

Time Related Factors of Data Accuracy, Completeness, and Currency in Multi-Channel Information Systems

Cinzia Cappiello, Chiara Francalanci and Barbara Pernici

Politecnico di Milano, Milano, Italy

Abstract. Multi-channel information systems involve data redundancy, which, in turn, requires the implementation of synchronization mechanisms among overlapping databases. This need for periodic data realignment causes data quality issues related to delays in data updates, which can vary across databases and corresponding channels and services. Out-of-date data reduce the overall currency of databases and, in turn, decrease their accuracy and completeness. This paper proposes time-related measures of accuracy, completeness and currency and discusses their mutual relationships in multi-channel information systems.

1 Data accuracy, completeness and currency

Critical data quality issues are related to the degree of data integration, which is a consequence of design choices on the overall architecture of a company's information system (IS). This paper focuses on multi-channel information systems in which services are typically provided through a variety of channels, including branches, agents, call centers, Web sites and mobile devices. Customers should be able to access consistent information through all channels and at all times, irrespective of the service that they request. This level of data quality requires a fully integrated multi-channel information system, which, unfortunately, is rarely implemented. More often, systems are composed of different applications with separate, redundant and possibly inconsistent databases. Data quality problems caused by redundancy and misalignments influence three fundamental data quality dimensions: accuracy, completeness and currency [4]. Different definitions of *accuracy* are provided by the data quality literature [5], [6]. This paper adopts the definition proposed in [5], where accuracy is associated with data values and is defined as a measure of the proximity of a data value v to some other value v' that is considered correct. A measure of accuracy can be associated with data sources. It can be defined as the ratio between the number of correct values and the total number of values available from a given source [5]. The definition of data *completeness* is consistent across research contributions. In [5], completeness is associated with data values and is defined as the degree to which a specific database includes all the values corresponding to a complete representation of a given set of real world events as database entities. According to this definition, it is possible to obtain an objective measure of the completeness of a data source

by adding up significant data values. *Currency* is not provided a standard definition in the literature. This paper adopts the definition proposed in [1], where currency is defined as a time interval that goes from the time when data are updated and the time when data are used. The multi-channel architecture with the highest degree of data redundancy and, correspondingly, lowest degree of integration is shown in Figure 1. Each channel-functionality combination is served by a different software application and each application has access to a separate operational database od_{ij} . Two strategies can be distinguished to improve the

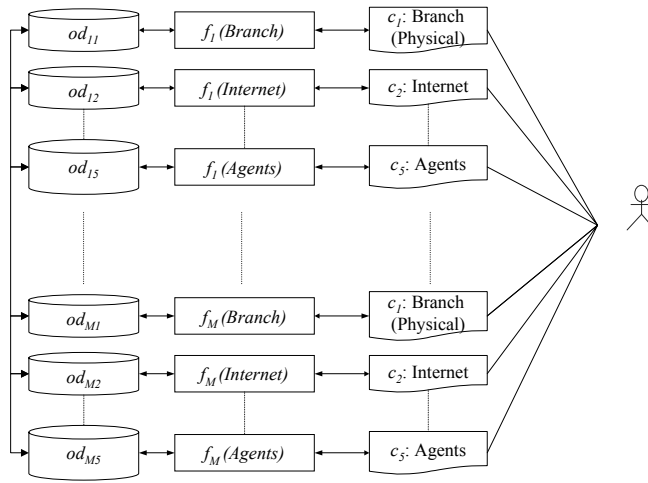


Fig. 1. IS architecture with the lowest degree of data integration

degree of integration of the architecture shown in Figure 1:

- *Channel integration strategy*: each channel has a private database which is shared among multiple functionalities.
- *Functional integration strategy*: each functionality has a private database that is shared among multiple channels.

In a multi-channel information system that is not fully integrated there is necessarily data redundancy among operational databases. Data sharing across operational databases raises a need for periodic data realignments. Each pair of operational databases od_{ij} and od_{mn} can be supposed to be aligned with a refresh period $rt_{ij,mn}$, which can be seen as the time interval before the data used by the m^{th} functionality of the n^{th} channel is updated with data created or modified by the i^{th} functionality of the j^{th} channel, and vice versa. To describe the behavior of customers, operational databases are associated with two parameters:

- *The operational frequency of f_{ij}* , defined as the average frequency with which users of the i^{th} functionality of the j^{th} channel change a data unit of operational database od_{ij} .

- *The combined frequency $cf_{ij,mn}$* , defined as the average frequency with which users of the i^{th} functionality of the j^{th} channel change a data unit in $od_{ij} \cap od_{mn}$

In order to guarantee a given level of accuracy, completeness and currency, refresh periods should decrease as the combined frequency of data changes increases. Based on these variables, accuracy, completeness and currency can be provided a mathematical measure consistent with the theoretical definitions provided in [5], and [1], respectively. A discussion of the mathematical method that has been used to calculate them can be found in [3].

2 Simulation results

Simulations compare the mathematical measures of accuracy, completeness and currency in architectures with different degrees of data integration. The following types of financial institutions are considered:

- Global, national and regional institutions, depending on their size and geographical presence.
- Physical or virtual institutions; the latter exclusively operate through technology based channels, while the former also operate through physical branches.

Users are classified into three categories depending on the average number of transactions that they complete over time: *Active users*, executing more than 200 transactions per year; *Moderate users*, executing more than 20, but less than 200 transactions per year; *Sleepy users*, executing less than 20 transactions per year. Different types of financial institutions have a different mix of active, moderate and sleepy customers, with different patterns of access to services across channels and, consequently, different values of the operating and combined frequencies. According to simulation results, global and national institutions show a trade-off between accuracy and completeness. While the former is maximized by the channel integration strategy, the latter increases with a functional integration strategy. A similar trade-off is also found between currency and completeness. As an explanation, accuracy and currency are significantly reduced if data changes from transactions performed by customers are not updated on all channels in real time; they are less sensitive to transactions that are not performed by a customer, but add new customer data, such as an inbound payment that must be credited to a customer's account. Customers will realize that they have received a payment simultaneously from all channels even if the architecture is not fully channel integrated. Conversely, completeness is highly sensitive to the creation of new customer data, even if they are a consequence of inbound transactions. Therefore, from a data quality standpoint, global and national financial institutions should choose a channel integration strategy. It should be noted that according to the professional literature both strategies seem to be followed in practice by global and national financial institutions. Note that the trade-off between accuracy and completeness disappears in favour of the functional integration strategy if customers are supposed to distribute

evenly among channels. Currency, accuracy and completeness are maximized by a functional integration strategy in regional financial institutions. The low frequency of operational transactions in regional institutions is a likely reason for these trends. Results confirm from a data quality standpoint the advantages of the functional integration strategy that is reported to be the followed in practice by most regional banks [2].

Virtual financial institutions show a trade-off between currency, accuracy and completeness, similar to global and national institutions. However, the trade-off disappears in favour of the functional integration strategy for high values of refresh time. These results indicate that virtual institutions can implement a channel integration strategy only if their architectures supports frequent realignments among channels. Conversely, if the refresh period is longer than 60 hours (3 days) a functional integration strategy should be implemented to guarantee maximum accuracy and completeness. This seems to raise data quality issues for virtual institutions sharing branches with multiple physical institutions.

3 Conclusions

Data integration across channels increases currency and accuracy more than integration across functionalities in the majority of cases. However, a functional integration strategy is more efficient when users' accesses are distributed evenly across channels and in all cases allows to increase completeness of databases involved in selected architecture. Future work will examine volatility requirements and will present a model to obtain consistent values of refresh periods based on required quality levels. Besides, future research will conduct a survey to collect real data on customer behaviour. Real data could be used to select the correct mix of channel and integration strategies depending for different patterns of customer behaviour and provide methodological design guidelines.

References

1. Bovee, M., Srivastava, R.P., Mak, B.: A Conceptual Framework and Belief- Function Approach to Assessing Overall Information Quality. Proceedings of the Sixth International Conference on Information Quality, 2001
2. Bracchi, G., Francalanci, C., Bognetti, A.: La Banca multicanale in Europa, Edibank 2002 (in italian).
3. Cappiello, C., Francalanci, C., Pernici, B.: A model of data currency in multi-channel financial architectures. In Proceedings of the Seventh International Conference on Information Quality (ICIQ'02), Boston, MA, USA, 2002.
4. Jarke, M., Lenzerini, M, Vassiliou, Y., Vassiliadis, P.: Fundamentals of Data Warehouse. Springer-Verlag Berlin Heidelberg 2000
5. Redman, T.C.: Data Quality for the Information Age. Artech House, 1996.
6. Wang, R. Y., Strong, D.M.: Beyond Accuracy: What Data Quality Means to Data Consumers, Journal of Management Information Systems, Spring 1996, Vol.12, No.4, pp.5-34