

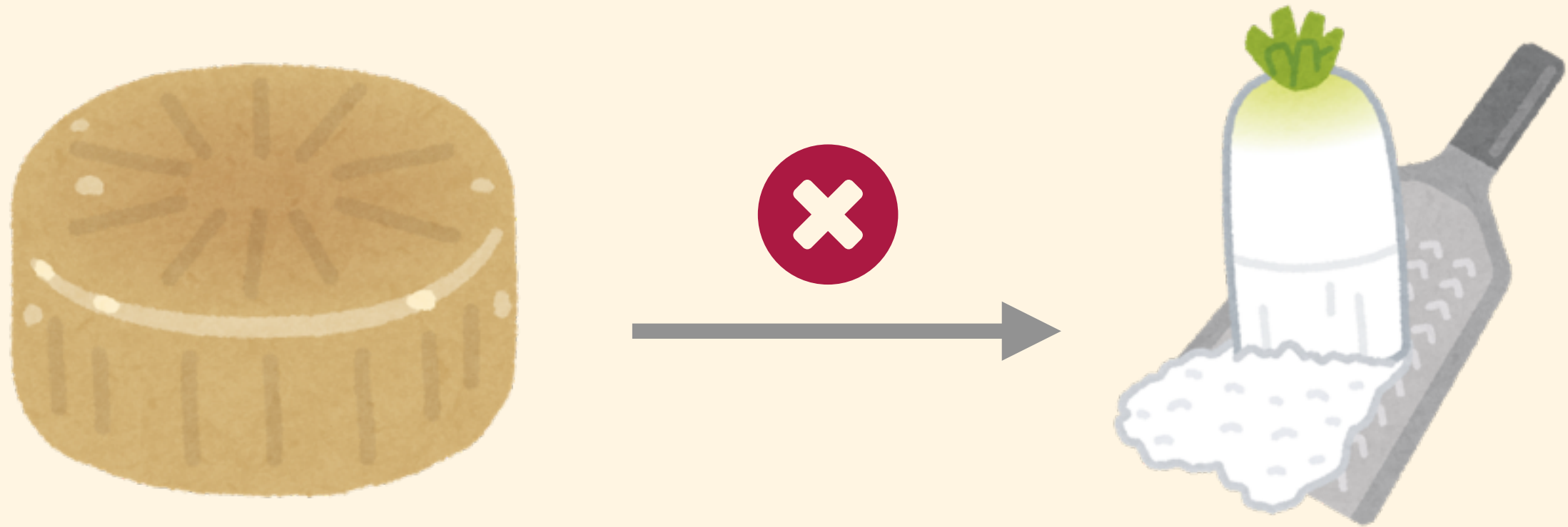
# マークダウンで作る 岩波新書

イースト株式会社  
デジタルドキュメント推進室

高瀬 拓史

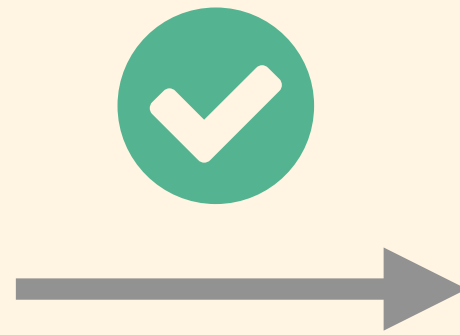
2019.7.31

突然ですが.....



**おでんの大根から  
大根おろしは作れません**

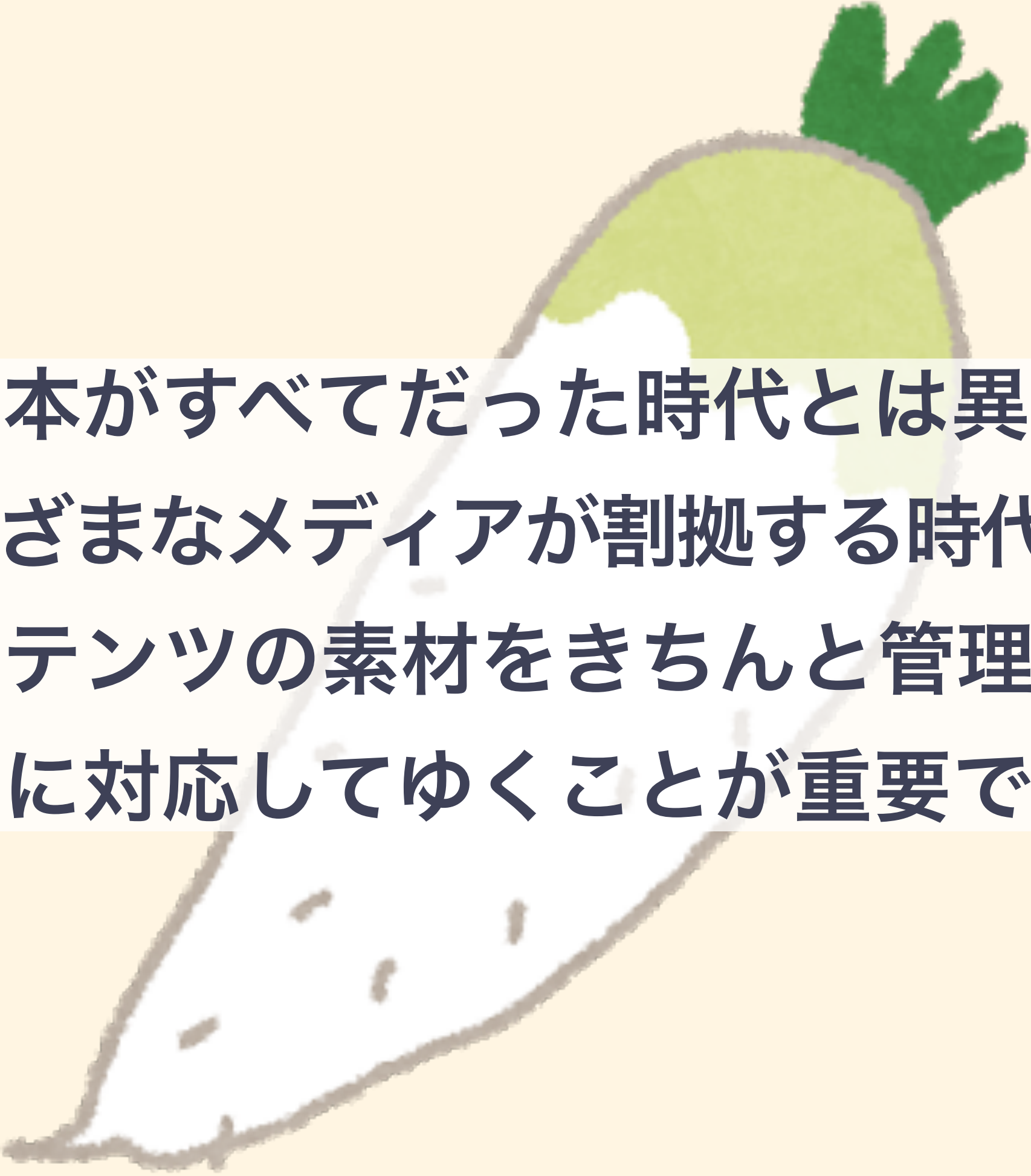
DTP界隈の一部で有名(?)な例え話



大根からはどちらも作れます



分類	素材	料理
例	テキスト、画像	DTPデータ、PDF・EPUBなど
汎用性	用途は特定しない	各メディアの用途に特化
再利用	加工して再利用できる	再利用には向かない
長期保存	高い	制作・閲覧ツールの影響を受ける

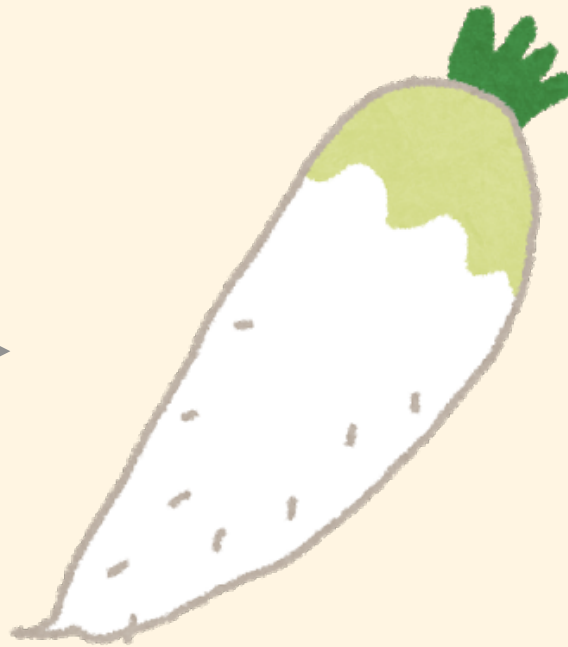


紙の本がすべてだった時代とは異なり、  
さまざまなメディアが割拠する時代には、  
コンテンツの素材をきちんと管理して  
変化に対応してゆくことが重要です

# 弊社 EPUBPack のサービス内容



印刷用PDF



テキスト・画像



EPUB

- 版元様から提供された印刷用PDFをからテキストと画像を抽出。弊社サーバーで素材として管理してゆきます。
- テキストと画像から各ストアに配信可能なEPUBを制作します。

# 岩波新書への導入

## 実現に必要な3つのポイント

### 1. 制作仕様を定める

電書協ガイド、固有ルール、テンプレート

### 2. PDFから構造化テキストを得る

pdf2md (弊社プログラム)

### 3. 構造化テキストを編集してEPUBを作成する

マークダウン、ddconv-core(弊社プログラム)、Git

# 1. 制作仕様を定める

どんなEPUBファイルを作るのか

岩波書店、印刷会社5社、弊社で協議しながら、**岩波新書EPUB制作仕様**を制定した。

**電書協ガイド**・**電書ラボ仕様**をベースに、出版社・レーベル固有のルールを加えて制作仕様とテンプレートを作った。

岩波新書EPUB制作仕様

固有ルール

電書ラボ仕様

電書協ガイド

EPUB 3.0.1



# 用語

## 電書協ガイド（電書協EPUB3制作ガイド）

日本電子書籍出版社協会が作成した一般書向けのEPUB3制作の手引きおよびフレームワーク。大手出版社が採用。

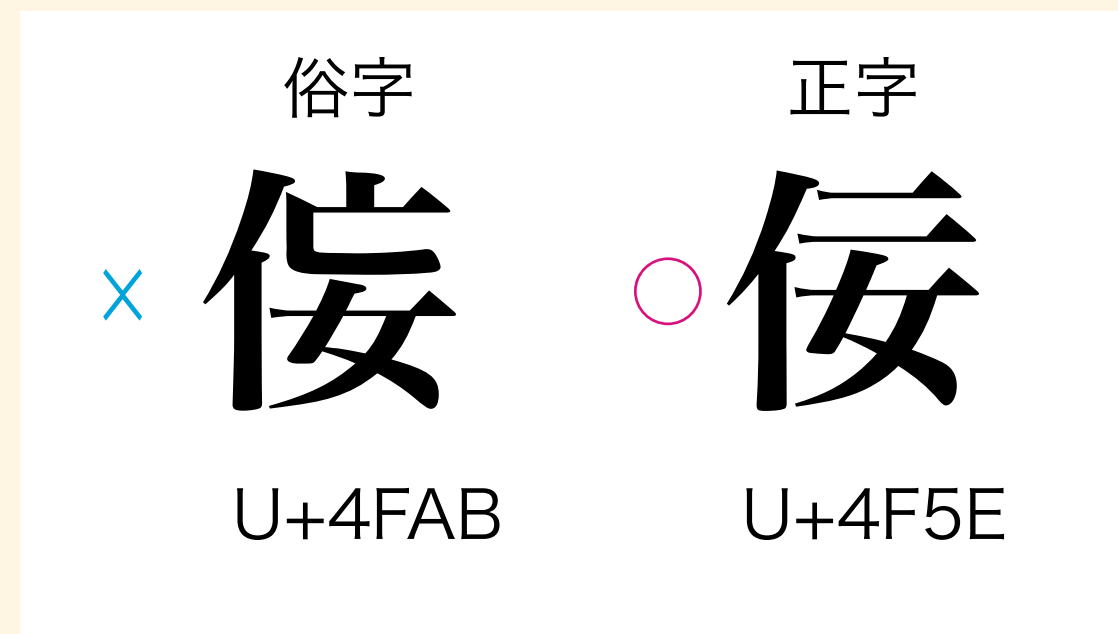
## 電書ラボ仕様

有志による電書協ガイドの詳細化。「緊デジ」のノウハウを継承。

# 固有ルールの例

出版社やレーベルの裁量で決める部分

- 電子化の対象範囲
- 使用する文字・避ける文字
- 画像のフォーマットやサイズ
- EPUBで不可能な表現への対処
- 扉・奥付などの様式
- 著作権表示や注意書きの文言など



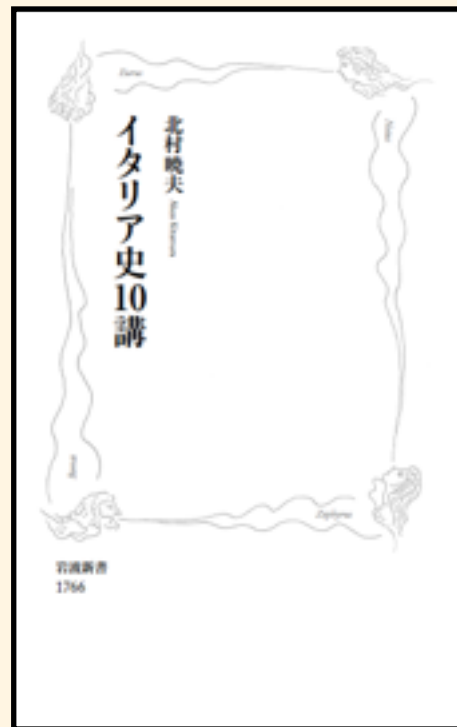
文字の使い方にもルールがある

# 岩波新書の内容区分 (抜粋)

構成を分析してテンプレート化する



表紙



本扉



章扉



奥付

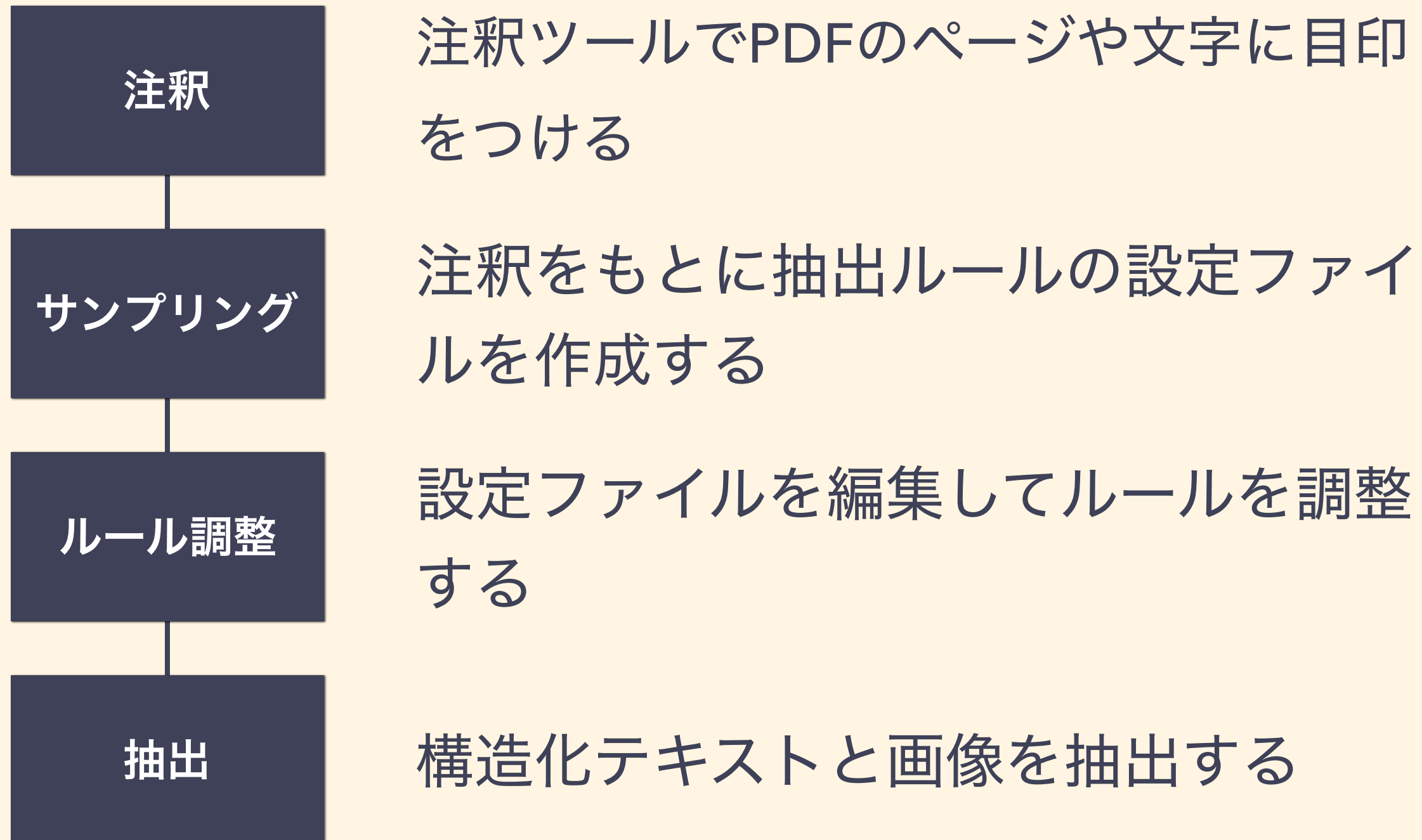


マニフェスト

電書協ガイドを参考に11種類を作成

# 2. PDFから構造化テキストを得る

弊社プログラム「pdf2md」を使用した作業フロー



# pdf2mdの機能

## テキスト抽出などにおける困難を解決

困難	pdf2md
テキストに柱・ノンブルが混じる	除去可能
行末で段落が途切れる	自動で段落を整形
ルビ文字が段落に混じる	構造化されたルビを抽出
見出しなどの文書構造情報が取れない	フォント情報をもとに構造化
圏点、傍線、縦中横などの表現が消える	表現を維持して抽出
図表の抽出が面倒	キャプション+余白ごと自動切り抜き

# PDFにAcrobat DCで注釈をつける

注釈機能とコメントを使って構造化の手がかりを与える

第1講 諸文化と古代ローマ

heading2

start

「イタリア史」を語るということ

1 諸文化の到来と生成

heading1

今日のおよそ半分である。

イタリア半島は、北はアルプス山脈、残り

今日のイタリア共和国は、  
ら主に構成されている。  
五分の四ほどの大きさで、

私が描くその戦争は、  
マ国民と繰り広げ  
したのが憎悪であ  
に戦争を仕掛けた  
支配が押しつけられ

—リウウィウス  
京都大学学

# サンプリング

PDFをpdf2mdに読み込ませて抽出ルールの設定ファイル(YAML)を作る

```
TextClass:  
- Class: Heading  
  Subclass: heading1  
  Rank: 2  
  Attributes:  
  - oo-midashi  
  Font:  
  - TsukuMinPr6N-D, 9.961
```

筑紫明朝Pr6N D、9.961ポイントの文を  
<h2 class="oo-midashi">として抽出するルール

# 抽出

再度PDFをpdf2mdに読み込ませて、  
抽出ルールに基づいたテキストと画像を得る

```
## 1 諸文化の到来と生成 {.oo-midashi}
```

```
### 「イタリア史」を語るということ {.ko-midashi}
```

今日のイタリア共和国は...

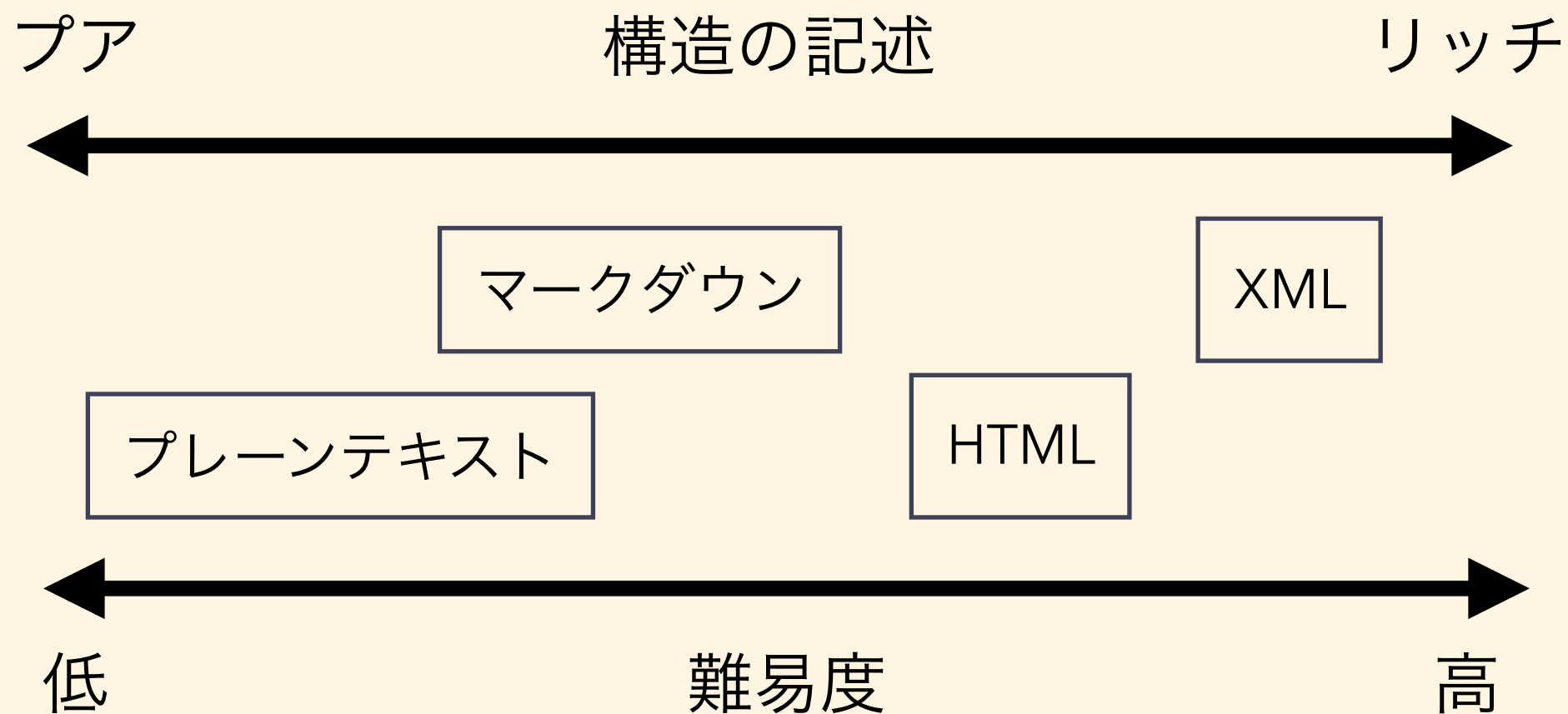
テキストはマークダウン形式（拡張あり）で出力される。



# 用語

## マークダウン

プレーンテキストからHTMLを手軽に生成するための軽量マークアップ言語。書きやすく読みやすい。ブログシステムなどでも使われる。弊社ではルビなどの機能を拡張したでんでんマークダウンを採用。



# 抽出した素材の管理

構成ファイルをバージョン管理システム(Git)に登録する

ファイル名	説明
_ddconv.yml	書誌情報、EPUB設定
01_body.txt	本文（マークダウンテキスト）
_custom.css	作品固有のCSS
cover.jpg	表紙画像
image-xxx.png	その他の画像

リポジトリ名には底本のISBN番号を使用：  
「9784004317661」

# マークダウンファイルの中身

## ローカル変数 + マークダウンテキスト

```
---  
layout: p-tobira  
spread: recto  
slug: 第1講  
---  
  
# 第1講 諸文化と古代ローマ  
<span class="gfont  
font-080per">黎明期～4世紀</  
span> {.tobira-midashi}
```

ローカル変数  
(YAML Frontmatter)  
ページのメタデータを記述

マークダウン  
(でんでんマークダウン)  
ページの内容を記述

# 主なローカル変数

電書協ガイドの要求する細やかなページ設定に対応

変数名	説明	省略可否
layout	テンプレートを指定する	可
stem	出力ファイル名を指定する	可
spread	見開きを指定する	可
slug	内部リンク用の短縮名を指定する	可
linear	spineのlinear属性を指定する	可
isVertical	縦書き／横書きを指定する	可

# 内容区分とローカル変数

## 岩波新書各ページのローカル変数の設定

区分	layout	stem	spread	出力ファイル名
表紙	p-cover	p-cover	—	p-cover.xhtml
口絵	p-image	p-fmatter	—	p-fmatter-xxx.xhtml
本扉	p-titlepage	p-titlepage	recto	p-titlepage.xhtml
注意書き	p-caution	p-caution	—	p-caution.xhtml
目次	p-toc	p-toc	—	p-toc.xhtml
章扉	p-tobira	—	recto	p-xxx.xhtml
本文	p-text	—	—	p-xxx.xhtml
画像ページ	p-image	—	—	p-xxx.xhtml
著者略歴	p-author	p-author	—	p-author.xhtml
奥付	p-colophon	p-colophon	—	p-colophon.xhtml
マニフェスト	p-manifest	p-manifest	—	p-manifest.xhtml

# 3. 構造化テキストを編集してEPUBを作成

## 自動化が難しい箇所を人力で補う

- 見出しランクの調整
  - 整形された段落の確認
  - 縦書きの数字や記号の向き of 補正
  - 抽出できない文字の補完（「■」で出力される）
  - 行アキの補完
  - 細かいスタイルの調整
  - ローカル変数の指定
  - 版元様の制作指示の反映
- など

# 素材とテーマを組み合わせせてEPUB化

弊社プログラム「[ddconv-core](#)」を使用

Gitで管理する素材

岩波新書用テーマ

岩波新書EPUB

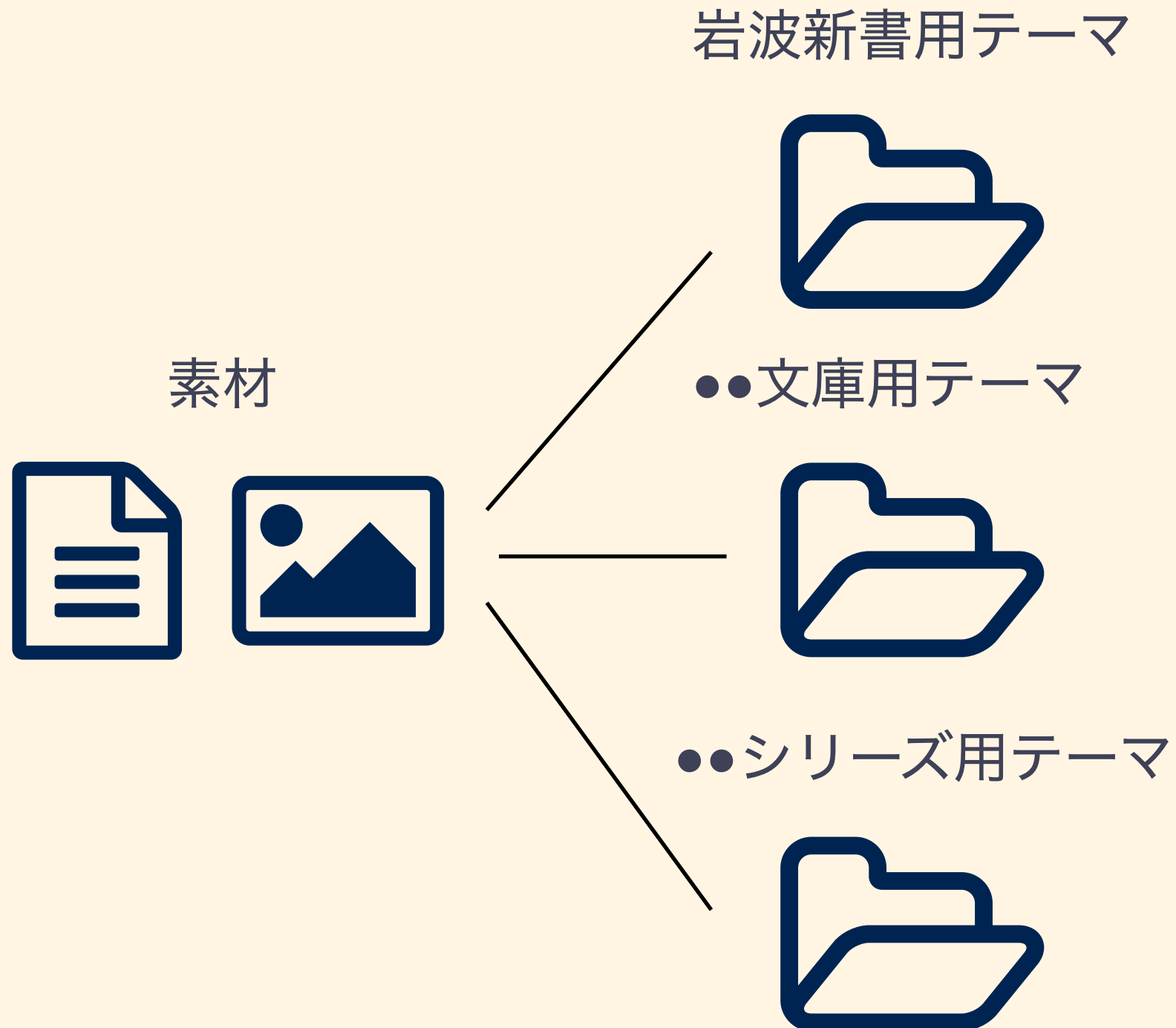


- 書誌情報、EPUB設定
- マークダウンテキスト
- 各種画像
- 作品固有のCSS

- 電書協ガイドCSS
- 岩波新書CSS
- 各種テンプレート
- フォルダ構成ルール
- ファイル命名ルール

# 素材とテーマの分離

さまざまな作品を同じ制作方法で作る。切り替えも自由。





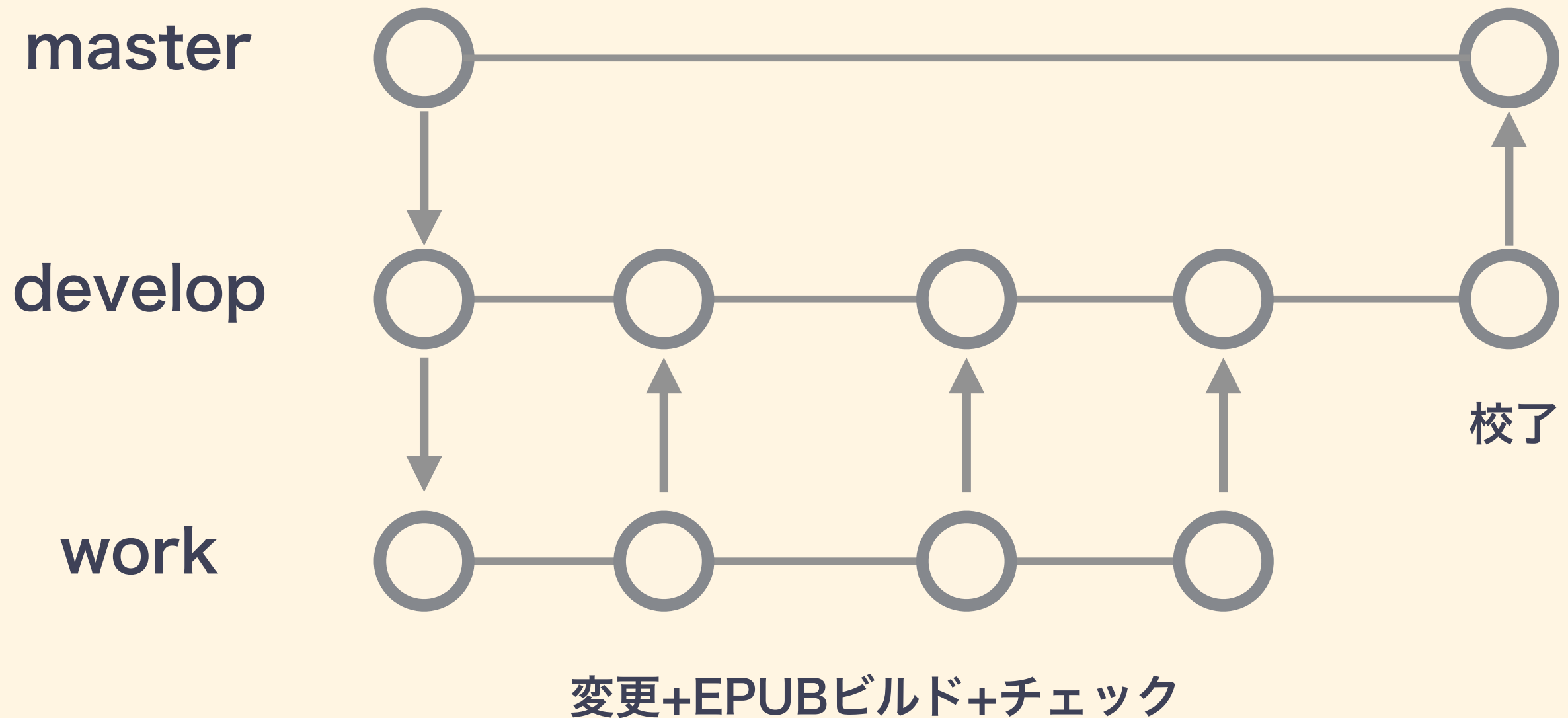
# ddconv-core 2.0

「でんでんコンバーター」のライブラリを大幅に機能強化

- 対応メディアタイプ
  - ビデオ・オーディオ、VTT・TTML（字幕）、PLS、SMIL、OpenType、Woff、Woff2
- テーマ機能
  - デフォルト／電書協ガイド
- マークダウン
  - YAML Frontmatter、拡張テーブル記法、後注記法、内部リンク記法
- アクセシビリティメタデータ
- 固定レイアウト＋リフロー混在
  - めがイラスト

# Gitによるバージョン管理

変更履歴を記録しながらコンテンツを制作する



# アーキテクチャ

## EPUB作成、EPUB送信、検証、通知を自動化

バージョン管理システム



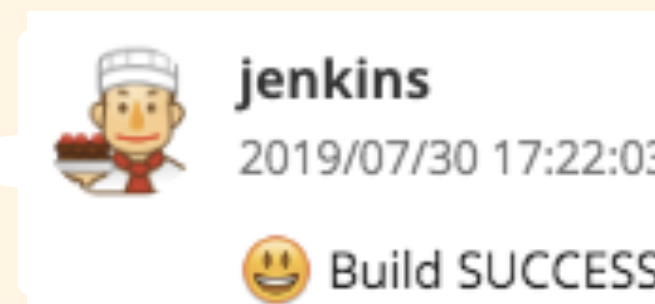
- ① リポジトリに変更が発生すると Jenkinsサーバーに通知を送信する

自動化ツール



- ② リポジトリのクローン
- ③ EPUBをddconv-coreでビルド
- ④ EPUBをBacklogのファイル共有に保存
- ⑤ EpubCheckによる検証
- ⑥ 一連の結果をBacklogに投稿

プロジェクト管理ツール



# まとめ

## EPUBPackのコンテンツ管理とEPUB制作をご検討ください！

- コンテンツは加工や再利用に適した形式で保存し、変化に備えることが重要。
- 岩波書店様の電子書籍への理解と的確な制作指示のおかげでスムーズな制作の流れを構築できた。
- pdf2mdによって印刷用PDFから精度の高い構造化テキストが抽出できることが確認できた。
- マークダウンを身につけて貰うことで、EPUB、HTML、CSSの専門知識がないメンバーでも制作が可能に。生産体制を拡大できた。
- 制作環境の整備とマニュアル化をすすめ、より多くの作品を電子化して送り出してゆきたい。