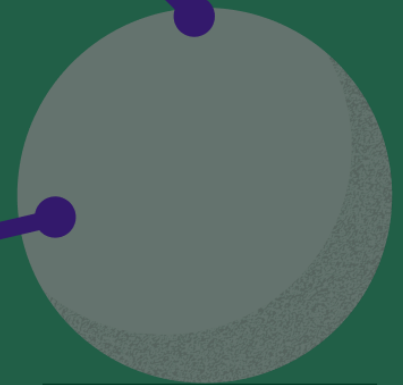




Jennifer Gupta

Get to Know DITA



Annette (and then & therefore Jennifer) started wondering how Large Language Models (LLM's) like ChatGPT parse and return data. What criteria does it "see". Is there a format that it prefers? Perhaps a format that allows for quicker response rates and/or more accurate data return?

Turns out, the answer lies within DITA compliance.

Introducing Darwin Information Typing Architecture (DITA)...
what we need to know...



Unstructured data



Structured data

Imagine you have a bunch of information that you need to add to a training/course/eLearning/etc. It came to you in a PowerPoint or word doc, and you plan on adding it to Rise, Storyline, or making a Quick Reference Guide out of it in PowerPoint or Word.

These formats, all of them, are inherently *unstructured*.... to a machine, the content looks like the image on the left.

Now, *structured* data, what machines can understand, looks more like the image on the right. It takes the same information but adds <META tags> so that the machine understands what is a:

```
<title> TITLE </title>
```

```
<body> Body text </body>
```

```
<image> Image.jpg </image>
```

and what the relationship between that data is.



RECIPE COURTESY OF JOSE ANDRES

Tichi's Gazpacho

★★★★★ 0 Reviews

Level: **Easy**

Total: **40 min**
(includes chilling time)

Active: **10 min**

Yield: **4 to 6 servings**

[Nutrition Info](#)

 Save Recipe



Ingredients:

- Deselect All
- 1 pound Campari tomatoes, seeds removed and reserved for garnish, optional
- 3 Persian cucumbers, roughly chopped
- 2 garlic cloves, peeled
- 1 green bell pepper, seeded, cored and roughly chopped
- 1/4 cup Spanish extra-virgin olive oil, plus more for drizzling
- 1/4 cup Oloroso sherry
- 2 tablespoons sherry vinegar
- Kosher salt
- Microgreens, for garnish, optional

Directions:

1 Working in batches if needed, place the tomatoes, cucumbers, garlic, bell pepper, olive oil, Oloroso sherry, sherry vinegar and salt in a blender and blend on high until very smooth (working in batches if needed). Season with salt to taste. Pour into a pitcher and refrigerate until chilled.

2 To serve: Pour the chilled gazpacho into “rocks” glasses or sippable glasses. Drizzle with olive oil and garnish with the reserved tomato seeds and microgreens if desired.

This recipe has been updated and may differ from what was originally published or broadcast.

Courtesy of <https://www.foodnetwork.com/recipes/tichis-gazpacho-19468827>

Lets take a look at a standard recipe. In it, you can expect to find the following items: Image, Title, Cooking Duration, Yield, Nutrition info, Ingredients, and Directions.

If we wanted to structure data for recipes in general, we might see:

```
<title> Tichi's Gazpacho </title>  
<image> gazpacho.gif </image>  
<duration> 40min </duration>  
<yield> 4-6 servings</yield>  
<nutrition> unknown </nutrition>  
<ingredients> 1 lb tomatoes </ingredients>  
<ingredients> 3 Persian cucumbers </ingredients>  
<ingredients> 2 garlic cloves </ingredients>  
<directions>Working in batches if needed... </directions>
```

***note** – I took the liberty of not listing all the ingredients, but I hope you get how it works

Now, in a *structured* data approach, a machine would know EXACTLY what each piece of the recipe is. We could duplicate this for every recipe we have, and then we can tell the machine what to do with each piece. As in ...

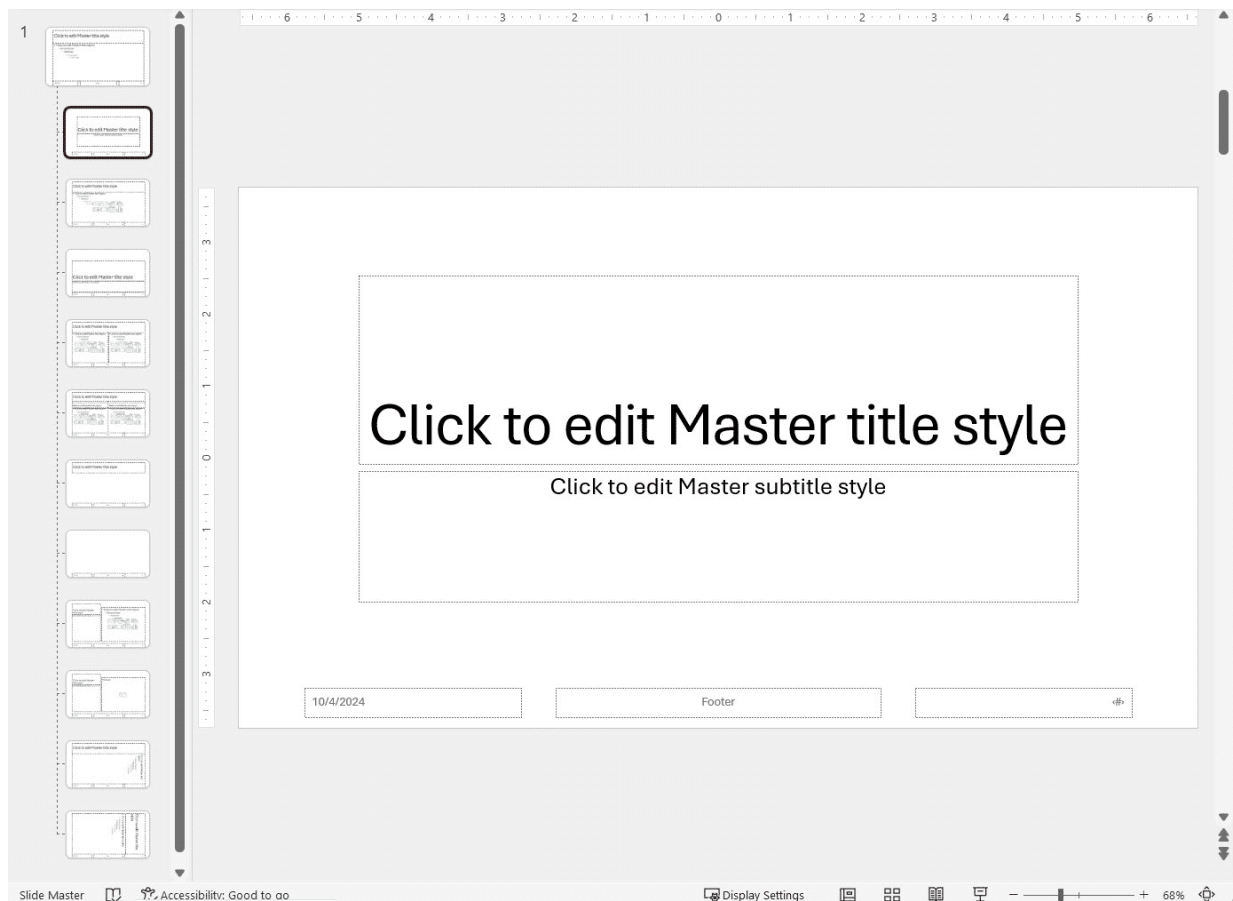
Display each title as Arial 24pt, black text at a specified location.

Display each image, centered, outline 1px, etc in a specified location...

Based on this format, it would recognize that THIS image is part of THIS recipe, these ingredients go with this recipe, etc. (This is the relationship of each tag to the other tags)

People with programming experience: Does this sound familiar? It might! DITA is an XML standard.

That's all well and good, but meta tagged data isn't very pretty... and we don't
build courses that way... so..?



PowerPoint

This is where Slide Masters come into play. LLM's, like ChatGPT, can actually read standard slide masters in our common authoring tools (PowerPoint, Storyline, etc). It can pull titles, body text, and images if we use them consistently. And guess what? These are actually the META tags that DITA uses!



We are still new to this, and our understand may evolve as we find out more information. However, based on our research, we have included a list of the things we can do in our courses that will help LLM's.

We have listed these below.

Storyline / eLearning

Organize your eLearning course into distinct sections or modules with descriptive titles —

Ensure each lesson or section follows a logical and recognizable structure.

Clearly label the module or lesson title, section headings, and sub-headings —

Use a consistent hierarchy (like H1, H2, H3) to help LLM's recognize the content structure.

Use predefined slide templates to help LLM's categorize the content (e.g., Title, Body Text, Image, Video) —

Maintain a consistent layout for instructional screens (title, text, image/video, interaction).

Include relevant keywords and concepts for each module or interaction —

These can help AI tools identify specific lessons or content when needed.

Add alt text to all images, videos, and interactive media —

For complex images, charts, or diagrams, provide a text description either as alt text or in a separate text field to ensure full accessibility and easy parsing by AI.

Clearly label interactive elements such as quizzes, simulations, or drag-and-drop activities —

Use descriptive text or alt text to indicate the purpose of each interaction.

PowerPoint

Use Consistent Slide Layouts —

Stick to predefined layouts like "Title Slide," "Title and Content," "Two Content," etc. LLM's or AI can recognize these layouts, which will help in parsing headers, body text, and images more accurately.

Using consistent layouts ensures that the content is organized and easy to follow. It also helps in maintaining a professional look throughout the presentation.

Apply Slide Titles and Headers —

Ensure every slide has a clear and unique title in the title field. LLM's will rely on titles to index slides, making the information easier to find.

Clear titles and headers help in navigating the presentation efficiently. They also provide a quick overview of the content covered in each slide.

Structured Text Formatting —

Use bullet points or numbered lists consistently in the content sections of your slides. Avoid using text boxes for key content that belongs in the predefined slide areas (like titles or body content). LLM's may have trouble identifying text placed arbitrarily in a slide.

Structured formatting makes the content more readable and easier to understand. It also helps in highlighting key points effectively.

Use Descriptive Alt Text for Images —

Adding descriptive alt text to all images will help AI understand what the image represents. This is essential if you include charts, graphs, or visuals that contain important data.

Descriptive alt text ensures that all users, including those with disabilities, can understand the visual content. It also aids in the accurate interpretation of images by AI tools.

Consistent Use of Tags and Keywords —

Insert relevant keywords in your slides and even in notes. This will help AI recognize essential terms, making it easier to find relevant content when asked.

Using consistent tags and keywords enhances the searchability of the content. It also helps in categorizing and indexing the information effectively.

Add Slide Notes —

Include more detailed explanations or important points in the slide notes section. LLM's might use these notes for additional context while summarizing or analyzing the presentation.

Slide notes provide extra information that can be useful for both presenters and viewers. They also serve as a reference for future revisions.

Use Table of Contents for Longer Decks —

Adding a table of contents with hyperlinks to various sections makes the presentation easier to navigate both for humans and AI tools.

A table of contents provides a clear structure and helps in quickly locating specific sections. It also enhances the overall user experience.

Meta Information —

Ensure file properties (title, subject, author, and keywords) are filled out correctly. AI might also pull from the metadata to contextualize information.

Properly filled metadata helps in organizing and retrieving the presentation. It also provides essential context for AI tools analyzing the content.

Rise

Clear, Consistent Structure —

Modular Structure: Rise courses are already organized into sections and lessons, which is ideal. Make sure each section has a clear and descriptive title so that AI tools can easily identify and categorize the information.

Lesson Titles: Break up the content into small, digestible lessons and use descriptive titles for each lesson, making it easier for AI to parse and categorize the material.

Standardized Lesson Layout —

Rise offers block-based content creation. Standardize the types of blocks (e.g., Text, Image, Quote, Accordion) you use across lessons, so AI can recognize patterns and content structure.

Use specific block types for intended purposes. For example, use quote blocks for key takeaways and accordion blocks for FAQs or detailed explanations.

Alt Text for Images and Media —

Rise makes it easy to add alt text for images. Ensure that every image has descriptive alt text, especially if it's a chart or contains critical information.

For videos, add transcripts directly into text blocks or use the captioning feature to make them searchable and accessible by AI tools.

Consistent Use of Interactive Blocks —

Rise includes interactive blocks like quizzes, flashcards, and knowledge checks. Ensure these elements have clear, descriptive questions and answers. AI can use this structured format to understand and respond to queries about the quizzes or interactions.

If you use interactive elements, such as labeled graphics or scenario blocks, ensure that labels and feedback are detailed and consistent.

Descriptive Tags and Keywords —

While Rise doesn't have a formal keyword tagging system like some other LMS platforms, you can still use descriptive keywords within lesson titles, headings, and body text. This will help AI tools recognize and index key concepts.

You can also use specific keywords in your learning objectives and throughout the text to make it easier for AI to extract course content.

Text Formatting for Emphasis —

Use consistent text formatting (bold, italics, and headers) to indicate important concepts or key points. This helps AI understand which content is emphasized and should be prioritized.

Make use of call-out blocks to highlight key information, as these can stand out and help the AI tool categorize important points.

Notes and Instructions —

While Rise doesn't have a traditional notes section like Storyline or PowerPoint, you can still add contextual notes for learners within the course using text blocks or "Continue" dividers. This extra detail will give AI tools more context when summarizing or analyzing sections.

Learning Management System (LMS) Integration —

If your Rise course is hosted on an LMS, ensure that it is categorized properly within the LMS, and metadata like keywords, course descriptions, and tags are used effectively. Many LMS platforms allow for detailed metadata, which helps AI tools navigate and understand course content more effectively.

Text Transcripts for Videos and Audio —

If you use video or audio in Rise, make sure you either include text transcripts as a separate block or provide subtitles directly in the media. This ensures that AI tools can understand and search through spoken content.

SCORM and xAPI Reporting —

Since Rise courses can be exported as SCORM or xAPI packages, make sure your reporting is detailed enough to provide useful data for AI tools to track progress and learning outcomes. Include

key metrics for quizzes and course completion to enable detailed analysis.

Consistent Alt Text for Interactions

For any interactive content (such as drag-and-drop, labeled graphics, or quiz questions), ensure that alt text or descriptions are added so that AI tools can interpret the purpose of these interactions.

Content Accessibility

Rise is designed to be mobile-first and accessible, which is beneficial for AI tools. Ensure that your content follows accessibility best practices, such as using meaningful headings, proper alt text for images, and clear navigation. This also improves AI's ability to parse and engage with the course.



The ultimate goal is to be able to load our courses into an LLM so that learners can ask a question. The LLM should be able to 'read' our courses, and return the correct answer in a timely manner.

Thank you!