

# デジタルドキュメント(10)

高久雅生

masao@slis.tsukuba.ac.jp

2013年6月20日(木)3・4時限

# (前回の復習 = ふりかえり)

- デジタルドキュメントとドキュメントフォーマット
- LaTeX
  - 文書例
  - LaTeX文書の構成要素
  - 利用例 / 組版
- PDF
  - 歴史, PostScript
  - 構成要素
  - 文書フォーマットの構造

# 前回の出席カードから(質疑、コメント)

## — LaTeXに関して —

- LaTeXは1年生の授業で少しだけ触れましたが、使いにくいという印象しか残っていないです。LaTeXの実用例なども紹介されると良いと思います。
  - やはり、実用ということで最も良く使われるのは学術分野でしょうか。
    - <http://www.acm.org/sigs/publications/proceedings-templates>
    - <http://www.springer.com/authors/book+authors?SGWID=0-154102-12-970131-0>
  - 他には、説明書、書籍などを、テキストファイルで執筆して、LaTeXを通じて組版するツールなどもあります。
    - 例: REView <https://github.com/kmuto/review/blob/master/doc/format.rdoc>
- LaTeXを利用するための環境設定はなかなか骨が折れると耳にしましたが、先生の場合、どうだったでしょうか。
- TeXは導入が面倒なんですよ...。
  - はい、その通りです。1) TeXファイルを変換する platex 環境、2) プレビューを行う dviware の環境、3) PDF変換を行う dvipdfmx の環境、4) フォントの設定など、複数のソフトウェアを個別にそれぞれセットアップする必要があるので面倒に感じますね。加えて、日本語と欧文の環境で違いがあるなど煩雑です。環境設定は面倒なので、マシンを乗り換えるたびに最初に導入してからはいじらないようにしています。昔(1990年代)に比べればだいぶマシになりましたが...
  - 参考: TeX wiki <http://oku.edu.mie-u.ac.jp/~okumura/texwiki/>

# 前回の出席カードから(感想)

1. XMLは今、マークアップ言語で取り扱っていて興味を持っている分野である。
2. PDFの構造がそんなに複雑なものとは知らなかった。
3. PDFがPortable Document Formatというのを知って、ひとつ賢くなりました。
4. どの形式も一長一短があることを再認識させられた。
5. PDFにセキュリティ問題なんてあったのかと驚いた。
6. PDFのフォーマットがとても複雑であるということに驚いた。実際PDFファイルをテキストエディタで開いてみたが、訳が分からなかった。
7. PDFのファイル内容、構造がどうなっているのかという解説があって良かったです。
8. レポートの×切は7/9がいいです。

# 前回の出席カードから(質疑、コメント)

## ー レポートに関して ー

- レポートに「読み込みが遅かった」と書いたら、「何で読んだの？」といわれていたが、特別な理由は無く、ただ読みたかったから読んだだけである。読み込みが遅いからといって読むのをやめようとは思わなかった。
  - ー レポートに付けた説明、コメントが悪く、申し訳ありません。
  - ー 「対象とした書籍は、具体的に、何のツールを使って読んだのか？」というコメントのつもりでした。たとえば、Web上で読める青空文庫の場合、ウェブブラウザやそのバージョンにより、その読み方、閲覧のUIが微妙に異なる場合があったりするので、それを具体的に示していただくことが重要と考えています(前提条件)。
- Aの評価とA-の評価をもらっていた場合、最終成績はAになることもあるのでしょうか？
  - ー はい、あります。

# 本日のお品書き

- オフィス文書とデジタルドキュメント
  - ワークフロー
  - 処理モデル
- ドキュメントフォーマットの事例
- Open Office XML
  - 標準化(独占規格から国際標準へ)
  - パッケージ形式による複合オブジェクトの埋め込み
  - 文書内容にあわせた文書要素

# オフィス文書

# オフィス文書の具体例

- 会議資料
- 議事録
- プレゼンテーション資料
- 広報資料
- 案内、お知らせ
- 会計書類：見積書、発注書、納品書、請求書
- 出張願、出張依頼、出張復命書
- 報告書
- 企画書
- 説明書、マニュアル
- 業務メモ
- 連絡文書
- レターカバー

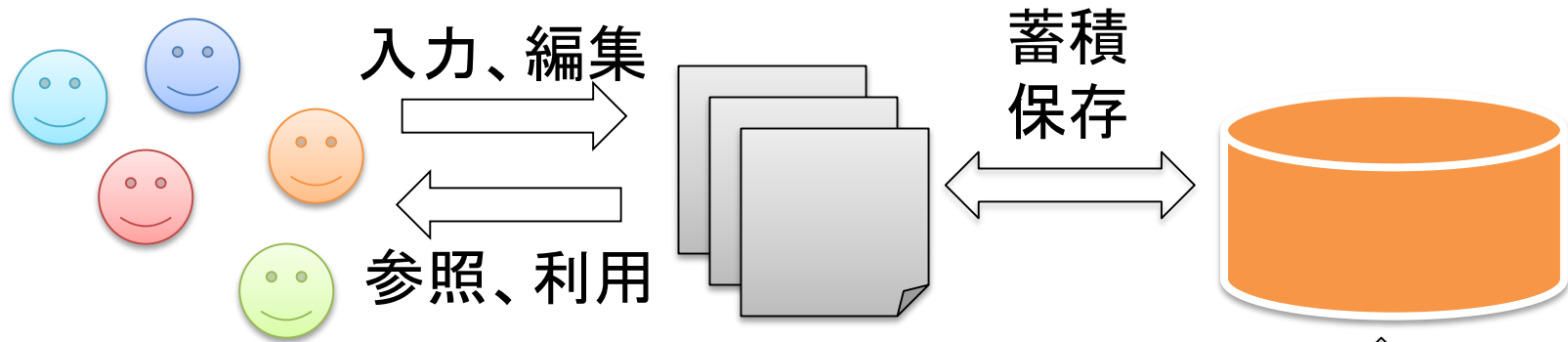


# オフィス文書とデジタルドキュメント

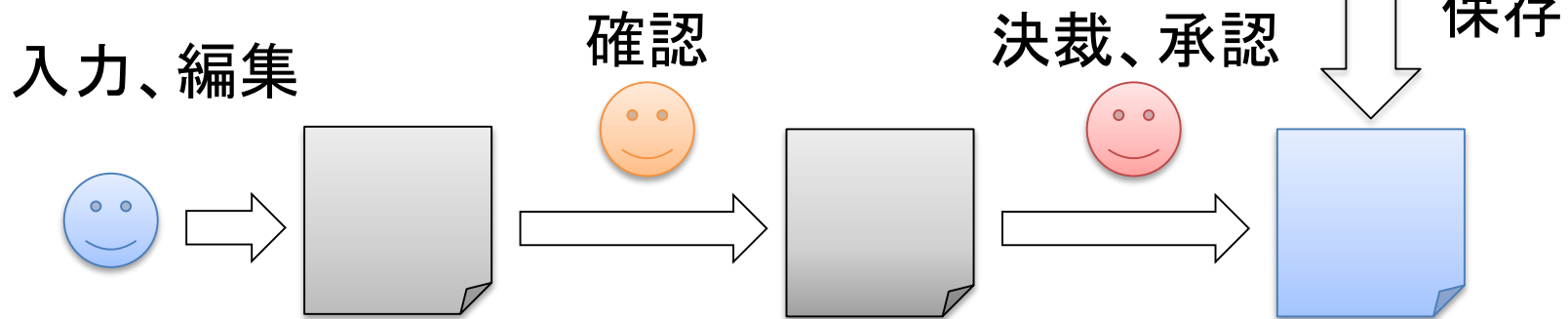
- オフィス文書はデジタルドキュメントの王道
- 文書のサイクルが出現
  - 作成
  - 編集
  - 承認
  - 管理
  - 配信・提供
  - 保存管理
- 関わるひとが多いことも特徴
  - ドキュメントの作成者
  - 共同入力者・編集者
  - 承認・決裁者
  - 利用者
  - 管理者
  - 2次的な利用者
- 紙の文書との併用も

# オフィス文書に関する処理の類型

## 1. 文書共有、保存



## 2. 文書入力、決裁



# オフィス文書の特徴

- 複数人による情報共有が前提とされている
  - 他者による引継ぎや検査
- 情報の保存
  - 日付(保存期間)、担当者など
- 定型形式
- 紙文書としての処理フロー
- 版管理機能

# 統合オフィスソフトウェアの事例

## (オフィススイート; Office suite)

名称	文書	表計算	プレゼン テーション	画像 処理
Microsoft Office	Word	Excel	Powerpoint	
LibreOffice	Writer	Calc	Impress	Draw
Google Docs	Docs	Spreadsheet, Form	Presentation	
iWorks	Pages	Numbers	Keynote	
JUST Suite	一太郎	三四郎		花子
Lotus (IBM)	WordPro	1-2-3		

# オフィスソフトウェアが取り扱う ドキュメントフォーマット

- それぞれの文書ソフトウェア毎に独自フォーマットが用いられることが多い。
- 1990年代後半からは、Windowsとともに、Microsoft Office製品が市場シェアを伸ばし、それらのソフトウェアがオフィス文書コンテンツの代表格となってきた。
- 2000年代以降、オープンフォーマットの重要性が高まり、標準化された文書ソフトウェアが伸張している。
  - OpenOffice.org (2000-2011)
  - → ドキュメントフォーマット ODF (Open Document Format)
- また、別の動きとしてオンライン上での文書ソフトウェアの提供の動きもある。
  - Google Docs

# Microsoft Officeのドキュメントフォーマット

- 長年の間、Microsoft Officeは独自形式(バイナリデータ)を用いていた
  - Microsoft Officeの初期バージョン以来、MS Office 2004まで
  - MS Office 2004以前の環境におけるドキュメントフォーマットは一社独占による文書形式 (proprietary)
    - 拡張子: .doc, .xls, .ppt
- 仕様の詳細が明らかでなかったため、MS社以外によるドキュメント利用が困難
  - 別のOS・機種における閲覧可能性が低い
  - 検索エンジンによるテキスト検索などでの応用
- 公的部門における調達への準拠(グローバル化への対応)

# オフィス文書における

## ドキュメントフォーマットの標準化 (ODF)

- 2000年にOpenOffice.orgがオープンソースソフトウェアとして公開されると、MS Officeの代替としての期待が高まった。同時に、データ交換のためにも、オープンフォーマットへの期待、ドキュメントフォーマットの互換性への期待も高まった。
- 2002-2005年にかけて議論され標準化された。
- Open Document Format (ODF)
  - 国際標準ISO/IEC 26300:2006

# オフィス文書における ドキュメントフォーマットの標準化 (OOXML)

- ODFの動きに呼応して、Microsoft側も、Office 2007リリースにあわせ、新しいドキュメントフォーマットを提案、標準化
- Office Open XML (OpenXML, OOXML) の提案
- 2006年：ヨーロッパの標準化団体ECMAにおける標準化
  - ECMA-376
- 2008年：国際標準化機構ISOにおける標準化
  - ISO/IEC 29500



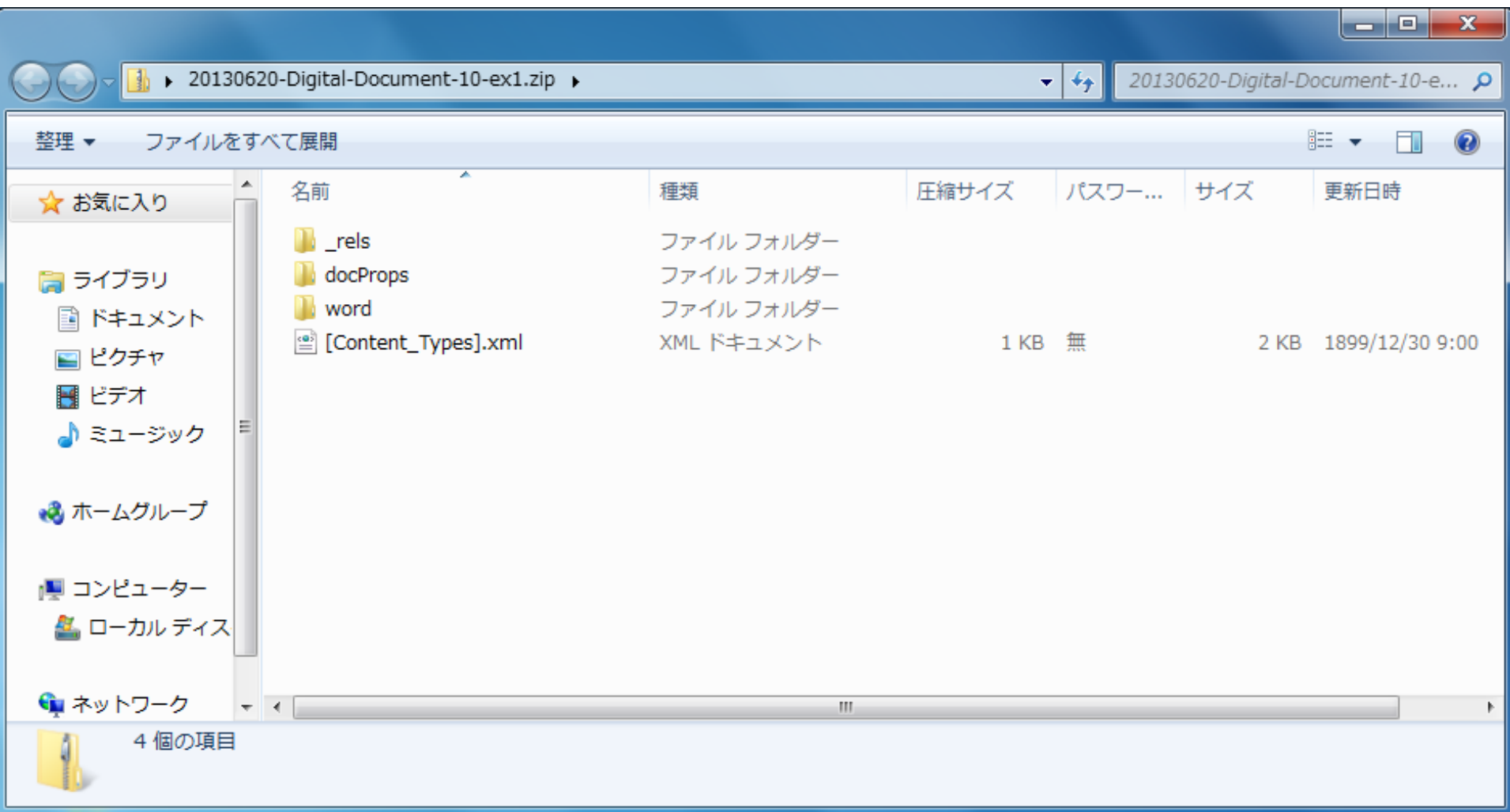
# OOXML と ODF

- 国際標準化に向けた競争の激化
- 両者のドキュメントフォーマットは基本的には、かなりの部分で類似のもの
  - (しいて言えば、OOXMLの方がより複雑: Microsoft Officeが持つ機能全体に対応)
- 3つの主要な文書種別をカバー
  - ワードプロセッシング、スプレッドシート、プレゼンテーション
  - それぞれ Word/Writer, Excel/Calc, Powerpoint/Impress に対応
  - ODFフォーマットの拡張子: .odt, .ods, .odp
  - OOXML新フォーマットの拡張子: .docx, .xlsx, .pptx
    - WordPressingML, SpreadsheetML, PresenationML
- 複合オブジェクトを許す汎用形式
  - パッケージ形式 (Zip)
  - XML (コンテンツ+メタデータ)
  - マルチメディアファイル

# OOXMLにおけるパッケージ形式

- Zip形式によるアーカイブ化＋圧縮
  - XML形式以外のファイル(画像等)の埋め込み、展開
  - テキスト形式であるXML文書の圧縮によるファイルサイズ縮小に貢献
- 任意のWord文書等の拡張子を .zip に変更してZipファイルとして展開、閲覧してみる
  - [Content\_Types].xml
  - \_rels/.xml
    - *word/document.xml* , ...
    - *docProps/core.xml* , ...

# OOXMLパッケージの展開例



# OOXMLパッケージの展開例

## \_rels/.rels

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<Relationships
xmlns="http://schemas.openxmlformats.org/package/2006/relationships
">
  <Relationship Id="rId3"
Type="http://schemas.openxmlformats.org/officeDocument/2006/relati
onships/extended-properties" Target="docProps/app.xml"/>
  <Relationship Id="rId2"
Type="http://schemas.openxmlformats.org/package/2006/relationships/
metadata/core-properties" Target="docProps/core.xml"/>
  <Relationship Id="rId1"
Type="http://schemas.openxmlformats.org/officeDocument/2006/relati
onships/officeDocument" Target="word/document.xml"/>
</Relationships>
```

# OOXMLパッケージの展開例

## word/document.xml

```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
- <w:document mc:Ignorable="w14 w15 wp14"
  xmlns:wps="http://schemas.microsoft.com/office/word/2010/wordprocessingShape"
  xmlns:wne="http://schemas.microsoft.com/office/word/2006/wordml"
  xmlns:wpi="http://schemas.microsoft.com/office/word/2010/wordprocessingInk"
  xmlns:wpg="http://schemas.microsoft.com/office/word/2010/wordprocessingGroup"
  xmlns:w15="http://schemas.microsoft.com/office/word/2012/wordml"
  xmlns:w14="http://schemas.microsoft.com/office/word/2010/wordml"
  xmlns:w="http://schemas.openxmlformats.org/wordprocessingml/2006/main"
  xmlns:w10="urn:schemas-microsoft-com:office:word"
  xmlns:wp="http://schemas.openxmlformats.org/drawingml/2006/wordprocessingDrawing"
  xmlns:wp14="http://schemas.microsoft.com/office/word/2010/wordprocessingDrawing"
  xmlns:v="urn:schemas-microsoft-com:vml"
  xmlns:m="http://schemas.openxmlformats.org/officeDocument/2006/math"
  xmlns:r="http://schemas.openxmlformats.org/officeDocument/2006/relationships"
  xmlns:o="urn:schemas-microsoft-com:office:office"
  xmlns:mc="http://schemas.openxmlformats.org/markup-compatibility/2006"
  xmlns:wpc="http://schemas.microsoft.com/office/word/2010/wordprocessingCanvas">
  - <w:body>
    - <w:p w:rsidP="00A770D1" w:rsidRDefault="00A770D1" w:rsidRPr="00A770D1"
      w:rsidR="000C7B86">
      - <w:pPr>
        <w:pStyle w:val="1"/>
      </w:pPr>
      - <w:r w:rsidRPr="00A770D1">
        <w:t>デジタルドキュメント文書の例</w:t>
      </w:r>
    </w:p>
    - <w:p w:rsidP="00A770D1" w:rsidRDefault="00A770D1" w:rsidR="00A770D1">
```

# (再掲)ドキュメントフォーマットの切り口 (1)

- テキスト (text) vs バイナリー (binary)
  - ビットデータ
  - 文字コードによる解釈
  - 外字
- フォーマットの指定・識別・判別
- シンプルコンテンツ vs 複合メディア
  - 埋め込みコンテンツ
  - ハイパーリンク
- メタデータ
  - 埋め込みメタデータ
  - 外部メタデータ記述
- 文書レイアウト
  - ページ概念
- 文書内の書式要素
  - 見栄え / スタイル
  - フォント
- ファイル vs ストリーム
  - データの保存・蓄積と配信

# (再掲)ドキュメントフォーマットの切り口 (2)

- オープンフォーマット
  - 移植可能性 / ソフトウェア独立性
  - Free / proprietary
- 標準化
  - デファクト標準とデジュール標準 (“de facto” vs “de jure”)
- 文書フォーマットのバージョン
- フォーマット変換
  - テキスト → HTML
  - LaTeX → PDF
- 圧縮
  - 可逆 vs 非可逆
- セキュリティ
  - パスワード
  - 電子署名
- 長期保存
- デジタルフォレンジック

# 本日のまとめ

- オフィス文書とデジタルドキュメント
  - ワークフロー
  - 処理モデル
- ドキュメントフォーマットの事例
- Open Office XML
  - 標準化(独占規格から国際標準へ)
  - パッケージ形式による複合オブジェクトの埋め込み
  - 文書内容にあわせた文書要素



# まとめのまとめ(授業全体のまとめ)

- デジタルドキュメントとは？
- 学術分野におけるデジタルドキュメント
  - 論文、オンラインジャーナル、プラットフォーム
  - 学術情報の利用、提供形態
- 電子書籍
  - 閲覧環境：ハードウェア、ソフトウェア、コンテンツ
- ドキュメントフォーマット
  - 文書構造、文書要素、書式・スタイル
  - プレインテキスト
  - HTML、XML
  - LaTeX、PDF
  - Office XML

# 第4回レポート課題

- 以下の11種類のドキュメントフォーマットから、1つを取り上げて、デジタルドキュメントとしての具体例を挙げながら、そのフォーマットがどのような領域分野で用いられているか、閲覧・作成環境と特徴、国際標準との関連等について文章で説明してください
- 対象とするドキュメントフォーマット
  - PDF, LaTeX
  - HTML, SGML, XML
  - EPub, Amazon Kindle, XMDF
  - Markdown
  - OOXML, ODF, RTF
- なお、取り上げたフォーマットに関する参考文献を一つ以上、必ず記載すること(SIST-02に準拠)

## 第4回レポート課題 (2)

- A4用紙:1ページ以上2ページ以内にまとめる
  - 2ページにわたる場合は裏面に記載のこと
- 課題番号(第4回レポート課題)、提出年月日、学籍番号、所属、氏名を提出用紙の一番上に必ず記入する
- 提出〆切:2013年7月9日(火)17:00
  - 春日エリア学務前のレポート提出BOXへ
- ※不明点等あれば、メールまたは研究室訪問のうえ、担当教員に問い合わせること

# 授業は以上です

- 以前にお知らせした通り、試験は行いません。
- 授業評価アンケートに記入、提出願います。
  - TWINSを用いた自由記入アンケートの入力期間は7月12日(金)までです。
- あわせて、出席票にも提出年月日、学籍番号、所属、氏名、感想コメント等(あれば)を記入のうえ、提出してください。
- なお、第4回レポート課題は成績登録後に返却する予定です。7月16日(火)前後を予定。