

標準化と多様化： 生成AIが橋渡しする「柔らかいメタデータ」の可能性

北本 朝展（KITAMOTO Asanobu）

国立情報学研究所

ROIS-DS人文学オープンデータ共同利用センター（CODH）

<https://researchmap.jp/kitamoto/>

kitamoto@nii.ac.jp

自己紹介

<https://researchmap.jp/kitamoto/>

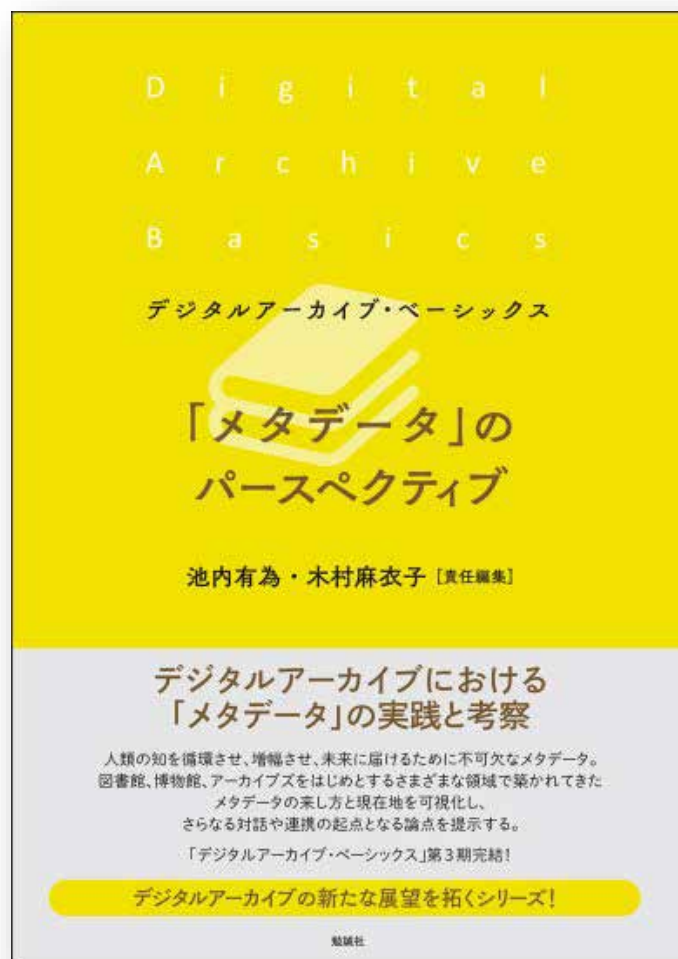


@kitamotoasanobu

気象予報士第5439号

- 北本朝展（きたもと あさのぶ）
- 国立情報学研究所 教授、ROIS-DS人文学オープンデータ共同利用センターセンター長
- 情報学、**人文情報学**、データ駆動型サイエンス（**地球科学**・防災等）
- **オープンサイエンス**：超学際的な共同研究の展開

メタデータとオープンサイエンス



巨人の肩とデジタルアーカイブ・オープンサイエンス運動 ーメタデータが円滑化する知の積み上げ

デジタルアーカイブ・ベーシックス「メタデータ」のパーспекティブ, 池内有為／木村麻衣子 責任編集 (編), pp. 339-360, 勉誠出版, ISBN 978-4-585-30306-0, 2025年6月

- **巨人の肩 (Shoulders of Giants)** : 人類の知の積み上げに関する有名な言葉
- **デジタルアーカイブ (DA)** : 肩の上に立つ手段を提供
- **オープンサイエンス (os)** : 肩の上に自分の貢献を積み重ねる手段を提供

メタデータの論点

- 知的生産物の粒度
- 利用者の類型
- 項目の類型
- 専門性の類型
- 付与の自動化
 - システムによる自動化
 - AIによる自動化
- 再利用性

利用者の類型

1. **自分（自組織）**：独自のメタデータ形式でもかまわないが、最初から相互運用性を想定しておいた方が技術的負債になりづらい
2. **一般的利用者**：流通を目的とするメタデータは、汎用的な事項を対象として、相互運用性を考慮した形式に
3. **専門的利用者**：より詳細な項目が対象、相互運用性については分野標準のメタデータ形式が存在すれば採用し、なければ独自に作成する

項目の種類

1. **基本メタデータ**：データに関する基本的な情報。タイトル、作成者や連絡先、ライセンス、データ形式など
2. **内容メタデータ**：データの内容に関する情報。説明、キーワード、要約、統計情報など
3. **来歴メタデータ**：データの品質や信頼性に関する情報。データの作成経緯や作成ツールなど
4. **管理メタデータ**：データの管理に関する情報。担当者、所蔵者、管理状態など
5. **利用メタデータ**：データの利用に関する情報。ユースケース、利用者、引用情報など

専門性の類型

1. **データライブラリアンの専門性**：客観的な基本メタデータや識別子をきちんと付与し、データに内在する基本価値を引き出す
2. **データキュレーターの専門性**：利用者視点から見たデータの新たな価値を発見し、データの潜在的な利用価値を引き出す
3. **データアーキビストの専門性**：現在のデータを未来に継承することで、データの長期的な価値を引き出す

付与の自動化

1. システムによる自動化

- **使い回し**：1回入力したメタデータを再利用（DOI連携等）
- **記録の自動化**：デジタルプラットフォーム上で作業推進

2. AIによる自動化

- **統制的な内容メタデータ**：AIにルールを教えることで、候補から選択するプロセスの負荷を軽減
- **非統制的な内容メタデータ**：AIが下読みして下書きを生成することで、最初から入力するプロセスの負荷を軽減
- **特にメタデータに「正解」がない場合は有効**

メタデータは目的か手段か？

目的

1. メタデータはデータの学術的説明であり、専門知が詰まっている
2. 解題などデータの周辺知識も含めた著作物として書く場合もある
3. ライブラリアンなど

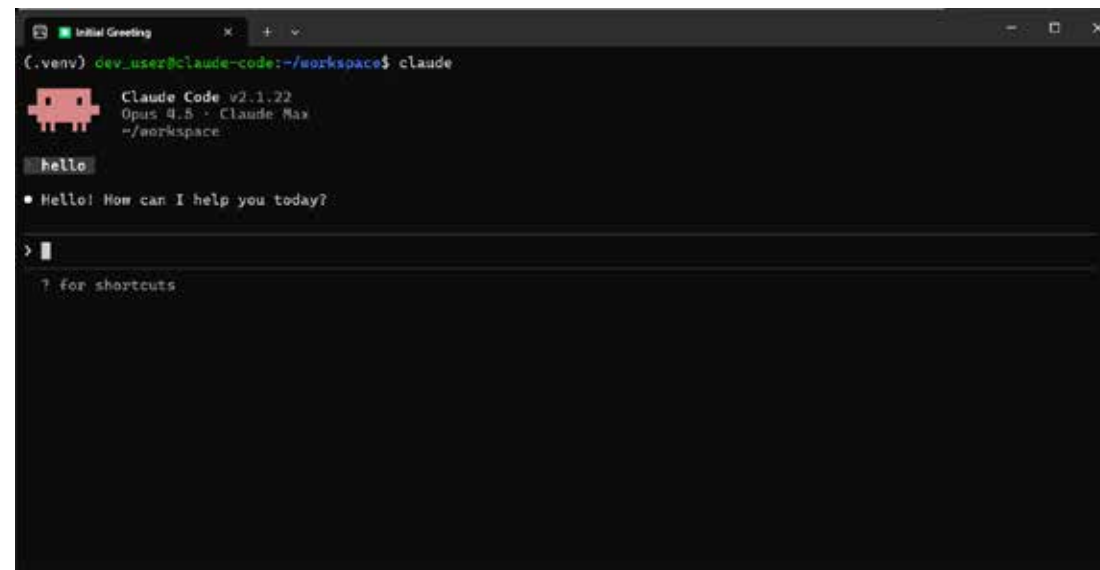
手段

1. メタデータはデータを探すための補助情報であり、単体では使わない
2. メタデータはあくまでデータの付属物であり、検索できれば構わない
3. 研究者など

生成AIの登場



- Anthropic Claude
- Webチャット
- 主に調査・計画



- Anthropic Claude Code
- コーディングエージェント
- 主にソフトウェア開発

堅いメタデータと柔らかいメタデータ

堅い（固定的な）メタデータ

Hard Metadata

1. 標準的な共通スキーマに基づき作成する
2. 標準化に基づき相互運用性を担保する
3. 標準化のための合意には多大なコストがかかる
4. 標準化できない情報は脱落してしまう


柔らかい（可塑的な）メタデータ

Soft Metadata

1. 目的によってスキーマを柔軟に使い分ける
2. 相互運用性はAIの橋渡しでなんとかする
3. 事前の合意は不要であり、利用者が最適化できる
4. 様々な情報を取り込み多様化を推進できる

つくし堂

AIによるメタデータ生成



国文学研究資料館 大型プロジェクト（DDHプロジェクト）

<https://lab.nijl.ac.jp/humanitiesthroughddps/>

**300,000点の日本古典籍
（1868年以前）を、デジ
タル化し、オープンデー
タとして公開**

日本文化のビッグ
データをどのよう
に活用するか？

ディープラーニング
によるくずし字認識
手法KuroNetを開発
1枚の画像の認識に約
1秒
条件が良ければ精度
95%

三てう殿に殿きたのかたならひておはし
ます御たいまいれりしううちよりまう
てたまへりくにくのしやうよりたうき
ぬぬのなともてまいれり御いそきのれう
にとてあやうす物かとりきぬなとお
ほく奉れたれはみくしけのする人御
まへまてはからひきたむそめくさ何くれの
としやうこのものともは一てう殿にもわかち
奉り給おはする事はなければ御かた
におほしなけきまくにきおとろかし



つくし堂

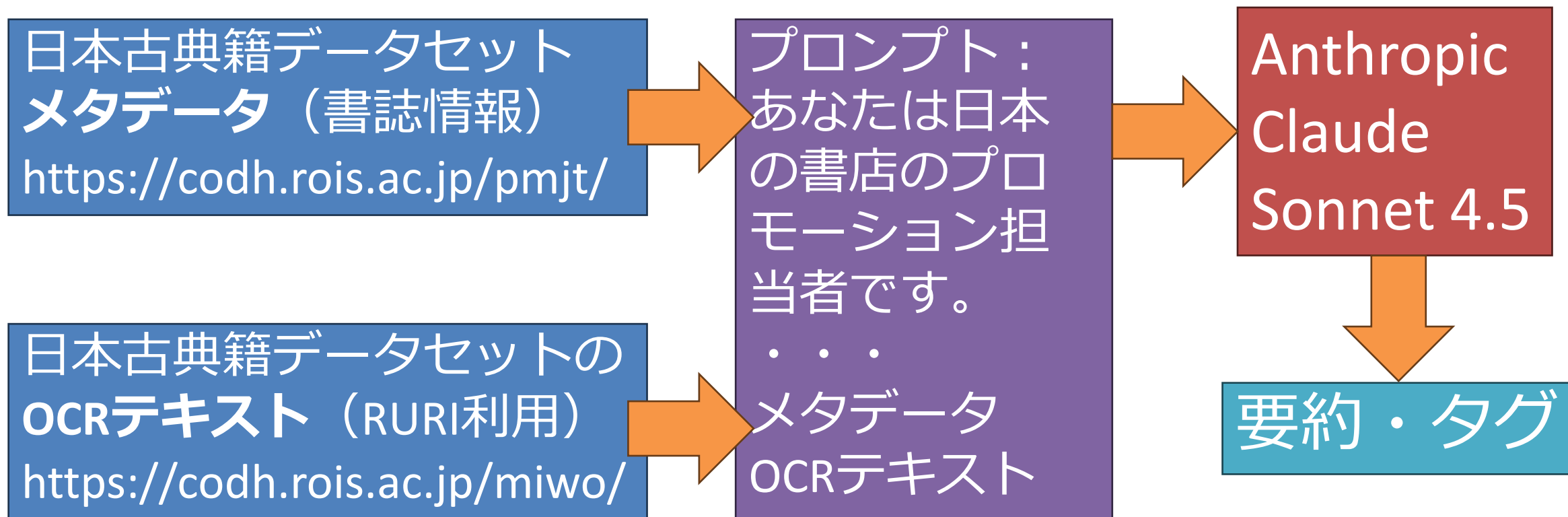
<https://codh.rois.ac.jp/tsukushi/books/>

2025年12月公開

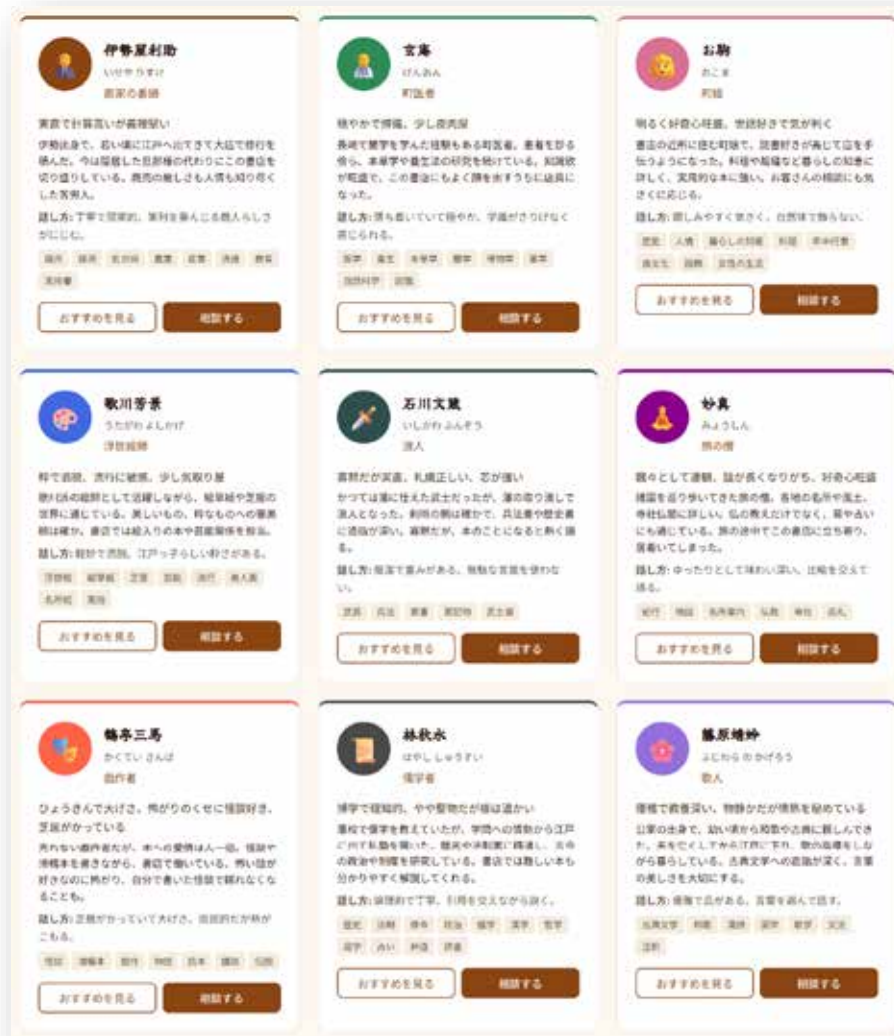


1. 古典籍を親しみやすく提供する「AI生成書店」
2. 古典籍の要約をOCRテキストとメタデータから生成
3. 古典籍のキャッチコピーを要約と店員ペルソナから生成
4. 現代日本語を読みながら、古典籍を発見できるサービス

要約・タグ生成



店員生成・担当割当



1. 書店の「店員」の名前、職業、ペルソナをAIが生成
2. 各店員の得意分野を表すキーワードをAIが生成
3. 割当が均等になるよう、AIがキーワードを調整
4. 人数を7人から9人に増やす作業にもAIは楽々対応

キャッチコピー生成

自讃歌／苔筵



藤原蜻蛉
歌人

中世の言の葉が織りなす、風情深き連歌の世界

天文5年に平雅胤によって書写された、中世和歌の自筆本です。「自讃歌」と連歌集「苔筵」の2作品を収めています。17×23センチの横本で、「月明荘」の印記が伝わります。心敬の句集や歌人たちの作品が豊富に含まれており、恋・四季・旅・人生の哀歓を詠んだ和歌や連歌の付句が多数収録されています。当時の文学表現や歌風を知ることのできる写本として、中世文学研究の重要な参考資料となる一冊です。

写本

和歌

連歌

中世文学

天文時代

👁 画像を閲覧する

🗨 つくしチャットで読む

📖 書誌情報を見る 🔗

1. 要約と店員のペルソナをプロンプトに入れ、**キャッチコピー**をAIが生成
2. **書店のPOP風表示**も、プロンプトに沿ってAIが生成

関係記述

以下の本もいかがですか？

壁草

天文5年書写の中世歌風研究資料と室町末期連歌集の比較により、中世和歌から連歌への文学表現の変遷が追跡できる

井蛙抄

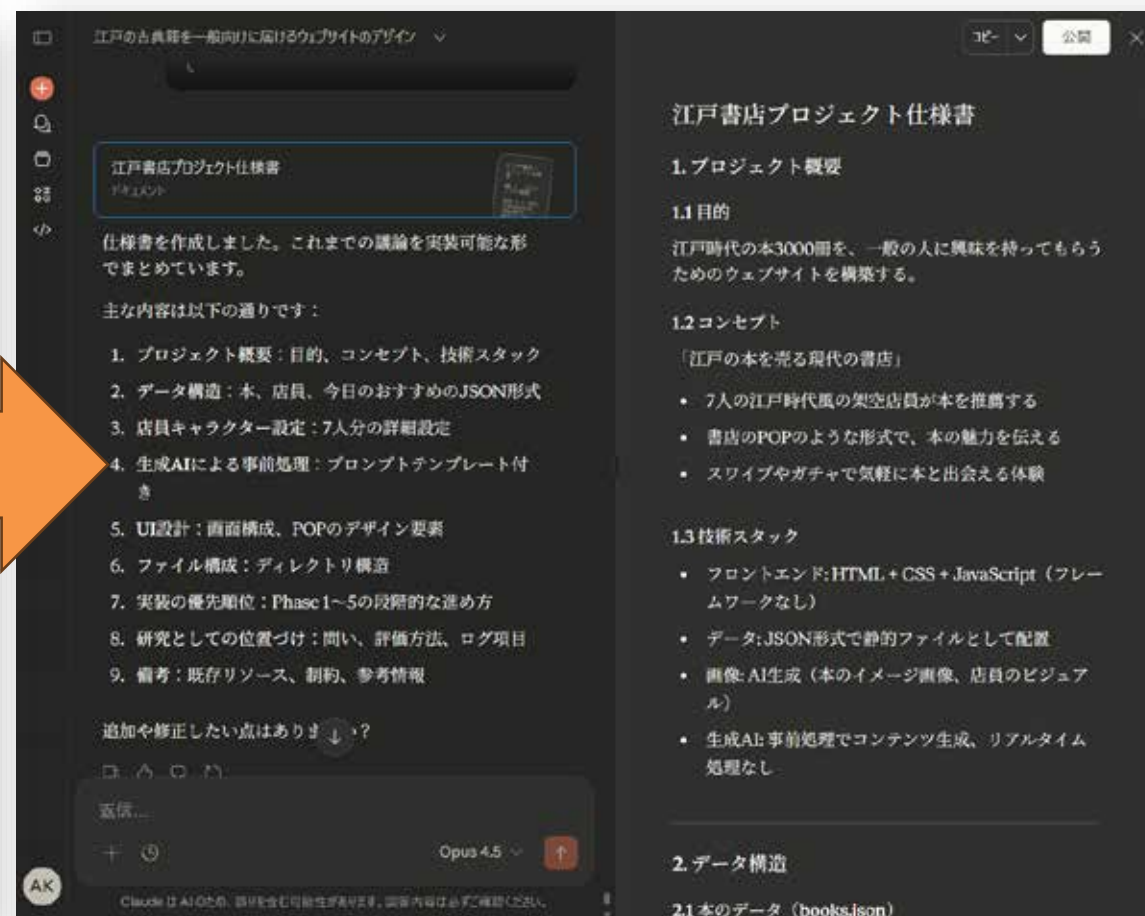
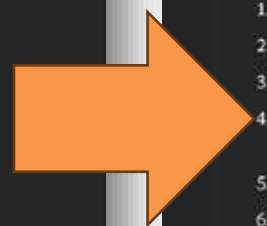
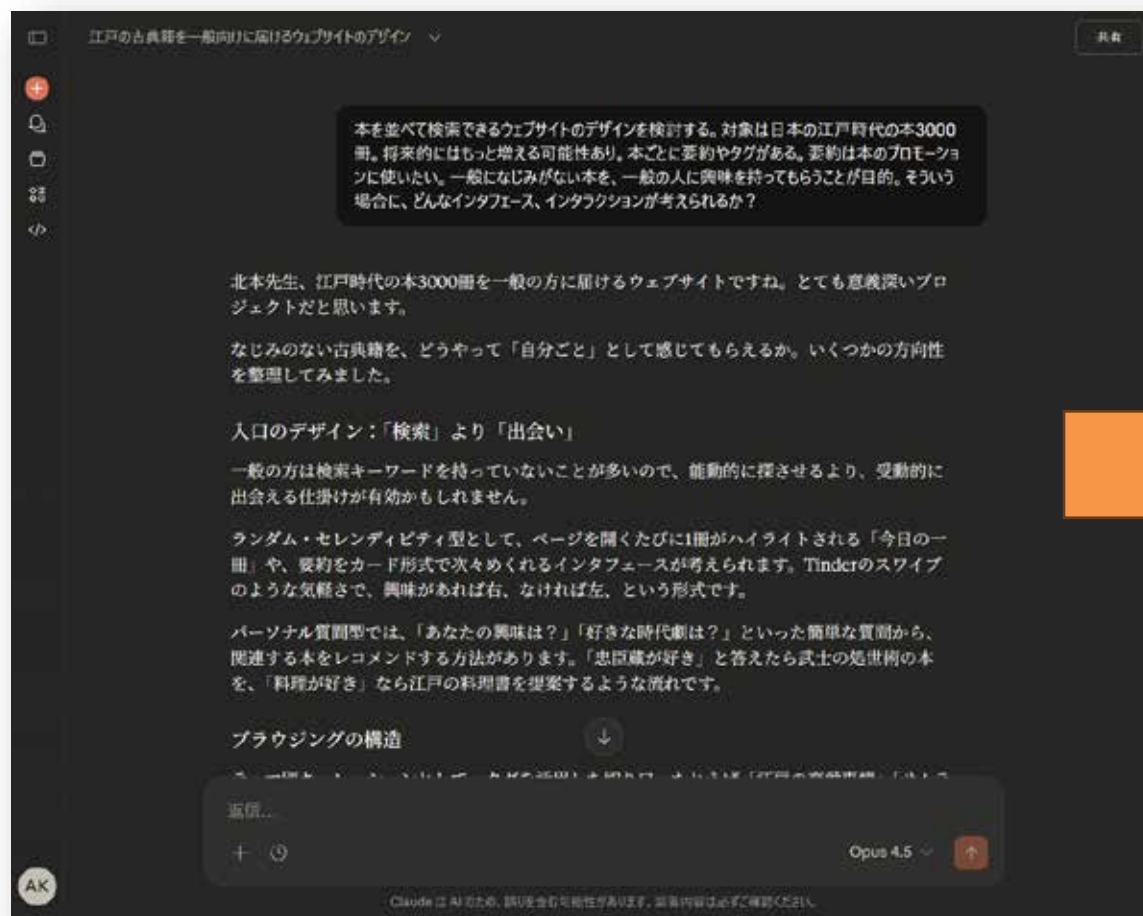
平雅胤による実作写本と宗祇による理論書を対比することで、中世連歌の創作実践と歌学理論の関係性が明確になる

徒然草

同時代の和歌・連歌実作集と兼好法師の文学観を照合することで、中世文学における韻文と散文の美意識の違いが浮き彫りになる

1. タグが共通する2点の要約をプロンプトに入力し、2点の関係記述をAIが生成
2. もっともらしい文章が生成されるため、もはや幻覚かどうかの確認も難しい

計画→仕様書 (Claude: Opus 4.5)



仕様書→実装 (Claude Code: Opus 4.5)

```
1  # 江戸書店プロジェクト仕様書
2
3  ## 1. プロジェクト概要
4
5  ### 1.1 目的
6  江戸時代の本3000冊を、一般の人に興味を持ってもらうためのウェブサイト構築する。
7
8  ### 1.2 コンセプト
9  「江戸の本を売る現代の書店」
10 - 7人の江戸時代風の架空店員が本を推薦する
11 - 書店のPOPのような形式で、本の魅力を伝える
12 - スワイプやガチャで気軽に本と出会う体験
13
14 ### 1.3 技術スタック
15 - フロントエンド: HTML + CSS + JavaScript (フレームワークなし)
16 - データ: JSON形式で静的ファイルとして配置
17 - 画像: AI生成 (本のイメージ画像、店員のビジュアル)
18 - 生成AI: 事前処理でコンテンツ生成、リアルタイム処理なし
19
20 ---
21
22 ## 2. データ構造
23
24 ### 2.1 本のデータ (books.json)
25 --- json
26 {
27   "id": "book_0001",
28   "title": "本のタイトル",
29   "summary": "要約 (約300字)",
30   "category_code": "国文研分類コード",
31   "category_name": "分類名 (一般向け表示用)",
32   "staff_id": "担当店員ID",
33   "catchcopy": "キャッチコピー (20-30字)",
34   "comment": "推薦コメント (約100字)",
35   "image_path": "images/books/book_0001.png",
36   "connections": [
37     {
38       "book_id": "book_0123",
39       "reason": "繋がる理由 (発見につながる説明)"
40     }
41   ]
42 }
43 ---
```



```
dev_user@a10680747f46:/wcr x + v
Welcome to Claude Code

CLAUDE

Claude Code can now be used with your Claude subscription or billed based on API usage through your Console account.

Select login method:

1. Claude account with subscription
   Starting at $20/mo for Pro, $100/mo for Max - Best value, predictable pricing

2. Anthropic Console account
   API usage billing

[0] 0:claude* "49eb466a7b2c" 02:12 01-Sep-25
```

全体のワークフロー

1. ウェブ版Claudeを利用して計画を開始。目標は仕様書を作成し、マークダウン形式で保存すること
2. 様々な視点でAIと議論してアイデアを洗練させ、十分に詳細な仕様書を作成。**時間配分40-50%**
3. Claude Codeに仕様書を読み込ませて実装を開始。数分から数十分で初期実装が完成。**時間配分10-20%**
4. エラー処理や文言の微修正を重ね、**システムの完成度を高めていく。時間配分40-50%**

AI生成メタデータ

1. 要約、キャッチコピー、タグなどのAI生成メタデータを書籍IDごとに付与
2. メタデータは標準化されていないが、現代人にとって読みやすいという明確な価値がある
3. 3000点以上の古典籍にキャッチコピーを付与することは、人間の専門家にとっては現実的でない
4. 正解がないメタデータに対しては、AI生成メタデータはゲームチェンジャーになる

Mahalo Button

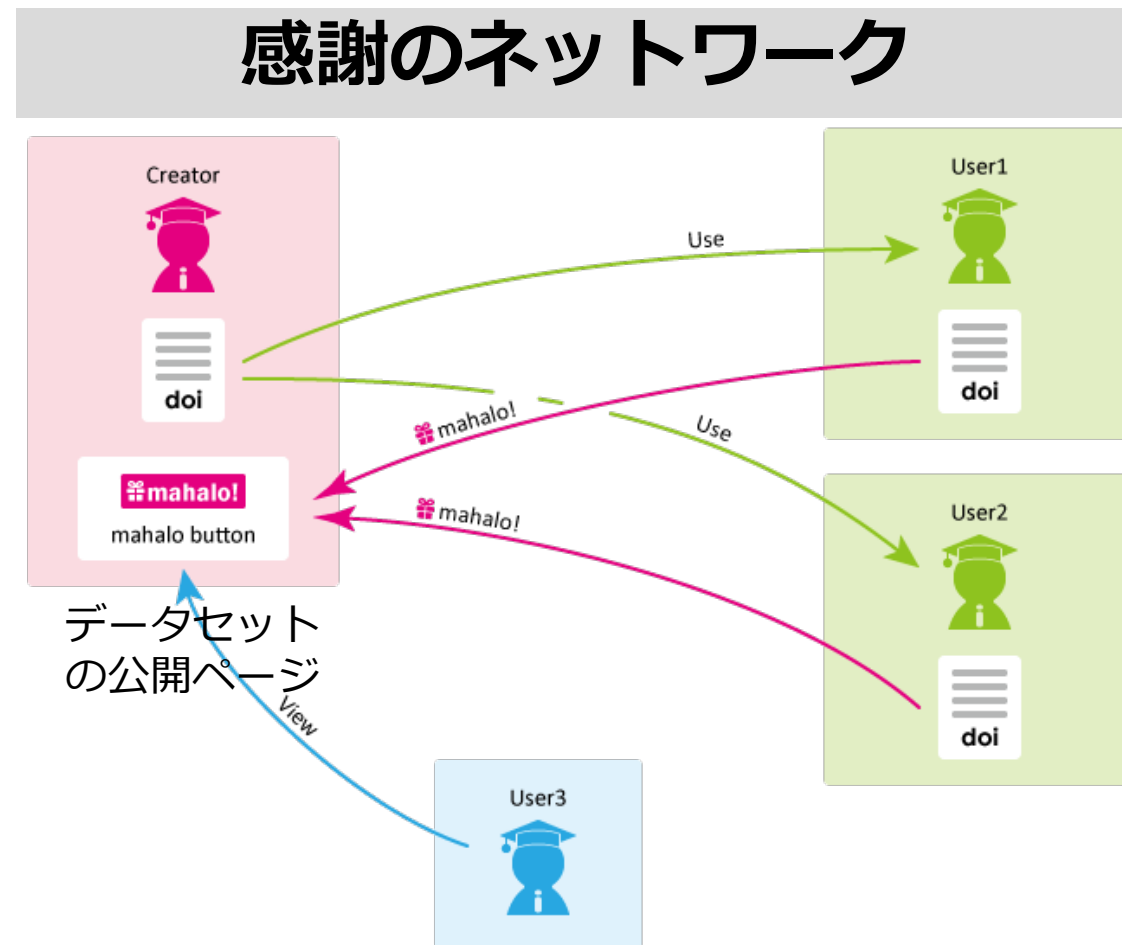
メタデータをデータセット利用事例に拡張

Mahalo Button

<https://mahalo.ex.nii.ac.jp/>



1. Mahalo Button = データセット公開ページに設置するボタン
2. **データ利用者**や**データ公開（管理）者**：データ利用事例を登録
3. **潜在的データ利用者**：データ利用事例から学ぶことで、新たなデータ利用者となる
4. ボタンで利用事例数を集計し、データ作成者の業績とする



DIASとMahalo Button

<https://diasjp.net/information/topics/20211116/>

Show Mahalo Cards

Global Soil Wetness Project Phase 3 Atmospheric Boundary Conditions (Experiment 1)

DOI: 10.20783/DIAS.501

URL: https://research.diasjp.net/dataset/GSWP3_FKPI_Forcings

Select All

Keyword

Filter

sort : Latest | Like

Identifying and quantifying the impact of climatic and non-climatic drivers on river discharge in Europe

DOI: [10.22541/essoar.171018485.57171785/v1](https://doi.org/10.22541/essoar.171018485.57171785/v1)

The study uses three climatic datasets to drive the Land Surface Model (LSM) for analyzing river discharge in Europe. The main dataset, GSWP3, spans 1901-2012 at a 3-hourly resolution and is bias-corrected using GPCC and CRU data. Two additional datasets, WFDEI-GPCC and E2OFD, covering 1979-2014, are used to test the robustness of the results and the sensitivity to different climate forcings.

LSM GSWP3 WFDEI-GPCC E2OFD river discharge

AI-Generated

Citation

3 months ago by DIAS Office dias-office@diasjp.net

🔍 ⏪ ⏩ 🌟 0

A synthesis of hydroclimatic, ecological, and socioeconomic data for transdisciplinary research in the Mekong

DOI: [10.1038/s41597-023-02193-0](https://doi.org/10.1038/s41597-023-02193-0)

Hydrological modeling in the Mekong River Basin (MRB) often relies on global meteorological products due to the lack of comprehensive gauge-based data. These products include ERA5, Princeton Global Forcing, WFDEI, GSWP3, and W5E5, which are derived from reanalysis datasets. Algorithms like SM2RAIN estimate effective rainfall from soil moisture data. However, the coarse spatial resolution of these products (typically 0.5° -50 km at the equator) limits their application to basin-scale modeling.

hydrological modeling Mekong River Basin meteorological data reanalysis datasets effective rainfall

AI-Generated

Citation

Give Card

AI Chat

Connecting Data Creators and Data Users as the "Network of Gratitude"

About | Contact | License | Privacy policy | Terms of use



データ俯瞰・検索システム
Dataset Search and Discovery

English

ホーム 使い方 このサイトについて

第3次 全球土壌水分プロジェクト 気象外力（実験1）

HTML PDF XML

このデータセットの引用文

金 炯俊. (2017). 第3次 全球土壌水分プロジェクト 気象外力（実験1） [Data set]. データ統合・解析システム(DIAS). <https://doi.org/10.20783/DIAS.501>

引用フォーマット: APA ▼

このデータセットを引用する際は、**mahalo!** ボタンを押してください。

生成AIによる データ登録支援

1. TitleとDescriptionを元の論文から抜き出して、**Summary**と**Tags**をAIが自動生成
2. メタデータを、人手で整理せず自動的に生成するため、**作業負荷が大幅に減る**

Mahalo Card

DOI of the citing work:

[Resolve the DOI](#)

The type of work:

Description (Citation of the dataset):

[Ask AI to generate summary and tags](#)

Summary [clear](#)

Tags [clear](#)

☐ AI-Generated

Your Name: [Use the top-right button in the header to update your name.](#)
Kitamoto

[Close](#) [Give](#)

生成AIによる データ利用支援

1. Mahalo Cardに登録した利用事例に**生成AIチャット**が使える
2. **プロンプト**を変えると、様々な内容の情報が取得できる
3. **データセットの潜在的利用**者を支援する

The screenshot shows a chat window titled "Mahalo Chat" with a "Hint" button. The chat history includes a light blue bubble asking for use cases and a pink bubble providing a list of 8 use cases for the GSWP3 dataset. Below this, another light blue bubble asks for more details on climate change applications, followed by a pink bubble providing 5 specific examples. At the bottom, there is a text input field with the placeholder "Ask a question about this dataset" and a send button.

このデータセットの利用事例をまとめてください

以下は、Global Soil Wetness Project Phase 3 (GSWP3) データセットの利用事例です：

1. 降水量と温度データを使用して、過去の気候の変動を調査するための再解析データセットとして利用されています [73]。
2. GRUNという観測に基づく世界規模の月次流出量の再構成データセットの作成に使用されています。これは、気温と降水量データに基づく機械学習アルゴリズムによって生成されました [97]。
3. 機械学習を使用した全球的な炭素循環のモデリング研究で、大気中の二酸化炭素濃度や降水パターンの変化を分析するために利用されています [54]。
4. 雪力バーの変動を分析するために、複数の気候モデルと共に気象データセットとして使用されています [76]。
5. 世界規模の流出量の予測において、観測データとモデルの出力を比較するための基準として使用されました [86]。
6. 地球温暖化に伴う降雨変動の将来予測において、気候モデルの精度を評価するための基準データセットとして使用されました [95]。
7. 土壌水分と流出の変動を分析し、異なる気候条件下での炭素吸収の変動を研究するために使用されました [94]。
8. 世界規模の炭素循環における気候変動の影響を理解するために、土壌水分と炭素吸収の関係を研究するために利用されています [90]。

これらの事例は、GSWP3データセットが気候変動や水循環に関するさまざまな研究で広く利用されていることを示しています。このデータセットは、気候モデルの評価、気候変動の影響評価、及び水文モデルの入力データとして重要な役割を果たしています。

気候変動への利用をもう少し詳しくまとめてください

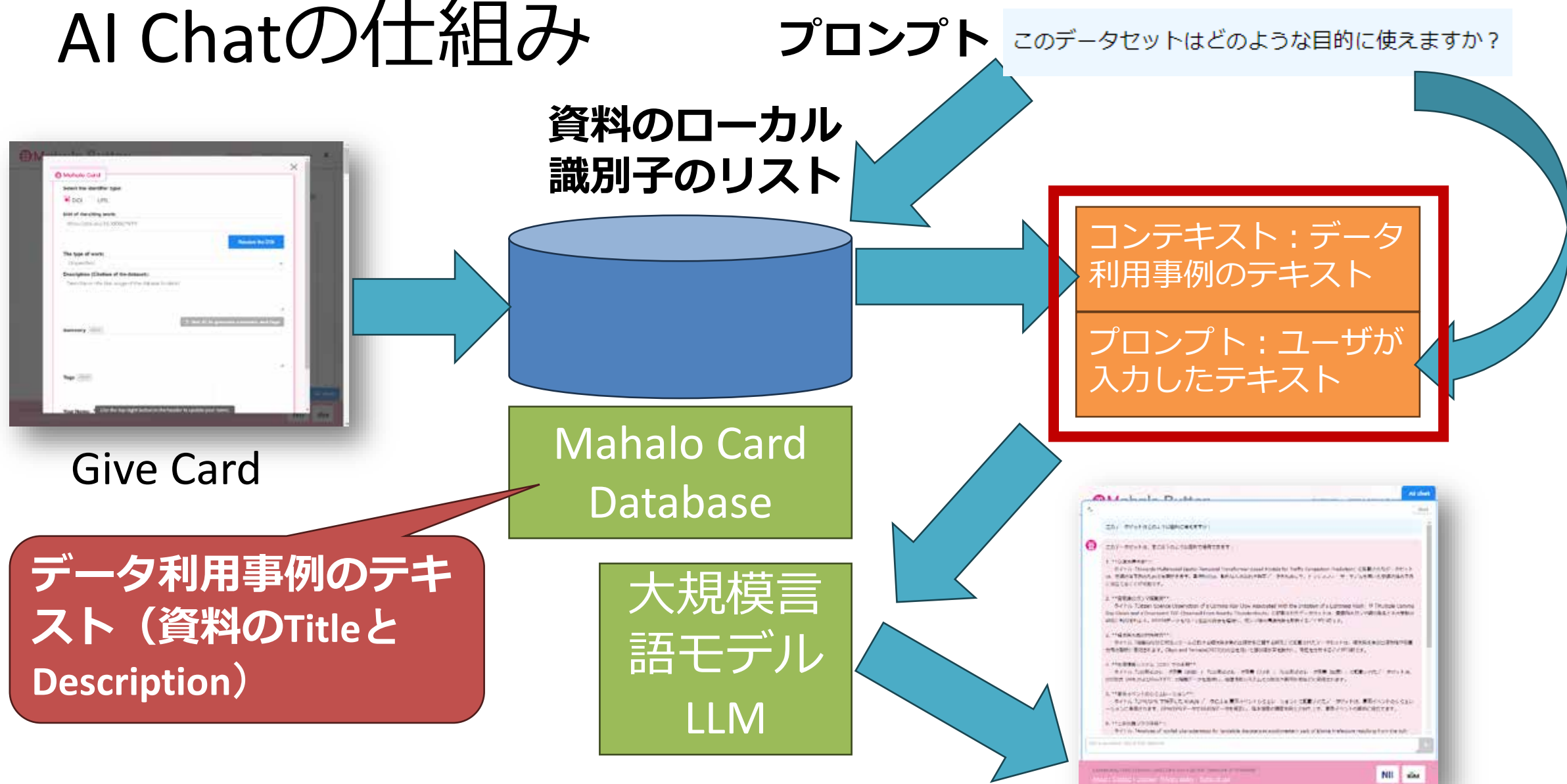
以下は、Global Soil Wetness Project Phase 3 (GSWP3) データセットが気候変動関連の研究でどのように利用されているかを詳述した事例です：

1. 気候変動による降水量の変動を分析するために、複数の気候モデルの比較において使用されています。これにより、将来の降雨パターンの変化を予測し、気候変動の影響を評価しています [49]。
2. 気候変動が土壌水分や流出量に与える影響を評価するために、GSWP3データセットが利用されています。これにより、気候変動によって引き起こされる水循環の変化を分析しています [86]。
3. 土壌水分と気候変動の関係を調査し、これが炭素吸収に及ぼす影響を研究するために使用されています。これにより、気候変動が生態系の炭素サイクルにどのように影響するかを理解する手助けをしています [94]。
4. 世界規模の降水量と気温の変化を分析するために、GSWP3データセットが基準データとして利用されています。気候変動シナリオの検証に役立っています [95]。
5. 土壌と大気のフィードバックメカニズムを研究することで、気候変動が生態系の炭素および養分循環に及ぼす影響を理解しています [90]。

これらの事例から、GSWP3データセットは気候変動の影響を多角的に研究するための重要なツールとして機能しており、特に水循環や炭素循環の変化の理解に貢献しています。

Ask a question about this dataset

AI Chatの仕組み

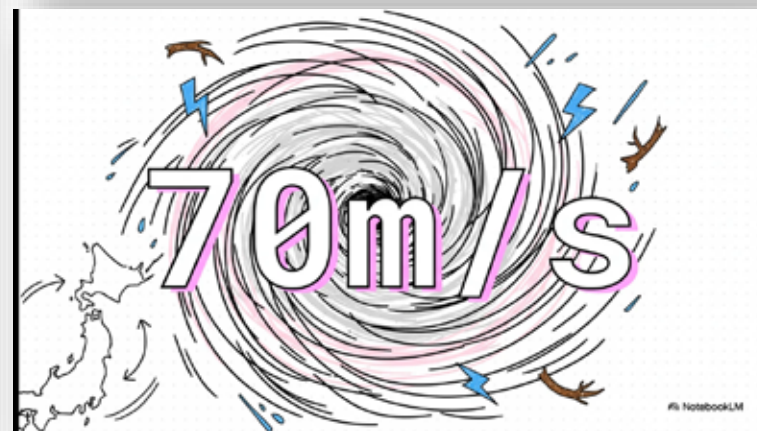
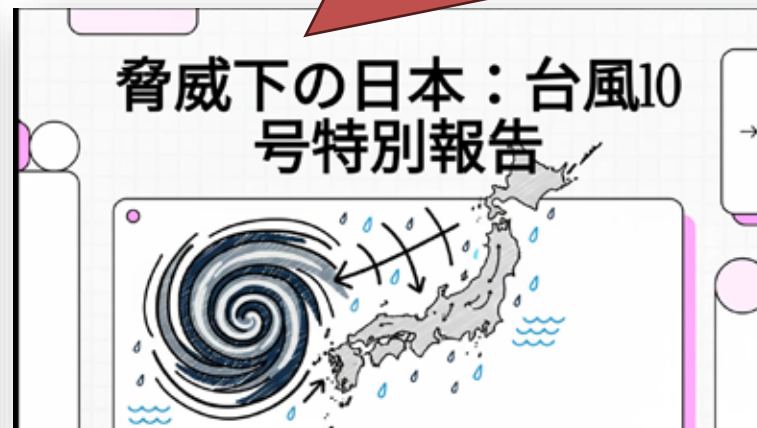


NotebookLM

<https://notebooklm.google.com/>

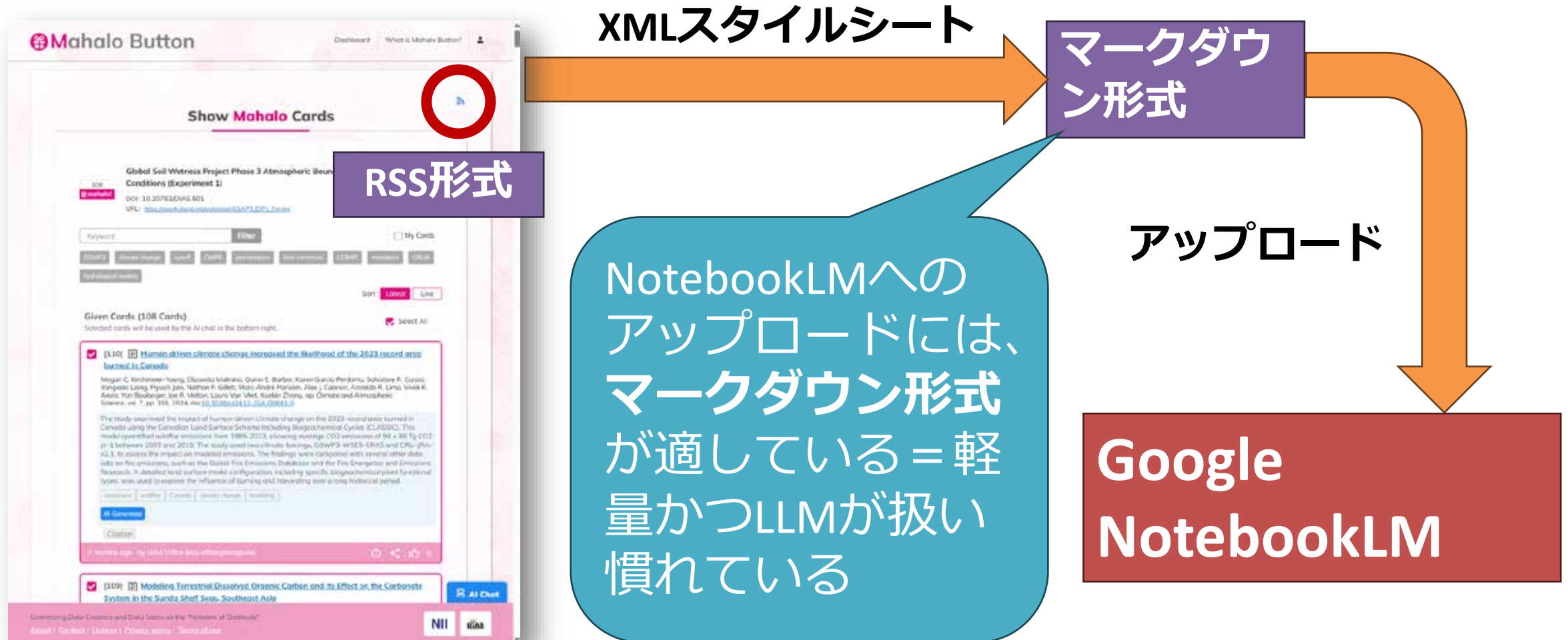
プロンプトなしで、自動的に音声や動画・レポートなどを生成

アップロードした資料を対象に、生成AIを活用する



Mahalo Button→画像・音声・動画生成

<https://notebooklm.google.com/notebook/929cdf7-d876-434e-8e6c-86305f00bbd5>



GSWP3：地球システム科学を支える気候データセット

GSWP3とは？

陸面モデルを駆動する 全球気候データ

気温、降水量、放射など、地球システムシミュレーションに不可欠な変数を提供します。



20世紀再解析データを基に開発



20世紀再解析データ
高解像度 (0.5度、3時間ごと) にダウンスケールされ、観測データでバイアス補正済みです。

100年以上にわたる 長期データ

1901 2010年代
1901年から2010年代までの長期間をカバーし、歴史的な気候トレンド分析を可能にします。

多様な研究分野での応用



水文学モデリング

全球の河川流出量、洪水リスク、水資源の評価研究に使用されます。



雪氷・気候科学

積雪モデルの駆動や、積雪が気候変動に与えるフィードバックの研究に貢献します。



生態系・炭素循環

陸域生態系の植物生産や炭素収支のシミュレーションに広く活用されています。

地球科学研究を支える全球気象データ：GSWP3とは？

地球の陸面をシミュレートする「駆動」データ

陸面モデルや水文モデルに入力する、基本的な全球気象データセットです。

1世紀以上にわたる 高解像度記録

1901年～2014年をカバーし、3時間ごと、0.5度格子で提供されます。

再解析データを観測値で 精密に補正

20世紀再解析を基に、GPCP降水量データなどでバイアス補正を実施。



水循環モデリング

河川流量、全球流出量、水資源の変動を再現・予測します。



炭素循環・生態系研究

土壌水分や植生の応答を計算し、陸域の炭素吸収量を評価します。



気候変動影響評価

干ばつ、洪水、熱波など、極端気象現象の研究に不可欠です。



国際研究の標準データ

CMIP6/LS3MIPなど、多くの国際共同研究で標準強制データとして採用。

地球科学研究の基盤：GSWP3気候データセット

GSWP3データセットの概要



陸面・生態系モデルのための
全球気象データ
1世紀以上にわたる気候変動研究の
基盤となる強制力データセット。



GPCPやCRU等の観測
データでバイアス補正



GSWP3の多様な活用事例：水循環モデリング

河川流量、全球流出量 (GRUN)、土壌水分、貯水量の
再現・再構築研究に利用。



GSWP3の多様な活用事例： 陸域生態系・炭素循環分析

純生態系生産量 (NEP)、植物の
一次生産、炭素循環の動態要因
分析に使用。



GSWP3の多様な活用事例：気候変動影響評価

CMIP6 (結合モデル相互比較プロジェクト) 等の国際的な
モデル比較プロジェクトで標準データとして採用。



地球科学研究を支える全球気象データ：GSWP3とは？

GSWP3の概要

地球の陸面を
シミュレートする
「駆動」データ
陸面モデルや水文モデルに
入力する、基本的な全球
気象データセットです。

1世紀以上にわたる
高解像度記録
1901年～2014年をカバーし、3時間ごと、0.5度格子で提供されます。

再解析データを
観測値で精密に補正
20世紀再解析を基に、
GPCP降水量データなどで
バイアス補正を実施。

GSWP3の主な活用分野

気候変動影響評価

干ばつ、洪水、熱波など、極端気象現象の研究に不可欠です。



国際研究の標準データ

CMIP6/LS3MIPなど、多くの国際共同研究で標準強制データとして採用。

水循環モデリング

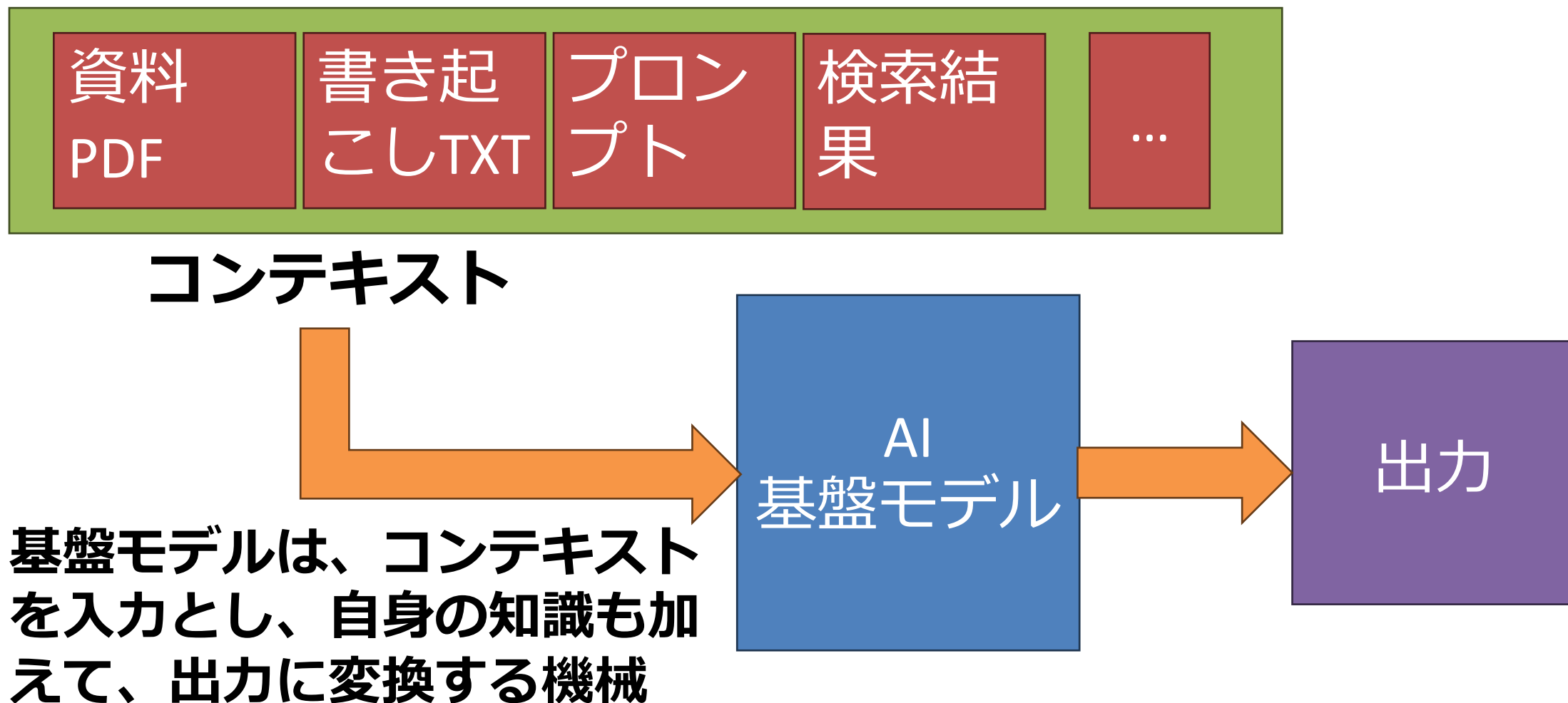
河川流量、全球流出量、水資源の変動を再現・予測します。

炭素循環・生態系研究

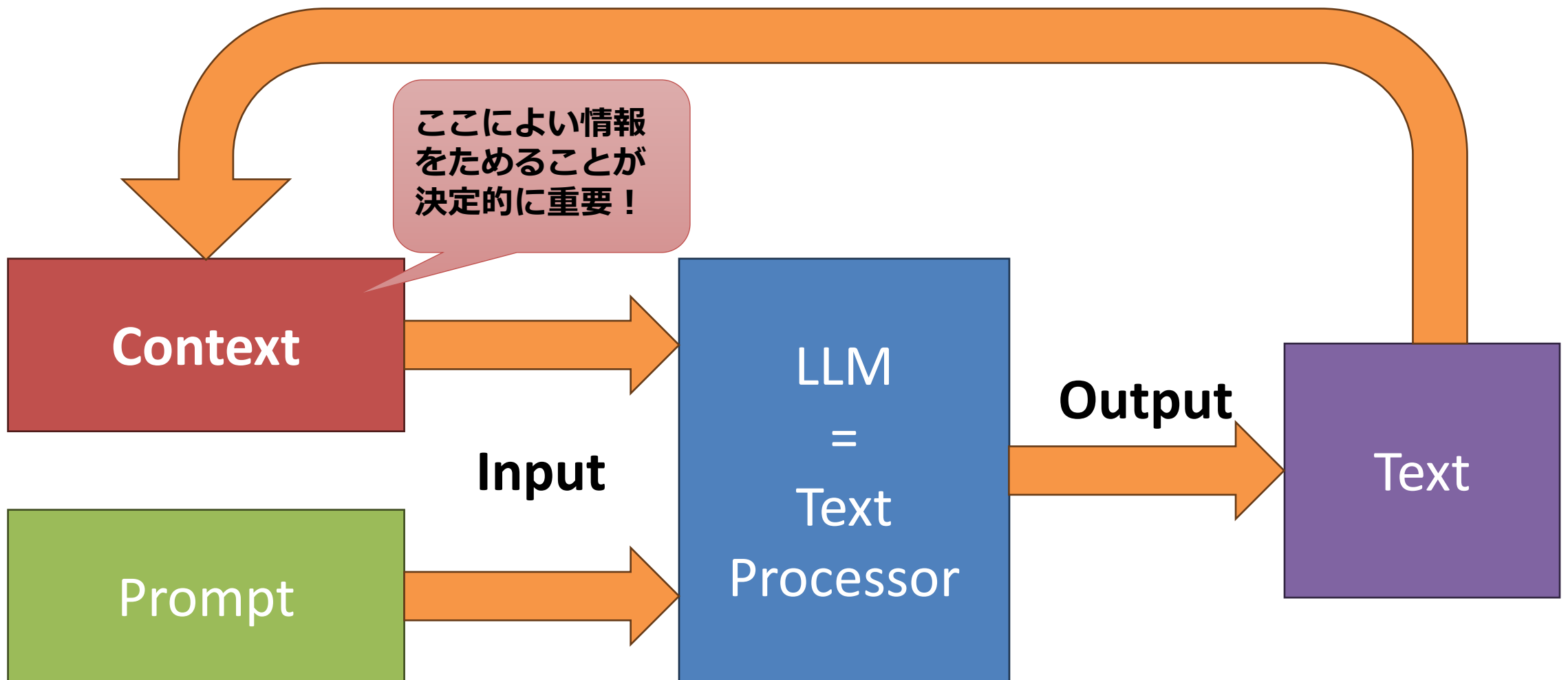
土壌水分や植生の応答を計算し、陸域の炭素吸収量を評価します。



Context Engineering



Text ProcessorとしてのLLM



Contextの重要性

1. 現在の最先端LLMは高い言語能力を備えており、**入力**
を出力に変換するText Processorとして非常に優秀
2. **優れた参考情報（Context）と優れた指示（Prompt）**
を与えれば、優れた出力が得られる
3. 参考情報を元に出力させることで、**幻覚（ハルシネー**
ション）の影響を軽減できる
4. **長さ制限のあるコンテキストに参考情報を詰め込む手**
法がホットな研究テーマ（例：検索ベースのRAG）

Mahalo Research

柔らかいメタデータの研究利用

Mahalo Research

<https://dias.ex.nii.ac.jp/mahalo/research/>

1. Mahalo Card（利用事例）
をもとに、**DIASデータ
セットの発見を支援する
サービス**
2. 要約、推薦、関連検索、
比較、プロフィールなどの
機能を、**生成AIをフル
活用して実現**



データセット要約

1. データセットのメタデータとして、**概要、価値、応用分野の3項目**を生成
2. どのデータセットもほぼ同じ文字数に統一可能
3. 利用事例があると、**特に応用分野の記述が充実**

The screenshot displays the JALC Data Commons interface. At the top, there are navigation links: 63 データセット, 527 利用事例, and 61 メタデータあり. Below these is a search bar labeled 'データセットを検索'. The main content area shows two dataset cards. The first card, titled 'Global Soil Wetness Project Phase 3 Atmospheric Boundary Conditions (Experiment 1)', includes a '108 件の利用事例' badge, a 'メタデータあり' badge, and a DOI: 10.20783/DIAS.501. It has sections for '概要' (Summary), '価値' (Value), and '応用分野' (Application Fields). The second card, titled 'MLIT XRAIN dataset', includes a '59 件の利用事例' badge, a 'メタデータあり' badge, and sections for '概要', '価値', and '応用分野'. At the bottom of each card are buttons for 'データセットページ', '全利用事例を見る', and '関連データセットを検索'.

データセット推薦

1. 研究計画を入力すると、それに適合するデータセットを推薦
2. 研究計画と類似する利用事例をベクトル検索し、ポイントごとに記述
3. Mahalo Card記載の参考文献も掲載

入力された研究計画

将来の温暖化シナリオにおける農業生産性への気候変動の影響を評価する。気候モデルの将来予測データと作物収量シミュレーションモデルを用いて、2050年までの生産性変化を予測・分析する。対象地域はアジア太平洋地域の農業地帯、21世紀（2020-2100年）を対象とする。気候予測データ（気温・降水量）、グリッドレベルの空間解像度、年次・季節別時系列データ、RCPまたはSSPシナリオに基づくデータを期待する。

研究計画を編集

推薦結果

中信頼度

1 The Japanese 55-year Reanalysis (JRA-55)

このデータセットは**長期の気象再解析データ**を提供し、将来予測の初期値・境界条件として活用可能です。

- **研究計画との関連:** Taniguchi (2016)の事例では、JRA-55とRCP8.5シナリオを組み合わせた擬似温暖化手法で将来気象イベントを推定
- **時間範囲:** 1958年から現在まで（～2024年）のデータで、将来予測モデルの検証・キャリブレーションに有用
- **活用方法:** 過去の気候データとして、将来予測シナリオの基準値や作物収量モデルの検証データとして利用可能
- **制約:** 将来予測データそのものではないため、GCM/RCMの将来シナリオデータとの組み合わせが必要

主な利用事例 (Mahalo Card)

最も関連性の高い1件を表示（すべての事例は「全利用事例を見る」から閲覧可能）

1

アンサンブルシミュレーションと擬似温暖化手法による特定の気象イベントの将来変化の推定

著者: Kenji TANIGUCHI

出版年: 2016

データセットページ

全利用事例を見る

関連データセットを検索

データセット比較

1. 2つのデータセットの共通点と相違点を比較
2. データセットのメタデータに加えて、**利用事例も入れることで、多角的に比較**
3. 使い方や選択基準も合わせて提示

比較結果

The Japanese 55-year Reanalysis (JRA-55)

[データセットページ](#)

[全利用事例を見る](#)

The Japanese Reanalysis for Three Quarters of a Century

[データセットページ](#)

[全利用事例を見る](#)

概要

JRA-55とJRA-3Qはいずれも気象庁が作成した長期再解析データセットで、全球の気象・気候変数を提供しています。JRA-55は1958年から2024年までの66年間をカバーする成熟したデータセット（利用事例55件）であるのに対し、JRA-3Qは1947年から現在まで継続中のより新しいデータセット（利用事例13件）で、最新の同化システムを採用し、より長期の気候研究を目指しています。

共通点

両データセットは気象庁による全球大気再解析データで、気温、風、降水量などの気象変数を提供し、気候変動研究、極端気象イベント解析、海洋大気相互作用研究など幅広い分野で利用されています。

データ提供機関と品質保証

両データセットとも日本気象庁（JMA）が作成した再解析データセットで、数値天気予報システムと観測データの同化により高品質な気候変数を提供

重要性: 同一機関による作成のため、データ処理手法や品質管理基準に一貫性があり、日本域での精度が特に高いことが期待できます

研究分野の適用範囲

プロフィール設定

1. 任意のウェブページを読み込み、研究キーワードや分野を分析
2. プロフィールを作成し、それに近いデータセットをベクトル検索
3. Mahalo Card記載の参考文献も掲載

[プロフィールを更新](#) [プロフィールに合うデータセットを表示](#) [プロフィールを削除](#)

あなたのプロフィールに合うデータセット

あなたの研究分野やキーワードに基づいて、関連性の高いデータセットを推薦します。

プロフィール概要

氏名: Asanobu Kitamoto
所属: National Institute of Informatics (NII)
キーワード: Digital Humanities, Machine Learning, Generative AI, Large Language Models, Digital Archives, Historical Big Data, Typhoon Research, Digital Typhoon, IIIF, Linked Open Data, Japanese Classical Literature, Historical Gazetteers, Toponym Platforms, Computer Vision, Explainable AI, Cultural Heritage, Geospatial Information, Open Science, FAIR Principles, Metadata
研究分野: Digital Humanities, Machine Learning and AI, Meteorology and Climate Science, Cultural Heritage Informatics, Geospatial Information Systems

推薦 1

関連性: 54.9%

Digital Typhoon Dataset

DOI: 10.20783/DIAS.664

利用事例数: 4件 | 平均関連性: 48.3%

関連性の高い利用事例:

1

54.9% 一致

Machine Learning for the Digital Typhoon Dataset: Extensions to Multiple Basins and New Developments in Representations and Tasks

This paper presents the Digital Typhoon Dataset V2, a new version of the longest typhoon satellite image dataset for 40+ years aimed at benchmarking machine learning models for long-term spatio-tempor...

著者: Asanobu Kitamoto/Erwan Dziku/Gaspard Faure
出版年: 2024

2

47.5% 一致

Digital Typhoon: Long-term Satellite Image Dataset for the Spatio-Temporal Modeling of

生成AIのためのHOWメタデータ

- 従来のメタデータは、何のデータセットか（WHAT）、なぜデータセットが作られたか（WHY）が中心
- 今後の生成AI活用を見据えると、データセットをどのように使うか（HOW）に関するメタデータも必要

1. Descriptive Metadata = WHAT
2. Contextual Metadata = WHY
3. Operational Metadata = HOW

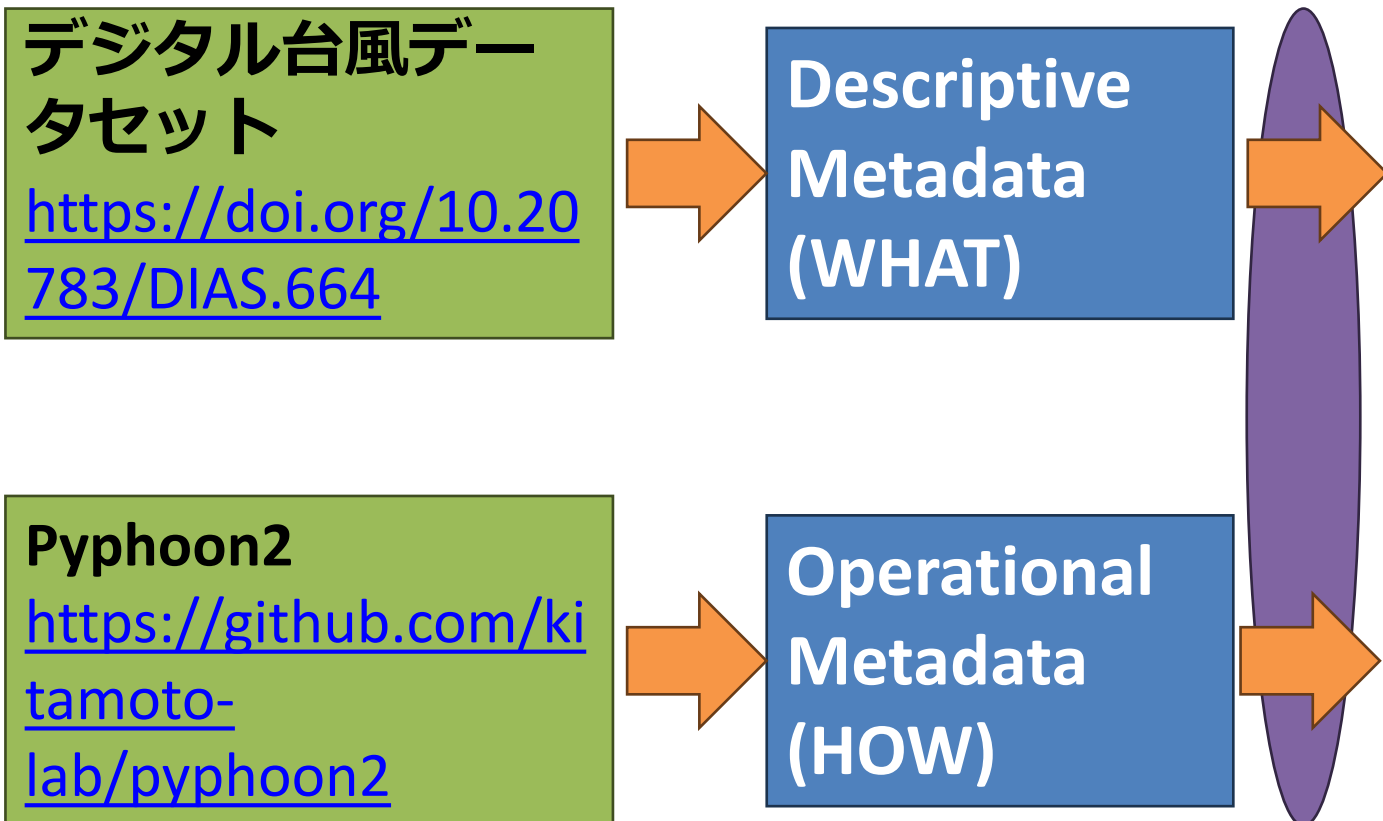
Operational Metadataの作成

```
! library_metadata.yaml X
C: > home > sync > presentations > jalc-2026 > ! library_metadata.yaml
 95 file_structure:
 97   components:
106     - name: "metadata_json"
110
111 files:
112   # --- 画像ファイル (HDF5) ---
113   - name: "image_file"
114     format: "hdf5"
115
116   filename:
117     pattern: "{datetime}-{sequence_id}-{satellite}.h5"
118     example: "2008041300-200801-MTS1-1.h5"
119     regex: "^((\\d{10})-(\\d{6})-([A-Z0-9-]+))\\.h5$"
120     components:
121       - name: "datetime"
122         position: "先頭10文字"
123         format: "YYYYMMDDHH"
124       - name: "sequence_id"
125         position: "2番目のセグメント"
126         format: "YYYYMM([年+番号])"
127       - name: "satellite"
128         position: "3番目のセグメント"
129
130   hdf5_spec:
131     default_dataset: "Infrared"
132     datasets:
133       - key: "Infrared"
134         dtype: "float64" # 温度データとして読み込まれる
135         shape: null # 画像サイズは可変 (実データ依存)
136         description: "赤外線画像データ (輝度温度)"
137     access_code: |
138       import h5py
139       import numpy as np
140       with h5py.File(filepath, 'r') as h5f:
141         image = np.array(h5f.get('Infrared'))
142
143   # --- トラックデータcsv ---
144   - name: "track file"
145     format: "csv"
146     encoding: "utf-8"
```

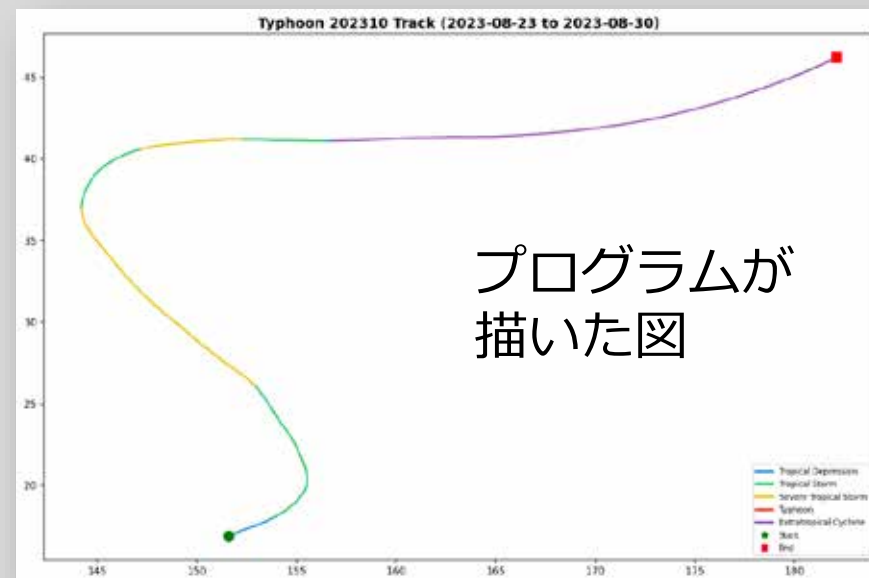
データの処理方法を記述するメタデータを、ソフトウェアから自動的に抽出（Claude Code Skillsなども活用）

1. ファイル形式と読み込み方法
2. 列定義（名前、型、単位、値の意味）
3. ファイル名パターン
4. API使用例
5. 注意点・制約

プログラムの自動生成



1. プロンプト：台風経路を描くスクリプトを作って
2. Pythonプログラムを生成
3. Pythonプログラムを実行



AI時代におけるメタデータ再考

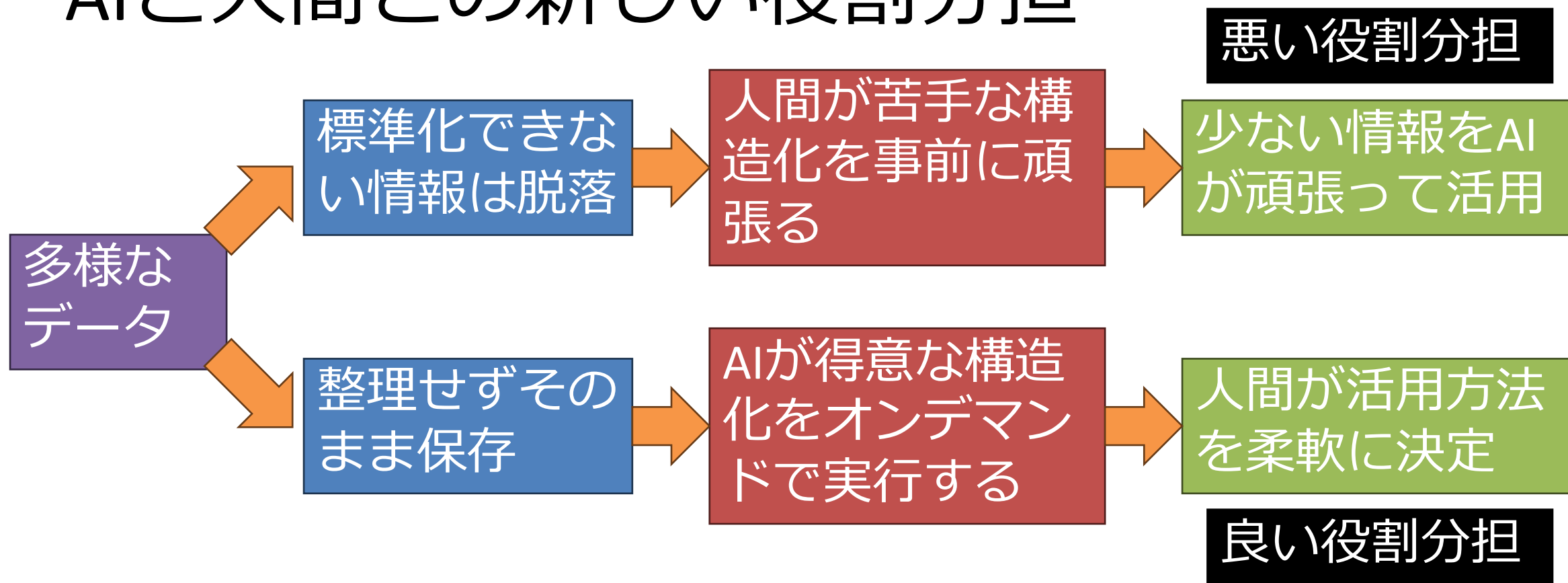
生成AI時代のメタデータ変容

1. **これまで**：厳密なスキーマに従った「堅い」メタデータでないと、相互運用性を確保できなかった
2. **これから**：生成AIは意味のギャップを橋渡しできるため、「柔らかい」メタデータの価値が増す
3. 流通を重視するシステム（DOIなど）は、**固定的なスキーマを用いて標準化を推進**
4. 発見（ファインダビリティ）を重視するシステムは、**可塑的なスキーマ＋生成AIを用いて多様化を推進**

柔らかいメタデータへ

1. 非構造化データを構造化データに変換することは、LLMが比較的得意とする分野
2. AIの入力データ（コンテキスト）を用意することは人間にしかできない
3. 利用者のニーズをプロンプトに指定することで、利用目的に最適な形式で生成できる
4. 多様なデータを保存し、AIが可塑的にメタデータを生成すれば、人間もAIも得意分野に集中できる

AIと人間との新しい役割分担



- 現在のメタデータ作成方式は「悪い役割分担」に近い
- 良い役割分担への移行にはメタデータの再定義が必要

まとめ

1. 標準化に基づく「堅いメタデータ」は、メタデータ流通においては依然として重要である
2. 多様化を後押しする「柔らかいメタデータ」は、生成AIの活用と橋渡しにより新たな価値を生み出す
3. コンテキストに優れた参考情報を詰め込むことが、生成AIの活用におけるキーポイントである
4. AIと人間との新しい役割分担に向けたメタデータの再定義が必要である