GS1 Trusted Source of Data

Proof of Concept Report
June 2011

Document Summary

Document Item	Current Value
Document Title	GS1 B2C Trusted Source of Data Proof-of-Concept Report
Date Last Modified	29-Jun-2011
Current Document Issue	Issue 1
Status	FINAL
Document Description	This document provides the technical details of the GS1 B2C Trusted Source of Data Proof-of-Concept.

Contributors

Name	Organization
Dipan Anarkat	GS1 GO
Cameron Green	GS1 GO
Joe Horwood	GS1 GO
Malcolm Bowden	GS1 GO
Ken Traub	Ken Traub Consulting, LLC
Christian Floerkemeier	Mirasense AG
Ramana Lavu	VeriSign

Log of Changes in Issue 1

Issue No.	Date of Change	Changed By	Summary of Change
1	29-Jun-11	Dipan Anarkat	FINAL

Disclaimer

Whilst every effort has been made to ensure that the guidelines to use the GS1 standards contained in the document are correct, GS1 and any other party involved in the creation of the document HEREBY STATE that the document is provided without warranty, either expressed or implied, of accuracy or fitness for purpose, AND HEREBY DISCLAIM any liability, direct or indirect, for damages or loss relating to the use of the document. The document may be modified, subject to developments in technology, changes to the standards, or new legal requirements. Several products and company names mentioned herein may be trademarks and/or registered trademarks of their respective companies.

Table of Contents

1. Executive Summary

1.1. Objective

GS1 aims to become the Trusted Source of Data (TSD), to support the communication of authentic product data provided by brand owners to consumers/shoppers, retailers, internet application providers, and government using internet and mobile devices.

The Proof of Concept (PoC) sought to test a basic architecture proposal for the system that would support this goal.

1.2. Approach

The architecture was devised to leverage existing infrastructure and meet key additional criteria identified by the B2C Project Board. Proof of concept participants were then recruited for each component of the architecture. Finally, the system was tested using real data and applications.

1.3. Results

The Proof of Concept enabled a working system to be developed. The system functioned correctly with appropriate response times.

1.4. Learnings & Recommendations

The Proof of Concept led to invaluable learnings that should be taken into account during the pilot phase of this project. Key recommendations are as follows:

- Develop data quality framework for B2C application for brand-owners, data pools and data aggregators.
- Develop implementation guide for B2C contextualization of GDSN master data and nutritional attributes.
- Develop a standardised web service interface (including corresponding data formats) for access to data aggregators.
- Implement Federated ONS as the foundation for GS1 B2C services.
- Develop and deploy a standard GS1 solution to address GTIN/GCP parsing issue.
- Consider JavaScript Object Notation (JSON) as an alternative mark-up to XML to simplify mobile app development.
- Provide detailed guidelines about data formats and mandatory/optional data elements

1.5. Acknowledgements

GS1 would like to express its thanks to all those who participated in the pilot. In particular, all members of the B2C Project Board, Robert Besford, Brad Fiegel, Christian Floerkemeier, David Hogan, Sanjay Kirtikar, Ramana Lavu, Cristina Macias, Rainer Pietschmann, Pascal Revenus and Ken Traub.

1.6. Next Steps

A multi-country pilot will take place from July to December 2011. The learnings and recommendations from the PoC should be actively integrated with the pilot.

2. Objective

GS1 aims to become the Trusted Source of Data (TSD), to support the communication of authentic product data provided by brand owners to consumers/shoppers, retailers, internet application providers, and government using internet and mobile devices (phones, laptops, etc).

The Proof of Concept (PoC) sought to test a basic architecture for the system that would support this goal.

The system used for the PoC was designed with the following considerations in mind:

- leverage existing standards based on brand data used by the Global Data Synchronization Network (GDSN) and other accredited data sources,
- focus on basic and nutritional information desired by consumers and available by brands in the GDSN.

The PoC was deemed successful if a mobile application designed by an Internet Application Provider (IAP) was able to send a request for information about a specific product and receive and display this information as authorised by the brand owner.

3. Methodology

The Global Data Synchronisation Network (GDSN) has proved to be an effective tool for sharing product data in the B2B/supply chain space. GS1 therefore intends to leverage the existing Global GDSN infrastructure.

However, GDSN today it is not sufficient for sharing product data for integration in consumer-facing applications for two main reasons:

- **Reliance on existing trading relationships**. GDSN follows a model where trading partners allow the sharing of product data with other trading partners. Consumer-facing applications require data to be shared with internet application providers who do not have a trading relationship with each brand owner.
- **Need to aggregate data that exists within the GDSN network with data that does not**. Digital product information or information beyond B2B master data, for consumers may need to be aggregated from various accredited sources including the GDSN and other third parties.

With this in mind, a system was designed where:

1. relevant product information is aggregated in from GDSN Data Pools and certified third-party sources
2. a registry function allows each product's Global Trade Item Number (GTIN) to be matched to the location of the aggregated product information

Below is a basic diagram of the architecture used.

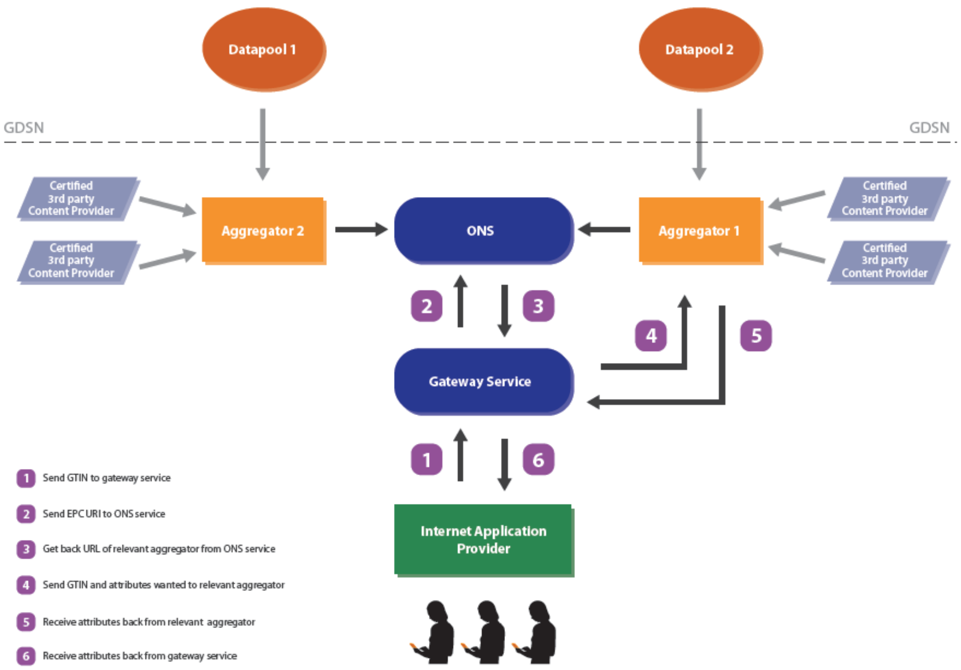

Figure 1: Proof of Concept Architecture and Data Flow

A simple way to describe the data flow is to imagine a consumer is scanning a barcode with the IAP application. The data flow then proceeds as follows:

- Steps 1 and 2: The product identifier (GTIN) is decoded from the barcode and passed to the Object Naming Service (ONS).
- Step 3: The ONS responds with the location of the information in the relevant data aggregator.
- Step 4 and 5: The GTIN is then passed with the request for information to the correct aggregator and is returned a packet of the information for the requested product.
- Step 6: This information is then rendered by the application for the consumer.

3.1. Pilot Components

To efficiently deliver the PoC, existing systems and implementers were selected to pass the information from source to consumer. Below is a list of the participants within each specific PoC role including the required actions to enable the PoC.

1. **Brand-owners**
   Coca-Cola, Hormel Foods, Nestle Vitality (via 1SYNC) and Nestle UK, Premier Foods, Unilever UK (via TrueSource)
   Action: Authorise basic product data and nutritional attributes (2 or more products by each brand) to be accessible to IAP.
2. **GDSN Datapools**
   SA2 World Sync & 1SYNC (Datapools powering Data Aggregators solution)
   Action: Send basic product data and nutritional attributes to aggregator.
3. **Certified 3rd Party Content Provider**
   Brandbank (since they were providing product images to TrueSource)
   Action: Authorise and send product image for specific products to aggregator.
4. **Data Aggregators**
   TrueSource & 1SYNC
   Action: Combine brand authorized data from GDSN Datapool with additional 3rd party data (i.e. product image). Load ONS with GTINs for products with information in aggregator.
5. **ONS**
   VeriSign Root ONS (until Federated ONS is developed and implemented)
   Action: Receive GTINs with location information (which aggregator), return location information to IAP requests.
6. **Gateway Service**
   North America B2C Alliance “Sandbox” since most of the work, excluding ONS, was already done
   Action: Enable interface/API for IAP to connect to TSD network. Exchange product information requests between IAP and ONS. Sends GTIN requests to ONS, receives location of product information (which aggregator), send request to aggregator, receive product data, return product data to IAP.
7. **IAP**
   Mirasense (solution provider with expertise in GS1 System integration, mobile barcodes and mobile app development)
   Action: Scan barcode on product to retrieve GTIN, send GTIN to Gateway Service, receive and render product data for consumer on mobile app

4. Results

4.1. Overall Architecture Approach

The goal at the outset was to allow a mobile phone user to scan a EAN/UPC barcode and receive trusted information for that GTIN as provided by a data provider. Each GTIN has a pointer to the data provider for that GTIN registered in ONS (a many-to-one relationship, as one data provider may serve data for many different GTINs).

One possible design is to have the phone itself perform the ONS lookup, and based on the pointer received from ONS contact the appropriate data provider directly. However, this is not a realistic design, for several reasons:

- The application programming environments typically available for handset software do not usually provide the ability to interface to ONS. Specifically, they generally do not provide a way to do a DNS lookup to obtain NAPTR records.
- Mobile phone applications are invariably designed to interact with a specific application back-end service provided by the handset application author. The back-end service then mediates any interaction with external data sources.
- Having an intermediate back-end service also provides for usage logging, additional services, and so forth, so it is a much more flexible application architecture.

For this reason, the PoC architecture had the handset application send the GTIN to a back-end application. The back end application performed the ONS lookup, and then queried the appropriate data provider, finally delivering the result to the handset.

In the PoC, we found that the GS1 US/Canada B2C Sandbox provided exactly the back-end application functionality that was needed. The B2C Sandbox was already designed to accept a query in which a GTIN is specified, and deliver information pertaining to that GTIN obtained by querying one or more external data sources. The B2C Sandbox includes a flexible data transformation mechanism, which allowed the PoC to adapt to slightly differing data formats made available by each data provider, and to transform these formats into a common format for delivery to the handset application. All that was necessary was to enhance the B2C Sandbox with the option to perform an ONS lookup to determine which data provider to use (the prior functionality was to query **all** data providers and aggregate the results).

4.2. Bar Code / ONS Integration

An ONS query is based on transforming an EPC identifier into a DNS domain name. In terms of GS1 Identification Keys, the ONS lookup requires that the portion of the key corresponding to the GS1 Company Prefix (GCP) be separated from the remainder of the key. A EAN/UPC bar code with a GTIN, however, does not indicate where to make the division between the GCP and the remainder of the key. This presents a challenge in doing an ONS lookup based on a GTIN barcode.

The solution adopted for the PoC was simply to take an iterative approach. The number of digits in the Global Company Prefix (GCP) is at least six and at most eleven. For any given GTIN, therefore, there are six different possible division points, corresponding to a GCP length of 6, 7, 8, etc. The ONS lookup software implemented for the B2C Sandbox simply tried each of these possibilities in turn, until the lookup succeeded. In the PoC, the actual GCP lengths were either 7 or 8, so in the PoC each GTIN lookup required two or three trials.

Despite the extra lookups, the overall latency of responding to a query was still acceptably low. End-to-end response times from the moment of scan until data was received from the TSD network and rendered on the mobile app provided a satisfying consumer experience without disengaging consumer interest.

4.3. Software Development Considerations for ONS

To do an ONS lookup, a software application must use the host operating system's DNS resolver to query for DNS NAPTR records. While all operating systems include a DNS resolver, looking up NAPTR records is a comparatively unusual operation, compared the usual DNS queries for A records (IP address lookup), MX records (mail server lookup), CNAME records (hostname alias lookup), and NS records (nameserver lookup).

General purpose programming environments typically provide a means for performing any kind of DNS lookup, but more specialized environments may not. As noted earlier, one such example is the programming environments available directly on mobile phone handsets. In the PoC, we encountered another example in trying to implement ONS lookup in the B2C Sandbox. The B2C Sandbox is implemented using the Google Application Engine for Java, which is a Platform-as-a-Service (PaaS) cloud-based platform for rapid application development. This environment provides a restricted set of networking primitives, which gives the platform greater flexibility to perform dynamic load management. However, it meant that an ONS lookup could not be directly performed within the B2C Sandbox's application code.

One possible solution was to port the B2C Sandbox to a less restrictive platform. For expediency, however, this approach was not taken. Instead, only the ONS lookup portion was implemented on a different platform, and this portion exposed a REST-style Web Service for use by the B2C Sandbox proper. Since the Google platform permitted access to Web Services (as virtually all development platforms do), this made it possible to perform ONS lookup with minimal software development effort.

The ONS lookup portion was implemented in Java running on the Amazon Elastic Computing Cloud (EC2), using the Elastic Beanstalk platform.

4.4. ONS Setup and Resolution

For the ONS-based resolution of GTINs to its corresponding web service URL to function, a typical ONS setup would consist of the Root ONS name server and multiple Local ONS name servers, one for each Company Prefix Owner.

The setup of the ONS would be as follows:

Root ONS	Local ONS
Contains “NS” (Name Server) DNS records	Contains “NAPTR” (Naming Authority PoinTeR) DNS records
One NS record for each GS1 Company Prefix	One NAPTR record for each *GTIN*
NS record has the internet address of the Local ONS name server	NAPTR record has the *Service-ID* and *URL*
Name Server maintained by GS1’s ONS service provider	Name Server maintained by Brand Owner’s ONS service provider
GS1 has authoritative control over the ONS root domain	Brand Owner has authoritative control over Local ONS domain.

The “typical” setup described above would require the brands participating in the pilot to setup an instance of a Local ONS within their enterprise network. To minimize impact to the production ONS system, and reduce Brand Owner involvement and investment, a test Root ONS service with “onstest.com” was setup by VeriSign. The Local ONS Name servers for all Brand owners involved in POC were setup and managed on the VeriSign Managed DNS service. Local ONS domains and NAPTR records are managed thru Managed DNS service UI. VeriSign Managed DNS service provides resolution services for ONS with 100% uptime and near realtime updates to domain data.

4.5. Mobile App

The PoC used a mobile app that featured the Scandit mobile barcode scanner to scan barcodes (www.scandit.com), retrieved data via the REST interface from the Gateway Service and rendered the XML formatted data in a native Android user interface. To access the Gateway Service the mobile application used an app key that was manually assigned a-priori. The data returned from the Gateway Service were formatted as key-value pairs in an XML document. The returned data was presented to the consumer in a format similar to the standard nutritional panel that appears on products.

Some sample screen shots of the mobile app are shown below.

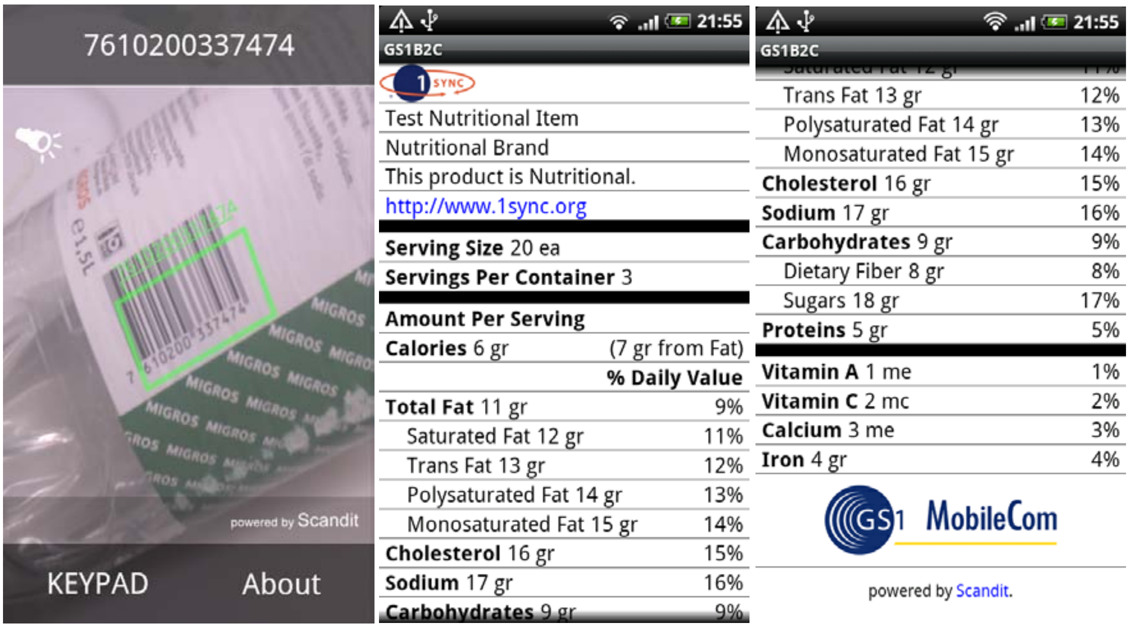

Figure 2: GS1 B2C Proof-of-concept Mobile App

4.6. System Interconnections

In order to test the PoC system architecture and validate the framework, desired results needed to be provided in real time by the live system components. Stakeholders participating in the PoC were interconnected using well defined web protocols and / or agreed upon web services interfaces. The table below provides an overview of the connections established in the PoC and the input and output results produced.

Connection Start	Connection End	Interface / Protocol	Data Format	Data Content
Mobile App	B2C Sandbox	Web service (HTTP REST)	HTTP GET parameters	GTIN
B2C Sandbox	Mobile App	Web service (HTTP REST)	XML	B2C Data in PoC Standard format
B2C Sandbox	Root ONS	DNS	DNS query	ONS Domain Name (from GTIN)
ONS	B2C Sandbox	DNS	DNS response	Data Aggregator Web Service URL
B2C Sandbox	1Sync	Web service (SOAP)	XML	GTIN
1 Sync	B2C Sandbox	Web service (SOAP)	XML	B2C Data in 1Sync format
B2C Sandbox	TrueSource	Web service (SOAP)	XML	GTIN
TrueSource	B2C Sandbox	Web service (SOAP)	XML	B2C Data in TrueSource format

Establishing the interface connections described in the table above required systems development and testing of all systems components. Additionally, B2C product data for the test products was authorized by the brands and made available to be used in the PoC via their GDSN Datapool. The custom mobile app developed for the PoC, was used to scan barcodes from the test products and retrieve data from the trusted source via the gateway service.

A successful end-to-end system test ensured that the connections worked for all the test products in the PoC. The following conclusions are observed:

- The trusted source corresponding to the GTIN scanned from the product was resolved properly via ONS.
- The trusted source provided authentic B2C product data coming from the GDS Network.
- Authentic B2C product data was provided to the mobile app to be displayed for the consumer.
- Response times calculated in real time, under normal operating conditions and from the time of the barcode scan until authentic product data is displayed in the mobile app, was under 3 seconds on average.

5. Learnings & Recommendations

The PoC was extremely successful and yielded a number of important learnings that need to be thoroughly considered for the future state GS1 Trusted Source of Data framework. These learnings have been categorized below from the perspective of the individual system components in the TSD framework.

5.1. Brand Owners

The PoC revealed the importance of data quality in GDSN. The following authentic B2C product data attributes from the brand owner's GDSN Datapool is provided to the consumer utilizing the TSD network.

- Basic Product Data: *GTIN, Product Name, Brand Owner Name, Product Description, Product URL, Product Image URL*
- Nutrition Data: *Attributes related to Vitamins, Calcium, Iron, Proteins, Calories / Energy , Carbohydrates, Sugars, Fat, Cholesterol, Sodium, Serving information*

The following data quality issues were discovered during the course of testing the system across the test products used in the PoC;

- Product Name and Product Description information was mostly not consumer friendly.
- Product URL which is used to provide a link to the product's website was mostly not available and when provided linked to the Brand's main website, rather than to a product-specific website.
- Product Image URL was mostly not available. When available, had varying image sizes and resolutions providing an inconsistent consumer experience when the images were displayed in the mobile app.
- Nutrition codes were not well defined in many cases leading to additional data harmonization and mapping across multiple product records. In some cases usage of nutrition codes was inconsistent across product records from the same brand. Some examples of different nutrition codes provided for the same nutritional attribute are:
  - Carbohydrates: CHO-, CHO, Carbohydrate, Carbohydrates
  - Sugars: SUGAR-, SUGAR, Saccharin, Sucralose, Sugars
  - Sodium: NA, NACL, Sodium
  - Fat: FATNLEA, FAT
  - Energy: Calories, ENER, ENER-
  - Vitamin A: VITA, VITA-
  - Vitamin C: VITC, VITC-

★**Recommendation**: Develop data quality framework for B2C applications for brand-owners, data pools and data aggregators.

5.2. GDSN Datapools

The following data mapping issues were discovered during the creation of the B2C Data canonical schema, while defining the mapping of B2C data attributes to their equivalent GDSN master data attributes

- Due to the presence of multiple candidate GDSN master data attributes, there was ambiguity on which GDSN attribute mapped directly to the Product Name, Product Description and Serving Size information B2C Data attributes. As a result different GDSN Datapools provided data from different GDSN master data attributes for the same B2C attribute.

- GDSN master data standards uses the UN INFOODS codelist as the standard for nutrition codes used for description of nutrition attributes. This is a very complicated and often misinterpreted codelist leading to ambiguities. The data quality issues raised above is likely a result of this complexity.

★**Recommendation**: Develop implementation guide for B2C contextualization of GDSN master data and nutritional attributes.

5.3. Data Aggregators

The Data Aggregators in the PoC provided customized and unique web service interfaces that allow for querying a GTIN and return of corresponding B2C Product data. The lack of a common web service interface and data format across the data aggregators implied multiple service integration efforts and data mappings. A standardized web service interface to the Data Aggregators would have provided a common service integration implementation and eliminated the need for data mapping to a canonical data format at the Gateway Service. This is not sustainable in a production environment. A standardized description of all web service interfaces and data formats is essential to reduce the implementation costs for service integration with multiple Data Aggregators that will participate in the TSD network.

★**Recommendation**: Develop a standardised web service interface (including corresponding data formats) for access to data aggregators.

5.4. ONS

The PoC utilized a technical setup of the ONS wherein the zone maintenance of all ONS records including the Local ONS was provided by the Root ONS operator. Ideally, the ONS system should be configured based on a nameserver delegation model wherein the Root ONS nameserver under GS1's (Global Office and Member Organization's) authoritative control provides resolution at the GTIN Company Prefix level and the Local ONS nameserver under the Brand Owner's authoritative control provides resolution at the GTIN Item Reference level. Although the current ONS setup works given the limited scope of the PoC, additional considerations need to be made for a commercial scale deployment of ONS in a future GS1 TSD system. The current setup would lead to number of DNS issues related to zone maintenance, domain control and such. If the current 'centralized' ONS model is used moving forward, it must fallback to a setup wherein the zone maintenance of the local ONS is under the control of the brand itself, either directly or provided by their respective local GS1 MO / DNS provider.

Additionally, based on concerns that are wholly outside the scope of the B2C pilot a Federated ONS approach is preferred. It provides for distributed control in a system designed for the wider Internet of Goods and Services, mitigating the issues concerning a single resolution root associated with the current 'centralized' ONS model

The GS1 Federated ONS initiative provides additional insight into the issues related to the current 'centralised' model of ONS and should be a consideration as a solution for future adoption.

★**Recommendation**: Implement Federated ONS as the foundation for GS1 B2C services.

5.5. Gateway Service

The Gateway Service performs the critical function of ONS lookups to resolve a GTIN to its corresponding data aggregator web service URL. This requires deriving the correct ONS domain name corresponding to the GTIN. However due to the now well-known "parsing issue" related to EPC-barcode interoperability, there is no direct way of deriving this reference from the GTIN. To get the correct ONS domain name reference for resolution, the Gateway Service has to adopt an effective but inelegant 'brute force' approach. This involves making multiple ONS queries with varying lengths of company prefix for the GTIN in consideration, until ONS returns a conformed successful or failed response. The need to make multiple ONS queries introduces additional latency in the TSD network and results in increased response times. This can be minimized if not eliminated, if a proper software solution was in place to address the 'parsing' issue.

★ **Recommendation**: Develop and deploy a standard GS1 solution to address GTIN/GCP parsing issue.

5.6. IAP

The PoC utilised a simplified XML data format consisting of key-value pairs that was provided by the Gateway Service to the mobile app. Future versions of the infrastructure should consider JavaScript Object Notation (JSON) as an alternative mark-up format, which is often used in mobile development and which simplifies data integration into mobile apps. The use of nested structures to group elements, e.g. Vitamin C total in grams and Vitamin C daily value in percent grouped in a Vitamin C data structure, might be an alternative to the "flat" key-value pairs. Guidelines about which data elements are mandatory and which are optional are also desirable

★ **Recommendation**: Consider JSON as an alternative mark-up to XML to simplify mobile app development.

★ **Recommendation**: Provide detailed guidelines about data formats and mandatory/optional data elements.

6. Next steps

The PoC can be deemed a success and the learnings from it are an opportunity to build the future foundations of a GS1 Trusted Source of Data framework.

A multi-country pilot will take place from July to December 2011. The learnings and recommendations from the PoC should be actively integrated with the pilot.

For more information about the PoC, contact Dipan Anarkat at dipan.anarkat@gs1.org.