Published June 1, 2003. Distribution restricted to Sponsors until September 1, 2003.

AUTO-ID CENTER

WHITE PAPER

The Electronic Product Code™ (EPC™) as a Meta Code

David L. Brock

AUTO-ID CENTER MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 400 TECHNOLOGY SQ, BUILDING NE46, 6TH FLOOR, CAMBRIDGE, MA 02139-4307, USA

ABSTRACT

The Electronic Product Code™ (EPC™) was conceived as a means to identify physical objects. These include not only retail products, but also containers, packages and shipments, as well as more general physical systems, assemblies and components. The EPC™ is a short, simple and extensible code designed primarily for efficient referencing to networked information. There are a variety of current and legacy identification coding systems that perform a similar function, such as the Uniform Product Code (U.P.C.) and its numbering superset, the Global Trade Item Number (GTIN) and the Vehicle Identification Number (VIN). Since the EPC™ is a general means of identifying objects and referencing networked information, we propose to use the EPC™ to reference other identification codes in other numbering systems. In other words, we propose to use the EPC™ as a sort of 'meta-code' to uniquely identify objects and correlate disparate coding systems.

Published June 1, 2003. Distribution restricted to Sponsors until September 1, 2003.

WHITE PAPER

The Electronic Product Code™ (EPC™) as a Meta Code

Biography

David L. Brock
Co-Director

Dr. David Brock received Bachelors degrees in theoretical mathematics and mechanical engineering from MIT, and his Masters and Ph.D. degrees from the Department of Mechanical Engineering at MIT with an affiliation to the Artificial Intelligence Lab. He is currently a Principal Research Scientist in the Laboratory for Manufacturing and Productivity and Co-Director of the MIT Auto-ID Center. Dr. Brock is also the Founder of Brock Rogers Surgical, a manufacturer of robotic medical devices. Dr. Brock has worked with a number of organizations including the Artificial Intelligence Laboratory, the Massachusetts Eye and Ear Infirmary, DARPA, Lockheed-Martin, Loral, BBN and Draper Laboratories.

WHITE PAPER

The Electronic Product Code™ (EPC™) as a Meta Code

Contents

1. Introduction ... 3
2. Electronic Product Code™ ... 3
3. Identification Codes ... 4
   3.1. International Standard Book Number (ISBN) ... 5
   3.2. The National Drug Code (NDC) ... 6
   3.3. Global Trade Item Number (GTIN) ... 7
   3.4. Vehicle Identification Number (VIN) ... 12
   3.5. Financial Security Identification Systems ... 13
   3.6. Social Security Number (SSN) ... 14
4. Connecting the EPC™ to Identification Codes ... 15
   4.1. Compressed Embedding ... 15
   4.2. Network Mapping ... 16
5. Conclusion ... 17
6. References ... 18

1. INTRODUCTION

The Electronic Product Code™ (EPC™) was conceived as a means to uniquely identify physical objects. This includes not only consumer products, but also physical assemblies, components and systems. The EPC™ contains very little information, and serves essentially as a reference to networked information. Together with the Object Name Service (ONS) and the Physical Markup Language (PML), the EPC™ connects the physical world and virtual world [1-4].

In addition to the newly proposed Electronic Product Code™, there is a vast array of coding systems in use today. From telephone numbers to license plates, coding systems serve to identify and classify objects. As with the EPC™, most coding systems also serve as a reference – explicitly or implicitly – to additional information. A license plate number, for example, can be used to 'look-up' information about the owner and vehicle.

There are, however, some important differences between the EPC™ and most coding systems. Most current and legacy numbering codes are application specific. In other words, these coding standards serve the needs of particular industries or products. The National Drug Code (NDC), for example, serves as universal identifier within the United States for all human drugs, and the International Standard Book Number (ISBN) as a book and media identifier worldwide. In order to serve the needs of their application domain, many coding system also include additional information about the identified object. Such information may include creation date, country of origin, packaging, pricing and expiration date.

The The Electronic Product Code™ was intended to be a somewhat broader object identification scheme – independent of application domain. As such, most information was removed from the EPC™ code and placed on the data network. Together with the Object Name Service and the Physical Markup Language, the EPC™ forms a standardized and tight coupling to networked information. Thus the requirement for embedded information or focused application is less critical with this proposed infrastructure, since additional data is immediately and easily available.

Extending this idea, it is clear that the EPC™ can reference not only data about an object, but also other codes used to identify that object. In other words, the EPC™ may serve as a **meta-code** linking together multiple, disparate coding systems. A single EPC™ code, for example, may reference a Vehicle Identification Number (VIN), as well as internal company numbers, industry coding systems and dealer number codes.

Since the EPC™ was designed at its inception to function as only one part of a tightly integrated information system, we can use this very infrastructure to tie together current and legacy coding systems and resolve potential conflicts with well established standards. Furthermore, such an approach smoothes the transition from highly function systems in use today to the revolutionary systems envisioned for tomorrow.

2. ELECTRONIC PRODUCT CODE™

The Electronic Product Code™ (EPC™) is a numbering scheme that provides unique identification for physical objects, assemblies and systems. Information is not stored in the code, but serves only as a reference to on-line – or Internet-based – information [3].

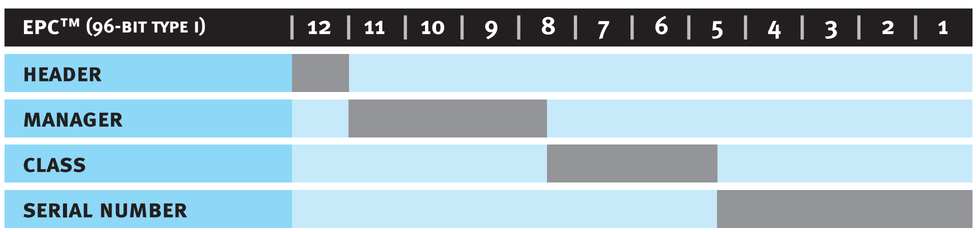
The EPC™ is embodied in small number of sizes, though all have similar structure and format. The 96-bit EPC™ Type I code is divided into four, fixed length partitions, as shown in Table 1. The first partition is an 8-bit **EPC™ Header**, which defines the number, type and length of subsequent data partitions. The purpose of the header is to provide extensibility for future, unanticipated data requirements.

The second partition identifies the **EPC™ Domain Manager**; that is the company, or entity, responsible for maintaining the subsequent codes. In other words, it is the EPC™ manager's responsibility to maintain both object type codes and serial numbers in their domain. The EPC™ manager ensures reliable operation of the Object Name Service and maintains and publishes the on-line product documents. The EPC™ manager partition spans a 28-bit section, encoding a maximum of 2^{28} = 268,435,456 companies.

The next partition, **EPC™ Object Class**, occupies the next 24-bits. When applied to retail products, the object class may be considered the skew or stock keeping unit (SKU). It may also be used for lot number or any other object-grouping scheme developed by the EPC™ manager. Since each organization is allowed more than 16 million object types, this partition could encode all the current UPC SKUs, as well as many other object classes. This is important as we expand beyond retail applications into general supply chain and manufacturing.

The final partition encodes a unique object identification number. For all objects of a similar type, the **EPC™ Serial Number** provides 36-bits, or 2^{36} = 68,719,476,736, unique identifiers. Together with the product code, this provides each company with 1.1×10^{18} unique item numbers — currently beyond the range of all manufactured products.

Table 1: The Electronic Product Code™ (EPC™) provides unique identification for all physical objects, assemblies and systems. The boundaries are shown here are in bytes.



3. IDENTIFICATION CODES

There is a wide variety of identification codes and coding schemes in use today. From the familiar bar-code to telephone numbers, each code system serves to identify and, in many cases, describe the associated entity.

Nearly every coding system follows a similar pattern – a series of numbers of fixed or variable length that encodes characteristics of the identified entity. The entity may be physical, as in the case of the Universal Product Code (U.P.C.), or virtual, in the case of patent numbers.

The numbers are represented in a wide variety of formats – from standard base-ten human language symbols to printed spatial patterns to electronic digital storage. In all cases, the numbers are the same; it is only their method of storage that varies. Furthermore, it is relatively easy to translate from one representation to another.

Although there are a great variety of codes, nearly every one contains a minimum of amount of information. Most contain a country or regional designation (at times represented implicitly in the length and structure of the code), a responsible party (i.e. a manufacturer or originator of the identified entity) and the entity identifier. Besides this basic information, most coding systems embed additional data, such as serial number, packaging, transaction data and other information pertinent to a target application.

This last point is critical. Most coding systems are designed for **specific** applications, and therefore are configured to optimize those applications. The brief review of some existing coding systems given in the following sections clearly shows the design based on the needs of the application.

Some information stored in a code is **implicit**. The length and structure of a coding system actually conveys information. The original U.P.C. code, for example, stored 12 digits, while the international variant stored 13. Information may also be stored in the number and location of partitions in the code sequence. The length, quantity and size of the number segments within the code may, for example, encode the country of origin.

In the following sections, we will briefly review some existing coding systems – particularly those that relate to object labeling and identification.

3.1. International Standard Book Number (ISBN)

The need for an international numbering system for books was introduced at the Third International Conference on Book Market Research and Rationalization in the Book Trade November 1966. The International Standard Book Number (ISBN) system that met this requirement was developed one year later in the United Kingdom in 1967 by J. Whitaker & Sons, Ltd., and in the United States in 1968 by R. R. Bowker [5].

The International Organization for Standardization (ISO) Technical Committee 46 on Information and Documentation set up a working group to investigate adaptation of the British system for international use. As a result of this effort the International Standard Book Number (ISBN) was approved as the ISO Standard 2108 in 1970. This was followed by a second edition in 1978 and a third in 1992.

The ISBN coordinates and standardizes international use of the standard and uniquely identifies the title or edition of a title. The original standard has been revised as book-like items began to appear in new forms of media. The basic structure of the ISBN, as defined in the first version, has not changed and is in use today in more than 150 countries.

ISBN Number Structure
The ISBN number consists of ten digits preceded by the letters ISBN. The ten-digit number is divided into four parts of variable length, which is separated by hyphens or spaces.

```
ISBN 0-852-00918-7
```

The ISBN number contains four partitions, as shown in Table 2. The number of digits in the first three partitions varies, but their sum totals 9 digits.

The first partition identifies the country, area or language area participating in the ISBN system. Some members form a language areas or regional units. For example the group number is '3' for German language group and '982' for the South Pacific region. The group identification partition varies from 1 to 5 digits.

The second partition identifies a particular publisher within the language group. The publisher identifier usually indicates a particular publishing house and their address. The ISBN publisher identification number may vary from 2 to 7 digits.

The third partition identifies a specific edition of a publication from the specified publisher. The title identifier may consist of up to six digits.

The final partition is the check digit used to validate both machine scans and manual data entry.

Table 2: International Standard Book Number (ISBN) is a compact 10-digit identification number that marks books and book-like material. The ISBN number is divided into four partitions that identify language group, publisher and title. The remaining partition is a check sum digit to validate machine or manual key entry. Here the grey shading denotes variable partition size.

CODE	10	9	8	7	6	5	4	3	2	1
GROUP IDENTIFIER										
PUBLISHER IDENTIFIER										
PACKAGE CODE										
CHECK CHARACTER										

3.2. The National Drug Code (NDC)

The National Drug Code (NDC) System was established as part of an out-of-hospital drug reimbursement program under Medicare [6]. The NDC serves as a universal product identifier for human drugs. The current edition of the National Drug Code Directory is limited to prescription drugs and a few selected OTC products.

The Drug Listing Act of 1972, amending the Federal Food, Drug, and Cosmetic Act, became effective February 1, 1973. Its purpose is to provide the Commissioner of the Food and Drug Administration (FDA) a current list of all drugs manufactured, prepared, propagated, compounded, or processed by a drug establishment registered under the Federal Food, Drug, and Cosmetic Act. The act requires submission of information on commercially marketed drugs and is used in the enforcement of the Federal Food, Drug, and Cosmetic Act.

Each drug product listed under Section 510 of the Federal Food, Drug, and Cosmetic Act is assigned a unique 10-digit, 3-segment number, as shown in Table 3. This number, known as the National Drug Code (NDC), identifies the labeler/vendor, product, and package size.

The first partition, the labeler code, is assigned by the FDA. A labeler is any firm that manufactures, repacks or distributes a drug product.

The second partition, the product code, identifies a specific strength, dosage form, and formulation for a particular firm.

The third partition, the package code, identifies package sizes. Both the product and package codes are assigned by the firm.

The NDC will be in one of the following configurations: 4–4–2, 5–3–2, or 5–4–1.

For consistency, other Government agencies may display the NDC in an eleven digit format. For example, the Centers for Medicare and Medicaid Services (CMMS) displays the labeler code as 5 digits with leading zeros; the product code as 4 digits with leading zeros; the package size as 2 characters with leading zeros.

Table 3: The National Drug Code (NDC) is a 10-digit, 3-segment number used in the United States to identify human drugs. Three segments of the NDC identify labeler, product and packaging in 4-4-2, 5-3-2 or 5-4-1 configurations.

CODE	10	9	8	7	6	5	4	3	2	1
LABELER										
PRODUCT CODE										
PACKAGE CODE										

3.3. Global Trade Item Number (GTIN)

The Global Trade Item Number (GTIN) is a family of item identification codes, which has for years identified products and their aggregations [9]. The GTIN is not really a new coding system, but a harmonized collection of existing codes within a single consolidating format, as shown in Table 4.

Table 4: The Uniform Code Council (UCC) and the International Article Numbering Association (EAN) Global Trade Item Number (GTIN) is a family of coding structures used to identify trade items across supply chains [7,8, 9].

TERMS	LEGACY ITEMS	DOMAIN
EAN/UCC-14	SSC-14 (Shipping Container Code)	Worldwide
EAN/UCC-13	EAN Code	Outside North America
UCC-12	U.P.C. (Universal Product Code)	Worldwide
EAN/UCC-8	EAN-8	Outside North America

The GTIN is a coding system for identifying trade items – both products and services – throughout the retail and commercial supply chains. The GTIN family of coding structures grew from retail checkout in the United States to generic trade items worldwide.

The GTIN began as a means for identifying grocery items and facilitating automatic checkout at the supermarket using bar-code laser scanners. The Universal Product Code (U.P.C.) has for years sped the tracking and identification of merchandise in stores, as well as in industrial and commercial applications across multiple supply chains.

Since the first scan in 1973, the grocery industry established the U.P.C. as the standard bar code system for product identification. The original U.P.C., however, grew to include three additional coding structures that expanded its capabilities – the EAN/UCC-13 for use outside the United States, the EAN/UCC-14 which used an indicator digit to identify similar trade units in different packaging configurations and the EAN/UCC-8 a short version developed for a specific bar code for small trade items.

By increasing the number of digits to 14, the various coding standards coexist under a common superset – the Global Trade Item Number, as shown in Table 5.

Table 5: The Global Trade Item Number (GTIN) system embeds four coding structures – EAN/UCC-14 (formally SCC-14), EAN/UCC-13 (EAN-13), UCC-12 (U.P.C.) and EAN/UCC-8.

CODE	14	13	12	11	10	9	8	7	6	5	4	3	2	1
EAN/UCC-14														c
EAN/UCC-13	0													c
UCC-12	0	0												c
EAN/UCC-8	0	0	0	0	0									c

The GTIN is organized and administered by the Uniform Code Council (UCC) and EAN International. The UCC and EAN International represent an alliance of organizations and numerous member companies worldwide. In the following sections, we will outline the code formats of the GTIN family in more detail.

3.3.1. EAN/UCC-14

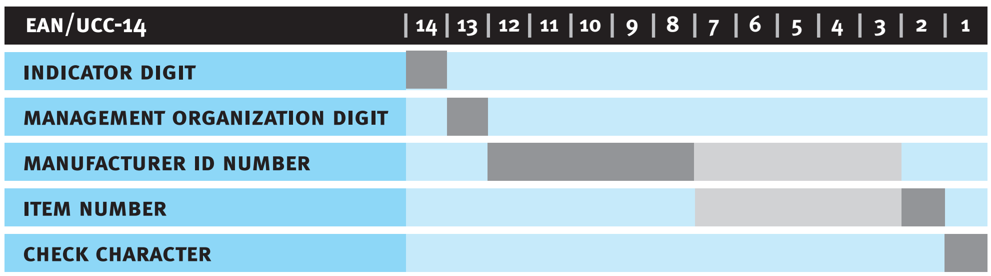
The EAN/UCC-14, previously known as the Shipping Container Code (SCC) or dispatch number outside North America, is a 14-digit code, which uses an indicator digit to increase the numbering capacity when identifying similar trade units in different packaging configurations. The bar code symbol used to encode the EAN/UCC-14 is the Interleaved 2-of-5 (ITF-14). The UCC/EAN-128 and the RSS symbologies also used for this purpose.

The EAN/UCC-14 contains four partitions, as shown in Table 6. The first partition is the Indicator Digit, which represents the level of packaging. Level '0'for unit level packaging (i.e. the lowest level of packaging). Numbers '1' through '8' indicate different packing configurations. A '9' indicates that the amount of product varies from container to container.

The second partition is a single digit representing the management organization. A '0' indicates that the structure is management by the UCC and a non-zero number means management by EAN. Essentially the remaining partitions match the UCC-12 and EAN/UCC-13 format.

The third partition is the EAN.UCC Company Prefix and the fourth the item number. The last digit, as with all EAN.UCC numbers, is the modulo checksum character.

Table 6: The EAN/UCC-14 uses an indicator digit to increase the numbering capacity when identifying similar trade units in different packaging configurations.



3.3.2. EAN/UCC-13

Growth of the North American U.P.C. code lead to the development of a European standard in 1976. This was quickly developed to include all countries outside North America. Aside from the extra digits used as part of the EAN.UCC Prefix, the EAN Code is nearly identical to the original U.P.C. standard. The Japanese Article Numbering (JAN) System for example uses the EAN.UCC Prefix 49.

The EAN/UCC-13 is a 13-digit code, which contains four partitions, as shown in Table 7. This first partition is the Member Organization (MO) Prefix containing 2 or 3 digits. The EAN.UCC Prefix represents the country code of the EAN Member Organization or the UCC assigning the EAN.UCC Company Prefix. For compatibility with the original U.P.C., the prefixes 00, 01, 03, 04 and 06 through 13 are assigned to the United States and Canada.

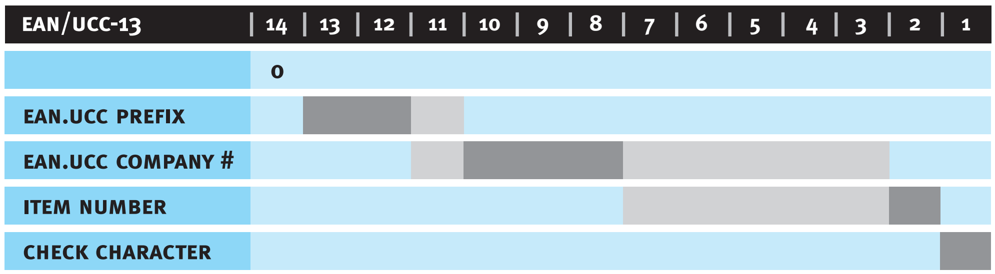
The second partition is the company number and is usually composed of 4 or 5 characters, but may have as many as 8 characters. For most EAN/UCC-13 numbers, the EAN.UCC Prefix together with the unique manufacturer identification number total 7 digits and is designated the EAN.UCC Company Prefix. Thus if the country code has 3 digits, for example Taiwan '471,' the company identification number has 4 digits.

The third partition is a 5-digit number representing the item identification – identical to the U.P.C. standard. This allows each manufacturer 100,000 unique product types. Given a variable length company prefix, the available product types are restricted to a number which when combined with the EAN/UCC Company Prefix totals 12 digits.

With 7 digits, the EAN/UCC-13 could, theoretically represent as many as 10,000,000 unique companies worldwide. However, the partitioning of Company Number into country and manufacturer reduces this theoretic maximum. Variable partitioning of the EAN.UCC Company Prefix yields many times this number. For example a single 10 digit prefix provides unique identification for 10 billion companies each having 100 products.

Finally, the last partition is the single digit encoding the modulo check sum for reader validation and key entry.

Table 7: The EAN/UCC-13 is a 13-digit code used to identify product in sold in countries outside the United States and Canada.



3.3.3. UCC-12

The EAN/UCC_12 or UCC-12 is essentially the original Universal Product Code (UPC), which is a 12-digit number that encodes the UCC Company Prefix (UCC Prefix and Company Number), item reference number, and check digit. The standard symbol, which represents the UCC-12, is a series of light and dark lines together with a human readable numeric equivalent [10].

The most common variant, U.P.C. Version A, has 12 digits; however, there is one other version, a zero-suppressed version, U.P.C. Version E, which is an 8-digit code used on items with packaging too small to accommodate the full bar code. Obviously to convert from a U.P.C. Version A to U.P.C Version E, there must be at least four zeros in the data set. The Version E, however, is always stores in its full expanded format.

The UCC-12 consists of (essentially) four partitions, as shown in Table 8. The partition is a single digit representing the UCC Prefix used to interpret the remaining digits. This digit is usually presented as a part of the UCC Company Prefix. A UCC Prefix of ‘0,’ for example, designates a regular U.P.C. Code, ‘2’ random weight item, a ‘3’ a National Drug Code (NDC) embedded into the UCC-12, and a ‘5’ a coupon. A complete list is given in Table 9.

The UCC Prefixes 1, 8 and 9 use variable length EAN.UCC Company Prefixes and Item reference partitions. This results in item numbers from 1 to 100,000 depending on the length of the EAN.UCC Company Prefix.

Table 8: The Universal Product Code UCC-12 (U.P.C.) is a 12-digit number representing a type (the UCC Prefix portion of the UCC Company Prefix), manufacturer (the UCC Company Prefix), Product and Check Digit.

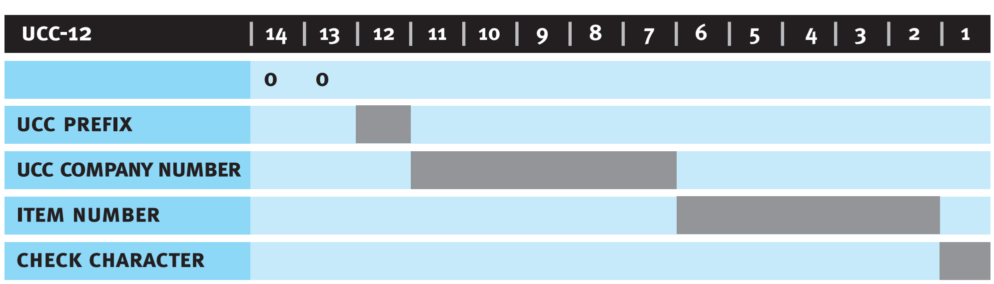


Table 9: The UCC-12 provides a single digit – the UCC Prefix – to represent the numbering system.

NUMBER	USAGE
0	Regular UPC Codes
1	Variable partition codes
2	Random weight items
3	National Drug Code
4	Internal company use
5	Coupon
6	Regular UPC Codes
7	Regular UPC Codes
8	Variable partition codes
9	Variable partition codes

The next five digits in the UCC Company Prefix designate the manufacturer, which correspond to digits 11 through 7 in the UCC-12. The following 5 digits encode the item number. The Item numbering is maintained by the manufacturer that must ensure unique numbers for each product type.

Finally, a single digit is added as a check character used to validate correct interpretation of machine scan or manual key entry.

3.3.3.1. RANDOM WEIGHT ITEMS

Not all UCC-12 Codes follow the regular structure given in Table 5. North American Retail Products sold after being weighted or measured have a different structure, as shown in Table 10.

Designated with a '2' in the first digit, the number is divided into five additional sections. The second partition is a 1-digit code called the Packager Code. Unlike fixed content products, the random weight identification number allows only one position to identify the supplier of the product. Traditionally digits '0,' ..., '3' indicate the retailer and digits '4,' ..., '9' their suppliers. The supplier and retailer must agree to the Packager Code assignment.

The next partition contains 4-digits indicating the Commodity Identification Number. The Commodity Identification Number allows 10,000 product types, which are divided among the trade associates for beef, veal, pork, chicken and others. The trade associate for beef, for example uses '1314' to identify ground round.

The fourth partition is a Price Check Digit, follow by a 4-digit price in dollars and cents. The final digit, as with all the other EAN.UCC Cods is the single modulo checksum digit.

Table 10: Random weight items use the UCC-12 number to encode packer, commodity and price.

UCC-12 RANDOM WEIGHT	14	13	12	11	10	9	8	7	6	5	4	3	2	1
	0	0												
RANDOM WEIGHT ITEM			2											
PACKER				■										
COMMODITY TYPE					■	■	■	■						
PRICE CHECK CHARACTER									■					
PRICE										■	■	■	■	
CHECK CHARACTER														■

3.3.3.2. NATIONAL DRUG CODE

The UCC-12, or more generally the GTIN, can also encode the National Drug Code (NDC). The UCC Prefix '3' is used to identify an NDC number or a Medical/Surgical Product Number using the National Health Related Item Code (NHRIC). The Food and Drug Administration (FDA) assigns the data portion of the NDC code The NHRIC Codes, however, are no longer being issued, as they now use either UCC-12 directly of the EAN/UCC-14.

As described in the previous section, the National Drug Code is composed of three partitions describing labeler, product and package. When displayed in an 11-digit format, the NDC partitions are configured as 5 digits for labeler, 4 for product and 2 for package type, as shown in Table 11.

Table 11: The National Drug Code (NDC) can easily be embedded into the UCC-12, which is part of the Global Trade Item Number (GTIN).

UCC-12 NDC	14	13	12	11	10	9	8	7	6	5	4	3	2	1
	0	0	0											
LABELER														
PRODUCT CODE														
PACKAGE CODE														
CHECK CHARACTER														

3.3.3.3. COUPONS

The UCC-12 Code allows the identification and description of coupons, as shown in Table 12. A UCC Prefix of '5' indicates a manufacturers' coupon for redemption. The identification number is formed by supplementing the U.P.C. Couple Code with one of five types of Coupon Extended Code Formats. Each of these formats varies in length and structure.

The U.P.C. Coupon Code, without the Coupon Extended Code Formats, is shown in Table 8. The main portion of the U.P.C. Coupon Code has the five partitions. The first partition is the single digit '5,' which indicates the UCC-12 is coupon.

The second is a 5 digit UCC Company Code. The third partition consists of 3 digits which indicate the Family Code, Summary Codes and Super Summary Codes, respectively. The Family Codes relate to individual products, while the others are assigned to larger groups of products. The Uniform Code Council details the Family Codes, as describe in the **U.P.C. Coupon Code Guidelines Manual.**

Table 12: Coupons can be identified quickly using standard bar-codes and the UCC-12 numbering system.

UCC-12 COUPON	14	13	12	11	10	9	8	7	6	5	4	3	2	1
	0	0												
COUPON			5											
COMPANY CODE														
FAMILY CODE														
VALUE CODE														
CHECK CHARACTER														

3.3.4. EAN/UCC-8

The EAN/UCC-8 is an abbreviated version of the EAN/UCC-13 and is used primarily outside the United States and Canada. The 8-digit code includes a 2 or 3 digit EAN.UCC Prefix, which correlates to a country code, a 4 or 5 digits of data and a single checksum, as shown in Table 13. The combined data digits specify a particular product within a particular country. A 3-digit country code, such as Brazil '789,' would for example allow 5 data digits or 10,000 individual product identification numbers.

Unlike the U.P.C. Version E symbol, which is a zero-suppressed variant of U.P.C. Version A, the EAN/UCC-8 directly identifies a combined manufacturer and product.

The EAN/UCC-8 is easily embedded within the GTIN. When right justified and zero filled to the left within a 14-digit GTIN format, the EAN/UCC-8 is unique.

The EAN/UCC-8 may include a 2-digit or 5-digit extension, though the primary purpose of this code is to use as little space as possible.

Table 13: The EAN/UCC-8 is an 8-digit used primarily outside the United States and Canada and on products with limited space.

EAN/UCC-8	14	13	12	11	10	9	8	7	6	5	4	3	2	1
	0	0	0	0	0	0								
EAN.UCC PREFIX							■	■	□					
PRODUCT ID NUMBER									□	■	■	■	■	
CHECK CHARACTER														■

3.4. Vehicle Identification Number (VIN)

Vehicle Identification Numbers (VIN) first appeared on American automobiles in 1954. The VIN was intended to identify the vehicle and provide a condensed description of the vehicle's characteristics. Early VIN numbers came in a variety of formats and configurations. The only way to interpret these numbers was through a VIN decoder.

In 1981, the U.S. National Highway Traffic Safety Administration (NHTSA) required all over-the-road vehicles to contain a standard, 17-character VIN, as shown in Table 14 [11]. In later years — in an effort to prevent theft — the Department of Transportation required VIN numbers to marked on major components.

Table 14: The Vehicle Identification Number (VIN) identifies the vehicle and provides a simple, condensed vehicle description.

VIN	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
MANUFACTURER CODE	■	■	■														
US MANUFACTURER CODE				■	■	■											
MODEL CODE							■	■									
U.S. MODEL CODE									■								
MODEL YEAR										■							
PRODUCTION PLANT											■						
MODEL TYPE												■					
BODY OR ENGINE TYPE													■				
SEQUENTIAL ID NUMBER														■	■	■	■

3.5. Financial Security Identification Systems

3.5.1. International Securities Identification Number (ISIN)

The International Securities Identification Number (ISIN) uniquely identifies securities and other financial instruments. The ISIN is designed for use in trades, management and assets accounting [12].

The ISIN code (ISO 6166) is a 12-character alpha-numeric code that identifies a security at trades and settlements on securities. Each country has a National Numbering Agency (NNA), which is a member of the Association of National Numbering Agencies (ANNA).

The ISIN is divided into two fixed partitions together with a single check character, as shown in Table 15. The first two characters are the country code, as issued in accordance with the international standard ISO 3166 of the country where the issuer of securities, other than debt securities, is legally registered or in which it has legal domicile.

Table 15: The ISIN code is a 12-character alpha-numeric code that identifies a security at trades and settlements on securities.

ISIN	12	11	10	9	8	7	6	5	4	3	2	1
COUNTRY CODE												
NSIN												
CHECK CHARACTER												

3.5.2. Committee on Uniform Security Identification Procedures (CUSIP)

The Committee of Uniform Security Identification Procedures (CUSIP) is a financial security identification system used primarily in the United States and Canada [13]. Every stock, bond, and other security has a unique, 9-digit CUSIP number chosen according to this system.

The CUSIP Number consists of nine alphanumeric characters consisting of two partitions together with a single check sum character. The first partition consists of 6 characters and is known as the issuer number. The first three characters are always numbers and the next three may be numbers or letters.

Within the CUSIP Issuer Number various blocks or gaps are reserved for particular applications. For example, the numbers 900 to 989 within each group of 1,000 numbers are reserved for overflow. Issuer Numbers 990 to 99Z in each group of 1,000 numbers are the user's internal use. This allows a user to assign an issuer number to any issuer that might be relevant to his holdings, but does not qualify for coverage under the CUSIP numbering systems.

The second partition consists of two alphanumeric characters and is the Issue Number and is also known as the CUSIP suffix. The Issue Number identifies each individual issue of the issuer. The Issue Number consists of two numeric characters when assigned to equity securities and two alphabetic characters or one numeric and one alphabetic character when assign to fixed income securities.

Table 16: The CUSIP Number consists of a 6 character Issuer Number. The first three digits are numeric and the last three are alphanumeric. The next two characters is the Issue Number for the Issuers and the last character is a check sum.

CUSIP	9	8	7	6	5	4	3	2	1
ISSUER NUMBER									
ISSUE NUMBER									
CHECK CHARACTER									

3.5.3. CUSIP International Numbering System (CINS)

The CUSIP International Numbering System (CINS) is closely related to the CUSIP system described above. For securities actively traded internationally, which are underwritten or domiciled outside the United States and Canada, the securities are identified by the CINS number.

The CINS number was developed in 1988 by Standard & Poor's and Telekurs (USA) in response to the North American securities industries need for a 9-character identifier. The CINS number is organized in the same way as the CUSIP number – 6 characters for the Issuer, 2 characters for the Issue and a single check character. The first character in the Issuer Number, however, is always an alphabetic character representing the country code.

Table 17. The CUSIP International Numbering System (CINS) is an alphanumeric code consisting of issuer number, issue number and a single check character. The Issuer Number contains a single character representing the Issuer's country code or geographic region.

CINS	9	8	7	6	5	4	3	2	1
ISSUER NUMBER – COUNTRY CODE									
ISSUER NUMBER									
ISSUE NUMBER									
CHECK CHARACTER									

3.6. Social Security Number (SSN)

In 1935, the U.S. established a permanent national pension system, financed through contributions by both the employee and employer. As a means of identifying the beneficiaries, unique Social Security Numbers were issued. In the United States, these Social Security Numbers (SSN) also serve as a common means of personal identification — from driver's licenses, to voter registration, to medical records.

The Social Security Number is a simple 9-digit number, divided into three sections of three, two and four digits, as shown in Table 18 [14]. The SSN can therefore provide only a total of 1 billion unique identifiers. Certainly enough for the U.S. population, but insufficient if applied globally.

The first partition, the area number, is three digits and is issued geographically beginning in the northeast and moving to the south west. The "geographical code," however, is loosely defined and was intended simply as a bookkeeping device for internal use and was never intended to convey accurate geographic information.

The second partition – the group number – is two digits ranging from 01 to 99 though not sequentially. The group numbers were allocated to odd numbers from 01 to 09 and even numbers after that. The group numbers are allocated to states. After all numbers in group 98 of a particular area are filled, then the even numbers from 02 to 08 and the odd numbers are issued.

The third partition is the serial number running consecutively from 0001 to 9999.

Table 18: The Social Security Numbers are used in the United States for personal identification.

SSN	9	8	7	6	5	4	3	2	1
AREA NUMBER									
GROUP NUMBER									
SERIAL NUMBER									

4. CONNECTING THE EPC™ TO IDENTIFICATION CODES

The Electronic Product Code™ (EPC™) has been proposed as a universal identification scheme for physical objects and a means of connecting those physical objects to networked information. As we have seen, there are a number of current and legacy numbering systems used for similar purposes. If the The Electronic Product Code™ – along with its associated elements – is to be effective, there must some means to correlate current systems, such as the GTIN and VIN, to the EPC™.

The Electronic Product Code™ fundamentally stores only three pieces of information – the domain manager, object class and serial number. All other information must be retrieved from the network. Although many other coding systems store this information, they also encode other types of data. It stands to reason that a simple mapping from the EPC™ to other codes is not possible if their specifications fundamentally differ.

There are however a number of ways we can solve this problem. We can (1) literally embed an existing code into the EPC™, (2) algorithmically embed a code or (3) perform a network "look-up" using the EPC™ as a reference number.

4.1. Literal Embedding

Perhaps the most straight forward way to "embed" an existing coding system into an EPC™ is to literally translate that code's data fields into a binary expression and store them in a portion of the EPC™ address space. In other words, literally translate the code into binary.

The advantage of this approach is that resolving an existing code, such as the Global Trade Item Number (GTIN), is trivial. It simply involves a binary to decimal conversion. Furthermore, this approach removes any network or data storage dependency and minimizes computation and algorithmic complexity.

A disadvantage, of course, is the wide disparity of current numbering systems that need to be represented in the EPC™ address space. Further is the potential fragmentation of the EPC™ into multiple, redundant subdomains.

There may, however, be a method that allows literal embeddings without removing the original intention of the EPC™ as a universal identification number. To this end, we propose that literal embeddings, if they are used, be stored in a **portion** of the EPC™ address space, while reserving another portion for universal identification. In other words, the EPC™ would include the generic **Universal Identification** number together with a set of **Domain Identification** codes that accommodate existing numbering systems.

Although, arguably not the most elegant approach, this method allows immediate adoption of the EPC™ – along with the ONS and PML – using existing numbering standards and provides a platform for future trans-industry universal identification. Further it removes any concerns regarding near-term adoption, while retaining the original philosophy behind the EPC™ – as embodied in the Universal Identification number.

4.1.1. GTIN

The Global Trade Item Number (GTIN) is perhaps the most widely used product identification system and the obvious first choice for any literal embedding. How might this be done? This section will outline a prospective approach to a binary encoded GTIN.

We first note the 14-digit GTIN includes a single check digit used to verify an optical scan. In our case, the verification of a correct scan takes place outside the scope of the EPC™. Therefore we are left with 13 decimal digits that must be encodes.

The second issue we are faced with is variable partition boundaries as specified for the GTIN. Specifically, the Company Prefix varies from 6 to 11 digits and the Item Reference (and Indicator Digit) from 2 to 7 digits. Together the Company Prefix, Item Reference and Indicator Digit total 13 decimal digits. Since we cannot include an implicit boundary, as is used in a printed label, we must explicitly encode the boundary size in some portion of the code itself.

The final issue is the unique serial number as required in the EPC™, but not included in the GTIN. For this, we simply append a fixed length serial number to the binary encoded GTIN.

Thus we now have the form of a binary representation of the Global Trade Item Number. The encoding should include (1) a partition that stores the location of the division, (2) a variable length Company Prefix, (3) a variable length Item Reference and (4) a fixed length serial number.

The exact divisions of a literal encoding of the various numbering systems within the fixed bit boundary of the EPC™, will be given in formal code specifications from the Auto-ID Center or its associated standards organization.

4.2. Compressed Embedding

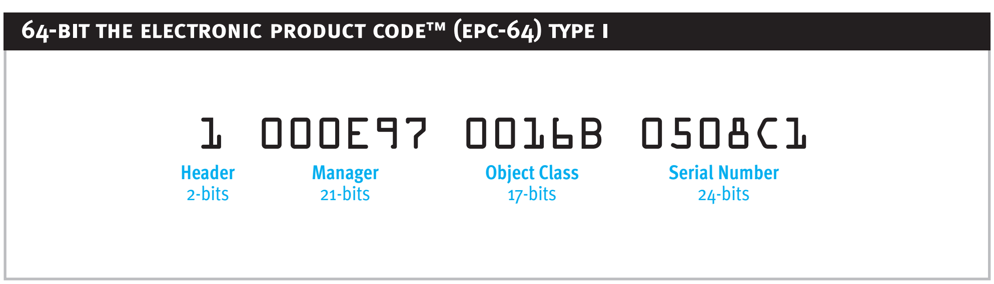
For identification codes that share a similar structure to the The Electronic Product Code™; that is, contain a product manager, item type and serial number, it is possible to embed that identification system into the EPC™. This embedding would not require the network and could be achieved at low cost and high performance.

4.2.1. GTIN

For example, let us consider the subset of the GTIN that encodes manager (i.e. EAN.UCC Prefix plus Company Number) and item number, together with the 64-bit EPC™ Type I code, as shown below in Figure 1. This variant of the compact EPC™ code can store up to 2,097,152 company identifiers, 131,072 item numbers and 16,777,216 serial numbers for each item.

Given that there are no more than 1 million member companies within the EAN.UCC, it seems there should be a method to easily embed GTIN into EPC™.

Figure 1: The 64-bit The Electronic Product Code™ (EPC-64) type I.



One method could be a block mapping, or **compressed embedding**, in which contiguous segments EAN.UCC Prefix Numbers are embedding into a continuous section of the EPC™ manager number, as shown in Figure 2.

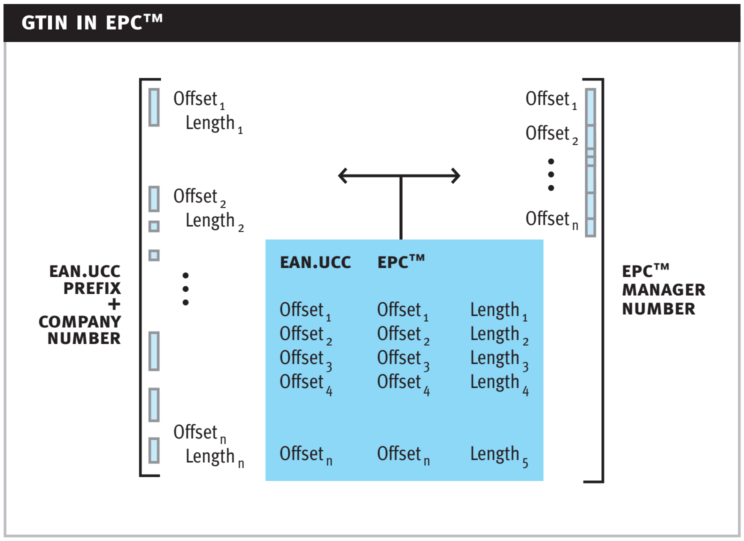


Figure 2: EAN.UCC Company Identifiers can be compressed into the EPC™ Manager Number using a simple, non-networked algorithm.

In order to achieve this mapping, we would need a table containing the beginning offsets from both the EAN.UCC Company Identifiers and the EPC™ Manager Numbers, together with the length of each contiguous segment. This triple of numbers would be stored in a data table.

The size of the data table is proportionate the fragmentation of the number space. For UCC managed numbers, there were 247,168 company identifiers containing 17,567 contiguous blocks. The resulting **embedding table** occupied 169KB of data storage, which was further reduced to 70KB using a simple binary compression.

Assuming a similar result for EAN managed numbers, the total embedding table file size would be approximately 210KB – well within the capacity of simple electronic devices and readers. (We are assuming approximately 12 _ cents per MB, which is the current price for non-volatile RAM at the time of this writing).

Finally, the EAN.UCC Item Number and EPC™ Object Class number would be embedded with the identity map. In other words, they are the same number.

4.2.2. VIN

The compressed embedding of the product identification subset of GTIN into 64-bit EPC™ type I Manager Numbers still leaves a substantial amount of room for other number systems, such as the VIN number. In this case, we can use a one-to-one data table embedding the VIN manufacture directly to EPC™ manager number, since there are so few VIN producers.

The Model Code, Year, Production Plant, Model, Body and Engine Type can be mapped to an EPC™ Item Number using a standard **hash table** approach. The VIN Sequential ID Number, of course, maps directly to the EPC™ serial number, as would be expected.

4.3. Network Mapping

The purpose of the The Electronic Product Code™ is to uniquely identify physical objects and serve as a reference to networked information. By its very nature the EPC™ exists along with the Object Name Service (ONS) and some standardized networked information, such as the Physical Markup Language (PML).

The most obvious, and in fact, the simplest means to connecting the EPC™ with current and legacy coding systems is to use network infrastructure. In other words, to use the EPC™ to reference – on the network – any number of existing identification codes.

Since the EPC™ stores a domain manager (i.e. a party responsible for keeping data about the identified object), an object class (i.e. a type of object, such as the Stock Keeping Unit used in retail) and a unique object serial number, this provides enough information to resolve all the data stored in any numbering system used today.

For example, with the EPC™ and network infrastructure, it is possible to determine lot number and expiration date, as well as country of origin, logistics information, service relations, packaging level and price. The **actual number** stored within the EPC™ is immaterial, since all relevant product information, as well as **numbering codes** can be retrieved easily from the network architecture.

4.3.1. From EPC™ to Identification Codes

We propose to map the The Electronic Product Code™ to existing coding systems using the current networking architecture – specifically the Object Name Service or the Object Name Service and Physical Markup Language. In this manner the EPC™ serves as a code for codes – or a **meta**-identification code.

It is important to note that the unique EPC™ Domain Manager Number and Object Class are sufficient to resolve the coding family, such as GTIN, NDC, ISBN, VIN, etc, and the actual code within that family.

Of course, given a certain application domain, there may be, in fact, multiple identification code families that correspond to a single EPC™.

4.3.2. From Identification Codes to EPC™

It is possible to use the Object Name Service to map other identification codes, such as GTIN, NDC, ISBN and VIN, as well as the typical EPC™ codes. In other words, the Object Name Service can translate a structured identification code into an EPC™ or – more directly –an Internet Protocol (IP) address.

Most codes, of course, will not be able to map to a complete EPC™, since most do not contain a unique serial number, but should be able to reference a unique EPC™ Domain Manager and Object Class.

One proposed solution is to append the suffix *.[code].obj.net to an identification number and use the ONS system to resolve an EPC™ and IP address. In this manner, the identification code mapping is contained within the general structure of the ONS infrastructure.

5. CONCLUSION

Therefore, we propose the method to connect the The Electronic Product Code™ to current and legacy numbering codes with stand-alone compressed embedding algorithms or the application of the proposed network architecture; that is, the The Electronic Product Code™ (EPC™), Object Name Service (ONS) and Physical Markup Language (PML).

We propose to connect the The Electronic Product Code™ to existing numbering systems using a combination of stand-alone compression algorithms and the application of the Auto-ID network infrastructure.

Not only does this approach resolve potential conflicts in numbering standards, it also provides additional benefits. Typically, different coding systems are used at different points in a product life cycle. With a single EPC™ code, an identified product may "assume" different identification numbers as it moves from producer to retailer to recycler. Internal company numbers, asset numbers, industry numbers, regulator numbers are often assigned to an identical product. The EPC™ can serve to "bridge the gap" among product identification systems.

Furthermore, the EPC™ – together with the network infrastructure – can serve as a translator or "middle man" between disparate coding systems. In other words, an object identification number used by one system can be correlated to another through the EPC™.

We propose to use the EPC™ as a sort of **meta-code**, which is generically applied to the identification of objects and a mechanism to reference networked information – independent of application domain. Furthermore, we propose to use the networking infrastructure currently under development to automatically and immediately link the EPC™ to any number of other current and legacy coding systems. In this way, effectively bridging the gap among industries and unifying the approach to object identification and description.

6. REFERENCES

1. **D. L. Brock, "Intelligent Infrastructure – A Method for Networking Physical Objects".**
   Presentation, MIT Smart World Conference, Apr 2000.
   http://auto-id.mit.edu/whatsnew/download/DB_Smart_World.pdf

2. **"The Networked Physical World – Proposal for Engineering the Next Generation of Computing, Commerce and Automatic-Identification".**
   Auto-ID White Paper, WH-001, Dec 2000.
   http://auto-id.mit.edu/pdf/MIT-.AUTOID-WH-001.pdf

3. **D. L. Brock, "The The Electronic Product Code™ – A Naming Scheme for Physical Objects".**
   Auto-ID White Paper, WH-002, Jan 2001.
   http://auto-id.mit.edu/pdf/MIT-AUTOID-WH-002.pdf

4. **D. L. Brock, "The Physical Markup Language – A Universal Language for Physical Objects".**
   Auto-ID White Paper, WH-003, Feb 2001.
   http://auto-id.mit.edu/pdf/MIT-AUTOID-WH-003.pdf

5. **International ISBN Agency, "The ISBN Users' Manual".**
   Berlin, Germany, 1999, ISBN 3-88053-075-0.
   http://www.bowker.com/standards/home/isbn/international/isbnmanual.html.

6. **The National Drug Code (NDC)**
   U.S. Food and Drug Administration (FDA)
   http://www.fda.gov/cder/ndc/database/default.htm

7. **The Uniform Code Council (UCC)**
   http://www.uc-council.org

8. **The European Article Numbering (EAN) International**
   http://www.ean.org

9. **The Uniform Code Council (UCC) Global Trade Item Number**
   http://www.uc-council.org/2005sunrise/global_trade_item_number.html

10. **Uniform Code Council (UCC), Uniform Product Code (U.P.C.), "Symbol Specification Manual".**
    January 1986.
    http://www.uc-council.org/reflib/01302/d36-t.htm.

11. **United States Department of Transportation (DOT),**
    National Highway Traffic Safety Administration (NHTSA),
    "Vehicle Identification Number Requirements".
    1995.
    http://198.17.75.65/fril/1995/19951025/95-26499.txt.

12. **Association of National Numbering Agencies (ANNA)**
    International Securities Identification Number (ISIN)
    http://www.anna-nna.com/

13. Committee on Uniform Security Identification Procedures (CUSIP)
CUSIP International Numbering System (CINS)
http://www.cusip.com/

14. Social Security Numbering Scheme, "From the Social Security Administration".
http://www.ssa.gov
http://www.ssa.gov/history/geocard.html