



SURFACE VEHICLE STANDARD

J2540™

JUN2019

Issued 2002-07
Stabilized 2019-06

Superseding J2540 JUL2002

Messages for Handling Strings and Look-up Tables in ATIS Standards

RATIONALE

This technical report is being stabilized because the technical committee cannot find users for the technical report.

STABILIZED NOTICE

This document has been declared "Stabilized" by the SAE V2X Core Technical Committee and will no longer be subjected to periodic reviews for currency. Users are responsible for verifying references and continued suitability of technical requirements. Newer technology may exist.

SAE Technical Standards Board Rules provide that: "This report is published by SAE to advance the state of technical and engineering sciences. The use of this report is entirely voluntary, and its applicability and suitability for any particular use, including any patent infringement arising therefrom, is the sole responsibility of the user."

SAE reviews each technical report at least every five years at which time it may be revised, reaffirmed, stabilized, or cancelled. SAE invites your written comments and suggestions.

Copyright © 2019 SAE International

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of SAE.

TO PLACE A DOCUMENT ORDER: Tel: 877-606-7323 (inside USA and Canada)
Tel: +1 724-776-4970 (outside USA)
Fax: 724-776-0790
Email: CustomerService@sae.org
<http://www.sae.org>

SAE WEB ADDRESS:

Full document details: SAE J2540_201906

SAE values your input. To provide feedback on this Technical Report, please visit http://standards.sae.org/J2540_201906

Documents
Document Sharing Hub

Foreword—This SAE Standard outlines the methodology and messages used to convey textual strings of messages to consumers in ATIS and other areas. It outlines how concepts originally developed for SAE J2369 can be expanded to cover the various types of ITIS messages used in US deployments in a consistent way and meet a number of goals. Among these goals is overcoming the historical problems of obsolete tables without requiring the mutual agreement and updating of all deployed users. It provides implementation details for deployments wishing to use tables in ways ranging from non-changing (“static”) embodiments, to highly flexible and changing table structures (“dynamic”), and provides a single uniform formatting standard for all (regardless of specific content). In the related documents are encodings of selected national level tables used in various deployments.

See also:

- SAE J2540-1 – RDS Phrase Lists
- SAE J2540-2 – ITIS Phrase Lists
- SAE J2540-3 – National Names Phrase Lists

TABLE OF CONTENTS

1.	Scope	3
2.	References	3
2.1	Applicable Publications	3
2.2	Related Publications	4
3.	Definitions.....	4
4.	Overview	8
4.1	Goals of the Document.....	8
4.2	Limits of This Approach.....	10
4.3	Summary	13
5.	Operating Concepts	14
5.1	Assignment of a Table in Use.....	14
5.2	Properties of a Table	14
5.3	Nesting of Tables	14
5.4	Supported Vocabularies and Alphabets	15

5.5	Table Expansion Values	16
5.6	Simple Text within Strings	17
5.7	Registering a table within ITS	18
5.8	Adding to a table's entries	18
5.9	Table Resorting	18
5.10	Considerations for end user equipment.....	19
5.11	Changes to existing standards	19
6.	Messages	19
6.1	The Table Message	20
6.2	The Partial Table Message.....	22
6.3	The Table Request Message	23
6.4	The Index Translate Message	23
6.5	The Index Translate Request Message	23
6.6	The String Translate Message.....	24
6.7	The String Translate Request Message	24
7.	Data Elements	24
7.1	Date Element: Table-Registration-Value	24
7.2	Date Element: Local-Table-Flags	26
7.3	Date Element: About-Flags	26
7.4	Date Element: Index-Start	27
7.5	Date Element: Index-Stop.....	27
7.6	Date Element: Index-Count	28
7.7	Date Element: Index-8	28
7.8	Date Element: Index-11	28
7.9	Date Element: Index-12	28
7.10	Date Element: Index-16	28
7.11	Date Element: Table-Entry	29
7.12	Date Element: Local-Number	29
7.13	Date Element: Included-Table-Flags	30
7.14	Date Element: Request-Type	32
7.15	Date Element: Revision	33
7.16	Date Element: Word-Count.....	33
7.17	Date Element: Entry-Type	33
7.18	Date Element: CRC-16.....	34
7.19	Date Element: SAE-String	34
8	Using This Document in Other Messages	35
8.1	General Comments for Use.....	35
8.2	Connecting Strings across Multiple Message Fields	36
8.3	Examples of use in SAE J2354	36
8.4	Examples of use in SAE J2369	37
Table 1	Typical Multi-Part Phrase Table	11
Table 2	Latin_1 Character Excluded By Current SAE J2369 Codings	13
Table 3	Types of Tables and Their Exception Tokens	16
Table 4	Example of Table Control Expansion Bits (From SAE J2369 Table 8).....	16
Table 5	Example of Numeric Table Control Expansion Bits (From SAE J2369 Table 9)	17
Table 6	Table Registration Value Ranges	25
Table 7	Reserved Table Registration Values	25
Table 8	Assigned Local Table Numbers	29
Table 9	Reserved Table Numbering Assignments	30
Table 10	Incident Flow Context Table Entries (From Table 24 of SAE J2369)	37

- 1. Scope**—This SAE Standard defines methods and messages to efficiently translate sequences of text and other types of data into and out of indexed values and look-up tables for effective transmission.

This document defines:

- a. Methods and Data Elements for handling indexes and strings in ATIS applications and message sets
- b. Message Sets to support the delivery and translations of tables used in such strings
- c. Tables of Nationally standardized strings for use in ATIS message descriptions

And examples of each in illustrative portions.

While developed for ATIS use, the methods defined in this document are useful for any textual strings in any Telematics applications found both in Intelligent Vehicles and elsewhere.

2. References

- 2.1 Applicable Publications**—The following publications form a part of this specification to the extent specified herein. Unless otherwise indicated, the latest version of SAE publications shall apply.

- 2.1.1 SAE PUBLICATIONS**—Available from SAE, 400 Commonwealth Drive, Warrendale, PA 15096-0001.

SAE J2313—On-Board Land Vehicle Mayday Reporting Interface, May 2000
SAE J2353—Data Dictionary for Advanced Traveler Information Systems (ATIS), October 1999
SAE J2354—Message Sets for Advanced Traveler Information Systems (ATIS), October 2000
SAEJ2369—Standards for ATIS Message Sets Delivered Over Reduced Bandwidth Media, November 2000
SAEJ2374—Information Report based on Location Reference Message Specification, Rev. B (MDI), May22, 1997
SAE J2540-1—RDS Phrase Lists
SAE J2540-2—ITIS Phrase Lists
SAE J2540-3—National Names Phrase Lists

- 2.1.2 ANSI PUBLICATION**—Available from ANSI, 25 West 43rd Street, New York, NY 10036-8002.

ANSI X3.4—American National Standard Code for Information Interchange

- 2.1.3 CEN PUBLICATIONS**—Available from CEN, 36 rue de Stassart, B-1050 Brussels, Belgium.

Traffic and Traveller¹ Information (TTI) – TTI Messages via Traffic Message Coding- Part 1: Coding protocol for Radio Data Systems – Traffic Message Channel (RDS-TMC) using ALERT-C, ISO/DIS 14819-1, August 1999

Traffic and Traveller Information (TTI) – TTI Messages via Traffic Message Coding- Part 2: Event and Information codes for Traffic Message Channel (TMC), PrENV 12313-2, Version 1.0, June 1996

- 2.1.4 GATS PUBLICATIONS**—Available from the GATS forum or from CEN TC278.

GATS Message Set; Main Part Main 2034.pdf 05.08.99
GATS Message Set; Annex 2 - Basic Information Elements A2_2033.pdf 05.08.99

- 2.1.5 ISO PUBLICATION**—Available from ANSI, 25 West 43rd Street, New York, NY 10036-8002.

ISO prEN 14819-1 (1999)—

1. European spelling of Traveler

2.2 Related Publications—The following publications are provided for information purposes only and are not a required part of this document.

2.2.1 SAE PUBLICATIONS—Available from SAE, 400 Commonwealth drive, Warrendale, PA 15096-0001.

SAE J2353—Data Dictionary for Advanced Traveler Information Systems (ATIS), October 1999

SAEJ2374—Information Report based on Location Reference Message Specification, Rev. B (MDI), May 22, 1997

2.2.2 FHWA PUBLICATIONS—Available from Lee Simmons, FHWA, HVH-1 Room 3400, 400 7th Street SW, Washington, D.C. 20590.

National ITS Architecture, Federal Highway Administration, U.S. Department of Transportation, Version 2.0, September 1998

2.2.3 ITE PUBLICATIONS—Available from ITE, 525 School Street SW, Suite 410, Washington, DC 20024.

Message Sets for External Traffic Management Center Communication (MS/ETMC2), TM 2.01, Rev. 2.0, February 1, 2001

2.2.4 W3C PUBLICATIONS—Available from W3C, <http://www.w3.org>.

"Hypertext Transfer Protocol -- HTTP/1.1," W3C - June 1999, RFC 2616, available at <http://www.w3.org/Protocols/rfc2616/rfc2616.txt>

"HTML 4.0 Specification," W3C - December 1997, revised 24 April 1998 available at <http://www.w3.org/TR/1998/REC-html40-19980424>

HTML 4.0 Guidelines for Mobile Access W3C Note - 15 March 1999 available at <http://www.w3.org/TR/NOTE-html40-mobile>

Character Model for the World Wide Web, World Wide Web Consortium Working Draft 29-November-1999 available at <http://www.w3.org/TR/1999/WD-charmod-19991129/>.

3. Definitions—The following terms and concepts are used in the body of this document.

3.1 Alert-C—A traffic message system using *phrase - index* encodings operating over RDS and established in 1990, see ISO prEN 14819-1 (1999) for current information. Along with a more advanced system called Alert-Plus (and typically a pay for service system) this system has been deployed in selected major cities in Europe. Typically under local government sponsorship. A number of early US deployments by Enterprise States have used derivative versions of the phrases established in this work for trial US deployments. A number of commercial vendors support derivative forms of it, both on RDS and on other media as well.

3.2 Compression—The process of replacing a sequence of simple text elements (or other strings) with a small index value for transmission is considered to be compression in this document.

3.3 Dense Table—A table where the index values increase by one each time without a break in the ordering. The index value can be implicitly computed and therefore need not be sent. Dense tables are preferred, and are easier to automatically generate than sparse tables.

3.4 Dynamic Table—A table which changes its contents periodically as a function of the textual information which is being transmitted and which attempts to reflect to achieve the best compression in the media. As contrasted with static tables which only change periodically. Strategies affecting the frequency of updating vary with implementation, media, and ISP operators' choices. Dynamic tables convey their changes over the same media as the messages using them. This is an essential aspect of their operation.

- 3.5 End Using Device**—A term used in this document to refer to the device which received the message and strings that are a subject of this document. The capabilities of such a device may vary widely from one user to another. The device is presumed to be mobile, but need not be. The end using device gets its data from the issuing ISP over the host media.
- 3.6 Exception**—A token which causes some change in the flow of control in processing the stream. In this document, exception tokens precede an index value and are used to select which table (and how) to expand from.
- 3.7 Expansion**—The process of replacing an index value found in a string with the corresponding string or text value which it represents.
- 3.8 Expansion Code**—The code which precedes an index value (also called an exception token). This code selects which of up to four tables is to be used in expanding the string back into text and provides some control flags for how the expansion is to occur.
- 3.9 Extending a Table**—The act of adding additional index entries to the end of an existing table. All tables can be extended. By this means new entries are added as needed while preserving the older index values. Periodically a *resort* may occur which reassigns the mapping of index values. The revision bits allow the end user to determine if a table has been extended and/or resorted.
- 3.10 Flags**—One or more bit values contained in some part of a message which effect the way a string is processed in converting it back into text. In the table message there are flags dealing with how the table itself, and how any included tables, are to be handled.
- 3.11 Flat Table**—A table whose entries do not contain indexes to any other tables, i.e., a table which is entirely self contained. A flat table may contain nested table entries to itself.
- 3.12 GATS**—Global Automotive Telematics Standard. A mobile phone system based on the GSM (Global System for Mobile Communication) phone system deployed in Germany and other areas of Europe. The GATS systems provides a message set which offers many of the same services in US ITS-ATIS. GATS is being advanced by the CEN TC278 process as a standard.
- 3.13 Group 8A**—The original message type used in RDS to carry TMC messages in a complete 37-bit message. Now such messages can be carried in any valid spare group by means of a soft assignment process called ODA.
- 3.14 Host Medium**—The host media are the transmission media that was used to transport the ATIS messages containing the string from the ISP to the end-using device (e.g., wireless data links, SubCarriers, digital phones, etc.). Often such media have a limited amount of bandwidth and it is for this reason that the string handling methods of this document were developed.
- 3.15 Included Table Flags**—Flags providing information about tables used in the entries of this table, found in the table header.
- 3.16 ISP**—Information Service Provider. In the context of this document the issuing source of data formatted into ATIS messages containing strings handled in accordance with this document. No presumption is made in this document regarding the business model or ownership of the ISP. The term ISP (as opposed to TMC) is often used to imply a private “for profit” venture. No such term is implied here.
- 3.17 ITIS**—International Traveler Information System. A term coined to indicate the area of common phrases for describing incident events.

- 3.18 Local Replacement**—A term for a table entry which “overlaps” another entry in another table with the same index value. This allow the creation of local tables which can redefine a previous entry. See also “overlapping” and 5.3 of the text.
- 3.19 Local Table Flags**—Flags providing information about the tables found in the table header.
- 3.20 Local Table Number**—A value assigned to any table used by the ISP in its messages. Used to relate specific strings to this table (and any underlying tables).
- 3.21 LRMS**—Location Referencing Message Set A body of work outlining a number of “profiles” used in the description of affected areas, typically roadway segments. The preferred methodology for use in ATIS is the ISP → Vehicle profile (SAE J1746) which provides a means to send Lat-Long values between the ISP and the vehicle. A highly compressed form of this is the GRID profile which is suitable for limited bandwidth applications. Indexing values, which typically require prior agreement between users, are called Geometry profiles in this body of work. The LMRS is expected to become an SAE Standard in the future and will reflect all the profiles then used, including indexes.
- 3.22 Modified ASCII**—This is a derivative of the ASCII character set specified in detail in SAE J2369. It uses the upper encoding values (beyond 127) to replace the most commonly occurring double letter pairs with a single value, effectively using all 8 bits of the octet and providing a compression of up to 25%. The upper range is also used to represent the exception callout token used to select an index table. See SAE J2369 for complete details.
- 3.23 Nested Table**—A table whose contents (entries) are made up of strings consisting of simple text and indexes to other tables. This is the most typical type of table (as opposed to a “flat” table).
- 3.24 Nested Table Entry**—A specific table entry, which is made up (in part) of other table entries from the same table. This type of recursion typically occurs when there is a need for a collection of related phrases.
- 3.25 ODA**—Open Data Applications, an improvement to the original RDS formatting which allows an application to register and use any unused message frame in a local RDS system. Developed to manage the use of the original 64 possible messages. In the US market one can reasonably expect broadcasting of traffic over RDS to use either ODA and the tables in this annex combined with index aspects of the LRMS system (as opposed to say Alert-Plus methods) or to use only Group 8A messages. ODA provides a way to send incidents in one channel and supporting table description data in another.
- 3.26 Overlapping**—Tables may overlap each other in entries. This occurs when a table (typically a local table) is mapped into an index space of another table. Because the search order of table is deterministic, this effectively replaces any table entry further in the search order with the local entry. This is often used to provide a local expression for an entry. Refer to 5.3 for additional information.
- 3.27 RBDS**—See RDS.
- 3.28 RDS**—Radio Data System, also called Radio Broadcast Data System (RBDS) in the US. A modulation and message framing standard for sending low speed (1187.5 BPS) data over the sub carrier area of commercial FM broadcast standards. It was developed the in mid-80's and used to send a variety of information packets. It allows a fraction of this capacity (~37 BPS) to carry traffic messages at the rate of about one per second. This portion, called Traffic Message Channel (TMC), employed a two part ATIS message wherein one part was an index table of location references and the other part was an index into a table textual description of the event. The precise table values used have varied significantly over the past 15 years. The location numbering has invariably been redeveloped by each deployment. The phrases, while also varying to some degree, have been stable and are being advanced in the CEN TC204 process. They are commonly called Alert C and Alert+. In the parlance of this document, the LRMS format used would be called an index type of profile, while the format used would be considered static one. A recommendation for using this media and the TMC format while keeping within the goals of this document is provided in Section 8.

- 3.29 Registration Table Number**—A national assigned and maintained number (by the data registry process) used to identify any table of strings. There is a prearranged range for nationally reused tables, and also one for local tables. There is a range of numbers assigned to dynamic tables, which are never registered.
- 3.30 Resorting a Table**—Tables may be extended by adding new values at the end. This process provides a means to add, but not to remove, entries. Periodically entries may be *re-sorted* as well. At this time entries may be removed or added. Resorting may change the mapping of index value to entries. Typically resorting results in alpha ordering; however this is not a requirement (and is typically of more value to the ISP in determining strings than to the end user re-expanding them). The revision bits allow the end user to determine if a table has been extended and/or resorted.
- 3.31 Revision Bits**—An 8 bit version control value used to detect extending or re-sorting a table. The lower 4 bits (providing 16 values) are incremented whenever a table is extended. The upper four bits are incremented (and the lower four are set to zero) whenever a table is resorted.
- 3.32 Search Order**—The order in which tables are processed to expand an index value. The search order is critical for cases where overlapping is used to insure that the proper table entries are used. The search order is implied by the order in which tables are added to the header sections, with the search following the order in which they appear.
- 3.33 Simple Text**—Text found in a string which is composed of basic characters. A string is made up of a sequence of one or more table index entries (typically preceded by an expansion code for each table) and surrounded by simple text as needed. Simple text can be viewed as the “glue” of sentences which did not get compressed into a table entry. Note that if the character set being used is modified ASCII, then the simple text may include the various two-letter compressions in it. Also called “raw text” at times.
- 3.34 Sparse Table**—A table where the index values are discontinuous and therefore must be explicitly sent with the table entries. Sparse tables allow the preservation of historical encoding rules at the expense of being able to automate the creation and maintenance of the tables and result in a larger size when transferred.
- 3.35 Static Table**—A table which does not change or only changes its contents very rarely. As contrasted with dynamic tables which may change frequently while in use. Static tables may convey their changes over the same media as the messages using them, but many media do not (or cannot) provide this ability. In many deployments it is expected that static tables which cannot be sent over the media will be obtainable by other means. For example, the tables used in an RDS-TMC system might be obtainable over an Internet site. Often static table are re-expanded into forms other than text (such as a prerecorded voice). The prior agreement that the index value representing the expansion does change is vital to ensuring that the alternative media will continue to work.
- 3.36 Strings**—A sequence of (multi-sized) data values, which can be converted to and from simple text. In order to do this the underlying character and/or string tables’ sets must be understood. The specific rules and resulting messages to achieve this form the primary content of this document. In this document the term *string* is used to refer to the encoded source material while the pre and post coded source is referred to as *text*. Strings are almost always made up a sequence of simple text and of one or more table entries, although this is not a requirement.
- 3.37 Table**—A table consists of two parts: *the table entries* and the *table header*. In devices receiving and processing ATIS messages, tables are used to replace index values with strings in the process of converting strings back into text. This reverses the compression process that was performed at the ISP before the string was transmitted over the host media.
- 3.38 Table Entries**—The actual string values related to an index value of a table. The table entries follow the table header in the construction of a table.

- 3.39 Table Header**—A multi-octet section at the start of a table which provides information on the registration of the table, the table content (other tables), the character set used and the number and length of entries.
- 3.40 Text**—A sequence of one or more octet long data values which, taken together, expresses a language concept such as a *name* or a *sentence*. Every occurrence of text must be interpreted with respect to the underlying encoding of the character set used to express it. In “English” this is typically the ASCII character set. In other (often non-Latin) languages this may not be true and the visible expression of each octet may vary. In this document the term *text* is used to refer to the pre-encoded source material or the post-decoded source material, while the term *string* is used to refer to the text when it is encoded. Text is almost always in a single character set when displayed although this is not a requirement (consider mixed English, Japanese Kanji and Kana in the same sentence).
- 3.41 TMC**—Traffic Message Channel. An RDS term for the message format originally developed to occupy a Group 8a-type RDS message, employing a 37-bit long message. This was formatted with 16-bits being used for the LRMS portion and 11 bits being used for the index into a string table. Current TMC message formats being advanced in CEN extend this in various ways to overcome its limits. Consult CEN prEN 14819-1 for further structural details. The term TMC is at times also used to refer to the information-issuing source (more properly called the ISP in this document). In this case the acronym means Traffic Management Center.
- 3.42 Token**—The smallest “atomic” element parsed up in processing a stream of input which may be operated on by the algorithms handling the stream.
- 4. Overview**—This section provides a summary of the design goals of the document and a summary of what design goals were not implemented. Various features not supported in this document but found elsewhere in the ITS world deployments using strings indexes are discussed. How these needs are accommodated in US ATIS are presented.
- 4.1 Goals of the Document**—Replacing strings of text (or other data objects) with a short index value representing the string has a history of use longer than ITS itself. The earliest systems for traffic event reporting ever devised made liberal use of this technique due to limitations in transmission bandwidth. Even when bandwidth was not a primary concern, implementers have found the use of index values in place of the complete string to be a helpful technique. Over the years a number of such systems have been deployed. In practice most such implementations suffer from a lack of extendibility, poor or nonexistent methods of updating, a diversity of implementation details which make re-use hard, and a total lack of standardization of the content contained in the actual tables. SAE ATIS itself has used table methods in existing standards (notably SAE J2369 and J2354), extending the techniques to make maintenance aspects of the tables more automated. The goal of this document is to support a rich diversity of user needs while overcoming these previous shortcomings with a single unified method of handling such strings.

There are a number of design goals that this document seeks to meet. Some of these are quite distinct from each other. Among these are the following key points:

- a. Support the use of tables in systems ranging from very dynamic to very static. That is, both systems which frequently and automatically update and issue new tables to those systems which only use string tables in a fixed and never changing static nature.
- b. Develop common methods for table use across such systems so that the same end receiver software can be used with a variety of media types and supported message data content. That is, regardless of the source of the ATIS messages, the same application software should be reusable to operate on it. This is a basic design goal of all SAE standards in this area, not only because it allows deployed end user systems to adapt to changes but because it also allows ISPs to evolve without obsolescence in the installed base.
- c. Develop a world-class standard that fully supports both English and non-English users with multiple language support. In other words, the string and table methods must be suitable for use in any language, any character set, and allow re-expansion of the transmitted string in other languages than

the one used for original transmission (as well as other media types). This is vital for successful consumer products in world markets.

- d. Support “custom” expansion where the text may vary within a common language (“Road” versus “Rd”) or where no text at all used (for example, reverting an index into digital speech codes).
- e. Overcome previous shortcomings that “table only” phrase systems have encountered. Prior experience has shown that systems that rely *only* on pre-determined phrases are cumbersome and never quite fit the real life operational needs. Attempts to merge the phrase lists and the exceptions at the operator interface have produced complex encoding rules that also have shortcomings. It is a goal of this document to allow a completely free mixing of pre-established phrases and free text. And to place the burden of reducing this to string encodings with the construction of the messages themselves (and hence be automatic) rather than involve human intervention.
- f. Provide a common methodology and message set for sending tables over media that can support it. That is, by establishing a means to distribute such tables, the automatic updating and maintenance of them will be transparent to most user applications. This is a key requirement for sustainable deployment.
- g. Provide a common methodology for extending and updating the entries in tables. That is, by establishing a means to maintain tables at the sender’s end, systems can evolve (both in complexity and in content richness) without obsolescing the installed base of user equipment. This is a key requirement for sustainable deployment.
- h. Provide support for tables of varying length and varying content meeting both local and national needs for standardized phrases. That is, the methodology must support both short and long tables without undue overhead, as well as national tables of traffic phrases and also local variants of such phrases – all in a consistent and interoperable way.
- i. Provide tables meeting all of the above features and goals by an automatic means which removes the requirement for any overt cooperation between the table issuer and the table user. That is, the previous need for mutually agreed upon tables is eliminated allowing a wider and open market for equipment makers at both ends.
- j. Finally, provide the utmost compatibility with selected installation legacy systems such as RDS where a reasonable number of end user equipment devices now exist. That is, almost every feature of these systems has been incorporated into this document; therefore deployment of these systems and the phrases they are now using can continue with only slight changes to become compatible with the path of US ATIS standards.

These goals are met by the concept of associating a compressed string with a local table number assignment (an octet value) which in turn maps to one or more actual tables that provides the expansion values and property needs to return the string to a simple text form. Systems using simple static legacy systems can provide the additional table information in a separate mapping table message which, once checked by the receiver, can in general be ignored. More complex dynamic systems will use the abilities provided in this document to implement tables changing on an “as needed” basis obtainable by the end user. The flag values associated with each table affect the processing to expand the resulting string.

The *string*, is a concept used in this document to refer to the resulting compressed set of octets, that becomes a complex data element in the terminology of the ITS registry which is uniformly applied in ATIS messages throughout SAE standards. Raw, or simple text can be viewed as being a simplistic case within the rules of a string. Single index look-up systems can be viewed as being a string composed of only one index value. In all cases a common processing recovery scheme can be implemented at the receiving end able to take advantage of any compression present (or of the lack of any compression).¹ Interoperability over a wide range of implementations is thus insured in this area.

1. A further goal of this document is to allow implementers who do not wish to perform any compression or table look up methods to avoid the issue and place complete strings of simple text into the same fields. As will be seen in the latter sections, this is accommodated. This operating mode is not of much practical value in wireless applications, but in wireline distribution can be useful from the simplicity in which systems can initially be deployed. Such systems can move to more elegant compression when conditions warrant at a subsequent date. Because of the extendibility of the tables in this document, such a change will not effect compliant receivers already deployed.

4.2 Limits of This Approach—The technical approach taken by this document accommodates a wide range of flexibility in use. Prior text has discussed a variety of these issues. This section considers some alternative approaches in string handling which are not supported by the document and provides both the rationale and alternatives to accommodate these practices within the document.

4.2.1 EXPANDED STRING VARIABLES—From the earliest operational days, it has been found that the most interesting aspects of an event were never quite sufficiently contained in any predetermined list.¹ This has led to the combining of pre-determined phrases with insertion variables to create more flexibility. For example, the descriptive phrase “Delay due to \$X on roadway” where \$X represents a string to be supplied at the time of use.² The inserted string might be chosen from a list such as [*breakdown, overturned truck, animal, wild dog, pink dancing elephants*, etc.] or from free text as conditions warranted. Examples of this problem have led to endless committee humor and debate concerning the merits of whatever the unusual description proposed by the pundit was.

Technically this problem has been solved most often by creating textual phrases with an embedded insertion point like the example above (as an alternative, other systems will either use the list or default to a total “simple text” textual description field). This is a scheme very similar to the use of “shell variables” found in Unix, DOS and other popular computing systems. Typically, the variable to be supplied is appended somewhere later in the message (resulting in a variable size message as well). Sometimes the inserted value comes from a defined list, at other times free text is allowed.³ Usually there is a finite short limit to the number of inserted variables - which to some degree still restricts the descriptive string.

There is no analogous concept in this document, although it provides this feature in other ways. Because string indexes and simple text can be fully mixed and interchanged in any way that the message creator desires, the previous case can be viewed as a very limited subset of the features provided by this document. This is not to suggest that the use of predetermined strings is not useful and encouraged in the design of operators’ consoles. This type of CAD assistance results in more consistent phraseology, a reduction in spelling errors,⁴ and an increase in operator productivity.

Another variation of the shell variable concept has to do with *numerical values* added to the descriptive phrase. An example would be “... expect 10 minute delay” where the value of the delay would be provided in a subsequent variable as per the above discussion. In this document, the pre-defined number variable table provides the same functionality. This allows an integer range of 0 to 65535 and the addition of the ordinal post value (*st, nd, rd, th* in local conventions). Some systems allow words for logarithmic values, that is accommodated in this document by use of a local table entry (although many such values are also presented in the first 255 entries of the National Names table). Again, this document allows any ordering of the combination of words and indexes.

4.2.2 NESTED MULTIPLE LIST INDEXES (MULTIPLE TABLES)—Some phrase systems were developed with the operator creating the phrases in mind, rather than with the methodology of handling the resulting octet of data in mind. These systems produce many short phrases (each with some assigned numbering value) which can then be strung together to form sentences in the natural language supported. An example would be the below where the final descriptive text is selected in a cafeteria style of one item from each column. See Table 1.

-
1. One ISP has stated that this type of variation has been found to in fact be a selling point to clients, saying that they seem to enjoy the odd details of events. They feel that the diversity of free text increases the marketability of their data.
 2. This example is slightly simplistic. A more typical example would be “\$1 due to \$2 on \$3” where \$3 might be roadway, street, overpass, etc., providing a range of roadway terms and \$1 might be delay, slow down, etc. In this example, \$1 and \$3 would likely be from prede termined lists only, while \$2 is likely to require free text.
 3. This also often resulted in ineffective nested index value schemes, discussed further in the text.
 4. Note that the technical approach of separating the phrase creation (and operator dispatch functions) from that of the phrase encoding (a message transmission handling function) has the effect that should the operator misspell a word, it would still be sent without a problem. If this is routinely performed, a dynamic system would even table-ize such an entry.

TABLE 1—TYPICAL MULTI-PART PHRASE TABLE

Table #1	Table #2	Table #3
Delay due to	breakdown	on roadway
Congestion from	overturned truck	on overpass
Slowdown due to	animal	by exit
Accident from	wild dog	near access road
Extreme caution	pink dancing elephants	

The problem with the previous is threefold. First, the index values (and typically a further index to select which column) are not efficiently encoded with respect to any other text strings, which may be present. This results in a number of short (typically under 50) table entries needing to be maintained (as well as any other strings used in the same message family). Encoding problems can occur when any sub-list gains one too many members (say a 33rd item). Avoiding this by having large index ranges (say 256) results in extra wasted bits being sent. This document tends to promote the fewest number of such tables for efficiency reasons. In use, the above could be encoded as the following: [table #1][item #x] [table #2][item #y], [table #3][item #4]. In some cases, the use of table 1,2,3 keywords is removed - but this limits the ability to construct complex phrases or to insert free text. A secondary problem which can occur with this system is that the software built to decode indexing of this sort is normally built to each specific case and little of it can be reused by ATIS implementers wishing to operate on strings from another source.¹

It should be pointed out that to a limited degree the creation of multiple tables has been done in establishing the ITIS phrase list. In this case, the need for nationally agreed upon list of incident features which could be used in almost all of the existing ITS standards for messages outweighed the need for the most efficiently encoded possible list. In the case of ITIS some pre-assigned ranges for adding additional subtypes have been established within the 16-bit indexing range (and are maintained by the SAE as part of the registry process). This is in addition to the allowance for large local list entries.

Finally, the existing collection of such tables takes control away from the data issuers (who are likely to be combining ATIS information with others multiple sources) and prevents them from generating tables which are optimized to both the content and frequency of use. Frequency of use is a vital concept; many historically existing tables have a vast number of phrases, which are not in fact used very often. Some, when used, are used frequently. Consider the frequency of use for the term SNOW in summer versus in winter. Also, many times several phrases are most often used together (e.g., Delay due to breakdown) and it may make better compression sense to combine such entries in transmission.² Few of these optimizations can occur when a pre-existing or fixed family of short lists is forced to be used.

In this document, this type of encoding should be avoided, and it is preferred that a “flat” collection of all possible strings be developed and encoding into the smallest number of tables possible (and employing nested table entries rather than nested tables). Nested table entries should also be used to concatenate sub-strings commonly used together as well. If the media allows dynamic table distribution, then the resulting table should be fairly short and messages not frequently used appended to it when their periodic usage warrants. The remainder of the time, such unusual phrases should be transmitted as simple text in the final string.³

1. Typically, the decoding logic is a collection of nested case statements where within each case the known collection of possible indexes is handled. By this method some very efficient but complex logic can be developed. However, code rework is normally required when a list changes. By contrast, string decoding in this document proceeds in a linear (as opposed to nested) fashion expanding one token at a time as the string is processed.

2. Such grouping also promotes automatic sorting and filtering to occur in client receiver devices.

3. Systems using phrase tables to assist the operator (rather than encode them for transmission) can ignore this issue as they are presumed to produce full text messages. Such full text would be compressed and converted into tables by automatic equipment downstream, typically at the point of transmission by the ISP.

If it is determined that such prior index tables are to be used then the resulting table should be contrived such that the lower bits of the index represent the pre-existing index values and the upper bits represent the preexisting table values. By this method, implementers can preserve the vast majority of the encoding previously used, and any implementers in compliance with this document can also process such strings as being fully compliant. The resulting table is likely to be a “sparse” table with unused entries wherever the prior system did not fully use the indexing range of the tables it created. The expansion of these tables into multiple languages would proceed as it did before without change, however implementers should be aware that language word order issues will exist.

- 4.2.3 **MULTIPLE EVENT OR LOCATION REFERENCE MESSAGES**—In some traffic messages using index methods (such as found in Alert Plus) there are provisions to report on an event covering multiple locations in the encoding. Typically, this is done with a range of numerically related index values. This subject is not covered by this document and in general SAE ATIS constructs messages with each event separately. In some broadcast messages (notably those of SAE J2369) an index to a single descriptive string will be used in multiple messages so as to conserve bandwidth. In such a case the affected roadway segments may all point to a common textual string for information (which in that standard is carried in a common table when reused in this way).

This document provides methods to transmit such strings both in tables (for shared use) and in-line in a specific message for which it is intended. A later chapter provides specific recommendations for existing SAE standards. It does not offer any recommendations on the LRMS methods used to refer to the affected area.

- 4.2.4 **TIME VALUES**—In some traffic messages using index methods (such as found in RDS AlertC) there are provisions for reporting the expected time or duration of an event as part of the string. This subject is not covered by this document and it is expected that such information would not be sent in the string. Existing SAE ATIS message sets offer a variety of standardized means to encode the same information.
- 4.2.5 **BAUDOT AND OTHER CHARACTER SETS**—In some traffic messages' textual fields (such as found in GATS) there are provisions using various 6-bit and 5-bit Baudot character sets¹ which encode as a sub-set of the ASCII and Latin_1 character sets (some of these provide only the upper case alphabet and number characters). These character sets are not supported in this document. The original rationale for using these codes was that a percentage saving is obtained by saving 2-bits or 3 bits out of every 8 needed by ASCII² or Latin_1. A similar level of savings can be obtained by using the modified ASCII outlined in this document and even more when using the string compression methods. Further, these codes (and their regional variants such as the Murray code used in the UK) do not allow a uniform mapping to world character sets. In rejecting support for these older codes, SAE is following the same decisions made by the WWW and other standards bodies concerned in providing support for world languages.

One exception to this is the provision to allow the insertion of characters into a sequence of indexes as per 5.5 Table Expansion Values. Among the table exception tokens defined in Table 3 there is a token which has an effect similar to the “shift-lock” character of Baudot. When processed in the stream, the effect of this character is that the bytes following are interpreted in the default character set of the string until the end of string character (for that character set) is found.

- 4.2.6 **MULTIPLE LANGUAGE SUPPORT**—The system of string handling proposed in this document overcomes a limitation found in SAE J2369 as it currently exists today. While the effort of producing SAE J2369 was instrumental in advancing the automatic handling of such strings, and while it did support ASCII, Latin_1 and Unicode character sets, it did not allow the use of the efficient string compression methods in any character set but ASCII and modified ASCII. This was sufficient for the US market but not for a world class standard.

1. Formally called the ITA2 code (International Telegraph Alphabet Number two).

2. American National Standard Code for Information Interchange (ANSI X3.4). All of which are now superseded by the work of Unicode. [Ed note: where it is referred to as International Alphabet No. 5 (IA5) - This is where the term IA5String comes from in the ASN.1 syntax].