

Creating structured documents with DITA and Teamcenter's content management solution

SIEMENS

White Paper

Publishing groups by necessity develop guidelines, processes and a common architecture to author and publish documents. Traditionally, this has led to detailed content structures, which impose significant overhead on document contributors. An alternative is now available called the Darwin Information Typing Architecture (DITA). This alternative enables publication groups to rapidly implement XML and content management, as well as dramatically improve the way source content is managed across a product document's lifecycle. With this in mind, Siemens PLM Software provides DITA support for the Teamcenter® suite's content management applications.

Contents

- Executive summary3**
- Introducing DITA.....4**
 - DITA concepts4
 - DITA structure5
- Using Teamcenter to create and manage DITA documents7**
 - Using DITA maps7
 - Creating DITA specializations in Teamcenter’s content management module 10
 - Publishing DITA documents from Teamcenter’s content management module 10
- DITA benefits11**
- Conclusion.....12**
- References.....13**

Executive summary

Publishing groups by necessity develop common guidelines and processes, as well as a common architecture, for their contributors. Organizations that deal with the interchange of content have two options.

One option enables them to build a very detailed, complex document type definition (DTD) that takes into account all possibilities and requirements for all of the different content and organizations involved in developing product information. S1000D is an example of this type of DTD. S1000D was developed to support the creation and publication of complex systems documentation, as well as the interchange of this documentation between organizations and content management systems. Unfortunately, this solution comes with a price. It is highly rigid, very large, and imposes significant overhead on authors and systems, especially if an organization directly uses it for authoring. In addition, changes to the DTD require careful monitoring and maintenance overhead because of the number of people impacted and the level of detail involved.

The second option is to use a more generic DTD that can be specialized to support specific functionality for different departments. The ability to create and interchange content in this instance is managed at a

higher level. Similarly, departmental requirements for functionality and content are handled using specializations specific to each group. Unlike the first option, this approach facilitates extensibility without as much impact to the different groups involved. This is because changes can be handled internally, without significant impact to the generic, high-level architectures involved. The Darwin Information Typing Architecture (DITA) provides this extensibility and functionality.

DITA documents are composed of self-contained topics. These topics are available for re-use and referencing, as are traditional content objects. However, because they are self-contained, there are changes involved in the way technical writers create specific portions of a document and compile the document; these considerations require adaptation from traditional document organization and writing styles.

Teamcenter's content management solution combines DITA with Teamcenter's ability to manage the relationships between engineering source materials and structured content. The resulting solution enables publication groups to quickly and easily migrate from traditional publishing methods to structured authoring.

Introducing DITA

From the beginning, the Darwin Information Typing Architecture (DITA) was developed to be more than “just another DTD” document type definition. Because optimization of the content development process is directly related to a company's specific product/service line and its internal document development processes, the number of universal DTDs is very small. However, the efficiencies that can be realized by publications groups have a direct relationship to the amount of DTD and schema customizations they can implement to support their specific needs.

By the same token, the delivery formats and systems used to deliver that information must be customized to a company's DTDs and specific customer-base, product and training needs. These requirements affect everything from writing methods to content management schemes and authoring and publishing application choices.

IBM had dealt with the issues that resulted from the implementation of a detailed, custom DTD that supported multiple departments. This effort focused on the second content interchange option, and DITA was the result. This standard became more than just a DTD or schema. It included a number of concepts and an architecture that provided specialization for department or targeted user functionality. It also provided the ability to create and manage topics for re-use that was directly in line with typical subject matter expertise document development used by many technical publications groups.

After testing and further development, IBM recognized the potential advantages of the standard for other organizations and proposed DITA to the Organization of the Advancement of Structured Information Standards (OASIS) in March, 2004. The OASIS DITA Technical Committee approved DITA as an official OASIS standard in June, 2005.

DITA concepts

DITA is more than another markup language defined in XML. To understand DITA's value, it is useful to have a look at the meaning and derivation of its name.

D The letter D was chosen for the British naturalist Charles Robert Darwin. Darwin achieved lasting fame for his “evolution theory,” which states that species develop by inheriting their experiences from previous generations. This focus on inheritance made Charles Darwin the eponym for the Darwin Information Typing Architecture.

DITA enables users to create new markup elements that are inherited from the basic DITA elements. This inheritance facilitates individual DITA applications for specific needs, called “specializations.” A new element is a specialized variant of its basic element.

IT Information Typing is an essential principle in DITA that facilitates modular structuring and functional marking of information.

In DITA, a complete document consists of many information modules, with each module containing content specific to one topic. These information modules are called “topics” in DITA.

A The letter A stands for “architecture,” which signifies the collaboration of many different components in one complex system, such as the construction of a house or the development of software.

DITA includes recommendations and concepts regarding information organization, publishing and exchange.

DITA structure

A structured document consists of at least three different structure levels: the entire document, the information modules within the document and the vocabulary available in the information modules. DITA comes with one solution for each of these three structure levels.

DITA arranges information in unique, self-contained information modules called "topics". Each DITA topic should only contain information specific to the subject covered by that topic. Topics have no internal hierarchical nesting. Instead, elements and attributes that define or directly support the type of information provided in that topic are available to authors as they write.

Similar to many structured content solutions, DITA can be used to reduce the copy/paste methods of re-use (where content from one document or section is copied into another). When authors wanted to perform copy/paste operation using traditional documents, all copies of the same content has to be maintained in each document where the content was pasted; traditional documents needed this extra effort to account for future content changes and ensure their consistency and quality. With DITA, re-use occurs at a topic level. Information designers can re-use a topic in a new information model and the architecture will process it consistently in its new context. This allows the topic to be maintained in a single location. Changes will be displayed appropriate to their use in any DITA map or document.

To enable users to store and create more specific types of information in typed topics, DITA offers three further information types: concept, task and reference. Each of these three information types applies to a specific DITA topic type, with relationships to other topic types as diagrammed in Figure 1. These information types represent specializations of topic and provide specific structures relative to procedural, conceptual and reference material. They also ensure proper contextual usage of the topic in multiple documents or DITA maps.

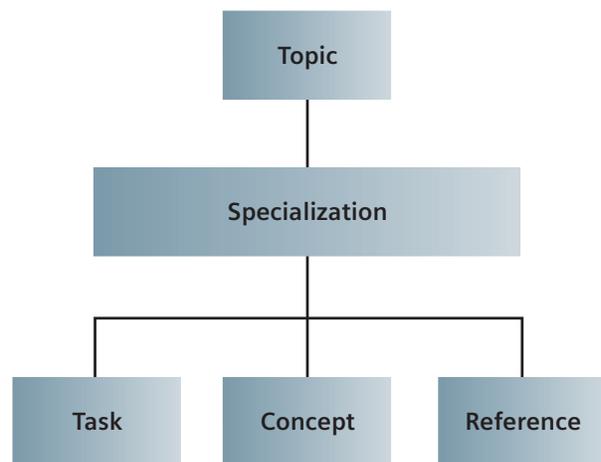


Figure 1. The four DITA information types.

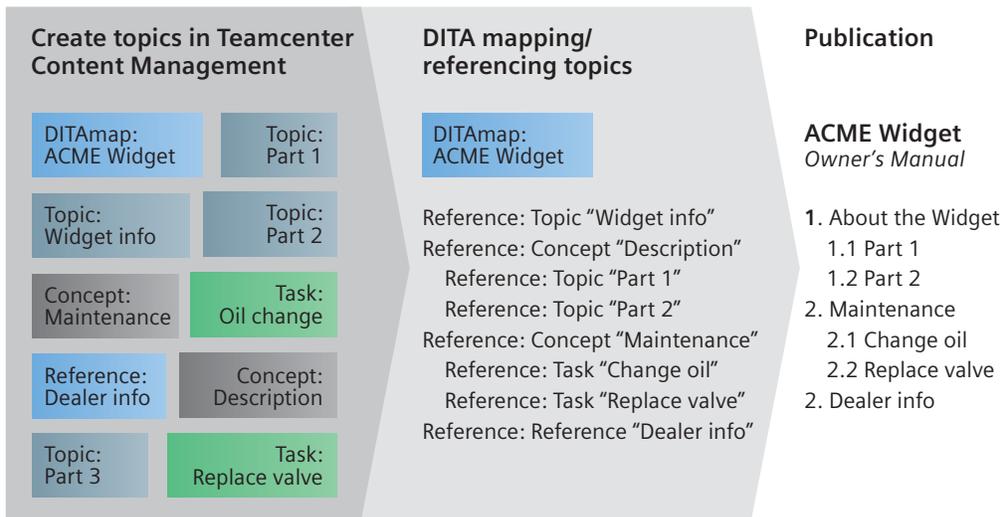


Figure 2. Principle of nested topic references in a DITA map.

DITA maps For structured documents that consist of nested information containers such as chapters and sections, DITA enables users to nest references to existing DITA topics in a DITA map. Figure 2 clarifies the principle of nested topic references in a DITA map.

DITA specialization DITA specialization allows customization of the existing DITA markup language for individual purposes without the need to change the DITA standard. As a result, it is possible to adopt DITA – which was originally designed for the software development sector – in manufacturing, aerospace and other industrial sectors by creating new elements via DITA specialization. The key to creating new elements specific to your documentation needs is the "class" attribute.

Any DITA element can be extended into a new element whose identifier gets added to the class attribute through its DTD. In this way, a new element

is always associated to its base element and to any element in its specialization hierarchy. Topics can be specialized from the generic topic into new information types and the new information types can be extended into more specific information structures. A maintenance task, training procedure and the steps involved in an operation sequence can all be extended from the generic topic.

In the same way, by the substitution or effectivity of domain-specific vocabularies, the element vocabulary within a topic can be specialized or extended specific to the information domain for which the topics are developed. Specialized vocabularies might include programming languages, mechanical terms or operation codes for a system, which can be hidden or displayed based on the user roles or context of the document.

Using Teamcenter to create and manage DITA documents

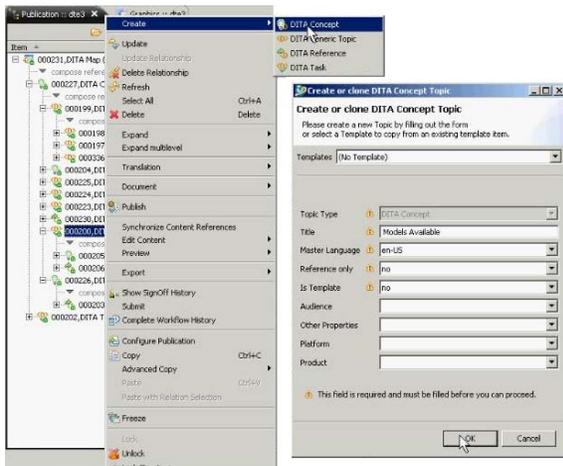


Figure 3. Creating DITA topics using Teamcenter's content management solution.

Teamcenter's content management solution enables you to create and manage topics and their relationships provided in the DITA architecture and XML schema or DTD. Topics can be created in both the Teamcenter content management rich client or by working directly in the integrated authoring applications.

Teamcenter's content management DITA module provides the four standard DITA information types (generic topic, concept, task and reference) out of the box. You can also create a DITA composite document that contains DITA topics of any information type.

Teamcenter's content management module provides two ways to re-use an existing DITA topic within other DITA topics: referencing or nesting content.

Teamcenter's content management module also enables you to assemble topics using topic references in DITA maps. Because nesting topic references within different information types is allowed in a DITA map, it is also possible to drop a topic onto a topic of a different information type, which results in a reference relationship between them.

It is possible to nest existing topics in Teamcenter by dragging a topic and dropping it onto another topic in Teamcenter's content management rich client. You

can nest DITA topics from both a generic topic type and from specialized topics to topics of the same information type. The result is the inclusion of the content from the nested topic in the parent topic. If you were to create a reference to the parent topic from a map, the nested topic would also be included in that map.

Using DITA maps

Creating DITA maps Teamcenter's content management module provides two different publication-types for DITA maps: editable and dynamic. Both publication types can easily be created in Teamcenter from the rich client with simple menu and dialog options.

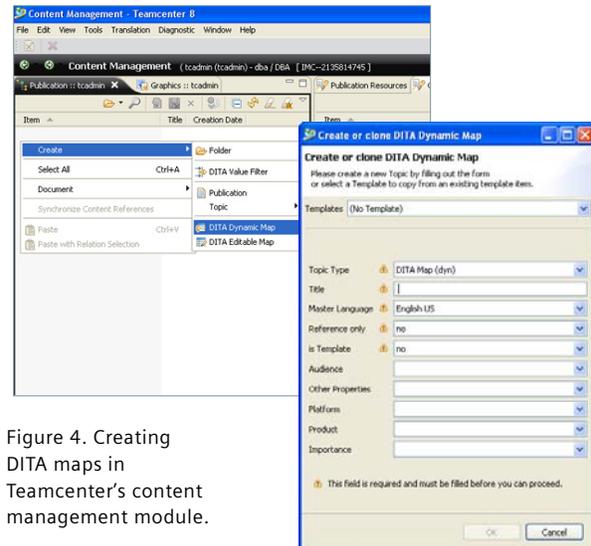


Figure 4. Creating DITA maps in Teamcenter's content management module.

The difference between the two map types is apparent when creating the content of the particular DITA map.

Editable DITA maps Because DITA maps are XML files, you can edit the content of an editable DITA map in any XML editor. Teamcenter's content management module delivers the DITA map document type definition (DTD) automatically to the XML editor, where the author creates valid DITA map

contents. Siemens PLM Software recommends the use of an XML editor with DITA support (such as XMetaL Author Enterprise Edition) because these applications will assist the author with DITA-specific WYSIWYG styling and timesaving features created especially for the creation of DITA maps.

To support the author in creating the XML elements for the topic references inside a DITA map,

Teamcenter enables authors to select DITA concept task reference and generic topics in the client and generate XML topic references into the clipboard. These XML topic references are pasted into a DITA map opened in an XML editor via the ordinary paste function.

Teamcenter's content management solution generates topic references for the selected topics.

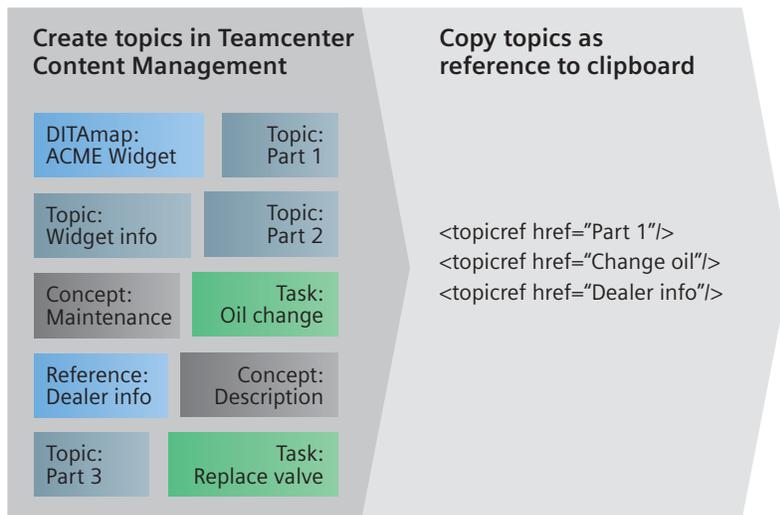


Figure 5. Copying Teamcenter content management topics as topic references.

Dynamic DITA maps Teamcenter also provides a way to add topic references to DITA maps directly in the Teamcenter content management client by dragging existing topics and dropping them onto a dynamic DITA map publication. Teamcenter then automatically inserts references between the topics and the dynamic DITA map.

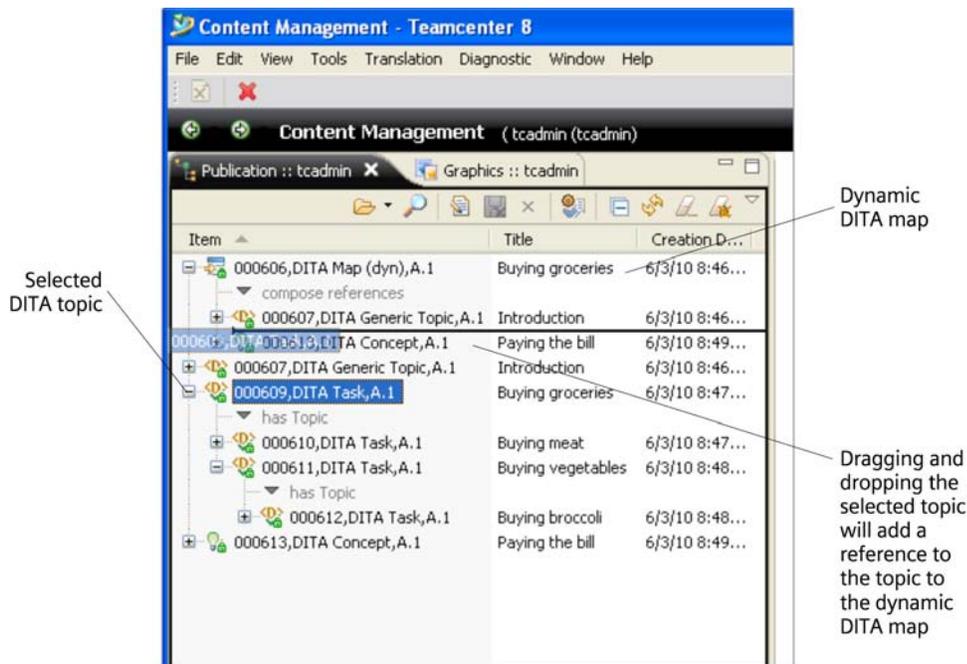
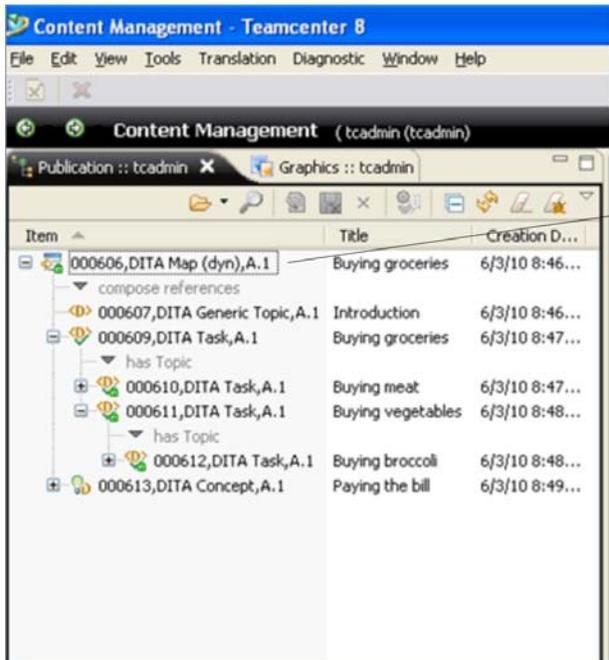


Figure 6. Dropping a topic onto a dynamic DITA map.

You can use Teamcenter's content management module to expand and collapse related topics and publications while browsing the structure of a dynamic DITA map. If a topic with references to other topics is dropped into a dynamic DITA map, the DITA map will get just one reference – but it will use the entire underlying structure of the dropped topic. In Figure 6, the drag and drop would cause one additional reference to the dynamic DITA map, but the whole underlying structure of the selected task would be used in the DITA map.



Structure of the dynamic DITA map after the drag and drop illustrated in Figure 6

Figure 7. Structure expand of a dynamic DITA map.

Creating DITA specializations in Teamcenter's content management module

Users can easily register new DITA specializations in Teamcenter's content management module. This task requires several basic Teamcenter content management administration skills (such as creating new document types and attaching schemas) and needs to be done in the Teamcenter content management administrator client. See the "Teamcenter Content Management Administration Manual" for further information.

Publishing DITA documents from Teamcenter's content management module

Teamcenter's content management module integrates the DITA Open Toolkit (available at <http://ditaot.sourceforge.net/>) and uses it for publishing DITA documents. Any further developments – e.g., in the

standard XSL styles – will be done on top of the Open Toolkit without changing the application. This guarantees full Open Toolkit support and an open interface to every DITA change or enhancement performed in the DITA Open Toolkit.

The DITA standard includes a solution for conditional publishing, also known as "effectivity" or "filtering". You can exclude specific contents in a DITA publication via "ditaval" files. Using Teamcenter's content management module, you can create, edit and manage ditaval filters and use them in publishing a DITA document. Teamcenter supports the standard attributes – including platform, product, audience, and otherprops as well as importance, rev and status – available in DITA.

Each ditaval filter object can be related to both editable and dynamic DITA maps in Teamcenter's content management module via drag and drop. While publishing a DITA document with related ditaval filter objects, Teamcenter composes the topics according to the DITA ditaval to determine what conditional text will be included in the document.

DITA benefits

The DITA architecture and its standard provide numerous benefits, especially to organizations that are migrating to structured content authoring and publishing for the first time. They also include various recommendations and best practices for implementation.

Traditionally, the development of DTDs and schemas involved information modeling, centered on existing and legacy writing and publishing processes and tools. This approach often reflects a "way we've always done it" attitude. Based on previous experience with complex DTDs, IBM indicated that "The ideal for future extensibility in DTDs for technical information (or any information that is continually exploited at the leading edge of technology) is to build the fewest possible presumptions about the top-down processing system into the design of the DTD."¹

Instead of requiring users to learn large, complex data structures, DITA's topic-centric design enables users to quickly become familiar with only those elements that are necessary in the topic types they need for their jobs. This aspect of the standard has benefits from training and implementation through the publishing and management phases, as supporting processes and transforms or templates need to only deal with structures specific to their needs.

Because the standard begins with a set of general structures, whose characteristics are inherited by any specializations, any specialized topic type can be treated as a more generic type for common processing.

The topic-oriented nature of DITA lends itself well to re-use and to both print and online delivery; this is one of the reasons it is being rapidly adopted for software product documentation. Keep in mind that moving from a linear document model to a topic-oriented model is a significant adjustment. Organizations that traditionally produce long linear documents and do not have a compelling reason to adopt a topic-oriented model would generally find it easier to use non-topic-oriented XML document types. If your customers require you to use an industry-specific DTD, such as S1000D, it would be easier to modify that standard to meet a technical publication group's writing requirements than to modify DITA and provide transforms to convert from DITA to the standard structure; these options should be analyzed carefully before implementing either option.

If you are already using XML and DTDs, and they meet your requirements, there may be no increased benefit in moving to DITA. If you are in a position of selecting DocBook versus DITA as your first structured authoring system, you will find greater tool support and momentum behind DITA than other recent standards, although almost all systems support traditional standards such as DocBook.

Conclusion

The DITA standard is rapidly gaining world-wide acceptance for implementing and exchanging XML-based content and supporting content management and publishing functions. Teamcenter's content management DITA module supports requirements for a DITA-compliant, complete CMS product suite. Teamcenter enables you to rapidly implement a productive solution for structured technical publications management based on a standard template. Users are provided with the ability to create, update, organize and version DITA topics without the need to open an authoring or editing application.

Teamcenter's content management solution will be maintained along with the OASIS roadmap of future DITA releases, ensuring ongoing conformance with future versions of the specification, as well as facilitating rapid access to new DITA capabilities. Inclusion of the DITA concepts and architecture directly within Teamcenter's content management rich client provides easy-to-use alternatives to creating and managing DITA maps in XML editors.

DITA is a comparatively easy-to-learn, significantly easier to implement, structured authoring alternative to cumbersome industry standards that focus on interchange. Siemens PLM Software combines DITA with Teamcenter's ability to relate product information topics directly to the engineering source product data, enabling users to store, version and manage those topics directly in the engineering product life-cycle management environment. This gives you an unprecedented opportunity to improve time-to-market and reduce the risk of inaccurate or outdated product information.

If interchange specifications are not required for specific product documentation, DITA is a strong candidate, especially for organizations that are just beginning to implement an XML writing and publication environment. For those environments that require interchange, but do not require compliance with the entire interchange specification, DITA with specializations may still be applicable, with transformations to the industry standard for interchanging product information.

References

1. "Introduction to the Darwin Information Typing Architecture: Toward portable technical information". March 2001; Don Day, Michael Priestley, David Schell, of IBM Corporation.

About JustSystems

JustSystems is a leading global software provider with three decades of successful innovation in office productivity, information management, and consumer and enterprise software. With over 2,500 customers worldwide, the company is continuing a global expansion strategy based on its XMetaL content lifecycle solutions. JustSystems is one of the 2009 KMWorld 100 Companies that Matter in Knowledge Management, a 2009 EContent 100 member, and was recognized on the 2009 KMWorld Trend-Setting Product list for XMetaL. Major strategic partnerships include IBM, Siemens and EMC.

About Siemens PLM Software

Siemens PLM Software, a business unit of the Siemens Industry Automation Division, is a leading global provider of product lifecycle management (PLM) software and services with 6.7 million licensed seats and more than 69,500 customers worldwide. Headquartered in Plano, Texas, Siemens PLM Software works collaboratively with companies to deliver open solutions that help them turn more ideas into successful products. For more information on Siemens PLM Software products and services, visit www.siemens.com/plm.

Siemens PLM Software

Headquarters

Granite Park One
5800 Granite Parkway
Suite 600
Plano, TX 75024
USA
972 987 3000
Fax 972 987 3398

Americas

Granite Park One
5800 Granite Parkway
Suite 600
Plano, TX 75024
USA
800 498 5351
Fax 972 987 3398

Europe

3 Knoll Road
Camberley
Surrey GU15 3SY
United Kingdom
44 (0) 1276 702000
Fax 44 (0) 1276 702130

Asia-Pacific

Suites 6804-8, 68/F
Central Plaza
18 Harbour Road
WanChai
Hong Kong
852 2230 3333
Fax 852 2230 3210

www.siemens.com/plm

© 2010 Siemens Product Lifecycle Management Software Inc. All rights reserved. Siemens and the Siemens logo are registered trademarks of Siemens AG. D-Cubed, Femap, Geolus, GO PLM, I-deas, Insight, JT, NX, Parasolid, Solid Edge, Teamcenter, Tecnomatix and Velocity Series are trademarks or registered trademarks of Siemens Product Lifecycle Management Software Inc. or its subsidiaries in the United States and in other countries. All other logos, trademarks, registered trademarks or service marks used herein are the property of their respective holders.

X3 22632 12/10 C

JUST.
SYSTEMS

Solution
Partner

Industry

SIEMENS