverage. Data sources with special topic related information (e.g. NCBS or panel data) have higher priority.

After talking about the general data situation in loyalty, we now want to look at the main tasks to fulfill. During his ownership, the customer is exposed to a variety of influences that can have an impact on his attitudes concerning our company, brands, or products. Such influences are called *triggering events*. There are two groups of triggering events we would like to point out, for reason of their high impact on the customer to business relationship. The first group contains events related to the *lifecycle* of the customer himself, such as marriage, death or unemployment. The second group encloses *product (car) or service* related events (e.g. accidents, car maintenance or mileage).

Product and serviced related events can sometimes be caused by lifecycle events. As shown in Fig. 8, all triggering events can be distinguished by expected and unexpected events as well.

A single triggering event or a combination of events can result in fundamental changes in the customer's needs and attitudes and therefore, his expectations and actions regarding the relationship to us may change, with serious consequences (see Fig. 8).

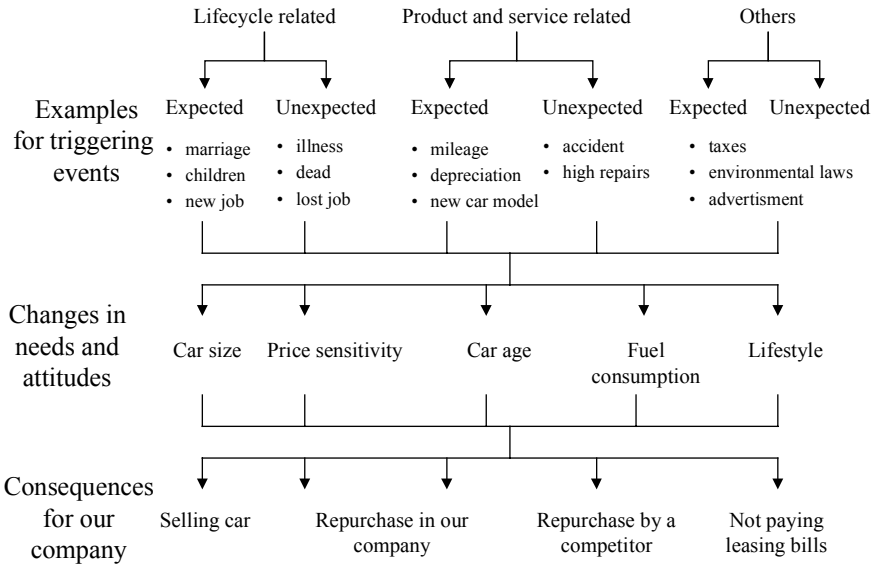

Fig. 8. Examples of triggering events and their consequences

One of the main tasks of aCRM for loyalty programs is the prediction (e.g. use of decision trees for predictive modeling) and recognition of triggering events (e.g. use of residual analysis for derivation recognition). Not only do we have to *prevent undesired consequences*, but additionally, triggering events often provide opportunities for *enhancing customer relationships*. The goal of prediction is hard to accomplish for expected events and barely possible for unexpected events (How to predict an accident?). For the latter we are mostly restricted to event recognition. Another approach could be the evaluation of likelihood classes as used e.g. for car insurance policies.

For the detection of triggering events we refer mostly to the use of internal data. This should be more detailed and have higher actuality than most external sources (Fig. 5). Examples for such triggering events are a sudden change of annual mileage that indicates different usage habits or a modified email-address that refers to a new company. External data might be used additionally for explaining such indicators (e.g. micro-geographical data can show new housing conditions).

Another important task for aCRM in loyalty is to provide customer insights for the tailored customer dialog. For that, we primarily use data mining applications from the fields of clustering (e.g. k-means) and description (e.g. association rules). In order to find customer segments we can use the external sources with topic related information, as described above. But to apply all customers to the appropriate segments we need to find corresponding selection criteria, which is available within our internal customer database.

4.4 Recovery

The customer becomes a target for the recovery program when he changes to another car company. Recovery pursues two goals. The main goal is to *win defected customers back*. But in order to do so and to avoid future defection, the second goal is to *learn the reasons for defection.*

The problem with recovering former customers is that when defection becomes known to us, the person has often already purchased a new car from a competitor. Considering the amount of time and emotions spent for choosing, as well as the money spent for acquiring the car, the customer won't sell the actual car in the near future. The average length of car ownership, e.g. in Germany, is approximately 3.5 years. That's why the goal of the dialogue cannot consist in winning back the customer immediately, but in establishing and keeping contact with him until his next repurchase. The final goal of the dialogue is to gain the top position in his evoked set again.

The reasons for defection must be questioned as one of the first steps of the recovery program. We need this information for planning the content and timelines of the recovering program and it is also helpful to prevent churn of actual customers in the loyalty program.

Because the recovery program deals with former customers, we generally have detailed internal information about the targeted subjects, at least more information than in acquisition. But this is true only for subjects who defected recently. The more time that has elapsed since people defected from our company, the less valuable their data becomes (see Fig. 5). More concretely, new information about car characteristics, as well as car usage and financial data emerge, but not in our enterprise. Unfortunately, such information is rarely available in public data sources. This implies that the lack of actual internal data cannot be compensated by external data. Additionally, the identification data becomes more and more obsolete and one day it is impossible to purchase external data because the customer's identification data is no longer valid. Regarding this debasing data situation, a recovery dialogue must be established as soon as the defection becomes clear. Only when the customer has been defected for a short time do we possess a lot of actual data about his contact address and his communication preferences. This is the precondition for a successful dialogue.

At this point we now know what do to when the defection is known already. In practice the customer rarely informs us of his defection. For that, it is necessary to find proper identification variables. There are two fundamentally different types of churn. People can churn *voluntarily* if they do not want to be customers anymore (e.g. because they are unsatisfied with the product or got a good offer from a competitor) or *involuntarily* if they are forced to stop consumption (e.g. because of unemployment or illness). Both types of churn can occur for many different reasons and can be recognized by different outcomes [7]. Involuntary churn is recognized more easily (e.g. people stop to pay their bills or the car is wrecked in an accident). For the recovery program we want to recognize mainly voluntary defection, because voluntary churners are more likely to return and stay good customers. For that we mostly have to use indirect indicators. Examples of such indicators are the abrupt stop of service bills or if loyalty measures get unanswered suddenly. In order to find these hints within our internal data we use the techniques of deviation recognition or basic statistics.

Before closing this section we want to point out that recognizing the reasons for defection is mainly part of the recovery program, but is important for the loyalty program as well. The behavior patterns and indicators discovered in recovery are the basis for churn prevention in loyalty.

4.5 Interactions between the CRM programs

As mentioned already within the preceding sections, there can be several interactions between the different programs. In this section, we want to examine them from different perspectives.

Looking at the interactions of the CRM programs we have to consider the *objects* of interaction. Naturally, subjects evolve from suspects to customers and therefore pass through the different program levels. These subjects are represented through their data and consequently the objects of the exchange are data. The exchanged data can differ concerning its *reference level*. First, data can be exchanged on an *individual level*. This means that data about a specific customer gained in the acquisition program is used in loyalty as well. Second, the data exchange can take place on an

aggregated level. This implies that insights gained about a certain group of people, for instance in loyalty, are used for marketing activities in the acquisition program.

Because external data is available for each program and can be bought any time, normally, we do not include it in the program interchange. But we must keep in mind that, when purchasing external data for one program, the external data needs in other programs can be taken into account. More concretely, it makes sense to consider data needs in loyalty and recovery programs when buying external data for the acquisition program and the other way around.

Individual internal data is exchanged mostly *forward oriented*, meaning that data gained in acquisition is used in loyalty and recovery. In fact the individual data is not transferred, but the customer file is enriched with all information gained about the subject during his CLC. Prospect data acquired during the acquisition program is retained and enriched with new customer data. If the customer churns, the inquired reasons for defection are stored in the same customer file as well. This type of data exchange is already assumed in the explanations within the programs and is not examined in detail here. Aggregated data can also be exchanged forward oriented. But the insights gained through the evaluation of aggregated data are especially advantageous in *backward direction* as we will explain.

For that, we want to consider the *reasons* for the data exchange. Of course, data is exchanged to gain further knowledge, but there can also be interaction to compare and adjust data. This means that individual data with similar content may be gathered in different programs. For example, we may inquire as to the preferred car characteristics of a prospect via a questionnaire in acquisition phase. When the prospect becomes a customer, we gain reliable knowledge about his chosen car characteristics. As discussed before, the latter are better than the volunteer non- proofed statements. We can update information about preferred car characteristics with data about chosen car characteristics. This update, again, is forward oriented on an individual level. Backward oriented we can analyze the differences between announced and true product preferences in order to make these insights available in precedent programs. This point refers to aggregated data.

Another example of backward oriented exchange of aggregated and individual data is the use of defection data to identify customers likely to churn. Therefore, we need information about the typical profile of churners. The corresponding data emerges within the recovery program and is transferred to the loyalty program. There, the descriptive variables about customers are used as input variables in order to generate a predictive model [4]. The output variable is the information regarding whether a customer has defected or not. The generated predictive model assigns a score to new customers that indicates their fit with the churners profile.

It becomes clear that backward oriented exchange of aggregated data often goes along with the use of data mining methods like classification, prediction, and clustering. Similar to the prediction methods just described, clustering permits distinguishing between different groups of responders, active customers or churners. Deviation analysis must also be established on an aggregated level in order to incorporate a certain standard deviation in the model.

5 Conclusion

5.1 Summary

So far, the selection and integration of data within the CRM programs was rarely mentioned and hardly ever treated in marketing or data mining research. But for both domains, combining the right internal and external data is a guarantee for marketing success and for the use of data mining techniques within analytical CRM.

Most research has been restricted either to CRM strategy without discussing the practical issues of deploying the strategy or to characterize data independent on the business problem. Criteria like actuality, accuracy, and completeness resulted. Although they are really important in evaluating data sources, they don't give evidence as to which categories of data should be used depending on the marketing goal and on the internal data situation. In this paper, we connected the CRM goals within acquisition, loyalty and recovery to the internal data situation. Based on that, we derived the need for external data and the need for data exchange between the CRM programs.

We showed several levels at which the CRM process can be described. First, we pointed out the CLC level showing the possible evolvement from a suspect over a prospect to a (former) customer. The amount and quality of internally available data grows when the subject becomes an active customer. Second, we illustrated the organizational level with its programs. We showed the modest internal data situation in acquisition, which make additional external data a must. Loyalty and recovery also require external data – but to a lower extent and with different focal points. Depending on the CRM goals we stated different requirements on the external data. While in acquisition market coverage is an important criterion, specific data about purchase history and behavior are much more important in loyalty and recovery. Furthermore, in order to reach the corresponding CRM goals the programs have to be coordinated closely especially for exchanging data and knowledge derived.

5.2 Further Research

Our paper has not focused on the question of how to distinguish between different data providers that offer the same or similar kind of external data. This problem arises after the needed data within the programs is determined. The selection of a wrong provider often can destroy the benefits of the precedent steps. We will discuss this issue in another paper.

Our thoughts were oriented on CRM programs in automotive industry. But the CRM process model and the data suggestions are applicable for other industries as well. Nevertheless, there are differences in goals and data between industries depending on their product (industrial goods, consumer or utility goods) and their customers (businesses or end-users). These differences will have an impact on the concrete design of the CRM process model, the CRM goals, the internal data situation and the resulting need for external data. Further research should examine the differences and discuss the resulting consequences.

6 References

1 Ahlert, H.: Enterprise Customer Management: Integrating Corporate and Customer Information. In: Henning-Thurau, T., Hansen, U. (eds.): Relationship Marketing, Springer, Berlin Heidelberg New York (2000) 255-264

2 Arndt, D.; Gersten, W.: External Data Selection for Data Mining in Direct Marketing. In: International Conference on Information Quality 2001, MIT, Boston (2001) (to appear)

3 Arndt, D.;Gersten, W.;Nakhaeizadeh, G.;Wirth, R.: eCustomers – How the Internet affects our relationship to our customers. In: DaimlerChrysler (ed.): eMagine: Journey into the eFuture, Stuttgart (2000), 23-27

4 Arndt, D., Gersten, W., Wirth, R.: Kundenprofile zur Prognose der Markenaffinität im Automobilsektor. In: Hippner, H., Küsters, U., Meyer, M., Wilde, K. (eds.): Handbuch Data Mining im Marketing. Vieweg, Braunschweig Wiesbaden (2001) 591-606

5 Becker, J.: Marketing-Konzeption. 5th end. Vahlen, München (1995)

6 Berry, L.L.: Relationship Marketing. In: Berry, L.L., Shostack; G.D., Upah, G.D. (eds.): Emerging Perspectives in Service Marketing1983, Chicago (1983) 25-28

7 Berry, M.J.A., Linoff, G.S.: Mastering Data Mining. Wiley, New York (2000)

8 Chojnacki, K.: Relationship Marketing at Volkswagen. In: Henning-Thurau, T., Hansen, U. (eds.): Relationship Marketing, Springer. Berlin Heidelberg New York (2000) 49-58

9 Diller, H.: Vahlens großes Marketinglexikon. Deutscher Taschenbuchverlag, München (1994)

10 ECCS: CRM Defining customer relationship marketing and management. In: http://www.eccs.uk.com/crm/crmdefinitions/define.asp, 27.01.1999 (1999)

11 English, L.P.: Improving Data Warehouse and Business Information Quality. Wiley, New York (1999)

12 Gersten, W., Wirth, R., Arndt, D.: Predictive Modeling in Automotive Direct Marketing: Tools, Experiences and Open Issues. In: Proceedings of the 6th International Conference on Machine Learning. ACM, New York (2000) 398-406

13 Muther, A.: Electronic Commerce. Springer, Berlin Heidelberg New York (1999)

14 Pyle, D.: Data preparation for data mining. Morgan Kaufmann Publishers, San Francisco (1999)

15 Schmidt, R.E.: Mit Customer Relationship Management zum Prozessportal. In: Bach, V., Österle, H.: Customer Relationship Management in der Praxis. XY, Berlin et al. (2000)

17 Schweiger, A., Wilde, K.D.: Database Marketing – Aufbau und Management. in: Hilke, W.: Direkt Marketing. Gabler, Wiesbaden (1993) 89-125
18 Skinner, S.J.: Marketing. Houghton Mifflin, Boston et al. (1990)
19 Wilde, K.D.: Date Warehouse, OLAP und Data Mining im Marketing – Moderne Informationstechnologien im Zusammenspiel. In: Hippner, H., Küsters, U., Meyer, M., Wilde, K. (eds.): Handbuch Data Mining im Marketing. Vieweg, Braunschweig Wiesbaden (2001) 1-19