Document: AIM IUIDC-1-2008
Product Number: X5-nn

DRAFT

AIM IUID Compaction Guideline

Document Type: AIM Guideline
Document Version: Draft Rev. 1.2
Date: 2008-10-14

Foreword

This document was prepared by the Technical Symbology Committee (TSC) of AIM, Inc. (AIM Global), the trade association for the Automatic Identification and Mobility industry. AIM Global is the source for technically accurate, unbiased, commercial-free, and up-to-date information on all Automatic Identification and Mobility technologies including:

Bar Code including 2-D Symbologies
Biometrics
Enterprise Mobile Computing
Machine Vision
Magnetic Stripe
Optical Cards
Optical Character Recognition
Radio Frequency Identification
Smart Cards
Touch Memory
Voice Recognition
WLAN

AIM Global is an accredited standards organization under the American National Standards Institute (ANSI).

The association serves members world-wide through a network of AIM chapters. Member companies represent a wide array of AIDC technologies either as manufacturers or as providers of equipment, systems, and services. A listing of AIM Global's membership and chapters can be found at (http://www.aimglobal.org).

Published by:

AIM, Inc.
125 Warrendale-Bayne Road
Suite 100
Warrendale PA 15086
Phone: +1 724 937 4470
Fax: +1 724 934 4495
Internet email: aidc@aimglobal.org
Web: http://www.aimglobal.org

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. AIM, Inc. shall not be held responsible for identifying any or all such patent rights.

Individuals or organizations that hold intellectual property that may be referenced in this document are encouraged to disclose their patents during the public review process.

NOTE: This is a committee draft and shall not be circulated except at the direction of the committee. It is not an official AIM, Inc. Guideline document.

Table of Contents

Introduction

The Item-Unique Identification (IUID) program is a US Department of Defense program that enables improved access to information about DoD property to make management of items faster and more accurate. A DoD Unique Item Identifier (UII) permanently identifies an individual item distinctly from all other individual items that the DoD buys and owns. The item supplier marks (or writes) the UII data elements within a Data Matrix ECC200 symbol on an item. In order to read the UII data elements in a receiving system, each of the data elements have to be described to the receiving system by a prefix known as a data qualifier. A “Data qualifier” is a specified character (or string of characters) immediately preceding a data field that defines the general category or intended use of the data that follows. Data qualifiers in the UII application are drawn from three industry-standard data systems: alphanumeric Data Identifiers (DI), numeric Application Identifiers (AI), or alpha Text Element Identifiers (TEI).

In order to unambiguously denote which data system is utilized within a given symbol, the symbol is encoded using a “message envelope” that is defined in ISO/IEC 15434. Within that message envelope, a two-digit Format Indicator is encoded. In this application, the Format Indicator will be one of three choices: “05” indicates the use of AI data qualifiers, “06” indicates the use of DI data qualifiers, and “12” indicates the use of TEI data qualifiers,

Although all three data systems are supported in ISO/IEC 15434, only the “05” and “06” formats have corresponding “macro” codewords in the Data Matrix Symbology that reduce the encoded overhead of the message envelope. That fact, plus the inherent encoded length of each TEI data qualifier, has resulted in Data Matrix symbols that are sometimes too large to fit within the allotted space on parts used in TEI applications.

This document defines a method for creating more compact Data Matrix symbols, when encoding TEI data for the purposes of the IUID application. It uses a “compaction overlay” approach, whereby the “compacted” data is still valid in the application domain, and therefore compatible with receiving systems that are unaware of the compaction process. This is in contrast to the option for a more aggressive compaction method, one that would possibly produce smaller symbols, but where the resulting symbols would not convey meaningful information to a receiving system that had not been upgraded to decipher the compaction.

Although the resulting symbol is intended to be fully compliant with the IUID program, as with all symbology and quality standards, it is the responsibility of the application specification developers to define the appropriate parameters of this guideline for use in conjunction with a particular application.

Note: while this document is specific to the DoD IUID application and, therefore, uses the Data Matrix symbology, this overlay compaction approach could be applicable to other data carriers employing the same set and sequence of TEIs.

1 Scope

This is a technical engineering document intended for application specification developers and suppliers of encoding and decoding equipment for use in the IUID application.

This document describes a method for converting a source message, using ISO/IEC 15434 Format 12 for TEIs, into a compacted message which contains sufficient information for back-conversion to the original Format 12 source message. When encoded, the compacted message usually results in a smaller Data Matrix symbol. The compacted message is compliant with IUID specifications, and is suitable for use without back-conversion.

This document fully describes the algorithms for converting the source message to a compacted message, and for converting back to the original source message. The compacted message is compatible with all Data Matrix encoding and decoding equipment. However, in order to take advantage of the compaction, the new source-conversion algorithm will need to be added to source-marking systems. If back-conversion to the original TEI-based message is required, the back-conversion algorithm will need to be added to receiving systems.

2 Normative references

ISO/IEC 16022 - Information technology — Automatic identification and data capture techniques — Data Matrix bar code symbology specification
ISO/IEC 15434 - Information technology — Syntax for high capacity ADC media

3 Term(s) and definition(s)

The terms and definitions given in ISO/IEC 19762, ISO/IEC 15416 and ISO/IEC 15415 apply, together with the following:

3.1
back-conversion
The translation of an ISO/IEC 15434 Format 06 message, if it contains the Translation DI, to a Format 12 message.

3.2
DI
Data Identifier as defined in ISO 15418

3.3
source-conversion
The translation of an ISO/IEC 15434 Format 12 message, if it contains only TEIs defined for the IUID application, to a Format 06 message that includes a Translation DI.

3.4
TEI
Text Element Identifier (ref. ATA Spec 2000), three capital letters followed by an encoded space.

3.5
Translation DI
A Data-Identifier whose associated data element defines instructions for back-conversion.

4 Symbols (and abbreviated terms)

IUID Item-Unique Identification
UII Unique Item Identifier

5 Overview of the IUID Compaction methodology

5.1 Original Source message

An original source message, using ISO/IEC 15434 Format 12 for TEIs, will contain a set of one or more data elements from which the concatenated UII can be derived. For items that are serialized within the enterprise identifier, the UII data set includes the data elements of enterprise identifier and a unique serial number (Construct #1). For items that are serialized within the part, lot or batch number within the enterprise identifier, the UII data set includes the data elements of enterprise identifier, the original part number or lot or batch number, and the serial number (Construct #2). In addition to the two constructs composed of multiple data elements, the UII data set may be represented by a single data element, such as would be defined by the use of the TEIs USN, UST or UID.

The compaction method described in this guideline will improve the encoded efficiency of any TEI-based IUID message, but the improvement is most evident when starting from a Construct 2 message. This is because the three TEI components of a Construct 2 message require additional overhead compared to other IUID formats.

A typical TEI-based Construct #2 UII to be encoded has three component data elements:

- an enterprise identifier, using a data qualifier such as “MFR “,
- the original part, lot or batch number, using a data qualifier such as “PNO “, and
- the serial number, using a data qualifier such as “SEQ ”

As an example, the following ISO/IEC 15434 message string represents a typical Construct 2 TEI-based IUID message:

[)>$^{R}_{S}$12 $^{G}_{S}$MFR 88236 $^{G}_{S}$PNO 15252P$^{G}_{S}$SEQ A630 $^{R}_{S}{}^{E}_{O_T}$

Since there is no “macro codeword” for Format 12, all of the ISO/IEC 15434 format characters shown above need to be explicitly encoded in the Data Matrix symbol, as well as the three 4-character TEIs and the two internal ASCII Group Separator characters.

5.2 Source-Conversion

This guideline defines a source-conversion algorithm, that takes as input an IUID-compliant Format 12 message (TEI-based, such as the example shown above), and translates that message to an equivalent IUID-compliant Format 06 message (DI-based, specifically, a Concatenated UII, using DI “25S”).

The rules for this source conversion algorithm are given in Clause 7.

The translated message can be encoded into a smaller Data Matrix symbol, for two reasons:

- Because the translated message uses ISO/IEC 15434 Format 06, it benefits from a Data Matrix “macro” codeword, which eliminates the need to encode the overhead characters at the front and back of the message;
- Because it uses a Concatenated UII format, it only needs to encode one data qualifier instead of three.

However, it must be noted that although one can always represent the three data components (of the Construct 2 form) as a Concatenated UII, one cannot normally examine the concatenated data and "parse" it back into the three data components of the equivalent Construct 2 form. That is true because:

- There are many data qualifiers that can represent the three components, but the choice of qualifier is lost when the data is concatenated, and
- The part number and serial number are both variable-length data elements, and after they are concatenated, the boundary between data elements is also lost.

Although the receiving IUID application itself does not need access to the three individual data components, many other receiving systems do make use of the individual Enterprise ID, Part Number, and Serial Number components, all properly labeled with their original data qualifiers.

Therefore, this guideline defines a secondary Data Identifier, a "Translation DI" identified by the data qualifier "6C", which is encoded immediately after the "25S" and its data. This Translation DI efficiently encodes the additional information that will allow a receiving system to reconstruct the source message in its original Construct 2, TEI-based, form. This Translation DI utilizes the IUID Compaction Table defined in Annex A of this guideline.

Each row of the Annex A IUID Compaction Table lists one of the valid TEI-based data element combinations that meet the IUID requirements. For each row, it also:

- indicates which Issuing Agency Code (IAC) must follow the "25S" when these data elements are concatenated (either "D" for CAGE or "UN" for DUNS),
- gives the row an Identifier (the row's index into the "D" or "UN" section of the table), and
- indicates which data element length(s), if any, need to be encoded in the secondary DI.

Using the example shown above, one can observe that the ninth row of the Table (ID 8, when using an IAC of "D") contains the example's TEI sequence of "MFR ", "PNO ", and "SEQ ". This ninth row also indicates that the length of the second data item (the Part Number) needs to be encoded in the "6C" Translation DI of the converted message. Therefore, the Translation DI and its data string are "6C85", where:

- 8 indicates the row ID from the Table (when "25S" is followed by "D"), and
- 5 indicates the length of the second data component, minus one (since the length of the first data component is fixed at length 5 when the IAC is "D", only one length needs to be encoded in order to determine all three component boundaries).

The resulting converted message, ready for efficient encoding into a Data Matrix symbol, is:

[)>$^{R}_{S}$ 06 $^{G}_{S}$ 25SD8823615252PA630 $^{G}_{S}$ 6C85$^{R}_{S}$ $^{E}O_{T}$

After applying the 06 macro, the message encoded in the Data Matrix symbol is:

<codeword 237>25SD8823615252PA630$^{G}_{S}$6C85

where <codeword 237> is in the first character position as an abbreviation for the header [)>$^{R}_{S}$06$^{G}_{S}$ and the trailer $^{R}_{S}$ $^{E}O_{T}$.

Data Matrix symbol showing example of uncompacted data message described in 5.1 above	Data Matrix symbol showing example of compacted data message described in 5.2 above

Figure 1 Example of uncompacted and compacted messages.

5.3 Back-Conversion

Depending on the requirements of the receiving system, a back-conversion step may or may not be desired. For example, if the receiving system is designed to meet the IUID requirements, then it will most likely simply store the data from the first data string of the message (the data associated with the "25S" DI), and discard the remainder of the message. If instead the receiving system wants to see the data in its original TEI form, then the back-conversion algorithm should be implemented in the receiving system. Within an ISO/IEC 15434 Format 06 message, the appearance of a "6C" data qualifier immediately after a "25S" and its data is the indication that this Data Matrix symbol's message can be back-converted from its current DI-based Construct 1 format to the original TEI-based Construct 2 format.

The back-conversion algorithm performs the following steps, to reconstruct the original TEI-based message:

- It finds the proper row of the Annex A table, using the combination of the IAC character(s) (the "D" or "UN" immediately after the first DI of "25S") and the Table ID character (the first character after the second DI of "6C").
- From the selected row of the Table, it determines the proper sequence of TEIs.
- If so indicated by the selected Table row, it determines an encoded data element length (from the second character after the "6C" DI).
- It parses the concatenated IUID data, using the TEIs and length information from the Table to reconstruct the original TEI-based message.
- It puts a Format 12 message envelope around the reconstructed message, resulting in the original message (such as that shown at the end of section 5.1 above) that was input to the source-conversion step.

6 The Translation DI

6.1 Assignment of the Translation DI

Data Identifiers (DIs) are maintained by the DI Maintenance subcommittee of the ANS MH10.8 committee. At the time of this writing, they have tentatively agreed to assign "6C" (from the "C" series of "Continuation" DIs) to designate the Translation DI that enables IUID Compaction in accordance with this guideline.

6.2 Translation DI Format

In a source-converted Format 06 message, the Concatenated UII data element (identified by Data Identifier "25S") shall be immediately followed by the Translation DI "6C" and its data element.

The Data associated with "6C" consists of an IUID Compaction Table ID, followed by zero or more Component Length indicators (as specified by the table row corresponding to the Table ID), then optionally followed by additional TEI data elements, as described in the following subsections.

6.2.1 IUID Compaction Table ID

A IUID Compaction Table ID is normally a single character (a digit or an uppercase Alpha in the range 'A' through 'Y' inclusive), providing up to 35 indices into a region of the Table, where all rows of a given region have the same Issuing Agency Code (IAC) specified. If in a future revision of this guideline, more than 35 IDs are ever required within a single region, then these will be identified by "Z0", "Z1", "ZA", "ZZ0", etc.

6.2.2 Component Length Indicators

A Component Length Indicator is usually a single character (a digit or an uppercase Alpha), providing up to 35 lengths (encoding lengths in the ranges 1..10 and 11..35 as '0'..'9' and 'A'..'Y' respectively). Component Lengths greater than 35 are not anticipated in this application, but if needed, they are identified by "Z0", "ZA", "ZZ0", etc.

The length encodings above use a self-identifying number of characters, to make it possible to encode optional additional TEI data items after the Table ID and Component Length information relating to the component TEIs that constitute the 25S data. This option is described in the following subsection.

6.2.3 Additional Option TEI data strings

Immediately after the last Component Length indicator, one or more additional TEI data strings may be encoded as a trailing portion of the Translation DI data string. If present, these shall be encoded exactly as they would be in a Format 12 message (a four-character TEI identifier ending with a space, followed by the corresponding text string), except that the GS characters normally used to separate multiple TEI data strings are omitted, as the use of a GS character in this context would prematurely terminate the Translation DI data string.

The decoding process uses the space character at the end of each TEI in order to parse these additional optional data strings. Therefore, a TEI data string that itself contains a space character cannot be encoded within a Translation DI data string.

7 Source-Conversion algorithm

Source-conversion is the process of taking an IUID-compliant ISO/IEC 15434 message in Format 12 (TEIs) and converting it to a minimum-length IUID-compliant message in Format 06 (DIs), with the addition of a Translation DI to provide the ability to reconstruct the original TEI-format message at the receiving system if desired.

The source-conversion process is implemented by the following steps:

7.1 Input Validation

Determine that the input message is in Format 12 (TEIs), and conforms to the IUID requirements (for details, see the US DoD's homepage for the IUID program at http://www.acq.osd.mil/dpap/pdi/uid/index.html).

7.2 Create a Format 06 Concatenated UII

The Concatenated UII is created from the original input message by performing the following steps:

1. Strip off the ISO/IEC 15434 header and trailer characters from the input message string.
2. Strip off each internal ASCII Group Separator character of the remaining message string.
    a. As this step is performed, record the sequence of TEI data qualifiers in the order that they are encountered, and record the length of each data string found between the TEI and the next internal Group Separator (or end of the message).
3. Find the row of the Annex A IUID Compaction Table that contains the exact set of TEIs that were found at the start of the input message (ignoring any optional additional TEIs that might follow the matching set). Abort processing if no match is found; otherwise:
    a. If the set is not in the same sequence as shown in the table in Annex A, reorder the data elements to match the sequence in that row of the table.
    Note: if this reordering is required, the sequence of the elements in the back-

conversion will match the sequence of elements in the table, not the original sequence.
   b. Record the IAC listed in the first column of that row
   c. Record the Table ID listed for that row
   d. Record the length of the data component whose length needs to be encoded (if any)
   e. Confirm that the lengths of any data components listed as fixed-length in the table are of the correct length. If any are incorrect, abort processing
4. preface the message string with "25S" followed by the IAC (either "D" or "UN") that was determined from the previous step.
5. append each of the input data elements from the matching set found in step 3.a, after first removing the leading four-character TEI from each of these data elements.

7.3 Create the Translation DI

The Translation DI is created using information derived from the previous steps, as follows:

1. initialize the Translation DI string as "6C"
2. Append to this string the Table ID character(s) determined from Step 3.b above.
3. Append to this string the (length – 1) recorded from Step 3.c above.
4. If an additional optional TEI data strings remain in the input after encoding the matching set into the "25S" data string, append each of these to the end of the Translation DI data string, including their leading four-character TEI data qualifiers, but do not encode any GS characters in between these strings.

7.4 Create the final Format 06 message

The final source-converted message is created by performing the following steps:

1. Initialize the final string to the ISO/IEC 15434 Format 06 header "[)>$^{R}_{S}$ 06 $^{G}_{S}$."
2. Append the Concatenated UII string (beginning with "25S") that was created in section 7.2
3. Append an ASCII Group Separator character
4. Append the Translation DI string (beginning with "6C" that was created in Section 7.3
5. Append the ISO/IEC 15434 Trailer "$^{R}_{S}{}^{E}O_{T}$"

7.5 Encode the final message into a Data Matrix symbol

The final source-converted Format 06 message is compatible with existing Data Matrix encoding equipment, and the encoding process should take advantage of the "Macro 06" codeword that is defined in ISO/IEC 16022.

8 Back-Conversion algorithm

Back-conversion is the process of taking as input an IUID-compliant ISO/IEC 15434 message in Format 06 (DIs), determining that the message contains a "25S" concatenated UII followed by a "6C" Translation DI, and converting the Format 06 message to an IUID-compliant message in Format 12 (TEIs), using the information contained in the Translation DI.

The back-conversion process is implemented by the following steps:

8.1 Decode the Data Matrix symbol

Using the procedures of ISO/IEC 16022, decode the Data Matrix symbol.

8.2 Validate the input to the back-conversion algorithm

Verify that the output of the previous step is an ISO/IEC 15434 message, beginning with a format "06" format envelope, containing only two data items with data qualifiers of "25S" and "6C". If this is not the case, abort back-conversion processing.

8.3 Create the temporary data strings

Create two temporary data strings:

1. Strip off the ISO/IEC 15434 header and trailer characters from the input message string.
2. Create an Interim Data string, using all of the characters following the "25S", up through but not including the Group Separator.
3. Skip the Group Separator and the following "6C", and create a Translation Data string from the remainder of the input message

8.4 Create the final result: a series of TEI data strings

Recreate the original component TEI data strings, as follows:

1. Initialize a Results string to "[)>$^{R}_{S}$ 12"
2. Find the appropriate row of the Annex A IUID Compaction Table, using the combination of the IAC (the first one or two characters of the Interim Data String) and the Table ID (the first one or more characters of the Translation Data string) as the search key.
3. For each TEI listed in that row of the table, perform the following substeps:
    a. Append a Group Separator character, followed by the listed TEI, to the Results string
    b. Append the next 'n' characters of the Interim Data String to the Results string, where 'n' is either a fixed number listed in the Table, or (if so indicated by the Table entry) a variable-length number derived by parsing the next Component Length from the Translation Data string
4. *If any data remains in the Translation DI's data string after all Component Length indicators have been processed, then this data represents optional additional TEIs that are not described in the Compaction Table, but were encoded verbatim. Process these by repeating the following steps until no more Translation DI data remains:*
    a. verify that the fourth character from the current point is a space
    b. append a GS character to the end of the results string, then append these four characters to the results string, and advance the "current point" past the four characters just copied to the results.
    c. identify the corresponding data string, which ends either four characters before the next space in the Translation DI's data,, or at the end of the Translation DI's data (if a next space character is not found).
    d. Append the corresponding data string to the results string, and advance the "current point" to the end of the data string that was just copied to the results.
5. Append to the Results string a message trailer string "$^{R}_{S}{}^{E}O_{T}$"

Annex A (normative) IUID Compaction Table

The table in this Annex lists all of the valid combinations of TEI-based components in the IUID application, and provides the necessary information to encode or decode the Translation DI when a given row of the table is utilized.

25S IAC	ID	1st comp	2nd Comp	3rd Comp	Length1	Length2
D	0	UID				
D	1	USN				
D	2	UST				
D	3	CAG	SER		5	
D	4	CAG	UCN		5	
D	5	MFR	SER		5	
D	6	SPL	UCN		5	
D	7	CAG	PNO	SEQ	5	Comp2
D	8	MFR	PNO	SEQ	5	Comp2
D	9	SPL	PNO	SEQ	5	Comp2
D	A	CAG	LOT	SEQ	5	Comp2
D	B	MFR	LOT	SEQ	5	Comp2
D	C	SPL	LOT	SEQ	5	Comp2
D	D	CAG	LTN	SEQ	5	Comp2
D	E	MFR	LTN	SEQ	5	Comp2
D	F	SPL	LTN	SEQ	5	Comp2
D	G	CAG	BII	SEQ	5	Comp2
D	H	MFR	BII	SEQ	5	Comp2
D	I	SPL	BII	SEQ	5	Comp2
UN	0	UID				
UN	1	DUN	SER		9	
UN	2	DUN	UCN		9	
UN	3	DUN	PNO	SEQ	9	Comp2
UN	4	DUN	LOT	SEQ	9	Comp2
UN	5	DUN	LTN	SEQ	9	Comp2
UN	6	DUN	BII	SEQ	9	Comp2

Bibliography

ISO 15418 Information technology -- Automatic identification and data capture techniques -- GS1 Application identifiers and ASC MH 10 data identifiers and maintenance [note: current version references EAN/UCC rather than GS1 -- this is in committee to change]

ATA SPEC2000, http://www.spec2000.com/50.html.