

DOCXの仕様検討について

日本製薬工業協会
医薬品評価委員会
電子化情報部会 タスクフォース3
ICH M2 サポートチーム

2016年2月18日

DOCXとは主にMicrosoft Word（以下、Word）で作られるファイルで、拡張子 .docx をもつファイル形式です。

ICH M2 サポートチーム：ICHの中で「医薬品規制情報の伝送に関する電子的標準」の検討を行うM2の活動を製薬協の立場からサポートするチームです。M2では、規制当局とのデータ・文書の授受を効率化する為にファイル形式等の標準化検討とその勧告を行っています。

電子化情報部会 タスクフォース3（以下、EI部会 TF3）

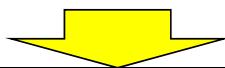
目次

1. “DOCX”勧告とその背景
2. ICH M2 サポートチームの検討目的
3. DOCXとは/その種類と特徴
 - ▶ そもそもDOCXって？
 - ▶ DOCX仕様検討上の課題
 - ▶ “Strict” と “Transitional” の見分け方 (参考)
4. DOCX仕様検討のまとめ
5. DOCXの将来展望と注意点
6. まとめ

▶ 2

1. “DOCX”勧告とその背景

- ▶ 2015年6月 ICH M2がDOCXを勧告 (recommendation)
- ▶ 勧告の主旨：規制情報を提出するファイル形式としてDOCXが利用可能（義務化では無い）
- ▶ 本勧告で「2バージョンを勧告」
 - ▶ バージョンの意味や違い、その影響などについて理解をしておくことが必要



本資料では、背景、本検討の目的、DOCXの種類と特徴、注意点等を紹介

▶ 3

2015年6月にICH M2（医薬品規制情報の伝送に関する電子的標準）から、ICHで利用可能なファイル形式としてDOCXが勧告（recommendation）されました。本スライドに、その抜粋を示しています。

この勧告は、「直ちにDOCXの利用を義務づける」という意味ではなく、「今後、ICHの場で規制情報を提出するファイル形式を検討する際、1つの選択肢としてDOCXが利用可能」であることを意味し、具体的な適用対象の文書ファイルや時期についてはregulatorsが今後検討します。

皆さんの会社でもDOCXは既に広く利用されていると思いますが、本勧告のConditions項には「2バージョンを勧告する」と書かれており、バージョンの意味や違い、その影響などについて十分に理解をしておくことが必要と思われます。

本資料では、DOCXの勧告とその背景、本検討の目的、DOCXとは何か/その種類と特徴、今後の利用・展開を検討する場合の注意点等を紹介します。

1. “DOCX”勧告とその背景

2015年6月 ICH M2 “DOCX”勧告 (原文抜粋 参考和訳)
(原文全文は <http://estri.org/recommendations/index.htm> 参照)

勧告:

具体的には、narrative documentsに対して、「DOCX」(ISO/IEC 29500のサブセット)として知られているファイル形式を勧告する。この勧告は（利用可能な）ファイル形式の追加であり、ISO-32000(PDF)またはPDF/Aの置き換えを意図するものでは無い。
(ICH M2) 専門家作業部会は、各規制当局が以下を規定する実施計画を作成することを勧める：

- DOCXファイルを受け入れができるかどうか
- 受け入れができる (DOCXの) **version(s)**
- **どの種類の規制文書に対して**、DOCXファイルを受け入れができるか

何に対していつから適用するか
は、各当局が今後検討

状況:

この勧告は、ISO 29500標準で規定される以下の**2バージョン**を対象とする：

- ISO/IEC 29500:2008(Transitional)
- ISO/IEC 29500:2012(Strict)

「2バージョン」って?
何が違う？ その影響は？

専門家作業部会は、**将来のある時点でISO/IEC 29500:2012(Strict)に標準化すること**を意図している。

▶ 4

DOCX勧告においては下記の様にバージョンが表記されています。

- ISO/IEC 29500:2008 (Transitional)
- ISO/IEC 29500:2012 (Strict)

この記載は厳密なものではありません。

ここで言う「Transitional」「Strict」は、いわば通称のようなものであり、それぞれ

• Transitional … ISO/IEC 29500-4

• Strict … ISO/IEC 29500-1

を指します（年号表記は改定版が発行されると変更になります）。

1. “DOCX”勧告とその背景

2015年6月 ICH M2 “DOCX”勧告（原文抜粋 参照）
(原文全文は <http://estri.org/recommendations/index.htm> 参照)

Recommendation:

Specifically, for narrative documents, the format known as “DOCX” (a subset of ISO/IEC 29500) is recommended. This recommendation is for an additional file format and is not intended as a replacement for ISO-32000 (PDF) or PDF/A.

The EWG recommends that regulators develop an implementation plan which specifies:

- if they can accept DOCX files
- the version(s) they are able to accept
- for which kinds of regulatory documents they are able to accept DOCX files

Conditions:

This recommendation is for the following 2 versions of the ISO 29500 standard as follows:

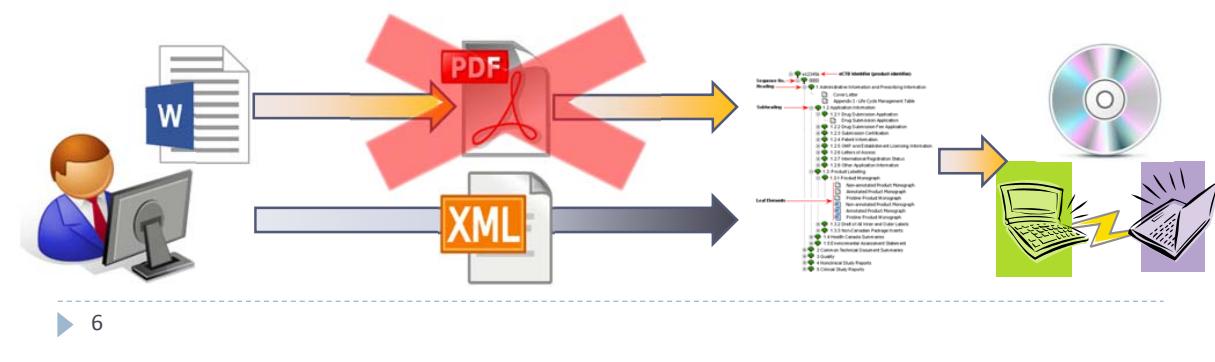
- ISO/IEC 29500:2008 (Transitional)
- ISO/IEC 29500:2012 (Strict)

It is the intention of the EWG to standardize on ISO/IEC 29500:2012 (Strict) at some point in the future.

▶ 5

1. “DOCX”勧告とその背景

- ▶ DOCXが利用できるとPDF化（PDFレンディション）の手間とコストを削減できる
- ▶ FDAはDOCXを既に受入れている
→今後更に拡充される可能性あり



PDFレンディション：Word等の文書を表示用にPDFに変換すること

FDAで受け入れているDOCX文書の事例： eCTD Module 1 : m1-14 (Labeling)

（下記リンクで開くファイルのp2に「DOCX」が記載）

<http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM347471.pdf>

2. ICH M2 サポートチームの検討目的

- ▶ “DOCX”勧告を見てみると
 - ▶ 「“DOCX”はISO/IEC 29500のsubset」って？？？
 - ▶ 「ISO 29500 の2バージョン」を勧告・・DOCXのバージョンって？
 - ▶ 「将来のある時点でISO/IEC 29500:2012 (Strict) に標準化」って？
- ▶ まずは「DOCXとはどんなものか」を知ること
 - ▶ DOCXの仕様・規格 (ISOとの関係など)
 - ▶ DOCXのバージョン：その意味や違い、我々への影響、要注意点
 - ▶ 将来的な展望：（例）文書構造化の進展

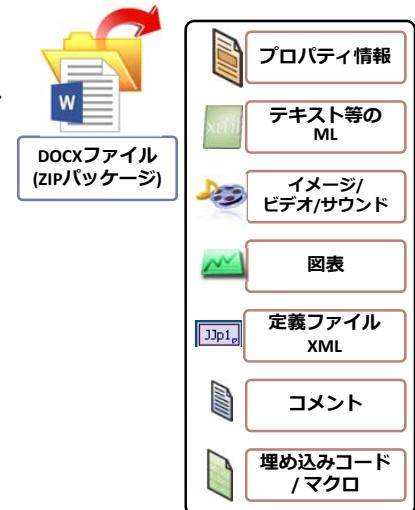
「これらの情報を調査・検討し資料としてまとめる」
ことを目的としてEI部会のICH M2 サポートチームで活動
→ 本資料を作成

▶ 7

3. DOCXとは/その種類と特徴 そもそもDOCXって？

<DOCXの特徴>

- ▶ ISOによって制定された国際規格として公開されている
- ▶ Word 2007から導入された新たなXMLベースのファイル形式
- ▶ パーツ（コンポーネント）化されたファイルをZIPでまとめたもの
 - ▶ 圧縮技術（ZIP）を使用しており、ファイル容量を軽減
 - ▶ Schemaの仕様は無償ライセンスの下で公開
 - ▶ 各パートにMicrosoft社製品を経由せずにアクセスが可能
 - ▶ Microsoft社製品以外のアプリケーションを使用してデータを閲覧・編集できる
 - ▶ パーツの再利用が容易



▶ 8

従来のMicrosoft Office中のアプリケーション（Word）のファイル形式(.doc)は、Microsoft社による非公開ファイル形式が採用されており、対応アプリケーション上で強力な処理能力を発揮することはできても、汎用的にファイルを扱うことが困難である、という難点がありました。

そのため、DOCファイルの場合、Microsoft社製品がないと見読性が保証できません。

3. DOCXとは/その種類と特徴

DOCX仕様検討上の課題

国際標準とWord (DOCX) のバージョン

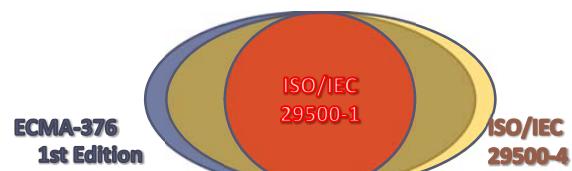
- ▶ 課題：DOCXには**複数のバージョン**がある
 - ▶ それぞれのバージョンが別々の国際標準として登録されている

国際標準として3バージョンが登録

- ▶ **ECMA-376 1st Edition** (Word 2007 で対応)
- ▶ **ISO/IEC 29500-4 Transitional** (Word 2010 で対応)
- ▶ **ISO/IEC 29500-1 Strict** (Word 2013 で対応)

「ISO完全準拠」と呼ばれている

ECMA-376 1ST EDITION ≈ ISO/IEC 29500-4 > ISO/IEC 29500-1



▶ 9

DOCXはOffice Open XML (OOXML) 標準としてECMAおよびISO/IECで標準化されています。

DOCXには、現在3つのバージョンが国際標準として登録されています。

- ECMA-376 1st Edition (Word 2007 で対応)
ECMA (欧洲電子計算機工業会) に登録された形式。
- ISO/IEC 29500-4 Transitional (Word 2010 で対応)
ISO / IEC (IEC : 国際電気標準会議) に登録された、暫定適合クラスの形式。互換性のためのレガシー形式 (過去の仕様。例えば、.docで用いられていた一部の技術等) の利用を許可。ECMA-376 1st Edition とほとんど同じ。
- ISO/IEC 29500-1 Strict (Word 2013 で対応)
ISO / IEC に登録された、レガシー形式を含まない完全準拠の形式。ECMA-376 2nd Edition に相当。

なお、実際の国際標準はWordを含むOfficeとして登録されていますが、本資料では便宜上「Word」に特化して表示しています。

実際の互換性については次のページを参照して下さい。

3. DOCXとは/その種類と特徴

DOCX仕様検討上の課題

DOCXとWordのバージョン（1）

Word	ECMA-376 1st edition	ISO/IEC 29500-4 Transitional	ISO/IEC 29500-1 Strict
Word 2007	Read/Write		
Word 2007+SP2*	Read	Read/Write	
Word 2010	Read	Read/Write	Read
Word 2013	Read	Read/Write	Read/Write

*サービスパック2

赤枠で示した範囲がICH M2勧告のDOCXです。

▶ 10

“Read”：表示・レイアウトが変らず表示が出来る。

“Read/Write”：表示・保存（読み書き）が出来る。

“空欄（ピンクセル）”：表示・保存が出来ない、作成出来ない。

上記の図はWordのバージョンとDOCXのバージョンの関係を示しており、赤枠で示したものがICH M2の勧告の範囲です。

Strict と Transitionalでファイル形式を変更したとしても、原則、表示には影響はありません。

ただし、レガシーな描画機能で作成したデータを含むTransitionalファイルをStrictに変換した場合は、表示を担保するものではありません。

3. DOCXとは/その種類と特徴

DOCX仕様検討上の課題

DOCXとWordのバージョン（2）

▶ DOCXの3バージョンのうち、何故2バージョンのみを勧告？

✖ ECMA：勧告しなかったのは、1st editionがOffice2007のみの対応であること、およびISO化されていないため

○ Transitional：レガシー形式との互換性を重視した仕様のため、現時点では勧告対象に加えることが必要と判断

○ Strict：一部の機能はレガシー形式と互換性を持たないが、仕様の厳密さを重視しており、将来のある時点でこれに標準化することを意図

	Transitional	Strict
特定のアプリケーションの依存	有り	無し
長期保存の展望	低い	高い
レガシーな描画機能（図）を使ったファイルの取り扱い	有り	無し

▶ 11

3. DOCXとは/その種類と特徴 "Strict" と "Transitional" の見分け方 (参考)

"Strict" と "Transitional" の「XML namespaces」の定義の違い

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<w:document xmlns:mc="http://schemas.openxmlformats.org/markup-compatibility/2006" xmlns:r="http://schemas.openxmlformats.org/officeDocument/2006/relationships" xmlns:m="http://schemas.openxmlformats.org/main/2006" xmlns:wp14="http://schemas.microsoft.com/office/word/2010/14/main" xmlns:wp="http://purl.oclc.org/ooxml/wordprocessingML" xmlns:w="http://purl.oclc.org/ooxml/wordprocessingML" xmlns:w14="http://schemas.microsoft.com/office/word/2010/14/main" xmlns:w15="http://schemas.microsoft.com/office/word/2011/15/main" xmlns:wne="http://schemas.microsoft.com/office/word/2006/main" mc:Ignorable="w14 w15 wne wp14" w:...
```

...
</w:document>

Strict

...
</w:document>

Transitional

文書の見た目には差は無くとも、記述様式（仕様）は大きく異なるので、データベース取り込みなど機械処理時には違いが生じる可能性あり

▶ 12

DOCXを展開するとXML表示になりますが、その中のnamespaceの定義（書き方）が違っています。StrictとTransitionalのファイル形式を見分けるには、現時点ではこのnamespaceを確認するしか方法はありません。

4. DOCX仕様検討のまとめ

DOCXの仕様の違いによる影響

▶ 影響なし

- ▶ Word での表示（見た目）：Word2007+サービスパック2以降なら担保される

▶ 影響あり

- ▶ DOCXの仕様の違いにより、編集や閲覧できるWordのバージョンが異なる：
Strictは利用できるWordのバージョンが制限されるため、標準として利用するには注意が必要。Word 2007では更に制限が多いので注意が必要
- ▶ .doc等で用いられていたレガシーな描画機能との互換性に違いがある
 - ▶ Transitional：互換性あり
 - ▶ Strict：互換性なし
- ▶ XMLの記述宣言文様式（namespace）に違いがある
 - ▶ XMLを利用してコンテンツの構造化を想定している場合、DOCXとWordとで対応するバージョンに注意する必要がある
(場合によっては、運用やツールなどを事前に検討しておく必要がある)

▶ 13

5. DOCXの将来展望と注意点

- ▶ 将来展望：（例）文書構造化の進展
 - ▶ 安全性報告や添付文書など、文書の構造化が進んできている
 - ▶ 標準化により、文書ファイルの一般化・長期保存が可能である
 - ▶ 文章やデータの引用・再利用が容易になる
 - ▶ 当局に集まる膨大なデータを効率的に構造化・データベース化できる可能性がある（見た目は文書だが、データベース等への入出力が可能）
- ▶ 注意点：構造化した際にもっとも注意すべき（strictとTransitionalの具体的な）違いは
 - ▶ 「XML namespaces」の定義が異なる
 - ▶ **Strict** (ISO/IEC 29500-1)におけるnamespaceの定義
Root: <http://purl.oclc.org/oxml/> · · · ·
 - ▶ **Transitional** (ISO/IEC 29500-4)におけるnamespaceの定義
Root: <http://schemas.openxmlformats.org/> · · · ·

▶ 14

将来展望として、文書の構造化に利用することが考えられます。

XMLベースのファイル形式であるため、タグ情報を使って文書内の情報を構造化することが可能です。

構造化された文書データは、他のファイルやデータベースからの引用や再利用、もしくはデータベースへのデータや文章の入力が容易になります。

すでに安全性報告や添付文書などで文書の構造化は進んでおり、これらの目的のためにはPDFよりもDOCXの方が適したファイル形式であると思われます。

そのため、将来的にはPDFからDOCXへと置きかえられていく可能性があります。

StrictとTransitionalのファイルは12ページに示したようにXML namespacesの定義に違いがあります。

これらの違いは、単にファイルを表示するだけでは、余り問題になりません。

しかし、構造化においてはこれらの違いが、データベース等への入力時に問題（エラー等）が起きる可能性があります。

6. まとめ

- ▶ 利用可能なDOCXのファイル形式として2バージョンが勧告されました
 - ▶ Transitional
 - ▶ Strict
- ▶ これらのバージョン間には、以下のような違いがあります
 - ▶ 編集や閲覧できるWordのバージョンの違い
 - ▶ レガシー機能への対応（互換性の有無）
 - ▶ XMLとしての書式の違い
- ▶ これらのバージョン間の違いによる影響
 - ▶ 見た目や通常のWordでの作業にはほぼ影響ないことがわかりました
 - ▶ XMLとしての書式は異なるので、例えば、構造化文書としての利用時には問題が起きる可能性があります

▶ 15

謝辞

本資料作成にあたり、

- ・日本マイクロソフト社 柳澤 政夫氏
- ・Microsoft社 Jim Thatcher氏

にご協力頂きました。心より感謝致します。

電子化情報部会 タスクフォース3 ICH M2 サポートチーム

部会長	吉本 克彦	日本新薬株式会社
副部会長	橋本 勝弘 (ICH M2 Topic Leader)	大日本住友製薬株式会社
運営幹事	井上 佳紀	参天製薬株式会社
運営幹事	工藤 稔	鳥居薬品株式会社
拡大運営幹事	佐久間 直樹 (ICH M2 Expert)	帝人ファーマ株式会社
	平松 理恵子	ブリストル・マイヤーズ 株式会社
	細井 礼子	ファイザー株式会社
	宮崎 恵	MSD株式会社

参考/用語説明

略号	
DOC	DOCファイルとは、Microsoft Word標準の文書ファイル、およびファイル形式。
DOCX	Word 2007から導入された新たなXMLベースのファイル形式（Office Open XML形式の一部）。バーツ（コンポーネント）化されたファイルをZIPでまとめたもの。
ECMA-376	Office Open XML形式は、当初、ECMAと呼ばれる標準化団体によって2006年12月に定義された（ECMAは、情報交換に関する国際標準を規定）。Office Open XMLを定義する規格は、ECMA-376標準。
ICH M2	ICHの中で「医薬品規制情報の伝送に関する電子的標準」を検討するグループ。
ISO/IEC 29500	Office Open XMLのISOおよびIECにおける規格。2008年4月にISOとIECの合同技術委員会においては初めて、ISO/IEC 29500として標準化された。
Office Open XML	情報の保存と交換を行うためにオフィスソフトウェアスイートに実装される、国際的に標準化されたファイル形式の規格。
Schema	文書内のデータ構造定義を記述したファイル。
Strict	「厳密な、精密な」の意。本資料ではISO/IEC 29500-1の通称として使用。
Transitional	「移り変わる、過渡的な、過渡期の」の意。本資料ではISO/IEC 29500-4の通称として使用。
XML	文書やデータの意味や構造を記述するためのマークアップ言語の一つ。

▶ 18