

質問回答

NO.	質問	回答
1	図表についての質問です。 仕様書のページ1 4. 業務の内容 (1) の口 表・図等の構造化対応 「Markdown化が困難な場合には、表番号・図番号、表題・図題等を明示した代替表現を記載すること。」 【質問】 この「代替表現」とは例えばどのようなものですか？ Markdownファイル内の該当箇所に、「代替表現 + 図表の番号や題名」を記載するのでしょうか？ それとも、図表の番号や題名が「代替表現」ということになるのでしょうか？	Markdown化を行うガイドラインについては、全ての図・表に図表番号及び題名が記載されておりますので、当該図表の番号や題名をMarkdown化することが「代替表現」になるというご認識で問題ありません。
2	系統図やフローチャート、数式など、通常のMarkdown（テーブル記法など）での表現が困難な図表について、代替表現の具体的な指定（Mermaid記法、LaTeX形式、画像+代替テキストなど）はございますか。 （将来的なRAG（AI検索）システムで利用する際、画像や数式の意味をAIが正しく理解できる形式に記述を統一したいため）	代替表現の指定はありません。画像や数式の意味をAIが正しく理解できる形式又は図題からどのような画像や数式が記載されているのかAIが正しく理解できる形式にしてください。
3	JIS外字や、単位（μg/L, m?など）などの特殊文字について、UTF-8でそのまま記述してよいですか。環境省指定の代替表記ルール（ゲタ記号「㍿」、実体参照など）はございますか。 （閲覧環境での文字化けを防ぎ、AIが表記ゆれを起こさずに数値を正確に認識できるようにしたいため）	環境省指定の代替表記ルールはありません。閲覧環境での文字化けを防ぎ、AIが表記ゆれを起こさずに数値や単位を正確に認識できるように記載してください。
4	Word内の章、節、項、9号および図表タイトル等には「見出し」等のスタイルが適用されている認識でよろしいでしょうか。 （AI活用しやすいデータ状況かどうか確認したいため）	適用されております。
5	提供されるWord形式データ（約1,200ページ）は、単一ファイルですか、章ごとに分割されていますか。また、図表等で外部ファイルが埋め込まれている場合、オリジナル高解像度画像データ（PNG/JPEG等）を個別ファイルとして別途ご提供いただくことは可能でしょうか。 （並行処理による制作体制の最適化と、図表データのスムーズな移行作業を計画するため）	提供するWord形式データ（約1,200ページ）は、章ごとに分割されており、表紙・目次を含めて34ファイルあります。 図表等のオリジナル高解像度画像データ（PNG/JPEG等）を個別ファイルとして別途提供することはできません。
6	Word内のイラスト（文字列含む）はグループ化されていますでしょうか。 （イラストとして抽出しやすいかどうか、テキストがWord本文に入力されていないかどうかを確認したいため）	Word内の画像はグループ化されております。
7	図タイトルの配置を、図の下から上に変えるような変更は問題ないでしょうか。 （画面で閲覧するユーザーの利便性を向上させるための処理の可否を確認したいため）	記載内容が変わるような変更でないのであれば問題ありません。提供データの記載内容をAIが正しく理解できる形式にしてください。
8	表で記述した方が体裁が整う箇所につきまして、見出し行の追加等は許容されますでしょうか。 （マークダウン形式への変更に伴い、レイアウトを保持するための処理の可否を確認したいため）	記載内容が変わるような変更でないのであれば問題ありませんが、その都度、環境省と協議をお願いします。

9	表を分割して、連結している部分を見出しにするなどのリライトは許可されますでしょうか。（マークダウン形式では表セルの連結ができないため、許可可否を確認したいため） 例）紙面 9ページ（PDF 20ページ）の表の場合 「分類」ごとに表を3分割し、分類列の内容を「分類1：第一種特定有害物質（揮発性有機化合物）」のような見出しにして、3分割した表の上に記述いたします。	記載内容が変わるような変更でないのであれば問題ありませんが、その都度、環境省と協議をお願いします。
10	「適切な情報粒度の確保」とは、「原文の要約・改変」ではなく、「見出しごとのMarkdown階層（#〜#）による構造化や、相互リンクの付加」という理解でよいですか。また、想定される粒度のサンプルはございますか。 （仕様書の「要約・再構成の完全禁止」と矛盾せず、RAGにおける最適な分割（チャンク化）が可能なデータ設計とするため）	「適切な情報粒度の確保」とは、「見出しごとのMarkdown階層（#〜#）による構造化や、相互リンクの付加」という理解で問題ありません。 サンプルはPDFデータであれば以下に公開しております。 https://www.env.go.jp/water/dojo/gl-man.html