Start with an identifier

Abstract

The GS1 system of identifiers is designed with barcodes in mind. The original barcodes that began to appear on products in the early 1970s could only encode digits, and even modern 1 and 2 dimensional barcodes are more efficient at encoding digits than any other characters. It's the drive for efficiency of encoding that leads to the payload of many barcodes to be mostly numeric and unintelligible to general purpose systems outside the world of Automatic Identification and Data Capture. GS1 Digital Link sacrifices some of that efficiency to encode identifiers in a URI that can be parsed without an online lookup, as well as operating like any other URL. Using a combination of ideas from Linked Data and HATEOAS, the standard further defines how GS1 identifiers can be resolved to multiple related resources. It is hoped that near-future work will standardize the approach such that it can be applied to any identification system.

Author/speaker

Phil Archer, Web Solutions Director, GS1 Global Office phil.archer@gs1.org

Introduction

On 31st March, 1971, 15 men – they were all men – met in New York to see if they could agree on way to eliminate the need to add a price sticker to every item in their stores. That led to the introduction of the barcode that first went beep at a checkout on 26 June 1974 [GTIN50]. The entire message from product to back end computer system was a string of digits.

Figure 1 The basic message when a 1D barcode is scanned – just the digits

The original barcodes from the early 70s could only encode digits. Even today, it is much more efficient to encode digits than any other character. Figure 2 shows two QR codes that have the same size modules and both encode 10 characters. However, it's clear that the one encoding purely digits is smaller than the one encoding Greek letters.

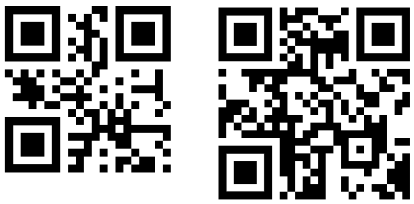

Figure 2 Two QR codes both encoding 10 characters. On the left, it's digits 0-9, on the right, Greek letters α to κ.

GS1 Element Strings

The number encoded in a familiar 1D barcode, known as a GTIN (Global Trade Item Number) is a class-level identifier that does its job well but there are significant efficiency gains to be made if the barcode can include things like batch/lot numbers, serial numbers, expiry dates, measured weights and more. GS1 is working with industry towards replacing the familiar 1D barcode with 2D barcodes that can encode much more data, such as that shown in the table below.

Table 1 Sample data to be encoded

Term	Value
GTIN	09506000134352
Batch/lot	ABCD
Serial number	1234
Expiry date	30 June 2023

Driven by the need for efficiency, terms like 'GTIN' and 'Batch/lot' are not encoded directly in the barcode. Rather, *Application Identifiers* are used [AI]. These are numeric so that a better representation of the data to be encoded is as shown in Table 2. The task then is to write all four pieces of data in a single message.

Table 2 Data to be encoded using Application Identifiers

Application Identifier	Value
01 (GTIN)	09506000134352
10 (Batch/lot)	ABCD
21 (Serial number)	1234
17 (Expiry date)	230630

Two of these data fields take fixed-length values and, by putting them first, there is no need to include a separator between them. Therefore the complete message becomes

01095060001343521723063010ABCD<gs>211234

The application identifiers have been highlighted in blue for legibility. The <gs> term is usually encoded as the non-printing ASCII character 29. It's only necessary because the last two elements – the batch/lot and serial number – are of variable length. It is the only character in this example that is not actual data. Known as GS1 Element String syntax, this is the kind of string that typically gets encoded in a Data Matrix as seen on the right [DM].

GS1 Digital Link syntax

GS1 Element String syntax is very familiar in the world of Automatic Identification and Data Capture but outside that it's unknown and likely to be dismissed as unintelligible. GS1 Digital Link [GDL] sacrifices some of the efficiency to offer an alternative but more familiar-looking syntax for the same data elements:

https://example.com/01/9506000134352/10/ABCD/21/1234?17=230630

The structure of the URI is defined precisely in the standard. Elements that identify the item – its GTIN, batch/lot and serial number – are encoded as path segments becoming more granular from left to right. This follows the typical pattern seen in Linked Data datasets. Other data elements, the example here is an expiry date but others might be a measured weight or a price, describe but do not identify the item and so can be in any order as query string parameters.

At the time of writing, major scanner manufacturers are testing their implementations for recognizing and parsing the various elements *without* making an online lookup – just as they do for the older GS1 Element String syntax.

Of course the major advantage of the GS1 Digital Link syntax is that it can be treated just like any other URL. A QR code such as that shown can thus be used *both* as an identifier for the item *and* as an entry point for digital information.

In terms of identification, the domain name in the URI is unimportant. It is the GTIN or other GS1 identifier that identifies the item, not the full URI. Two important facts flow from this:

1. It is the item that is identified, not the electronic information associated with it.
2. Applications can easily swap the domain name for another one and look up the same item.

Resolution

The GS1 Digital Link standard makes use of the ideas both in Linked Data and [HATEOAS] to define a GS1 conformant resolver that can link an identified item to multiple related information resources. Imagine a product for which several links are available.

Table 3 Sample links for a fictitious product

Link type	Lang	Target URL
gs1:pip*	en	https://dalgiardino.com/risotto-rice-with-mushrooms/index.html.en
gs1:pip†	es	https://dalgiardino.com/risotto-rice-with-mushrooms/index.html.es
gs1:sustainabilityInfo	en	https://dalgiardino.com/about/index.html.en
gs1:sustainabilityInfo	es	https://dalgiardino.com/about/index.html.es

The resolver will redirect to the default target URL, marked with a * character in Table 3, unless it has information that can be used to make a better choice. The link here is to the *product information page (gs1:pip).* GS1 defines a set of link relation types (abbreviated to link types) as part of its Web Vocabulary [Voc].

If the user's language is provided through the HTTP Accept-Language header, as smartphones typically do, and if the user's language is Spanish, they will be redirected to the Spanish version of the product information page. This is one of the 'default link multi' options in the resolver. No specialized app is required for these interactions.

The behaviour of the resolver can be controlled directly through the `linkType` parameter. Continuing the example, if an app appends the GS1 Digital link URI with the `linkType` parameter set to `gs1:sustainabilityInfo`, it will redirect to that target URL instead (again, taking into account the device's language). A request for a link type that is not available in the resolver results in it following the default.

Finally, if the `linkType` parameter is set to `all`, or the HTTP Accept header is set to `application/linkset+json`, the resolver will not redirect but will return the full set of links available for the item following the Linkset proposal at IETF [Linkset].

Resultant services and relevance to Digital Product Passports

Using the GS1 Digital Link standard, services can be built that utilize the ubiquitous identifiers seen in barcodes as the basis for a Linked Data Node. Links to related data can be found by rendering the linksets as JSON-LD and querying them using Linked Data techniques, or they can be accessed directly with simple RESTful queries. If the linked resources are themselves Linked Data, the resolver is a means of accessing the wider knowledge graph. By ordering the identifiers to be more granular from left to right, a query for the linkset at a granular level will inherit links from higher up the hierarchy.

For emphasis, it is the item that is identified by the GS1 Digital Link URI, not the information resources to which it is linked. This makes it particularly powerful in, for example, the provision of Digital Product Passports that can be reposed *anywhere* – because the starting point is the product and its identifier.

URI Design and Ownership

The IETF's Best Current Practice 190 [IETF190] makes it very clear that it is the server – the domain – that has full control over its URI space. The statement above that, "the domain name in the GS1 Digital Link URI is unimportant" is clearly at odds with this Best Practice document. GS1 Digital Link offers two ameliorations:

1. Resolvers are sovereign. There is no expectation that different resolvers will return the same results (this is in contrast to the Handle system used by DOIs where the expectation is that the same DOI will resolve to the same document, whichever resolver is used)
2. Conformant resolvers MUST include a Resolver Description File at a well-known location. If absent, the domain should not be considered a resolver.

Future standardization

The GS1 standard, unsurprisingly, is based on GS1 identifiers. However, the *principles* can be applied to any existing identification standard. A proposal is being submitted to ISO Steering Committee 31, the one responsible for AIDC, to create a generic standard that will codify:

1. Identifier types and their values should be encoded as path segments, becoming more granular from left to right.
2. The use of the `linkType` parameter to request redirection to a specific type of link if available, or to request the full Linkset.

If successful, services based on this generic standard will be more interoperable than at present. Dataspaces should be accessible across different systems as there will be a commonality of foundational API that starts with an identifier.

[AI]
The full set of GS1 Application Identifiers is listed at https://www.gs1.org/standards/barcodes/application-identifiers

[DM]
ISO/IEC 16022:2006 Information technology — Automatic identification and data capture techniques — Data Matrix bar code symbology specification

[GDL]
GS1 Digital Link standard https://www.gs1.org/standards/gs1-digital-link

[GTIN50]
GTIN 50th Anniversary Video https://youtu.be/IRhsloxAZUw

[HATEOAS]
Hypermedia as the Engine of Application State (HATEOAS) See https://en.wikipedia.org/wiki/HATEOAS for an overview

[IETF190]
URI Design and Ownership. Mark Nottingham, IETF June 2020. https://www.rfc-editor.org/rfc/rfc8820.html

[Linkset]
Linkset: Media Types and a Link Relation Type for Link Sets. Erik Wilde, Herbert van de Sompel. Expected to reach RFC status imminently https://datatracker.ietf.org/doc/draft-ietf-httpapi-linkset/

[VOC]
The GS1 Web Vocabulary https://gs1.org/voc/