

# The Application of MatML to Contaminant Emissions Data

BY ED BEGLEY AND CYNTHIA HOWARD-REED

**S**tandard data formats are generally employed because they save time and money by promoting interoperability — that is, by facilitating data exchange and preserving information for re-use — and are available for a wide variety of business, scientific, and technological domains. MatML,<sup>1</sup> for example, is an extensible markup language (XML)<sup>2</sup> used for the management and exchange of materials property data.

During the October 2004 ASTM Committee D22 Conference on Indoor Emissions Testing, MatML was introduced as a format suitable for contaminant source emission rate data. This article describes how these data are composed into MatML documents and provides guidance to anyone who might want to use MatML to manage and exchange other forms of materials property data that are at present stored in legacy relational databases. Data on volatile organic compound (VOC) source emission rates are used to illustrate the method.

## BACKGROUND

Most indoor air quality, or IAQ, models that estimate building contaminant concentrations require the user to provide data related to contaminant source strengths and other contaminant transport mechanisms. Many of the required model inputs are available in the literature; however, these data have generally not been compiled in a readily accessible source, thereby requiring users to furnish their own model input parameters. To facilitate the IAQ modeling process and allow for assessment of data quality and completeness, the National Institute of Standards and Technology (NIST) is exploring two ways to store and manage published model input data.

These approaches may be viewed as complementary and reflective of the general need to link data aggregations with similar content but different storage formats. The first method uses a relational database management system, known as an RDBMS, to compile model input data into searchable databases. The second effort employs MatML to demonstrate how 1) the RDBMS data may be alternatively stored in a repository of searchable MatML documents; 2) MatML might be used as an exchange mechanism between disparate storage formats, especially in the context of providing a convenient way to access data via the Internet for any IAQ model; and 3) value may be added to the data by using an extensible stylesheet language transformation, or XSLT,<sup>3</sup> to provide online links to the U.S. Census Bureau's North American Industry Classification System (NAICS) and the NIST Chemistry WebBook via the NAICS and Chemical Abstract Services Number (CASN) specifications contained in the MatML documents.

## RDBMS FORMAT

A well-designed database stores a collection of information in a readily accessible format and also allows for the assessment of data quality, trends in observations, and data gaps. While there is an abundance of VOC source emission rate data available in the literature and other sources, no comprehensive database exists. There are, however, several abridged databases, two of which were used by NIST to provide the foundation for a good database design.

The first database, assembled by the National Research Council of Canada (NRC),<sup>4</sup> includes VOC emissions data from tests conducted in their Indoor Environment Program

laboratory chambers. This database represents a collection of data from a single testing facility and contains product manufacturer information, emissions testing conditions, chemical information, emission factors, emission profiles, and comments. Its design was used as a basis to build a database of emission rates from the published literature.

The second database, a collection of peer-reviewed source emission rates, was compiled by the Indoor Environment Management Branch of the U.S. Environmental Protection Agency (EPA).<sup>5</sup> The data were stored in a spreadsheet containing information regarding emission source classification, emission testing conditions, chemical information and analytical methods, emission factors, emission modeling parameters if available, and comments.

Drawing from the existing NRC and EPA database designs, as well as the recommended data reporting requirements from several testing guides,<sup>6,7,8,9</sup> a VOC source-emission rate database was created. The new design includes nine tables: emission rate category (Category), type of material within category (Types), literature reference (Reference), material properties (Material), contaminant properties (Property), environmental test conditions (Testcond), material test conditions (ETest), source model equation (Equation), and contaminant emission rate factors (Contaminant). The tables are linked using a "one-to-many" relationship (Figure 1). While populating this database with available records, a further consideration was highlighted, the need for a consistent data reporting format. A constrained format serves to ensure the availability of all parameters necessary for the purposes of IAQ modeling and other analyses.

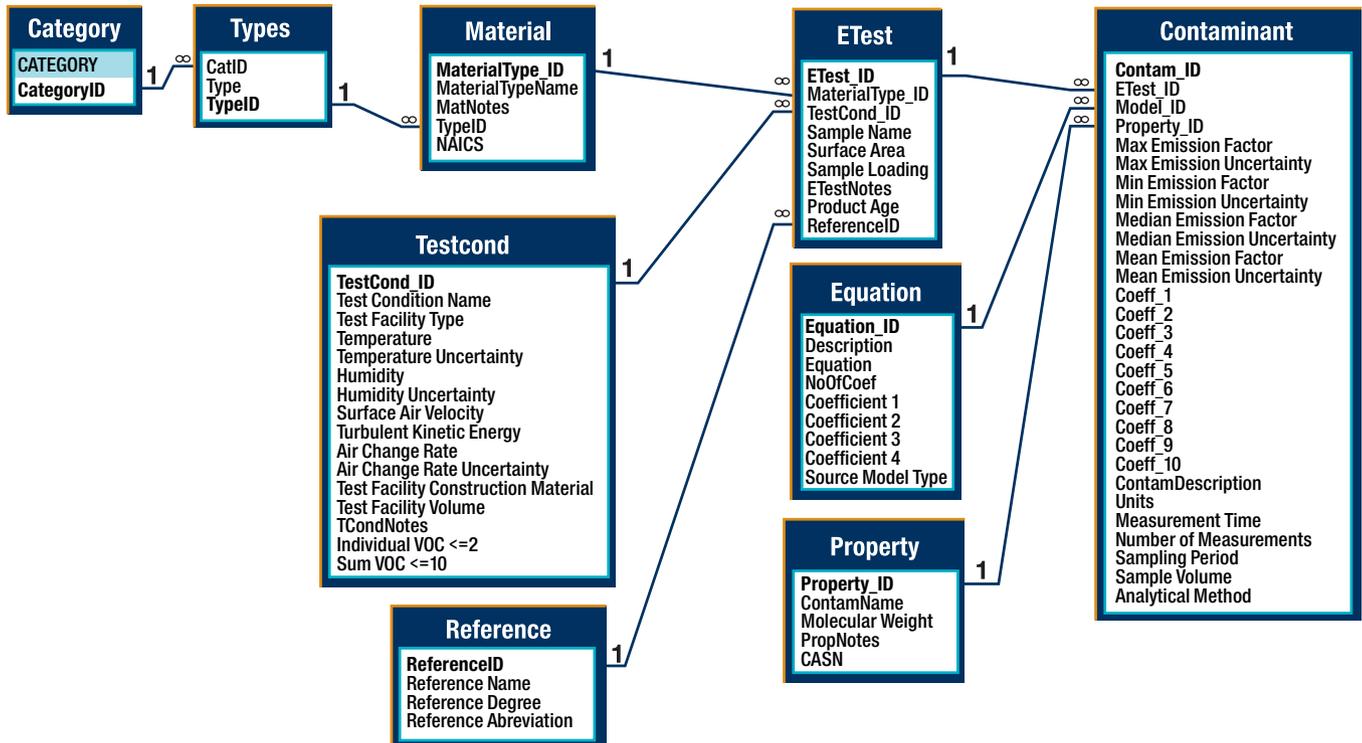


Figure 1 — Relationships for tables in VOC source emission rate database.

**MatML**

To maximize availability, emission rate data should be accessible on the Web and in a format that can be understood by any indoor air quality modeling program. Until recently, most data have been distributed on the Web using hypertext markup language (HTML), which only specifies how data are to be formatted for display and conveys no description of the data themselves. XML was developed in response to that situation so that communities could define their own domain-specific markup languages for data management and exchange and thereby permit efficient parsing and interpretation of those data via software. The XML developed to manage and exchange materials information is called MatML.

Initiated in October 1999 by NIST, the development of MatML was an international collaboration of industry, academia, government laboratories, and standards organizations. The effort leveraged a number of pertinent materials data resources including several ASTM

**Table 1 — ASTM guides leveraged by the MatML Development Effort.**

E 1308	Guide for Identification of Polymers in Computerized Material Property Databases
E 1309	Guide for Identification of Composite Materials in Computerized Material Property Databases
E 1313	Guide for the Development of Standard Data Records for Computerization of Material Property Data
E 1338	Guide for the Identification of Metals and Alloys in Computerized Material Property Databases
E 1339	Guide for Identification of Aluminum Alloys and Parts in Computerized Material Property Databases
E 1434	Guide for Development of Standard Data Records for Computerization of Mechanical Test Data for High-Modulus Fiber-Reinforced Composite Materials
E 1471	Guide for the Identification of Fibers, Fillers, and Core Materials in Computerized Material Property Databases
E 1484	Guide for Formatting and Use of Material and Chemical Property Data and Database Quality Indicators

guides (Table 1). Following its publication,<sup>10</sup> MatML was transferred to a commercial development group, which has been conducting trial applications and pursuing certification through the Organization for the Advancement of Structured Information Standards (OASIS).<sup>11</sup>

Since MatML was designed to

address any materials property data, it should be applicable to IAQ data as a means of managing and exchanging contaminant source emission rates. Because of its standard format, this representation of the data could be utilized by any IAQ computer model and is ideally suited for distribution on the Web.

**Table 2 — Mapping artifact. Note that units are treated implicitly in the database.**

Table	Table Field	Data
Category	category	FLOOR MATERIALS
Types	Type	Carpet- synthetic fiber
Reference	Reference name	Hodgson, A.T., Wooley, J.D., Daisey, J.M. (1993). Emissions of Volatile Organic Compounds from New Carpets Measured in a Large-Scale Environmental Chamber
Material	MaterialTypeName	Carpet 3_HOD93
	MatNotes	Residential; Propylene primary backing; Polyvinylchloride secondary backing; 100% nylon-6; textured loop construction; solution dyed; Scotchgard and antimicrobial fiber treatments; 44 x 46 cm tiles; glue down installation
	NAICS	31411
Property	ContamName	2,2,4-Trimethylpentane
	Molecular weight	114
	CASN	540-84-1
Testcond	Test Facility Type	Large chamber
	Temperature	23
	Temperature Uncertainty	1
	Humidity	50
	Humidity Uncertainty	5
	Surface Air Velocity	0.09
	Air change rate	1
	Air change rate uncertainty	0.1
	Test facility construction material	stainless steel
	Test facility volume	20
	TCondNotes	Range of values for all tests: T = 22.8 - 23.5; RH = 46.5 - 50.2; ACH = 0.98 - 1.0; air velocity = 6.5 - 9. Background concentrations of individual VOCs and Sum VOCs not measured.
		Individual VOC <= 2
	Sum VOC <= 10	False
ETest	Sample name	Carpet 3_HOD93
	Surface area	8.76
	Sample loading	0.44
	ETestNotes	Only one measurement for formaldehyde and acetaldehyde; at least two measurements for other chemicals; product age = 3 weeks after manufacture + 24 h sample time
	Product Age	528
Equation	Description	Steady-state emission rate
	Equation	S = G G = generation rate
	NoOfCoef	1
	Coefficient 1	G
	Source Model Type	Constant Coefficient
Contaminant	Mean emission factor	60
	Mean emission uncertainty	7.7
	Coeff_1	60
	Units	ug/m <sup>2</sup> /h
	Measurement time	24
	Number of Measurements	2
	Sample Volume	0.0013 - 0.01
	Analytical Method	GC/MS

**Table 3 — Mapping of RDBMS table fields onto MatML.**

Table	Table Field	MatML
Category	category	Material/BulkDetails/Class
Types	Type	Material/BulkDetails/Subclass
Reference	Reference name	MetaData/DataSourceDetails/Name
Material	MaterialTypeName	Material/BulkDetails/Name
	MatNotes	Material/BulkDetails/Notes
	NAICS	Material/BulkDetails/Specification
Property	ContamName	Material/ComponentDetails/Name
	Molecular weight	Material/ComponentDetails/Notes
	CASN	Material/ComponentDetails/Specification
Testcond	Test Facility Type	MetaData/TestConditionDetails/ParameterValue/Data
	Temperature	MetaData/TestConditionDetails/ParameterValue/Data
	Temperature Uncertainty	MetaData/TestConditionDetails/ParameterValue/Uncertainty/Value
	Humidity	MetaData/TestConditionDetails/ParameterValue/Data
	Humidity Uncertainty	MetaData/TestConditionDetails/ParameterValue/Uncertainty/Value
	Surface Air Velocity	MetaData/TestConditionDetails/ParameterValue/Data
	Air change rate	MetaData/TestConditionDetails/ParameterValue/Data
	Air change rate uncertainty	MetaData/TestConditionDetails/ParameterValue/Uncertainty/Value
	Test facility construction material	MetaData/TestConditionDetails/ParameterValue/Data
	Test facility volume	MetaData/TestConditionDetails/ParameterValue/Data
	TCondNotes	MetaData/TestConditionDetails/Notes
	Individual VOC <= 2	MetaData/TestConditionDetails/ParameterValue/Data
	Sum VOC <= 10	MetaData/TestConditionDetails/ParameterValue/Data
ETest	Sample name	MetaData/SpecimenDetails/Name
	Surface area	MetaData/SpecimenDetails/Notes
	Sample loading	MetaData/SpecimenDetails/Notes
	ETestNotes	MetaData/SpecimenDetails/Notes
	Product Age	MetaData/SpecimenDetails/Notes
Equation	Description	MetaData/MeasurementTechniqueDetails/Notes
	Equation	MetaData/MeasurementTechniqueDetails/Notes
	NoOfCoef	MetaData/MeasurementTechniqueDetails/Notes
	Coefficient 1	MetaData/MeasurementTechniqueDetails/Notes
	Source Model Type	MetaData/MeasurementTechniqueDetails/Notes
Contaminant	Mean emission factor	Material/ComponentDetails/PropertyData/Data
	Mean emission uncertainty	Material/ComponentDetails/PropertyData/Uncertainty
	Coeff_1	MetaData/MeasurementTechniqueDetails/Notes
	Units	Material/ComponentDetails/PropertyData/Units
		MetaData/MeasurementTechniqueDetails/Notes
	Measurement time	Material/ComponentDetails/PropertyData/ParameterValue/Data
	Number of Measurements	MetaData/SpecimenDetails/Notes
	Sample Volume	Material/ComponentDetails/PropertyData/ParameterValue/Data
Analytical Method	MetaData/MeasurementTechniqueDetails/Name	

### MAPPING THE RDBMS FORMAT INTO MatML

The mapping is illustrated using an artifact, a mapping table, and a listing of a MatML document.

#### Mapping Artifact

The artifact (Table 2), derived

from an inner join of the database tables using their ID fields, is an aid for understanding the mapping (Table 3) and the corresponding MatML document (Listing 1, see page 57). Each record of the join contained the data associated with a single contaminant. An alternative join associating all

contaminants for a single material could have been created but the single contaminant approach was chosen to illustrate the mapping in the clearest possible manner.

#### Mapping

Using MatML requires an un-



Figure 2 — XSLT rendering of MatML document in a Web browser.

Understanding of the language's vocabulary (tags, formally called elements) and grammar (order of tags). Since thorough discussion of MatML is beyond the scope of this article, interested readers will find helpful the documentation contained in the MatML schema.<sup>12</sup> The following high-level overview is provided as an aid for interpreting the mapping.

- ▶ The root element for a MatML document is named MatML\_Doc, which contains one or more Material elements and a Metadata element.
- ▶ The Material element contains a description of the material and its properties and is compartmentalized as follows:
  - ◆ BulkDetails describes the bulk material;
  - ◆ ComponentDetails describes each component of the bulk material, which, in this mapping, is a volatile organic compound;

- ◆ Graphs encodes two dimensional graphics, which is not needed in this mapping;
- ◆ Glossary contains definitions of terms found in the document, which also is not needed in this mapping.

The Metadata element contains information pertinent to a material encoded within the MatML document and is compartmentalized as follows:

- ▶ AuthorityDetails describes an authority;
- ▶ DataSourceDetails describes a data source;
- ▶ MeasurementTechniqueDetails describes a measurement technique;
- ▶ ParameterDetails describes a parameter;
- ▶ PropertyDetails describes a property;
- ▶ SourceDetails describes the source of a component;
- ▶ SpecimenDetails describes a

sample specimen;

- ▶ TestConditionsDetails describes the test conditions.

The mapping relates the table fields to the MatML tags. ID fields, fields containing no data, and other fields only useful to the database designers have not been mapped. The shorthand used for the MatML tags is read hierarchically. "Material/BulkDetails/Class," for example, indicates the Class element within the BulkDetails element within the Material element. All tags are contained within the MatML\_Doc element, which, for ease of reading, is not redundantly listed in the mapping.

#### Listing

The listing on the next page presents the MatML document resulting from the mapping using the data from the artifact. The document was generated by a program written to produce a repository of MatML documents from the inner

```

<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet href="epa2matml.xsl" type="text/xsl"?>
<MatML_Doc xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="http://www.matml.org/downloads/matml31.xsd">
  <Material>
    <BulkDetails>
      <Name>Carpet 3_HOD93</Name>
      <Class>FLOOR MATERIALS</Class>
      <Subclass>Carpet- synthetic fiber</Subclass>
      <Specification authority="NAICS">31411</Specification>
      <Notes>Residential; Propylene primary backing; Polyvinylchloride
secondary backing; 100% nylon-6; textured loop construction; solution
dyed; Scotchgard and antimicrobial fiber treatments; 44 x 46 cm tiles; glue
down installation</Notes>
    </BulkDetails>
    <ComponentDetails>
      <Name>2,2,4-Trimethylpentane</Name>
      <Specification authority="CASN">540-84-1</Specification>
      <PropertyData property="pr4" technique="m1" source="ds1" specimen="s1" test="t1">
        <Data format="float">60</Data>
        <Uncertainty>
          <Value format="float">7.7</Value>
          <Units name="ug/m^2/h">
            <Unit>
              <Name>ug</Name>
            </Unit>
            <Unit power="-2">
              <Name>m</Name>
            </Unit>
            <Unit power="-1">
              <Name>h</Name>
            </Unit>
          </Units>
        </Uncertainty>
        <ParameterValue parameter="p10" format="float">
          <Data>24</Data>
        </ParameterValue>
        <ParameterValue parameter="p12" format="float">
          <Data>0.0013 - 0.01</Data>
        </ParameterValue>
      </PropertyData>
      <Notes>Molecular weight is 114 amu. </Notes>
    </ComponentDetails>
  </Material>
  <Metadata>
    <AuthorityDetails id="NAICS">
      <Name>North American Industry Classification System</Name>
    </AuthorityDetails>
    <AuthorityDetails id="CASN">
      <Name>Chemical Abstract Services Number</Name>
    </AuthorityDetails>
    <DataSourceDetails id="ds1">
      <Name>Hodgson, A.T., Wooley, J.D., Daisey, J.M. (1993). Emissions of
Volatile Organic Compounds from New Carpets Measured in a Large-
Scale Environmental Chamber</Name>
      <Notes>Data were abstracted from the publication.</Notes>
    </DataSourceDetails>
    <MeasurementTechniqueDetails id="m1">
      <Name>GC/MS</Name>
      <Notes>Model description is Steady-state emission rate; the model
equation is S = G where G = generation rate; 1 coefficient(s), coefficient 1
(G) is 60 (ug/m^2/h); and the source model type is Constant
Coefficient</Notes>
    </MeasurementTechniqueDetails> ...
  </Metadata>
</MatML_Doc>

```

**Listing — Partial illustration of MatML document derived from mapping artifact. To download a document containing the full listing, go to [www.astm.org/SNEWS/OCTOBER\\_2005/begreelist\\_oct05.html](http://www.astm.org/SNEWS/OCTOBER_2005/begreelist_oct05.html) or contact Ed Begley at NIST, 100 Bureau Dr., Stop 8630, Gaithersburg, MD 20899-8630 (phone: 301/975-6118).**

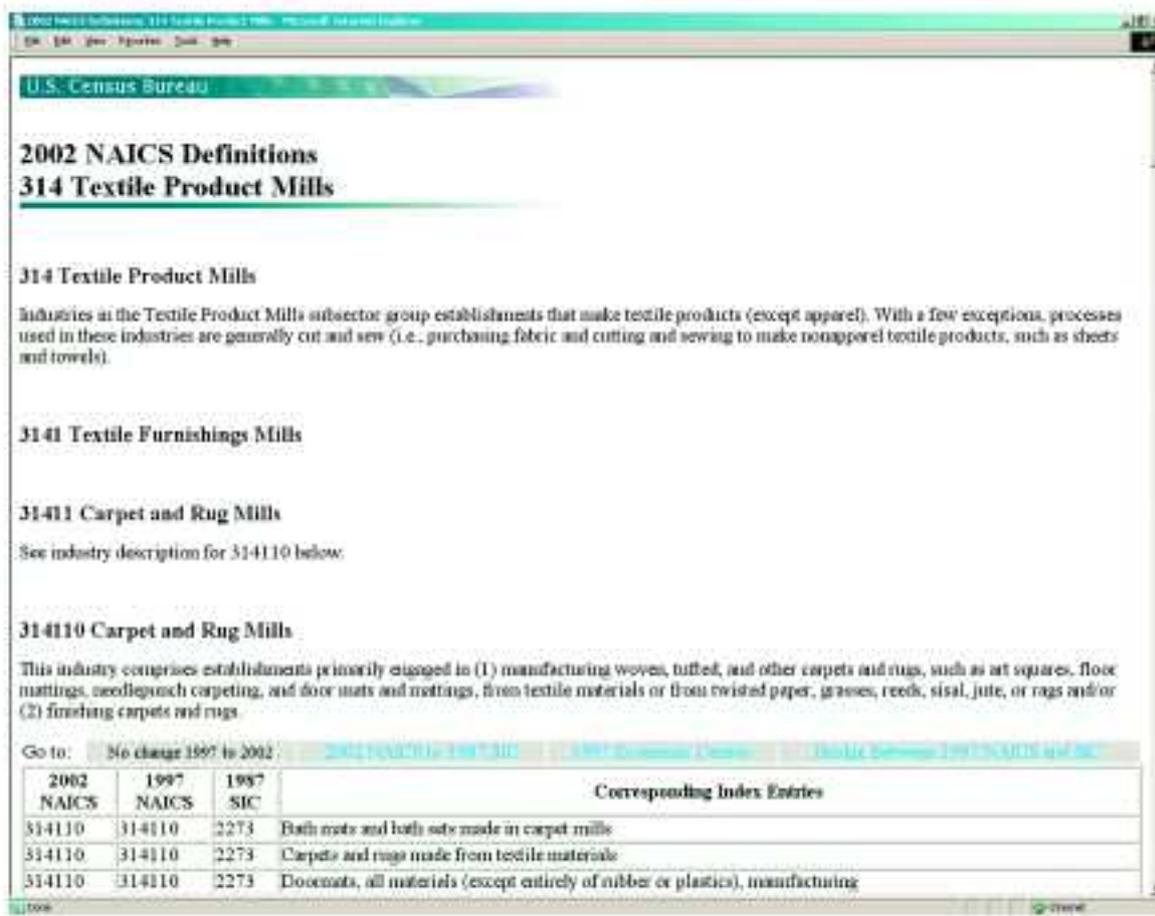


Figure 3 — Link from MatML document to U.S. Census Bureau using NAICS specification.

join of the database tables using the mapping's rules.

#### ADDING VALUE

As with all extensible markup languages, MatML separates information content from its display format. This feature essentially renders the data "future-proof," since documents are written using simple ASCII text rather than a proprietary encoding technique. Conversion to whatever data storage formats arise in the future, therefore, becomes a straightforward exercise. Moreover, MatML documents may be processed according to need, such as extracting specific data and formatting them for import into IAQ computer modeling software, formatting the data for publication in e-journals, or formatting the data for Web browser display.

As an illustration (Figure 2), the data contained in the listing were rendered for browser display using

**“The ultimate goal of NIST’s efforts to store VOC emission rate data is to help promote the standardization needed for consistency and reliability in reporting, accessing, and manipulating these data.”**

an XSLT with processing rules for also adding value to the original data; the NAICS and CASN specifications were identified and appropriate hyperlinks were created within the displayed document to the U.S. Census Bureau’s NAICS (Figure 3)

and NIST Chemistry WebBook (Figure 4), respectively.

#### CONCLUSION

The ultimate goal of NIST’s efforts to store VOC emission rate data is to help promote the standardization needed for consistency and reliability in reporting, accessing, and manipulating these data. MatML provides a suitable format for data management and exchange, especially on the Web, and associated XML technologies offer powerful tools for streamlining data access by any IAQ model. In the near future, user-oriented applications such as an editor will be developed that will simplify the preparation of MatML documents and lead to the compilation of searchable document repositories. These repositories, in turn, may provide broad and easy accessibility to pertinent materials information for communities with a specific interest, including

**Pentane, 2,2,4-trimethyl-**

- Formula: C<sub>10</sub>H<sub>22</sub>
- Molecular weight: 142.28
- IUPAC International Chemical Identifier:
  - InChI=1/C10H22/c1-7(2)6-8(3,4)5/67H, 8H2, 1-5H2
  - [Download the identifier as a file.](#)
- CAS Registry Number: 540-84-1
- Chemical structure:

This structure is also available as a [2d Mol file](#) or as a [compressed 3d Mol file](#).

- Other names: Isooctane; Isobutyltrimethylmethane; 2,2,4-Trimethylpentane; (CH<sub>3</sub>)<sub>2</sub>CHCH<sub>2</sub>C(CH<sub>3</sub>)<sub>3</sub>; Isobutyltrimethylmethane; UN 1262
- Information on this page:
  - [Notes / Error Report](#)
- Other data available:
  - [Gas phase thermodynamic data](#)
  - [Condensed phase thermodynamic data](#)
  - [Phase change data](#)
  - [Reaction thermodynamic data](#)
  - [Henry's Law data](#)
  - [Gas phase ionization energies data](#)
  - [IR Spectrum](#)
  - [Mass Spectrum](#)
  - [Gas Chromatography](#)
- Data at other NIST sites:
  - [Get phase kinetics \(in kinetics web site\)](#)
- Options:
  - [Switch to calcsite-based mode](#)

**Notes / Error Report**

Figure 4 — Link from MatML document to NIST Chemistry WebBook using CASN.

indoor air quality and material selection based on a range of environmental impacts. //

#### References

- 1 <http://www.matml.org>
- 2 <http://www.w3.org/XML>
- 3 <http://www.w3.org/Style/XSL>
- 4 Zhang, J.S.; Shaw, C.Y.; Sander, D.; Zhu, J.P.; Huang, Y. *MEDB-IAQ: A Material Emission Database and Single-Zone IAQ Simulation Program – A Tool for Building Designers, Engineers and Managers*. National Research Council Canada. 1999.
- 5 U.S. EPA. *Sources of Indoor Air Emissions*. U.S. Environmental Protection Agency. 1999.
- 6 ASTM. *Standard Practice for Full-Scale Chamber Determination of Volatile Organic Emissions from Indoor Materials/Products. D 6670-01*. American Society for Testing and Materials. 2001.
- 7 ASTM. *Standard Guide for Small-Scale Environmental Chamber Determinations of Organic*

*Emissions from Indoor Materials/Products. D 5116-97*. American Society for Testing and Materials. 1997.

- 8 European Communities. *European Concerted Action Indoor Air Quality & Its Impact on Man (EUR 13593), Guideline for the Characterization of Volatile Organic Compounds Emitted from Indoor Materials and Products Using Small Test Chambers. Report No. 8*. COST Project 613.

Luxembourg: Office for Publications of the European Communities, 1991.

- 9 Matthews, T.G. *Atmospheric Environment*. 1987, 21, 321 – 329.
- 10 Begley, E.F. (2003). "MatML Version 3.0 Schema," *NISTIR 6939*, National Institute of Standards and Technology, Gaithersburg, MD.
- 11 <http://www.oasis-open.org/committees/materials/charter.php>
- 12 <http://www.matml.org/downloads/matml31.xsd>

**ED BEGLEY** is a computer scientist in the Building and Fire Research Laboratory at NIST. His research has included extensive work in materials informatics and from 1999 to 2002 he was the project leader of the international effort to develop MatML, an extensible markup language for the management and exchange of materials property data. For more information please see [www2.bfrl.nist.gov/profiles/profiles.asp?lastname=begley](http://www2.bfrl.nist.gov/profiles/profiles.asp?lastname=begley).

**CYNTHIA HOWARD-REED** is an environmental engineer in the Indoor Air Quality and Ventilation Group at NIST. Before coming to NIST, she completed a post-doc with the National Exposure Research Laboratory of the U.S. Environmental Protection Agency. Her research interests include experimental and computational studies of contaminant transport in residential buildings.