

Information as a Strategic Asset

White Paper - Version 1.2

July 2008

Helmuth Gümbel - Managing Partner

Strategy Partners International

Chasa Ursa Major

CH-7550 Scuol

helmuth.guembel@strategypartners.com

Text and Illustrations were compiled with utmost of care. However, the author will not take responsibility or be held liable for any possible remaining incorrect information and the resulting consequences thereof.

This piece of work is protected through copyright. All rights concerning the translation, reprinting, the use of illustrations, radio broadcasting, the reproduction through photographic or any other means, as well the storage in data processing facilities, even if only excerpts are utilized, are therefore reserved.

SAP, R/3, mySAP and NetWeaver are brands and trademarks of SAP. WebSphere is a brand and trademark of IBM Corp. .NET is a brand and trademark of Microsoft.

Table of Contents

Why you need to read this	1
The Core of the Problem	3
Trusted Data	5
Organizational Aspects.....	10
Customer Cases.....	11
Hornbach (German DIY market chain).....	11
City of Tacoma.....	11
Lenovo	11
Bottom Line	13
Table of Figures	14
Recommended Additional Reading	15

Why you need to read this

The original idea of suites like SAP was to control all data. It was expected that the suite architecture would guarantee quality and cohesiveness throughout. The past 10 years have taught us that this is fundamentally unattainable for a number of reasons:

- SAP has not been able to create a uniform data model that spans all SAP applications
- Very few (if any) customers are able to rely exclusively on SAP. Heterogeneous application portfolios are the norm.
- Many larger customers operate multiple SAP installations that frequently differ substantially
- Uniformity within a customer's SAP application portfolio is frequently negatively impacted by multiple SAP versions that have to be operated in parallel
- While SAP is attempting to converge data models, this convergence is slow and not well aligned with customer needs.
- SAP's goal is to ultimately consolidate all data and all processes into SAP provided solutions. As business relevant data is spread over many applications including mail and imaging systems, it is not likely that this goal is attainable lest SAP would become the one and only supplier of applications.

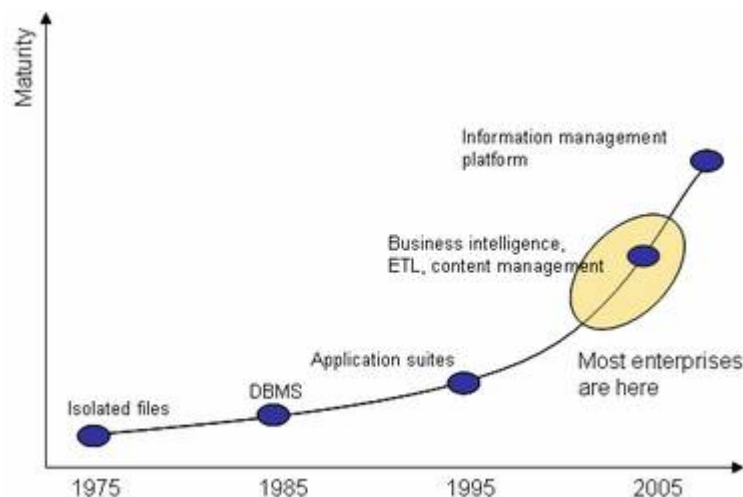
Even in the rare cases where enterprises have managed to unify their transactional systems under a standard SAP umbrella, the solution was rarely sustainable. Mergers and acquisitions frequently and unexpectedly made these houses of cards collapse.

Like other suite vendors, SAP was never able to fully displace other applications, including pre-existing legacy applications. Most applications are designed to operate with data sources that were specifically created for them. Flexible interchange of data was not part of the original design. In the days when the focus was on accounting applications, there was little need to integrate applications and data. Today, businesses have extended usage of IT to virtually all areas including connecting with suppliers, customers, and partners. Enterprises have started to invest significantly into business intelligence and content management to better understand their stakeholders. They also began to embrace unstructured information. Data that was once designed to exist in dedicated, stove-pipe-like environments is now exploited across the board. Data quality, originally an issue only within the bounds of a single application or a suite, is now of interest across the whole portfolio.

Enterprises have to rely on the information legally available to them regardless of their source. Universal access to data requires quality standards that make this access safe and productive.¹

We have serious doubts that enterprise-wide data quality can be achieved from within the applications portfolio given that such a portfolio is heterogeneous by choice or need. It embraces many applications that were never designed to be orchestrated around a single, consistent, and manageable data definition.

To make sure that you can rely on your data, a trusted data concept is required. Such a concept has never been pursued by SAP. Apart from forming the base for good data quality, trusted data allows for universal data access, latency reduction, information integration, data lineage, and the introduction of governance and data control features. This requires an application external information management platform that implements the functions required for an enterprise wide data strategy. It constitutes the logical next step in information management maturity (see).



• Figure 1 Next level in maturity: enterprise information management (Source: SPI)

Responsibility for the business processes of an enterprise cannot be delegated to a vendor. Without data you can trust, the outcome of any business process is and will be questionable.

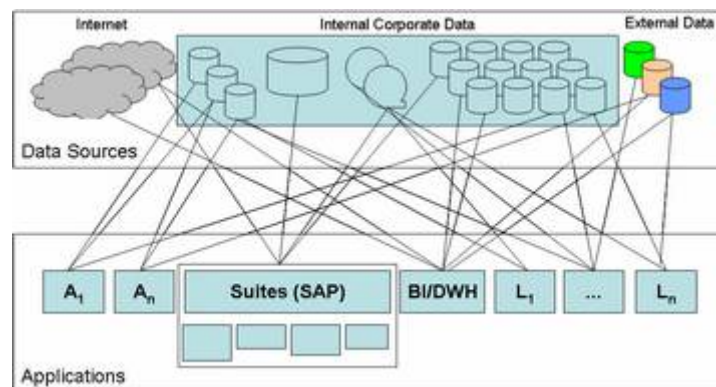
In this paper, we describe why cross-application data control is necessary and how to get to trusted data.

¹ In everyday life, the most direct analogy is in the food and beverage industry. Originally, there were no quality standards. Quality once was determined by trial and error, sometimes with fatal consequences. Today, we can safely consume e.g. bottled water in all civilized nations without any health risks. This was achieved by imposing quality standards across vendors in total independence of their production methods.

The Core of the Problem

When enterprises started to base their business processes on IT, they largely automated manual procedures to gain speed and to tackle volume. Data was designed for specific applications and quality control was their responsibility. Free recombination was not part of the design.

Application suites with their databases were meant to reduce the application portfolio down to a single set of applications from one vendor. All corporate data would be in one database. Installing such a suite would eliminate all previous application interface and data quality issues.



• Figure 2 Relation between data and applications, strongly simplified (Source: SPI)

Reality, however, is different (see Figure 2). The suites were never able to displace other applications nor were they successful in embracing corporate data in total. Today, enterprises use more applications and more diverse data sources than ever.

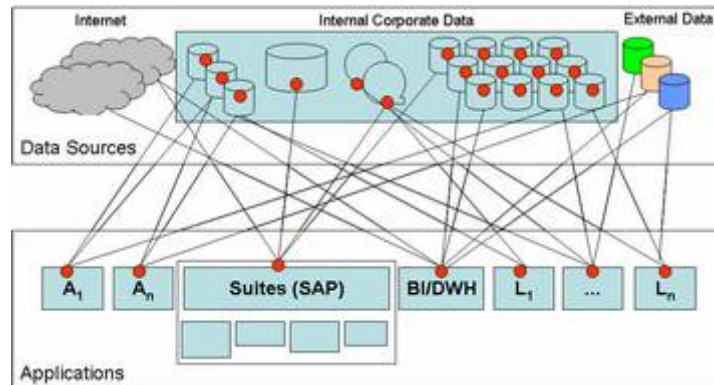
Achievements such as

- the Internet
- usage of voice, image, and video data
- integration of office- and e-mail applications
- usage of external data sources from vendors, customers, or commercial data providers
- data warehousing

have greatly increased both opportunities and risks.

Data quality problems can spread like a disease. To avoid this, multiple checks are implemented in each application. Even then, care must be applied when recombining data. Hence, data frequently is deliberately limited in its usage to a particular set of applications where it is believed to be usable.

When reaching beyond these limits, cumbersome and hard to maintain transformation procedures are used. In practice, comprehensive, reliable, and sustainable data quality control has remained largely an illusion. Figure 3, though greatly simplified, gives some idea of the issue. In reality, most enterprises have thousands of data sources and hundreds of applications. These are spread over many installations and organizational entities. As applications change through updates, the problem is further compounded.



• Figure 3 Red dots indicate data quality check points (Source: SPI)

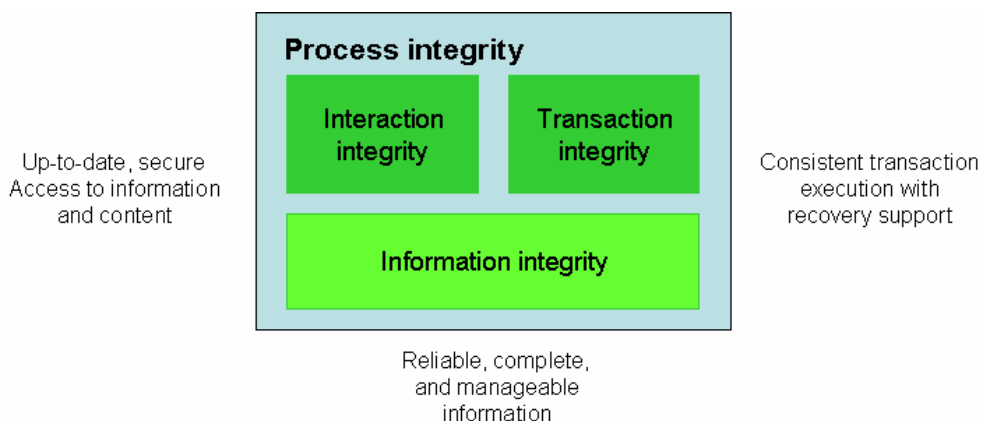
As a result, enterprises cannot maximize the usage of one of their key assets, their data. They are forced accept rather unpleasant disadvantages:

- reduced flexibility in data exploitation and integration
- high costs for establishing minimal data quality
- high interdependency when integrating data and applications
- high migration costs when upgrading to newer technology
- redundant and costly maintenance procedures
- isolated data exploitation resulting in information silos
- inability to track the source of data and the history of changes
- high costs and huge impediments when implementing data governance
- increased latency.

Trusted Data

This results in a striking imbalance: management is centrally responsible for its decisions and, hence, also for the tools and the input employed for management actions. With tactically oriented data quality fixes that have no common base such responsibility cannot be assumed.

Information is one of the key ingredients of business process integrity (see Figure 4). Without reliable, complete, and manageable data even correctly defined business processes will fail.

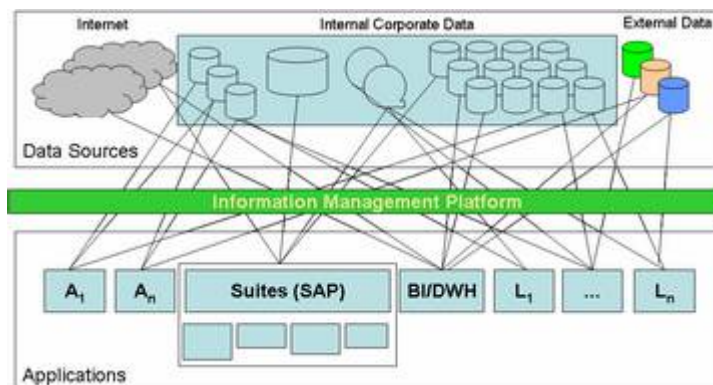


• Figure 4 Quality information is the foundation of business process integrity (Source: IBM)

To establish trusted data, the control has to happen outside of the application portfolio. Rules and procedures that are common to all affected data sources have to be defined and maintained. All this needs to be transparent, flexible, traceable, and auditable.

To achieve this, an information management layer is inserted between applications and data. This information management platform uses common metadata for all data sources under its control. It provides master data management and it is the base for implementing data governance.

Trusted data has to be sustainable – a one time check and transformation is not enough. Permanent supervision and data cleansing is required to ensure ongoing quality. This does not weigh lightly on processor power requirements. Hence, parallel processing capabilities are required as well as support for scaleable hardware.



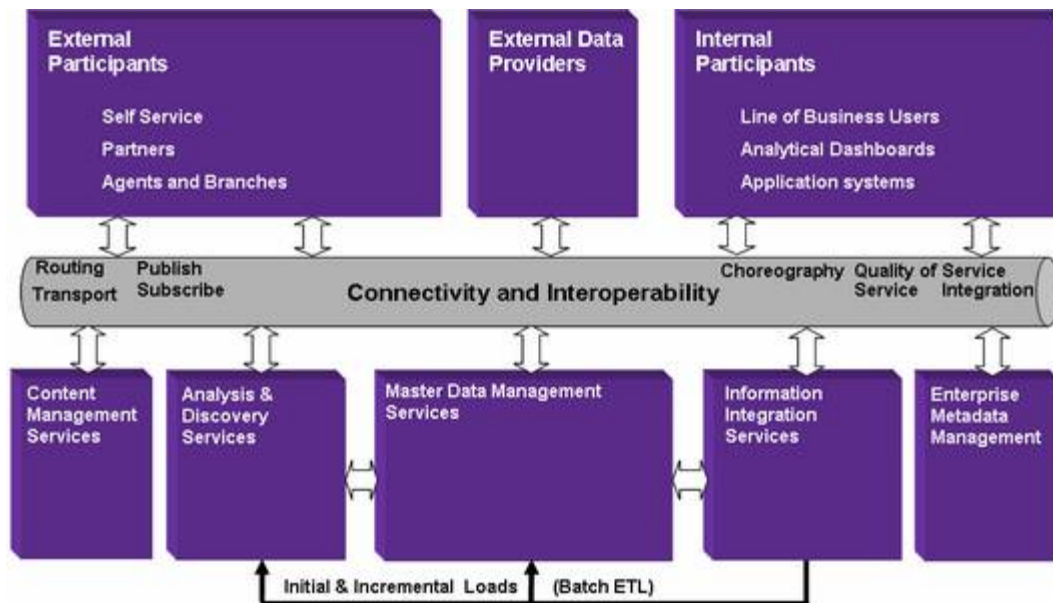
• Figure 5 Common information management layer (Source: SPI)

The information management layer tightly controls all information definition and access. Any request by applications for data is routed to this layer. The functions of the information management layer are:

- **Unified data access.** Data can be accessed in a standard way from any application that is authorized to do so, regardless of the access mode selected (real time, batch, replication for changes etc.)
- **Information integration.** Information sources can be combined regardless of location and technical implementation. Although independent of Enterprise Application Integration (EAI), it is an essential complement to application integration which can fail without information integration.
- **Common meta data.** The structure of all sources is commonly described. Any changes to these structures are immediately effective allowing applications to adapt automatically.
- **Data lineage.** The origin of data is tracked and any changes are recorded allowing detailed back tracking.
- **Implementation of governance and compliance related rules.** This allows for central control even in very federated and decentralized environments.
- **Master data management.** This provides joint management for data that is shared between multiple applications throughout the lifecycle of the data. It assures that the CRUD cycle² is handled consistently using common meta data. Ideally, such management is transaction based.

²Create, read update, delete

There is no standard product structure or architecture that is common to all vendors supplying products for these functions. A very good approach, however, is the usage of a reference architecture to illustrate the relationship between the entities involved.



• Figure 6 Logical System Architecture (Source: IBM/SPI)

The reference architecture (see) represents the functions as services extending the concept of SOA³ to information management (often referred to as “Information as a Service”). Thus, information can be integrated easily within the standards of the SOA environment.

In this reference architecture, the master data management services provide the following services⁴:

- **Interface services** for a consistent and standardized access regardless from where and in what technological guise the access request comes
- **Lifecycle management services** providing CRUD support and embracing data quality management services that are indirectly invoked to enforce data quality. Any quality issues will trigger alerts based on pre-defined rules to ensure prompt and coordinated action.
- **Hierarchy and relationship management services** can optionally detect interdependencies between data items.

³ Service-Oriented Architecture (SOA) is a software architecture where functionality is grouped around business processes and packaged as interoperable services (from Wikipedia).

⁴ See “An Introduction to the Master Data Reference Architecture” by Martin Oberhofer and Allen Dreibelbis, April 2008, IBM, pp10

- **Event management services** trigger operations based on events detected such as governance violations or changes to critical data to request authorization.
- **Authoring services** support collaborative work styles based on workflow providing services to author, approve, manage, customize, and define master data.
- **Data quality management services.** They enforce data quality and standardization rules for both data structures and values. If required, interaction with integrity services takes place.
- **Base services** provide security, privacy, search, audit, logging, and workflow support as common elements of the other services.
- **The master data repository services.** The master data repository holds the metadata for the MDM system. It also records any changes to the master data and the associated meta data.

Although not part of this reference architecture, it is essential to monitor the performance of the information management layer by defining key performance indicators and measuring their fulfillment.

The benefits of the introduction of such a layer are substantial and essential:

- Efficient and sustainable data governance across all data sources and all applications is now a real possibility. The central responsibility management has to assume in all governance and compliance matters is paralleled by IT.
- Globalization of enterprises requires a world-spanning view on business opportunities. To support this, data utilization must be elevated to a global level, too. The information management layer makes this possible without risking data quality issues.
- Data quality goes up and costs are greatly reduced since common metadata and centrally controlled cleansing processes replace overlapping, redundant, and often patchy isolated measures⁵.
- Liberal usage of all data becomes safe: no need for restricting data usage to information silos. The danger of data misinterpretation by applications is reduced and can be ultimately eliminated.
- Decision processes based on analytics become faster and safer – no need for extracts that are “magically” transformed in obscure and intransparent procedures. Thus, investments into BI will become much more valuable and justifiable.

⁵ More on the importance of data quality can be found in: “Data Quality – the Base for Your Enterprise Applications” by Helmuth Gumbel, Strategy Partners Internationalsl

- Technical management of data and applications becomes much easier and more efficient. Reductions in downtime and easier (less costly) application upgrade paths lower cost and provide for better alignment of IT with business.

Organizational Aspects

Although trusted data is impossible without modern tools and powerful hardware, it needs rigorous organizational support. The CIO must create awareness on all levels for the trusted data concept, create a data governance organization, define and install policies and processes for trusted data management.

Trusted data requires cooperation between line of business to define

- Business requirements for the data
- Data ownership and accountability
- Data lineage requirements
- Controls and usage
- Quality criteria and associated KPI
- Operational procedures for data collection and maintenance

These definitions cannot be enforced without establishing accountability in both IT-departments and the LOB organization.

Accountability has three levels:

- **Operational accountability.** Applying corporate standards, data entry is performed.
- **Tactical accountability.** The data steward assumes responsibility for the data that is produced in his domain (e. g. customer data in a sales territory)
- **Strategic accountability.** The data asset owner has overall responsibility on corporate level a master data category. He defines the quality level, the standards for achieving this quality, the KPIs to measure standards compliance, and the maintenance processes to assure ongoing reliability.

Trusted data is a holistic concept subject to constant monitoring and improvement. To avoid setbacks from new application and integration projects, trusted data requirements have to be an integral and enforceable part of their definition. CIOs have to set up review and sign-off processes to maintain conformance and integrity.

Customer Cases

Although data migration, data integration, and data quality have been issues with IT since its inception, some more recent cases may help to illustrate the need for a more profound and sustainable solution.

Hornbach (German DIY market chain)⁶

Hornbach wanted to roll out SAP in all 120 markets in a 33 million Euro project. The project came to a halt after the 48th market was done. Inconsistent data and missing reports were cited as reasons for revenue declines in the 48 markets. Meanwhile, the project has been restarted. With a concept for trusted information in place, data quality could have been achieved while implementing SAP. Moreover, the causes for the nagging inconsistencies could have been uncovered and sustainable improvements would have been possible.

City of Tacoma⁷

Here, over 100 legacy systems with their data were consolidated into one SAP system with very significant savings expected. These savings, however, came only significantly later and much lower than expected. Data migration tasks and data quality issues were severely underestimated. “This is not a resource for my job,” one employee wrote. “I fight getting accurate data from the system.” It is, however, not even certain that the costly data migration has solved the problems for sure: when users discover “impossible results”, confidence in the system erodes quickly. A trusted data concept enables root cause analysis for data inconsistencies identifying areas requiring improvements.

Lenovo⁸

Lenovo runs several versions of SAP’s suite. After the separation from IBM, order data was fed from IBM’s systems into Lenovo’s SAP systems. Faulty application integration prevented timely and correct data transfer impacting Lenovo’s capability to comply with delivery obligations in Australia and New Zealand. As customer pressure built up, Lenovo finally resorted to manually re-entering the data.

⁶ <http://www.computerwoche.de/index.cfm?pid=254&pk=1853318>

⁷ <http://www.thenewstribune.com/news/local/story/5900224p-5219989c.html>

⁸ <http://www.computerworld.com.au/index.php/id;715414843;fp;16;fpid;0>

To ensure ongoing data quality, a more central solution is still required. Monitoring data as it is created and processed for quality standard compliance and backtracking changes is required to deliver ongoing quality. A trusted data concept would ensure that ongoing controllable quality that can adapt to changing requirements is provided.

Bottom Line

In many ways, IT resembles process manufacturing. You need to control the input (ingredients), monitor processes that are applied, and check the quality of the output. Compliance requirements are introducing the need for backtracking and audits - very much like lot tracking and FDA requirements for food and drug production. Application such as ERP or CRM can be seen analogous to production machinery involved in manufacturing. This analogy makes it clear: data and processes have to be managed and controlled on a level that is separate from the applications to arrive at a holistic approach.

The analogy goes further. Manufacturing processes and the associated management are refined as needs arise. Modifications are the norm. To obtain trusted data and to keep it trustworthy requires ongoing and cohesive efforts. You need an information agenda to start with; a platform that will allow you to exercise application independent information management on all data sources, and the tools to support you enterprise wide when new requirements and insights mandate further actions. As the pace of business increases and regulatory requirements may change any moment, it is essential to be able to implement new measures without having to wait for application changes or tricky implementation adjustments that may cause hard to predict ripple effects.

Using an information management layer that is separate from data sources and applications does not only help with compliance and increases transparency, it also dramatically improves flexibility and significantly reduces costs. Information now can be looked upon as a freely accessible asset that can be managed uniformly. It becomes a service. This is a huge step forward.

Very few enterprises, if any, will be able to establish trusted data in a single, 360 degree step. In most cases, other events such as version upgrades, application migrations, SOA-projects, CRM introductions or business intelligence projects (to name a few) will be opportunities that can and should be leveraged to enter the rout towards trusted data. There is urgency, though: delaying efforts increases exposure and the burden for later corrective actions as all parameters (such as data volume, speed of change, business dynamics and regulatory requirements) develop into a more menacing direction potentially turning an asset into a liability. Since the responsibility for enterprise data is ultimately with corporate management, it is essential that you take action and define the rules.

Table of Figures

- **Figure 1 Next level in maturity: enterprise information management (Source: SPI).....2**
- **Figure 2 Relation between data and applications, strongly simplified (Source: SPI)3**
- **Figure 3 Red dots indicate data quality check points (Source: SPI)4**
- **Figure 4 Quality information is the foundation of business process integrity (Source: IBM).....5**
- **Figure 5 Common information management layer (Source: SPI)6**
- **Figure 6 Logical System Architecture (Source: IBM/SPI)7**

Recommended Additional Reading

Enterprise Master Data Management: An SOA Approach to Managing Core Information.

This book describes issues related to Master Data Management in much greater detail. It was written by Allen Dreibelbis, Eberhard Hechler, Ivan Milman, Martin Oberhofer, Paul van Run and Dan Wolfson and appears in June 2008 published by Pearson Publishing (ISBN-10: 0132366258, ISBN-13:9780132366250). The book covers many of the key aspects for understanding what is meant by Master Data Management, the business value of Master Data Management and how to architect an Enterprise Master Data Management Solution. It provides a comprehensive guide to architecting a Master Data Management Solution that includes a reference architecture, solution blueprints, architectural principles, patterns and properties of MDM Systems. The authors explain the relationship between MDM and Service Oriented Architectures and the importance of data governance for managing master data. The architectural model is vendor and software product agnostic focusing on the principles and methodologies to design the right architecture for an MDM Solution.

Enterprise Knowledge Management: The Data Quality Approach by David Loshin, ISBN13: 9780124558403 and ISBN10: 0124558402. The author, a data quality consultant, focuses on the business value of data quality and rules on how to improve and measure data quality. It also contains some great examples illustrating the business impact of data quality deficiencies. The style is both educating and entertaining.

Data Quality – the Base for your Enterprise Applications by Helmuth Gumbel, Strategy Partners International. This White Paper focuses on the importance of data quality, a key part of the trusted data concept. Available by ordering from orders@strategypartners.com free of charge.