

Institute of Information Management



University of St.Gallen



Global Data Synchronization Current Status and Future Trends

Jan Schemm, Christine Legner, Boris Otto

Table of Contents

| | |
|---|-----------|
| Introduction | 3 |
| Global Data Synchronization | 3 |
| Data collection and sample | 4 |
| Survey results | 5 |
| Data pool customers | 5 |
| Service portfolio | 6 |
| Information services | 6 |
| Integration services..... | 7 |
| Supplementary services | 9 |
| Revenue models | 10 |
| Conclusions and future trends | 10 |
| The role of data pools in multilateral master data synchronization..... | 10 |
| The emerging GDSN landscape | 10 |
| Implications for future development..... | 12 |
| References | 13 |

Abstract

Retail and consumer goods industries are considered one of the most advanced value chains with regard to electronic collaboration. Recent studies have underlined the importance of master data alignment between retailers and their suppliers in the implementation of Efficient Consumer Response (ECR).

This study explores the multilateral exchange of master data in the retail supply chain using the Global Data Synchronization Network (GDSN). Based on interviews with ten data pool providers, it analyzes the current state of multilateral master data exchange as well GDSN implementation. Issues under examination include the current customer base, the service portfolios offered, and revenue models of the participating data pools.

The characteristics of the sample suggest that the current GDSN landscape leads to an increasing specialization of service providers in either *Mega-Pools* or *Local Specialists*. The study characterizes these emerging types of data pools and discusses some implications on the future development of strategies for data pool providers as well as the preference of users.

Introduction

Since the 1980s, retailers and their suppliers have been a prominent example for electronic collaboration and subsequently developed the vision of Efficient Consumer Response (ECR) in order to align their activities more closely. ECR encompasses supply-side concepts (i.e. cooperative logistics) for improving procurement and distribution as well as demand-side concepts (i.e. cooperative marketing) which aim at increasing turnover and improving customer service by selective pricing, promotion, and placement policies (Reyes 2005). A growing number of studies report that poor data quality, in particular outdated or wrong product master data, negatively impacts the benefits retailers and their suppliers pursue by implementing tighter forms of collaboration (Grocery Manufacturers of America, Food Marketing Institute et al. 2003; Global Commerce Initiative and Capgemini 2005; Accenture 2006). Consequently, the timely exchange of master data is viewed as the key to maintain the data consistency among business

partners and to realize electronic process integration along a global supply chain.

While industry associations such as Global Standards One (GS1) and the Global Commerce Initiative (GCI) have been issuing product data standards since the 1970s, B2B intermediaries have subsequently strengthened their support for the multilateral exchange of master data by offering data pools for the retail and consumer goods industries. Their success in achieving critical mass however has been limited given the large number of data pools with specific local or industry focus. In 2004, the Global Data Synchronization Network (GDSN) was launched with the objective to establish federation between the data pools and thereby promote many-to-many relationships between retailers and their suppliers independent from their home data pool. Expectations toward the GDSN initiative are high, given the recent consolidation of data pools as well as announcements in favour of GDSN by leading retailers and their suppliers.

This study aims at assessing to what extent data pools have adopted the GDSN concepts. Based on interviews with 10 data pool providers and additional experts, which have been conducted in April and May 2007, the study analyzes the current state of multilateral master data exchange and GDSN implementation. It concludes with trends and recommendations related to the future role of data pools.

Global Data Synchronization

According to ECR, the supply and demand chain relies on electronic communication of transactional data such as forecasts, orders, shipping notifications or invoices. In order to interpret the exchanged data correctly, the different actors need to pre-align additional contextual information. For instance, an electronic order message contains at least a trading partner identification number, an identification of the product to be delivered, as well as the amount ordered and the accepted price. Since no further description of the customer or the product is included, manufacturer and retailer are only able to process the order message correctly if they maintain consistent partner and product information in their information systems. Inconsistencies in the contextual information might e.g. occur if a manufacturer discontinues a certain product variant or changes the package size with-

out disseminating the change to all trading partners.

The contextual information which forms the basis of supply and demand chain collaboration between retailers and manufacturers is also referred to as master data. Important master data objects include trade items, business partners or prices and conditions. Data synchronization or data alignment denotes the process of timely updating master data to maintain the data consistency between the business partners (Bowling, Licul et al. 2004; Nakatani, Chuang et al. 2006), thereby supporting efficient supply and demand chain processes.

In practice, the bilateral exchange of master data based on message standards has proven to be costly and complex to realize with a larger number of business partners. As a consequence, data pools have been promoted. As a specific type of B2B exchange (Corsten and Hofstetter 2001; Sparks and Wagner 2003), the data pool assumes the role of an intermediary which collects and redistributes master data. By offering multiple interfacing technologies (e.g. EDI using PRICAT, XML or web interface), it eases integration and reduces entry barriers to electronic data synchronization. Since the number of data pools exceeds 20 exchanges worldwide, network effects come into play, and

data pools face start-up problems (Shapiro and Varian 1998) which often prevent them from attracting the critical mass.

The Global Data Synchronization Network (GDSN) targets at establishing interoperability and federation between the data pools (Bowling, Licul et al. 2004; Nakatani, Chuang et al. 2006): When master data is published at a GDSN compliant data pool, it can be accessed from all other certified data pools based on the GS1 Global Registry, which acts as a global directory for the registration of items and parties (cf. Figure 1). GDSN federates data pools via the Global Registry and ensures interoperability by defining standards and protocols for pool-to-pool-synchronization.

Data collection and sample

The survey has been conducted in 2007 in the form of semi-structured interviews with data pool representatives. Data pool providers were selected based on a listing of certified data pools provided by GS1 (GS1 2007). This sample of 23 data pools represents the large majority of data pools providers for the retail and consumer goods industry. 10 data pool providers participated in the survey, representing roughly 98% of all products registered in the GS1 Global Registry (cf. Figure 2).

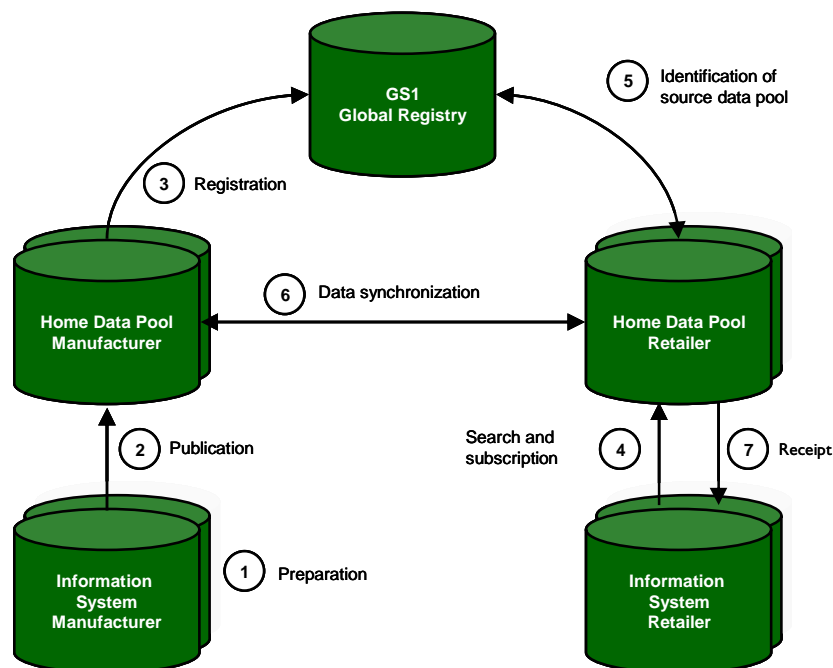


Figure 1: The Global Data Synchronization Network (GDSN)

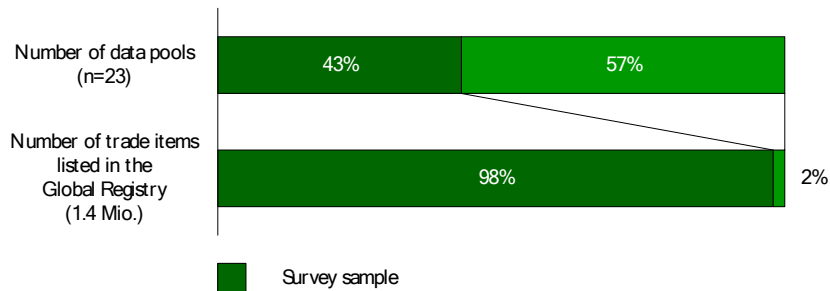


Figure 2: Survey sample

The majority of the participating pools is based in Europe (6) followed by North (3) and South America (1). The interview partners were responsible for the data synchronization business at the management level: 4 interviewees were CEOs, 4 area managers and 2 product managers for data synchronization. Additional interviews with representatives from GDSN Inc. as well as infrastructure service providers validated the survey results.

Survey results

Data pool customers

The majority of the survey participants offer their synchronization services to customers from the fast moving consumer goods (FMCG) sector, including food and non-food products (cf. Figure 3). Only data pools A to D show a significant sector diversification to apparel & sports, consumer electronics and hardware, whereas data pool I

concentrates on office supplies.

The value of using a data pool is primarily determined by its reach, i.e. the number of business partners connected to the provider. The customer base of the data pools in the sample varies significantly (cf. Figure 4). From the data pools represented in the survey, A has by far the largest number of customers, with an emphasis on North America. Whereas suppliers dominate in most other data pools, A has an exceptionally high fraction of retail customers with on average approximately 4 suppliers per retailer. The following pools B and D, which have a regional scope, serve 2200 and 1200 customers respectively. With a significantly lower customer base, C is the only data pool that operates on three continents. The remaining data pools provide services in merely one country or a small number of neighbouring countries and attract a significantly smaller customer base.

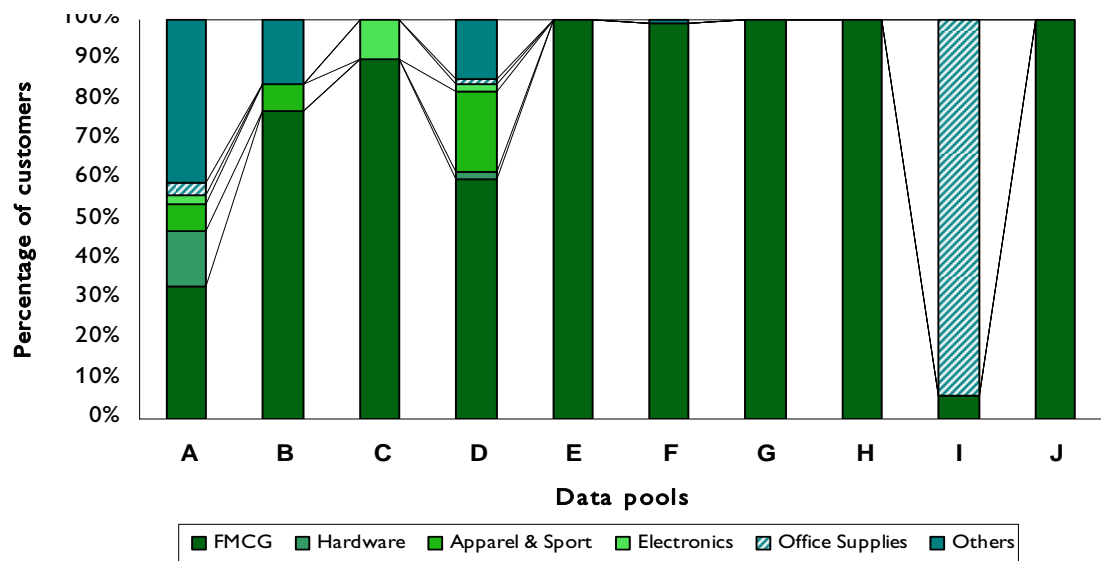


Figure 3: Sector distribution of data pool customers

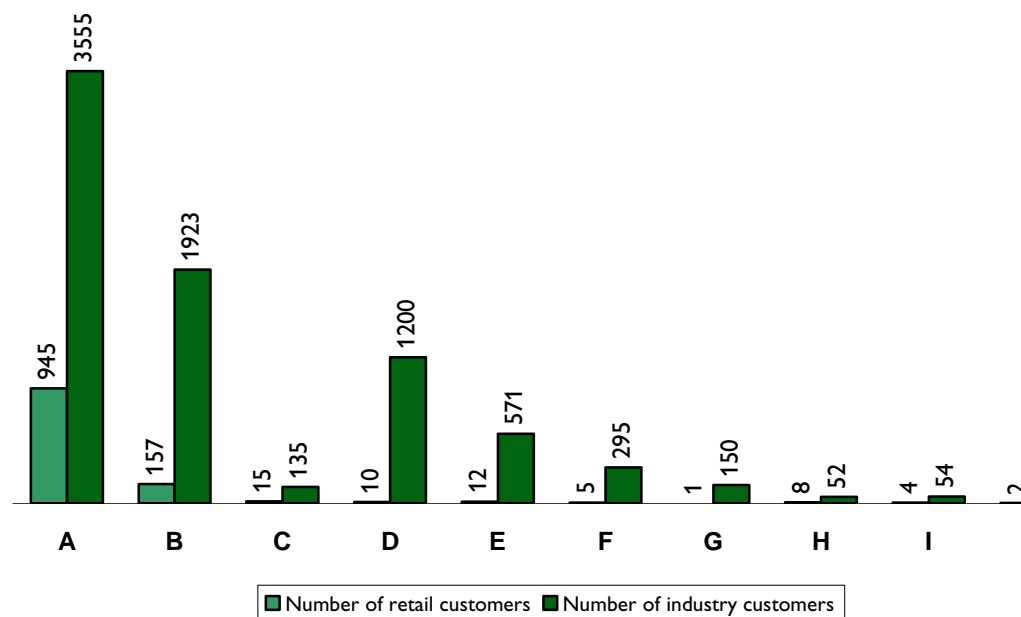


Figure 4: Customer base of the data pools

Service portfolio

The core services of a data pool are (1) information services for collecting and disseminating master data information and (2) integration services which establish electronic connectivity among business partners. There are three directions in which data pools complement their portfolio via (3) supplementary services: data quality services, educational and training services and consulting and implementation services.

Information services

The synchronization of article master data is the essential service offered by all examined providers (cf Table 1).

All pools support the approximately 200 standardized GDSN core attributes, and additionally provide sector- or country-specific extensions. The overall number of product attributes managed by a pool exceeds 1000 in some cases, although the survey revealed that an average customer only uses approximately 30-50 attributes, mainly basic data such as identification and classification attributes and logistics data such as trade item dimensions.

Seven data pools of the sample complement their product data services with additional business partner data such as contact or location information. In addition, the support for the exchange of price information and conditions is steadily growing. Five interviewees already provide this service, in some cases still using proprietary technology. Further data pool providers stated that they are currently implementing price synchronization and

| Service | | Data pool | | | | | | | | | |
|--|-----------------------|-----------|---|---|---|---|---|---|---|---|---|
| | | A | B | C | D | E | F | G | H | I | J |
| Data synchronization | Articles | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| | Business partners | ● | ● | ● | ● | ● | ● | ● | ○ | ○ | ○ |
| | Prices and conditions | ● | ● | ● | ● | ○ | ● | ○ | ○ | ○ | ○ |
| | Documents | ● | ● | ○ | ● | ○ | ○ | ○ | ○ | ● | ○ |
| ● Service offered, ○ Service not offered | | | | | | | | | | | |

Table 1: Information services

will pursue certification in the near future.

In addition to structured product, party and price information, four data pools offer services for the exchange of additional semi-structured data. The current emphasis is on product images in different resolutions and safety sheets for dangerous goods.

Integration services

The primary scope of integration services is to electronically connect multiple customers with their suppliers. The core connectivity services provided by all data pools establish and manage electronic connections between the attached business partners. Additionally, data pools provide different communication channels and standardize or translate data messages (cf. Table 2).

| Service | | Data pool | | | | | | | | | | |
|------------------------|---------------------------------|------------|---|---|---|---|---|---|---|---|---|---|
| | | A | B | C | D | E | F | G | H | I | J | |
| Connectivity | | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | |
| Communication channels | Human-Human (e.g. paper-based) | ○ | ● | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | |
| | Human-Maschine (e.g. Portal) | ● | ● | ● | ● | ● | ○ | ● | ● | ● | ○ | |
| | Machine-Machine (e.g. EDI, XML) | ● | ● | ● | ● | ● | ● | ○ | ● | ● | ● | |
| Standardization | Identification | GTIN | ● | ● | ● | ● | ● | ● | ● | ● | ● | |
| | | GLN | ● | ● | ● | ● | ● | ● | ● | ● | ● | |
| | | DUNS | ● | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ● | ○ |
| | | Others | ○ | ○ | ● | ○ | ○ | ○ | ○ | ○ | ● | ○ |
| | Classification | GPC | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| | | CCG | ● | ● | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| | | Others | ● | ○ | ○ | ○ | ● | ● | ○ | ○ | ○ | ○ |
| | Message Content | GS1 XML | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| | | PRICAT | ○ | ● | ○ | ● | ● | ● | ○ | ● | ○ | ● |
| | | Others | ● | ● | ○ | ● | ○ | ● | ○ | ○ | ● | ○ |
| | Message Syntax | XML | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| | | EDI | ● | ● | ○ | ● | ● | ● | ○ | ● | ● | ● |
| | | CSV | ● | ● | ● | ● | ○ | ○ | ○ | ● | ● | ○ |
| | | Excel | ● | ● | ● | ● | ○ | ○ | ○ | ● | ● | ○ |
| | | Others | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ● | ○ |
| | Transport | HTTP | ● | ● | ● | ● | ● | ○ | ● | ● | ● | ○ |
| | | SMTP | ● | ● | ○ | ○ | ○ | ○ | ○ | ○ | ● | ○ |
| | | EDIINT AS2 | ● | ● | ● | ● | ● | ● | ○ | ● | ● | ● |
| | | X.400 | ○ | ● | ○ | ○ | ● | ○ | ○ | ○ | ● | ○ |
| | | Others | ○ | ● | ○ | ○ | ○ | ○ | ○ | ○ | ● | ○ |

● Service offered, ○ Service not offered

Table 2: Integration services

The communication channel relates to the integration level on which customers can access the informational services of the data pool: human-human (e.g. paper-based), human-machine (e.g. access via portal) and machine-machine communication (e.g. EDI- or XML-based electronic data exchange) (McAfee 2005). B is the only data pool that provides a non-automated human-human interface. The extra-charged data processing service is an offering for customers that would like to participate in article master data synchronization but try to avoid the master data maintenance as well as the programming and converting of interfaces. The customers send article forms to the service provider, which checks, processes and uploads the data to the pool application.

With regard to the other types of interfaces, human-machine-interaction still dominates: most interviewees reported that the majority of customers transfer master data via the portal interface, leading to manual data re-entry. The low level of automation indicates that the majority of retail and industry companies are still struggling with integrating the external data synchronization interfaces with their internal product information systems.

Data pools foster industry-wide standardization by defining de-facto standards on the relevant levels of agreement i.e. semantics, communication, syntax and transport. All data pools interviewed passed the GDSN-certification and thereby proved to support the defined message choreography for global data synchronization. They build on the semantics defined by the core data model from the GS1 Global Data Dictionary (GDD). Specifically, they all agree on the mandatory standards for identification (GTIN, GLN) and classification (GPC) of master data objects. Complementing

these data standards, content extensions are realized for certain sectors or countries. In addition to attributes specific for distinct product groups (e.g. nutrition information for food products or warranty information for hardware), a common extension concerns local product classifications such as the German CCG classification.

Regarding message structure, XML is the commonly used syntactical representation form. Additionally, EDI EANCOM messages such as PRICAT still play a major role in connecting to customers on the retail side. In order to ease customer connections, the majority of data pool providers additionally supports less standardized message representations such as spreadsheets or structured text.

The transport level defines the way messages are transmitted between sender and receiver. The survey indicates that internet protocols dominate the transport level. Most of the interviewed pools use EDIINT AS2 for the automated transmission of EDI or XML messages and provide a portal frontend via HTTP. Some interviewees support additional internet protocols such as SMTP and FTP or message transmissions over value added networks using X.400.

With Global Data Synchronization, the reach is not limited to the direct customers of a pool, but can be extended by federation to customers of connected pools. Participation in the GDSN thereby enhances the value of the integration service for the customer base of the data pool. Our survey reveals significant differences in the ability to provide pool-to-pool connectivity via GDSN (cf. Figure 5): Whereas data pool A has established linkages to 13 other data pools, all other pools have not established more than 4 linkages. Two pools have not adopted GDSN yet.

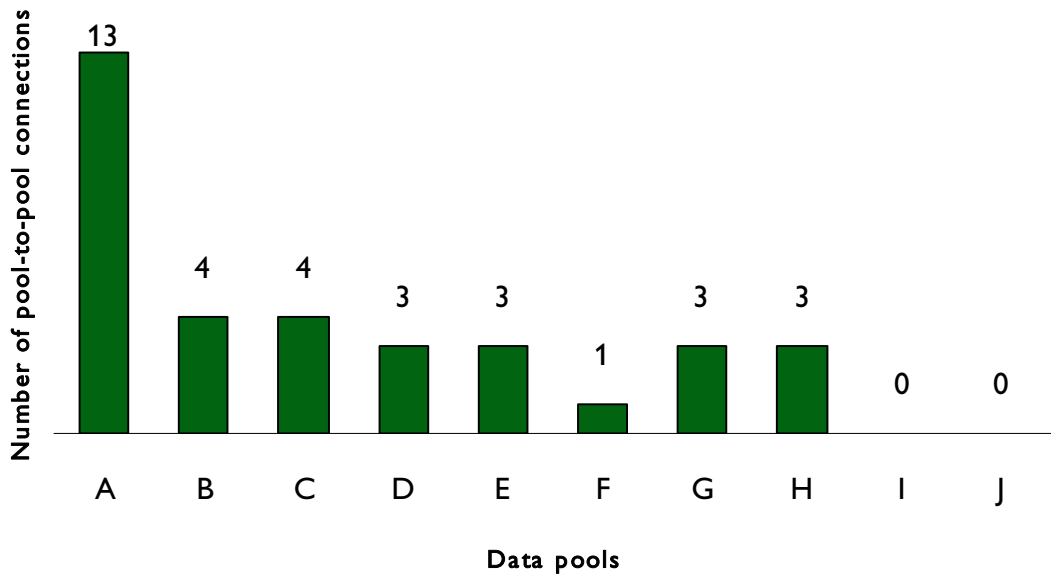


Figure 5: Active pool-to-pool connections per data pool

Supplementary services

Data quality services aim at enhancing the quality of information that is exchanged via the data pool. All data pool providers implement validation rules (cf. Table 3) that can detect syntactical errors and check simple content issues such as data types, domain ranges, or dimension and product hierarchy consistency.

In order to guarantee correct data, some data pools additionally offer measurement services: pool employees physically check product samples (e.g. for barcodes or dimensions) or conduct on-site inspections in order to test whether the product information is correct. A minority of data pools additionally checks data quality from the customer perspective: they offer a project-based analysis of

data quality issues at the retailers' systems and subsequently try to resolve the issues and root causes in collaboration with the data suppliers.

Alongside trainings for using the data pool interfaces and applications, educational services are offered by the majority of interviewees. They include training courses regarding the application of standards used by the data pools.

Onboarding services support customers which are realizing data pool connections. They consist of consulting and technical implementation services. The majority of pools provide these services themselves or in cooperation with partners.

| Service | | Data pool | | | | | | | | | |
|-------------------------------|-------------|-----------|---|---|---|---|---|---|---|---|---|
| | | A | B | C | D | E | F | G | H | I | J |
| Data quality management | Validation | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| | Measurement | ● | ○ | ○ | ● | ● | ● | ○ | ○ | ○ | ○ |
| | Projects | ● | ● | ● | ● | ○ | ○ | ○ | ○ | ○ | ○ |
| Education and training | | ● | ● | ○ | ○ | ● | ● | ● | ○ | ○ | ● |
| Consulting and implementation | | ● | ● | ● | ● | ● | ● | ● | ○ | ○ | ○ |

● Service offered, ○ Service not offered

Table 3: Supplementary Services

Revenue models

Most of the data pools interviewed charge their customers an annual fee, which depends on the customers' size in terms of annual revenue (cf. Table 4)

Interestingly, the value proposition of data pools remains still very narrowly focussed on product master data with only few of them expanding their business to more sophisticated product data related services. From the planned extensions of the informational services (to additional product information, e.g. pricing or product images, or business partner information) there is however some

| Factor | Data pool | | | | | | | | | |
|---|-----------|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H | I | J |
| Customer revenue | ● | ● | ● | ● | ● | ○ | ● | ● | ○ | ● |
| Number of transactions | ○ | ● | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Services used | ● | ○ | ● | ○ | ● | ○ | ○ | ○ | ○ | ○ |
| Number of data objects | ○ | ○ | ● | ○ | ○ | ● | ○ | ○ | ● | ● |
| Number of business partners | ○ | ○ | ○ | ○ | ○ | ○ | ● | ○ | ○ | ○ |
| Interface technology used | ○ | ○ | ● | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| ● Part of the revenue model, ○ No part of the revenue model | | | | | | | | | | |

Table 4: Factors in data pool revenue models

This annual fee is in some cases accompanied by a one-time setup fee. Two pools offer an item-based price model, i.e. they charge fees depending on the number of data objects managed. Typically the fee per item is based on a diminishing scale. Other data pool providers implement multi-dimensional price schemes. They combine for instance the revenue-based with the item-based model or include other factors such as the number of transactions, the number of business partners or the level of electronic integration (portal vs. direct connection) used.

evidence that data pools will form the core of a more comprehensive product data exchange infrastructure in the future.

Conclusions and future trends

The role of data pools in multilateral master data synchronization

With regard to the role of data pools, our survey supports the view of intermediaries as providing an integration or collaboration infrastructure within an industry (Christiaanse and Rodon 2005). This is underlined by the fact that most data pool providers focus on the dissemination of structured product data and are very closely aligned with the industry-wide standardization initiatives by GS1.

From a transaction theory perspective, data pools are expected to lower the coordination costs, but to also facilitate efficiency of transactions. In this context, it is interesting to notice that most companies have not (yet) established backend integration with data pools, but rather use human-machine frontends (e.g. portals) for entering and collecting product data. This rather limited level of electronic integration is error-prone and restricts the immediate propagation of changes in the entire value chain which is considered a major advantage of Global Data Synchronization.

The emerging GDSN landscape

From the survey results Figure 6 deduces a segmentation of the existing data pools into either "Mega Pools" or "Local Specialists" (the cluster diameter illustrates the number of industry sectors supported by the data pool).

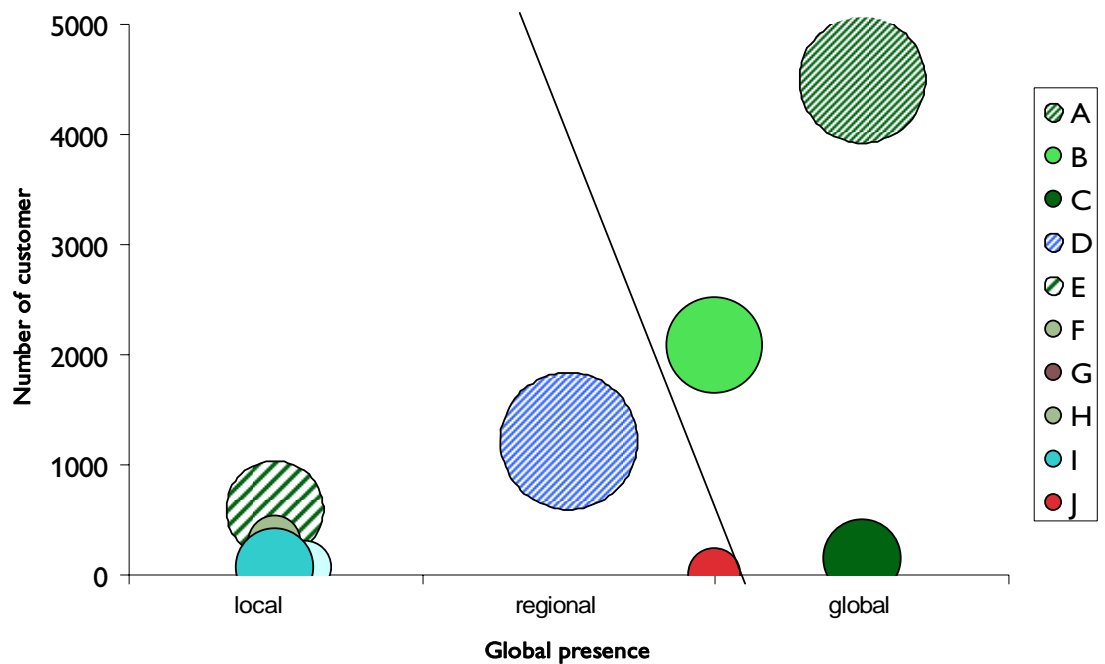


Figure 6: Data pool segmentation

Mega Pools are actively present in multiple continents and countries. They have diversified their services in multiple sectors and product groups and have – with exception of pool C – succeeded in acquiring a comparatively high number of customers. Local specialists service only a geographically bounded area. They focus on one country or a region of adjacent countries, subsequently reaching a lower number of customers. Additionally, they tend to focus on one sector or product group.

GDSN aims at solving the issue of “start-up-problems” which are typical for e-business scenarios in which positive network externalities prevail. From the analysis of the data pools, adoption of GDSN has started. Although in principle all data pools in the sample provide the ability to exchange data with each other, the survey shows a distinct pattern of the emerging active connections, which strongly supports the introduced data pool segmentation (cf. Figure 7).

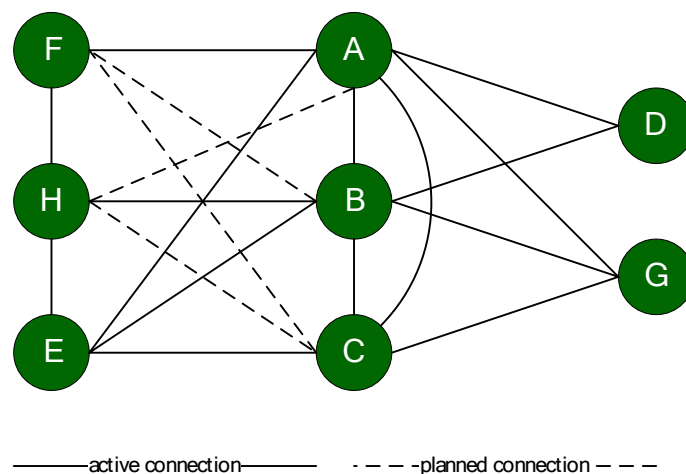


Figure 7: Active GDSN connections

The Mega Pools A, B, and C have distinctively more connections to other data pools than the rest of the sample. They constitute the core of the emerging GDSN network: almost all local data pools have implemented a master data exchange with them or are planning to do so in the near future. Apart from connections to the Mega Pools, interconnections between the Local Specialists are almost non-existent. Exceptions can be observed for regionally adjacent pools: F and H as well as H and E are based in countries with direct borders for instance.

Implications for future development

The indicated evolution of the GDSN imposes some implications on future development strategies for data pool providers as well as the preferences of users. Mega Pools face a tough market

consolidation and need to position themselves as global players in multiple countries. They will attract multinational retail and industry companies that need to distribute their master data across an international value chain. Mega Pools should further diversify into product sectors with global reach such as apparel & sports or electronics with the aim to realize first-mover advantages in these sectors and to obtain the critical mass of multinational customers. Local specialists on the other hand need to focus on locally operating companies such as SMEs or on sectors with national or regional particularities such as e.g. media or fresh groceries. In comparison to the Mega Pools, they are able to tailor services for their local markets and industries, e.g. in the form of specific data models or supplementary services. By focusing on specialized markets and industries, they will play a major role in onboarding smaller suppliers and retailers to the GDSN.

References

- Accenture (2006). Synchronization - The Next Generation of Business Partnering, Accenture.
- Bowling, T., E. Licul, et al. (2004). Global Data Synchronization: Building a flexible approach. Somers, IBM Institute for Business Value.
- Christiaanse, E. and J. Rodon (2005). "A Multilevel Analysis of eHub Adoption and Consequences." *Electronic Markets* 15(4): 355-364.
- Corsten, D. S. and J. S. Hofstetter (2001). "After the hype: The emerging landscape of B2B exchanges." *ECR Journal* 1(1): 51-59.
- Global Commerce Initiative and Capgemini (2005). *Global Data Synchronisation At Work in the Real World - Illustrating the Business Benefits*. Paris.
- Grocery Manufacturers of America, Food Marketing Institute, et al. (2003). *Action Plan to Accelerate Trading Partner Electronic Collaboration*.
- GS1 (2007). *List of GDSN Certified Data Pools*.
- McAfee, A. (2005). "Will Web Services Really Transform Collaboration?" *MIT Sloan Management Review* 46(2): 78-84.
- Nakatani, K., T.-T. Chuang, et al. (2006). "Data Synchronization Technology: Standards, Business Values and Implications." *Communications of AIS*(17): 2-60.
- Reyes, P. M. (2005). "Efficient consumer response: literature review." *Integrated Supply Management* 1(4): 346-386.
- Shapiro, C. and H. R. Varian (1998). *Information Rules - A Strategic Guide to the Network Economy*. Boston/Massachusetts, Harvard Business School Press.
- Sparks, L. and B. A. Wagner (2003). "Retail exchanges: a research agenda." *Supply Chain Management* 8(3): 201-108.

Institute of Information Management



University of St.Gallen



Institute of Information Management
Competence Center Corporate Data Quality
Müller-Friedberg-Strasse 8
CH-9000 St. Gallen

Phone +41 (0)71 224 38 00
Fax +41 (0)71 224 21 89

iwi2-cdq@unisg.ch
<http://cdq.iwi.unisg.ch>

