

# WEKO3 simple loader

～表形式メタデータをJSONパスにマッピング～

2021年9月code4lib japanライトニングトーク

東京大学情報システム部・前田 朗

WEKO3 simple loaderとは？

# WEKO3 simple loaderの意義

- WEKO3ユーザにとって
  - WEKO3のアイテム一括登録の補助ツール
- 図書館系技術者にとって
  - 表形式メタデータをJSONパスにマップするオープンソースコード
  - この仕様でどの程度の表現力があるか論理的に説明できれば論文にできるかも

# コードと詳細解説

- GitHubでコード配布
  - [https://github.com/maedaak/WEKO3\\_simple\\_loader/](https://github.com/maedaak/WEKO3_simple_loader/)
  - MITライセンス
- 詳細解説
  - [https://github.com/maedaak/WEKO3\\_simple\\_loader/blob/main/README.md](https://github.com/maedaak/WEKO3_simple_loader/blob/main/README.md)

# WEKO3とは

- 機関リポジトリソフトWEKOの新バージョン
  - <https://rcos.nii.ac.jp/service/weko3/>
- WEKOは機関リポジトリサービスJAIRO Cloudで採用
- JAIRO Cloudでは、WEKO2からWEKO3にリプレイス中
  - 東京大学を含む9機関で先行移行(2021年9月現在)

# WEKO3の一括登録をより簡単に

- WEKO3ではツリー構造をもつメタデータに対応
  - 一括登録メタデータは表形式(TSV)で記載
  - ツリー構造に対応すべく表のカラムがJSON形式のパスに
- WEKO3 simple loaderでは簡易な表形式に
  - WEKO2(旧)の一括登録フォーマットをイメージ
  - メタデータの表現力は若干落ちる

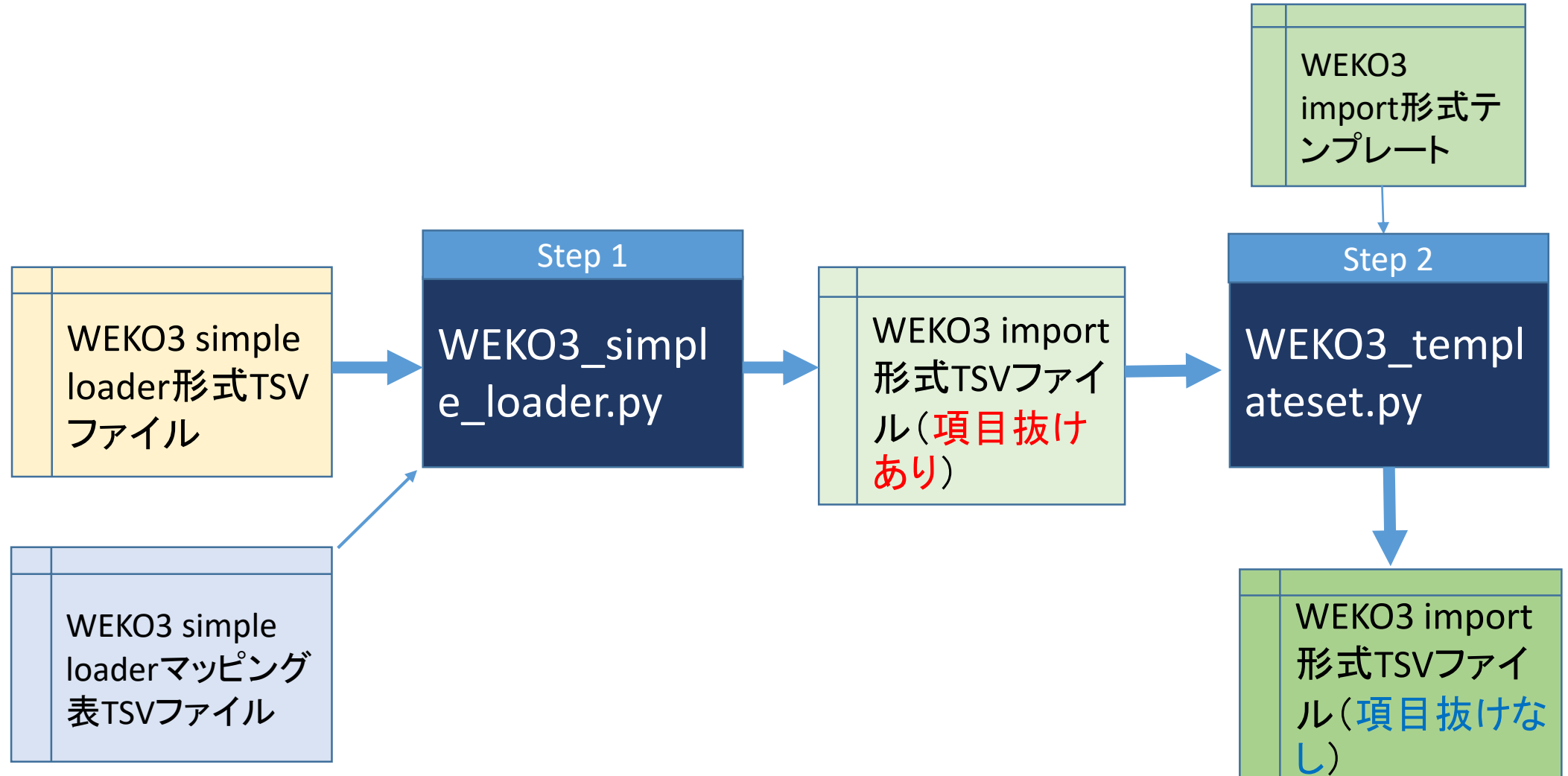
# プログラムの実行

# 事前準備

- Python 3の実行環境
- WEKO3 simple loaderのダウンロード
- テンプレートファイルの用意
- マッピング設定の用意



# 全体フロー(2つのプログラムを実行)

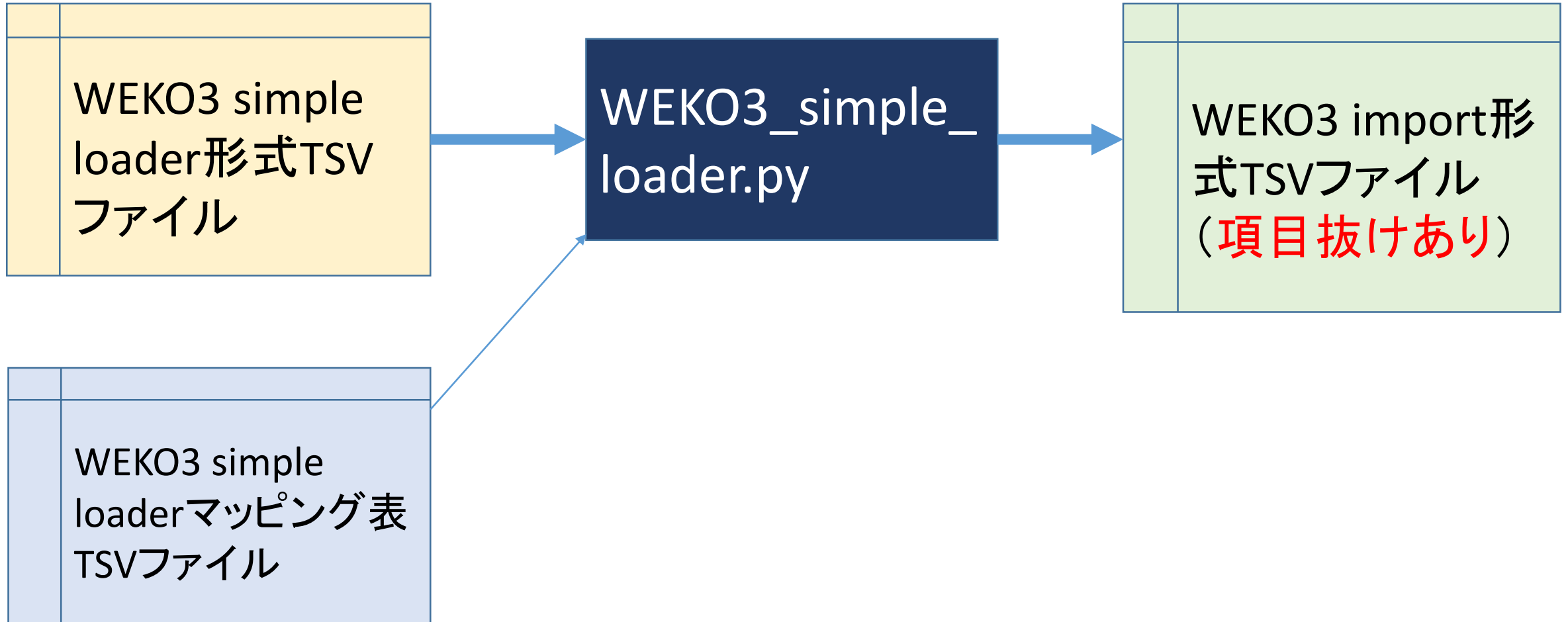


# 2つのプログラム

- WEKO3\_simple\_loader.py
  - 簡易な表形式(TSV)メタデータをJSON形式パスにマッピング
  - マッピングルールはマッピング表で定義する必要がある
- WEKO3\_templateset.py
  - WEKO3\_simple\_loader.pyの出力を、WEKO3の一括登録テンプレートにセット
  - 一括登録テンプレートには一括登録で必要とするすべてのJSONパスを記載しておく
  - WEKO3\_simple\_loader.pyでは値がない項目は出力されないが、WEKO3一括登録時には項目不足のエラーがでるため処理を追加

WEKO3\_simple\_loader.py

# WEKO3\_simple\_loader.py



# WEKO3 simple loaderのメタデータ様式(1部)例

カラム名	データサンプル
タイトル	Mastring Open Refine
タイトル_言語	en
タイトル	OpenRefineをマスターする
タイトル_言語	ja
本文言語	jpn
資源タイプ	departmental bulletin paper
資源タイプURI	
著者名	山田, 次郎 田中, 三郎
著者名_言語	ja ja
著者名	Yamada, Jiro Tanaka, Saburo
著者名_言語	en en
著者所属	東京大学情報基盤センター図書館電子化部門

# メタデータマッピング表の基本

カラム名	説明文
name	簡易一括登録フォーマットのカラム名を記載します。
config	「Allow Multiple」を設定すると繰り返し項目として扱われます。
path1	WEKO3のJSON形式パスを指定します。末尾の値の指定は削除ください。繰り返し項目の場合は繰り返しの起点までです。その際は繰り返しの添え字記述(例"[0]")を削除ください。
path2	path1の繰り返しの起点のあとにさらにJSON形式パス(値の記述を除く)があれば、記載します。
value	JSON形式パス中の値を示す部分パスを記載します。path1、path2で完結する場合は、#を記入ください。
attrib1	JSON形式パス中の属性1を示す部分パスを記載します。形式は「属性名="属性の固定値"」とします。
attrib2	JSON形式パス中の属性2を示す部分パスを記載します。形式は「属性名="属性の固定値"」とします。

# WEKO3の一括登録用TSVカラム(一部)例

JSON形式パス	メタデータ項目名
.metadata.item_titles[0].subitem_title	タイトル[0].タイトル
.metadata.item_titles[0].subitem_title_language	タイトル[0].言語
.metadata.item_titles[1].subitem_title	タイトル[1].タイトル
.metadata.item_titles[1].subitem_title_language	タイトル[1].言語
.metadata.item_language[0].subitem_language	言語[0].言語
.metadata.item_keyword[9].subitem_subject_uri	キーワード[9].主題URI
.metadata.item_resource_type.resourcetype	資源タイプ.資源タイプ識別子
.metadata.item_resource_type.resourceuri	資源タイプ.資源タイプ
.metadata.item_creator[0].creatorAffiliations[0].affiliationNameIdentifiers[0].affiliationNameIdentifier	著者[0].作成者所属[0].所属機関識別子[0].所属機関識別子
.metadata.item_creator[0].creatorAffiliations[0].affiliationNameIdentifiers[0].affiliationNameIdentifierScheme	著者[0].作成者所属[0].所属機関識別子[0].所属機関識別子Scheme
.metadata.item_creator[0].creatorAffiliations[0].affiliationNameIdentifiers[0].affiliationNameIdentifierURI	著者[0].作成者所属[0].所属機関識別子[0].所属機関識別子URI
.metadata.item_creator[0].creatorAffiliations[0].affiliationNames[0].affiliationName	著者[0].作成者所属[0].所属機関名[0].所属機関名
.metadata.item_creator[0].creatorAffiliations[0].affiliationNames[0].affiliationNameLanguage	著者[0].作成者所属[0].所属機関名[0].言語

# 値と属性をひとつのカラムに

- 属性は「固定値」で2つまで
- 例：
  - 簡易フォーマット項目:「抄録」
  - JPCOARスキーマ項目:「内容記述」、属性1:「"Abstract"」



# カラムの値繰り返しが可能

- 同名のカラム名を使うと繰り返しになる
- セル内で繰り返しを表現できる
  - 繰り返し項目の区切り記号は“|”（変更可）を使用

# 「桁合わせ」により項目同士の関連付け

- 次の例では「用語」に"ja"が、「Word」に"en"が関連付けられます。「???'は桁合わせにより「関連付けなし」になります。

キーワード	キーワード_言語
用語 ???' Word	ja  en

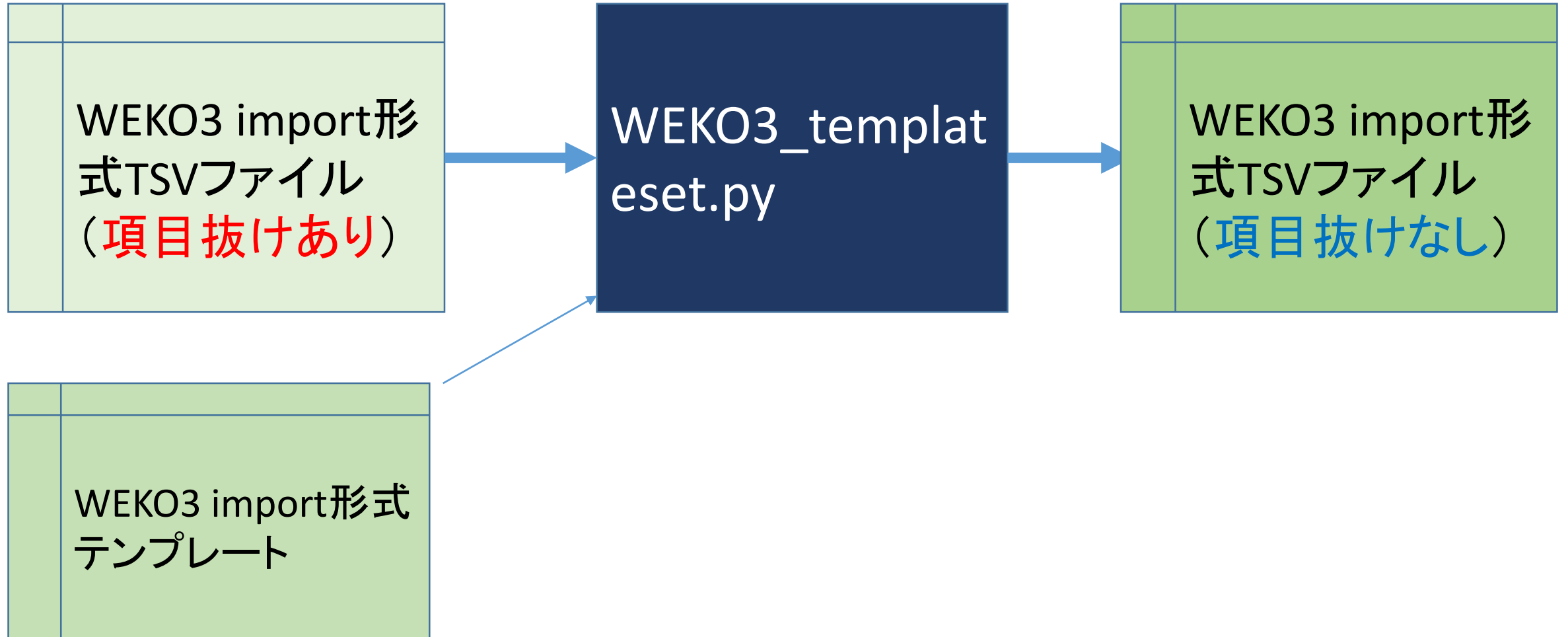
# repeat() 関数

- 直前の項目に関連し、すべて固定値をセットするには、repeat()関数
- 次の例では「用語」「名詞」「形容詞」「副詞」「動詞」が全て“ja”に関連付けられる。

キーワード	キーワード_言語
用語   名詞   形容詞   副詞   動詞	repeat("ja")

WEK03\_templateset.py

# WEKO3\_templateset.py



# WEKO3 import機能で一括登録

- 出力TSVを必要に応じ修正
  - WEKO3 simple loaderではセットできない値を入れることもできる
- TSVを本文ファイルと合わせzip圧縮
- WEKO3管理画面からimport機能で一括登録

ご清聴ありがとうございました