Community-Driven Information Quality Standards: How IBM Developed and Implemented Standards for Information Quality

Bob Vitas

Abstract

Purpose: Standards for information quality can help content developers within a company or enterprise create high-quality, high-value content, as well as an excellent user experience for its clients. This article explains how the content development community at IBM created meaningful standards, as well as the metrics to track their impact, as part of a closed-loop information quality process.

Method: Instead of having one group dictate what the standards will be, the content developers at IBM worked together as a community to identify key requirements from internal and external sources, tested the standards with a set of key products, and then put the standards to use in increasingly larger numbers of products.

Results: A community-driven approach to information quality standards allowed the IBM content developers to create standards that were meaningful to a variety of teams, ensuring that key aspects of information quality were addressed throughout the corporation. The use of metrics to track implementation and compliance allowed the IBM community to see when the standards were working, and when the standards needed to be updated to meet the changing needs of their clients.

Conclusions: Implementing information quality standards is an admirable goal, but it should not be the end of a company's information quality journey. This should be considered a closed loop or wheel, with continual analysis of compliance data, client feedback, and other metrics driving continual improvements in information quality.

Keywords: standards, metrics, information quality, user experience, community

Practitioner's Takeaway

- Information quality is an admirable goal, but it should not be considered the end of a journey. Instead, it should be viewed as a closed loop or wheel, in which improvement never stops.
- Engaging the content development community in developing and maintaining information quality standards helps to ensure buy-in and understanding.

- The use of metrics to understand the quality of information can provide valuable insight into what kinds of standards are needed, when standards are having an impact, and when standards need to be altered or removed.

Community-Driven Information Quality Standards

Introduction

At IBM, we are developing our technical communications body of knowledge with a focus on content as a business asset. Figure 1 shows the results of a survey that we have on our Web site; the results consistently show that a high-quality technical information experience is important to customers. As described in "Telling the Right Story: Proving the Business Value of Content" (Riley, Ames, & Jones, 2013), we at IBM have data from 179 respondents (none of whom are IBM employees), surveyed from 2010 to 2013, which says that they rely on having high-quality technical content for the products they purchase, for everything from planning and installation to using and maintaining. These data show that customers look for content on configuration, troubleshooting, and administration. They expect the content they find to be accurate, they expect it to be role-appropriate, and they expect it to provide a consistent, high-quality information experience.

Figure 1. Results of IBM "Importance of High-Quality Technical Information" Survey
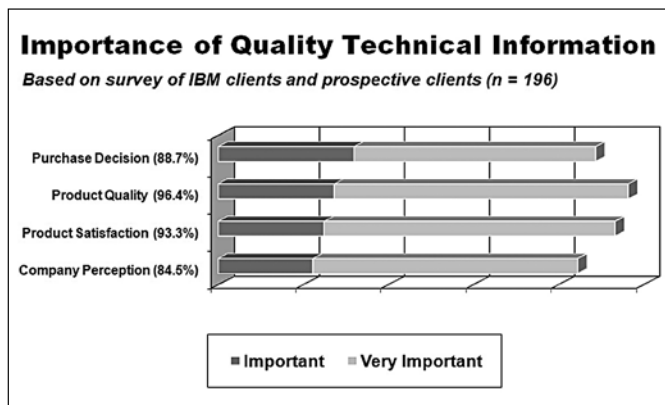
Given these types of requirements, corporations must find ways to ensure that their technical content meets certain criteria. These criteria come from internal sources—for example, corporate standards that all teams must comply with, guidelines that describe how to implement a standard, and the collateral and templates that help teams become compliant—as well as external sources—including customer requirements and industry standards. Having a consistent approach to meeting these criteria, regardless of their source, helps ensure that the customer's total information experience is the best that it can be.

Defining which of these criteria are most important can be a daunting task, especially considering that technical content can come from a variety of sources within a corporation. Most customers are familiar with "product documentation"—the user manuals, installation and configuration guides, and other content that is shipped with a software or hardware product. However, there is also a wealth of other content produced by product support teams, marketing and sales teams, and customer education materials, just to name a few. These various sources of content often have different customer expectations and delivery mechanisms, creating conflicts when applying criteria to the breadth of technical content that is available.

Customers looking for technical content for IBM's software and hardware products can find support content, educational materials, instructions that are specific to a given scenario or solution, and marketing and sales literature, all in addition to the product documentation. This means that focusing on the product documentation alone is only addressing one aspect of the content that is part of what we at IBM call the Total Information Experience. The Total Information Experience has to be consistent across the entire set of technical content, no matter how many content creators exist within the corporation. To bring some level of consistency, we at IBM needed to find a way to implement information quality standards for the Total Information Experience, regardless of who is creating the content.

There is a good deal of content in the technical communication literature about standards for writing and information quality. As an example, the Society for Technical Communication (n.d.) maintains an entire Web page on the various industry standards for writing, and the STC's archives contain many articles and contributions about developing content strategies and creating common practices for style and strategic goals. These sources are excellent starting points, but we at IBM needed something more suited to our particular needs, both internally and externally.

At IBM, we initially approached the requirement for information quality standards within the Information Development (ID) domain, which encompasses the product documentation. The ID community at IBM is mature, having evolved over the last ten years or so into a worldwide group of ID professionals. The journey toward having community-defined information quality

standards began by first defining what it meant to have a high-quality information experience, and then identifying the standards and supporting guidelines and collateral that should be used by content creators to meet this definition. Next, we put a compliance mechanism in place to track how well teams were able to meet the requirements of the standards, and we identified metrics and key performance indicators (KPIs) that allowed us to track our progress. The standards were reviewed regularly for applicability and currency so that our product documentation continues to evolve and improve.

However, rather than having one group dictate what the information quality standards should be, the Corporate User Technologies Team within IBM engaged the information development community in these activities, building the standards with a grass-roots approach that ensured everyone had a chance to voice their opinion. Using this community approach gave everyone a sense of ownership of the standards.

Defining a High-Quality Information Experience

The Corporate User Technologies Team at IBM determined that there were three key aspects of a high-quality information experience: high-quality technical content, a high-quality user experience (which includes the delivery mechanism or application, navigation systems, and the search technology, among other things), and high-value content.

Defining High-Quality Technical Content

Several years ago, IBM Press published *Developing Quality Technical Information: A Handbook for Writers and Editors* (Hargis et al., 2004), which outlined an approach to achieving quality information. The book identified the following aspects of high-quality content:

- **Accuracy**—Freedom from mistake or error; adherence to fact or truth
- **Clarity**—Freedom from ambiguity or obscurity; the presentation of information in such a way that users understand it the first time
- **Completeness**—The inclusion of all necessary parts—and only those parts

- **Concreteness**—The inclusion of appropriate examples, scenarios, similes, analogies, specific language, and graphics
- **Organization**—A coherent arrangement of parts that makes sense to the user
- **Retrievability**—The presentation of information in a way that enables users to find specific items quickly and easily
- **Style**—Correctness and appropriateness of writing conventions and of words and phrases
- **Task orientation**—A focus on helping users do tasks that are associated with a product or tool in relation to their jobs
- **Visual effectiveness**—Attractiveness and enhanced meaning of information through the use of layout, illustrations, color, typography, icons, and other graphical devices

At IBM, we continue to use these nine aspects of high-quality technical content to drive consistency across our Information Development community.

Defining a High-Quality User Experience

While it is important to have high-quality technical content, we at IBM believe it is equally important to have a high-quality user experience that provides the right content to the right person, at the right time, and in the right way. While there are many ways in which users consume technical content, some things are common enough to the overall user experience. We have learned over the years that the user experience should address (at a minimum) these key pain points:

- **Content must be easy to find**—In today's environment, there is technical content everywhere. Within a corporation as large as IBM, we have tens of millions of Web pages, providing all kinds of content about the thousands of software and hardware products that are in service. It is important that we make this content easy to search so that customers can quickly find the piece of content they need to solve a particular problem or question. Search results should provide enough detail about the content choices so that the user can narrow down the results and identify the content they need.

Community-Driven Information Quality Standards

- **Content must have a consistent look and feel**—Customers who use IBM's technical content expect it to be consistent in terms of quality, but also in terms of how it is presented. This is especially true in large enterprises such as IBM, where we have several different software and hardware brands that produce products that work together in solutions, but it is also true when comparing content for products within the same brand, or even content for different releases of the same product. If one piece of content presented to a user is different than another, the user experience can suffer. Navigation guides on a Web site might have different icons or placement on a page, forcing a customer to waste time figuring out how to get to the next piece of content. Content from the same company might be presented on a common Web site, but one piece may be viewable as an HTML file, while another might be presented as a PDF file. These are just two examples of how an inconsistent look and feel can impact the user's experience.

- **Content must be relevant to the person and task at hand**—Having perfectly accurate content is an admirable goal, but it is only useful if it meets the customer's needs at the time it is needed. For instance, a set of installation instructions might be highly detailed and accurate, but if a UNIX customer is presented with the installation instructions for a Windows operating system, it is of no use to the customer.

- **Content must be provided when and how a user needs it**—While the Internet has become a common place where users can find content, they don't always use a browser to view the content. The user may be a support technician who is working in a remote location, or a network administrator in a "dark shop" that does not have Internet connectivity. The support technician might need to access content for a product on a tablet, rather than a computer, and they may need access to troubleshooting instructions more often than they need a product installation guide. The network administrator requires a disconnected environment that can operate locally, without access to the online content source, and they might require the installation guide more than the troubleshooting instructions. It is important that the user's experience can be detected and understood so that they get the right content at the right time, and in the right format.

Defining High-Value Content

In addition to having high-quality technical information and a high-quality user experience, we must be able to provide high-value content to our users. This goes beyond the idea of having technically accurate content that is easy to find, and encompasses the ideas of information architecture and taxonomy, in an attempt to provide information that is valuable to our users. First, however, this content must be considered as a valuable to your company. Content should be considered a business asset, something that doesn't just explain how to use a product, but augments and enhances the overall product experience.

There are many different aspects of high-value content. The following definition and characteristics of high-value content were described in the "Point:Counterpoint" column in the February 2012 issue of *Intercom* (Riley, Bailie, & Ames, 2012). At its core, the authors said that high-value content is focused on users, content, and context, and exhibits the following characteristics:

- Leverages intelligence about users, their environment, the subject of the content, and context, and uses techniques such as minimalism to ensure appropriate choice of information to present and when and how to present it
- Incorporates a deep understanding of users, their business and task domains, and the products and solutions in those domains
- Is the product of research and analysis of users and experiences, which drives decisions about delivering information, such as answers to questions about "where" and "how"
- Uses modeling to understand complex information relationships
- Utilizes taxonomy and metadata to classify content for more efficient searching and customization
- Employs organization structures (such as navigation) and signposts (such as labels) that guide users to browse content and improve retrievability across chunks of information
- Takes advantage of information design methods to improve scanning within a chunk of information

- Synthesizes competing requirements to deliver innovation and excellence to users and readers
- Communicates effectively through all of the various dimensions of the information experience, such as content, context, interaction, algorithm (code), organization and structure, format, and visual design

In summary, high-value content makes the complex clear through all of the various dimensions of the information experience.

Our Information Development Community

Before going into the process we used to create our standards, let me describe our Information Development (ID) community. Our community is a world-wide organization made up of Information Development professionals in various roles, including technical writers and editors, people and project managers, build and tools specialists, and translation planners. The Information Development teams come from all parts of IBM, creating product documentation for both software and hardware products. Volunteers from the community work together on standing councils and in short-term workgroups, developing the collateral, best practices, guidelines, and standards that are used by every team to improve the quality of their information experience. The activities of these teams are governed by an advisory council that is made up of ID managers from each brand, and who work together to refine and implement our Information Development strategy. The overall strategy—which includes a common ID process, common tools and technologies, and common metrics for measuring success—is defined by the Corporate Information Development Team, with input from the advisory council and community leaders. In this way, the community is involved in all aspects of Information Development at IBM.

Community Development of Standards

With the definitions of high-quality, high-value content and user experience in place, we began the process of identifying the standards that would be required to ensure our content continued along its journey toward

becoming high-quality. Experts from across the ID community volunteered to be part of an Information Development Quality Council that was established to help define standards and metrics for information quality. This team worked several months to outline what standards would be needed, gathering input from the various councils and workgroups that were part of our ID community to identify the guidelines and collateral—templates, samples, best practices, and so forth—that were available to the ID community. These guidelines and collateral became the tools and techniques, or knowledge base, that would be used by all ID teams to understand and implement the various information quality standards. This team also looked at corporate requirements (such as legal notices and accessibility) and industry standards (such as *ISO/IEC FDIS 26514:2008(E) - Systems and software engineering — Requirements for designers and developers of user documentation* and the Darwin Information Typing Architecture standard) in addition to the standards that were needed for content quality.

In addition to identifying corporate and industry standards that we needed to comply with, the Information Development Quality Council used a variety of metrics and key performance indicators (KPIs) to evaluate the current state of our information. These metrics come from a variety of sources, including direct customer feedback from surveys and customer advocacy groups, problem reports and defect analysis, and satisfaction data. The data from this analysis helped to identify problem areas and customer pain points that needed to be addressed at a corporate level. For instance, defect metrics showed that we needed to improve the technical accuracy of our technical content, so a standard for performing information testing and technical reviews was created.

Once the ID standards were defined, they were broken into the following three major themes or buckets:

- **Corporate Standards**—This group included any requirement that had to be done by all Information Development teams, regardless of their products or the types of content they were producing. These corporate standards included legal requirements, translation and globalization requirements, and accessibility requirements. In many cases, the guidelines and collateral for these standards were

Community-Driven Information Quality Standards

developed outside of the ID community, and were referenced by the ID standards.

- **Information Quality Standards**—This group of standards was specific to the ID community, and focused on the different aspects of high value content, high-quality content, and a high-quality user experience.

- **As-Required Standards**—This group of standards was developed to address situations that were not required by all ID teams, but necessary for certain types of information. For example, if an ID team was going to produce a video-based tutorial, there might be specific standards for that type of content that would not apply to another team that was not producing video-based tutorials. These standards are not optional, but apply on an as-required basis.

With these definitions in place, the Corporate User Technologies Team decided that compliance with all the standards—regardless of which theme or bucket—was mandatory. Those standards that were "as-required" were mandatory if a team was producing content to which these standards applied. Because the idea of having information quality standards was new at the time, most standards were developed and written with a "crawl—walk—run" approach, where ID teams could move along a continuum and become progressively more compliant, and therefore, progressively improve the quality of their content. This approach required us to define compliance criteria that were built up, layer by layer, adding more and more rigor to teams as they advanced toward high quality.

As an example, the information testing and technical review described earlier was written so that, at a minimum or crawl level, ID teams were required to use checklists to show when they met certain aspects of high-quality content. The standard required more advanced team (those that were ready to "walk") to create and use an information test plan that could be used by the ID team to track their testing progress. For the well-established teams (those that were ready to "run"), their information test plan had to be incorporated into the overall test plan for the software or hardware product so that the content could be tested alongside the product itself.

Each standard was then assigned an owner, usually one of the ID councils that helped to govern our community. We then put the standards in place with

plans to revisit them every six months so that the owners could review them for continued applicability and see when new standards needed to be created. This semi-annual review is performed with the assistance of the many ID councils and workgroups, so that the latest versions of collateral and guidelines can be incorporated into the standards and made available to the larger ID community as quickly as possible.

Measuring Compliance

With the standards in place, we needed to determine how best to track compliance. With a corporation as large as IBM, we opted to start small and track compliance for a subset of key products that are important to the business. These products and their technical content were usually translated into non-English languages, so it was important for the base English content to have high quality prior to translation. This allowed us to experiment with compliance reporting, and gave us the opportunity to see how well the ID teams for these products could comply with the standards. Compliance was measured by standard, but each product team was also given an overall assessment that was based on the number of compliances and non-compliances against the entire set of standards.

At IBM, we started with twenty-eight information quality standards, which were tracked against approximately one hundred key products.

We actually ended up using two mechanisms for determining the overall assessment of a product's compliance. We initially used an algorithm that weighted certain standards heavier than others. The relative weighting was based on the need to improve information quality in areas where it was lagging behind at a corporate level, and so standards for these areas were given more weight than standards that focused on aspects of quality that all teams were readily compliant with. For instance, a standard requiring all ID teams to comply with the IBM Style Guidelines was given a lighter weight than a standard for adhering to information testing and technical review requirements because most IBM teams understood and complied with the IBM Style Guide, but many teams were struggling with ways to properly review and test their content for accuracy. In this way, we tracked an overall score for each product ID team. ID teams with higher scores were the teams that became

Bob Vitas

compliant with the higher-weighted standards faster than their counterparts. While this helped us to see which products achieved a "perfect score" and were fully compliant, it did not allow us to track products along the "crawl—walk—run" continuum.

Each software and hardware product in our initial group responded to a questionnaire that asked them to rate their compliance against the standards. Whenever a team was not compliant, they were given the opportunity to explain why they were not compliant. This compliance information was relayed to the owner of the standard, who then met with the non-compliant team to discuss a plan for becoming compliant. In some cases, a non-compliance was due to a legitimate business reason, and so this was noted in the questionnaire and the compliance data for that standard was not counted in the team's overall assessment of compliance. Whenever a non-compliance could not be attributed to a valid business reason, the team was required to become compliant within one product release cycle.

Once the ID teams for the software and hardware products in our initial group had matured and were complying with the standards at a high rate, we switched to a mechanism that assigned up to three points for compliance with a given standard, based on how far along the "crawl walk run" continuum a team was. A non-compliance was assigned zero points. If a product ID team met the minimum requirements ("crawl"), they were assigned one point. Two points were assigned for meeting the next level of compliance requirements, and three points were assigned to an ID team that was fully compliant at the highest levels. In this way, a product ID team could report that they were fully compliant, and at which level: crawl, walk, or run. This process allowed us to identify where there might be issues with moving to the next level of compliance, and to help struggling teams achieve further improvements.

For both of these compliance mechanisms, our initial reports to management and executives were scoped to a high level—scores for all products in a given brand were collected into a brand-level overall score—to show general trends toward improving information quality. We did not point out specific product ID teams as "problems" when they were not compliant, but instead chose to work with those teams to identify the ways that they could become more compliant. We used the teams with high compliance scores as examples, gathering best practices and success stories to augment

our collateral. With this process in place, we saw steady increases in compliance among our key products. We were able to correlate the changes in compliance with trends in defect rates and problem areas, and were able to continually improve the standards and collateral to meet challenges and improve information quality.

Our initial hope was to have all of the ID teams from our key products to have at least a 70% overall compliance with all the standards. We also looked at each individual standard, and hoped to have at least 70% of the product ID teams comply with it. The first reports showed that, at the corporate level, the overall compliance number was about 69%. Within a year after creating the standard and helping the ID teams learn how to become compliant, the overall compliance was at 75%. Within two years, overall compliance was at 85%. Since then, the key products have consistently been over 90% in overall compliance.

We have also monitored our metrics and KPIs over time, and have seen improvements in key areas, such as technical accuracy. As an example, after creating the standard for information testing and technical review, we monitored the percentage of product defects that were attributed to inaccuracies in the content. This percentage has been tracked for nearly eight years, and the percentage of defects that were due to technical inaccuracies has consistently dropped each year. The overall percentage of inaccuracy defects has dropped by half since the standard was implemented.

We have also used metrics to show when things aren't going so well, and have used this data to drive improvements to the processes and collateral we use. For instance, while the technical accuracy of our content was improving, customer feedback showed us that there were specific issues with the samples that were made available our customers to show them how to use our products. In order to improve the quality of samples, we created a standard with specific guidelines on the requirements for various types of samples. We created clear definitions of the types of samples (from small "technology samples" that showed how to use a specific widget or function, to large "showcase samples" that showed how to use many product functions to together to address a specific business scenario), and then provided specific criteria for documenting what the sample was for, how to validate that the sample was correct, and ensuring the usability of the sample once the customer started to use it.

Community-Driven Information Quality Standards

Standards for the Total Information Experience

At IBM, we began this journey toward having information quality standards for the corporation by engaging the Information Development domain and establishing an IBM-wide community of product content creators. We drew from the experience and expertise of the ID community, allowing the subject matter experts within the ID community to work together to improve information quality. This process worked well for the ID domain, but we also needed to consider the Total Information Experience. While each domain has its own quality measures, we need a way to standardize as much as possible at the corporate level so that the Total Information Experience has the same high quality, value, and user experience that we've developed in the product content. We are still in the early stages of this effort, but we have already started developing information quality standards for the Total Information Experience.

The challenge with these kinds of standards is that they have to apply to information creators who plan, develop, publish, and maintain their information using widely different practices and tools, not to mention different customer expectations. Thus, we had to develop standards that could be applied at a broader level, yet were focused on key aspects of information quality. It would also be ideal to have common processes and tools, where appropriate, so that everyone was working from the same playbook.

Working with experts from the various domains—including ID, Support, and Learning, among others—we identified the core standards that all teams must comply with. These standards included ensuring that content meets certain corporate requirements, ensuring that the content is technically accurate, ensuring that content is easy to find, and ensuring that current is kept current and up-to-date. These standards were written at a high level, often without requirements for tooling or process, and with specific metrics that could be used to track when compliance was met. This allowed each domain to determine whether they already had a process in place for ensuring compliance, and if they did, they were allowed to continue using that process, as long as it showed continual improvement over time.

For many aspects of the information quality standards that applied to high-quality content, IBM chose to use the Acrolinx IQ system to identify potential quality problems within pieces of content. Acrolinx IQ is available for a variety of source formats, and the system allowed us to write linguistic rules to flag problems with terminology, style, grammar, and other information quality characteristics, regardless of the format of the content. In this way, we can use a common tool across the Total Information Experience to help improve the quality of our content.

This movement toward a high-quality Total Information Experience led to a dilemma in terms of how to track improvements over time across a wide range of content, from product documentation to learning materials to support documents. Using a questionnaire approach was initially considered, but the amount of work to create and maintain a compliance tracking system—let alone the time it would take all teams to report on their compliance—was deemed too burdensome. Instead, we decided to use metrics and key performance indicators (KPIs) to track compliance. By assigning specific metrics and KPIs to each standard, we plan to see how our information quality is trending over time.

In many cases, we can track the changes in the metrics at a domain level—for instance, looking at the client feedback data for product documentation versus client feedback data for support documents—but often the metrics are at a more general or corporate level. Regardless of the level of granularity, we can see how the metrics change over time. If the metrics for a specific standard begin to trend in the wrong direction, we can work with the Total Information Experience community to investigate what is happening, and why compliance with a given standard is lagging. We can update the collateral for a domain or brand, or help them improve their processes, and work as a single team to improve the quality of the Total Information Experience.

While there is not a wealth of information about corporate trending on the scale of the Total Information Experience, we were able to gather historical data for the content produced by our ID teams, and have begun moving the ID community to this metrics-based approach for tracking compliance.

Conclusion: Take the Round Trip

This article describes how the teams at IBM approached the creation and tracking of information quality standards. Other corporations may have similar or different approaches. The key, however, is that the journey toward information quality standards is not a straight path from Point A to Point B. It should be thought of as a closed loop or wheel, where the work toward improvements never stops.

For instance, a group may implement a standard for information testing to improve the technical accuracy of your content. One may create test plan templates or quality checklists, and track product teams to specific goals and measurements. One day, you may see that these efforts have paid off, and the content is getting fewer defect reports from your customers. Even though the goal of improving technical accuracy has been reached, that should not be the end of the journey. The content might be technically accurate, but is it appropriate for the customer's role? Is it presented to them at the appropriate time? Is it relevant to the customer's situation?

Creating and implementing standards should be seen as the starting point on a never-ending journey. Our experience at IBM suggests taking an outside-in approach and always using customer feedback to understand their needs, and implementing standards that improve their experience. Keep track of what your peers and competitors are doing in the marketplace, such as when they update or improve their Web site to be more useful, or when they begin making content available for mobile devices. Watch for changes in industry standards that might affect the content you create. It is easy to get insulated and focus solely on your own organization, but in the end, you should strive to continually improve your customers' information experience. Drawing input from a variety of sources will help guide you to the next wave of quality improvements, and keep your content vital throughout its lifecycle.

References

Ames, A., Bailie, R., & Riley, A. (2012). Point:Counterpoint. Retrieved from http://intdev.stc.org/2012/02/point-counterpoint/

Ames, A., Riley, A., & Jones, E. (2013). Telling the right story: Proving the business value of content. *Intercom*, *60*(5), 33–39.

Eberlein, K., Anderson, R., & Joseph, G. (2010). Darwin Information Typing Architecture (DITA) 1.2 specification. Retrieved from http://docs.oasis-open.org/dita/v1.2/spec/DITA1.2-spec.html

Hargis, G., et al. (2004). *Developing quality technical information: A handbook for writers and editors* (2nd ed.). Upper Saddle River, NJ: IBM Press.

International Organization for Standardization. (2008). ISO/IEC FDIS 26514:2008(E) - Systems and software engineering — Requirements for designers and developers of user documentation. Retrieved from http://www.iso.org/iso/catalogue_detail?csnumber=43073

Society for Technical Communication. (n.d.) Writing Standards. Retrieved from http://intercom.stc.org/write-for-intercom/writing-standards/

About the Author

Bob Vitas is currently the Operations Manager for the corporate Information Development team at IBM. A computer science major in college and a project manager in practice, he has taken his knowledge of software testing and verification and applied them to the technical writing field, helping to develop processes, standards, and metrics for improving information quality. He has been a part of IBM's information development community for over a decade, working with professionals around the world to improve the quality of IBM's technical content, as well as the overall client technical content experience. Connect with Bob on LinkedIn (http://www.linkedin.com/in/bobvitas).