

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

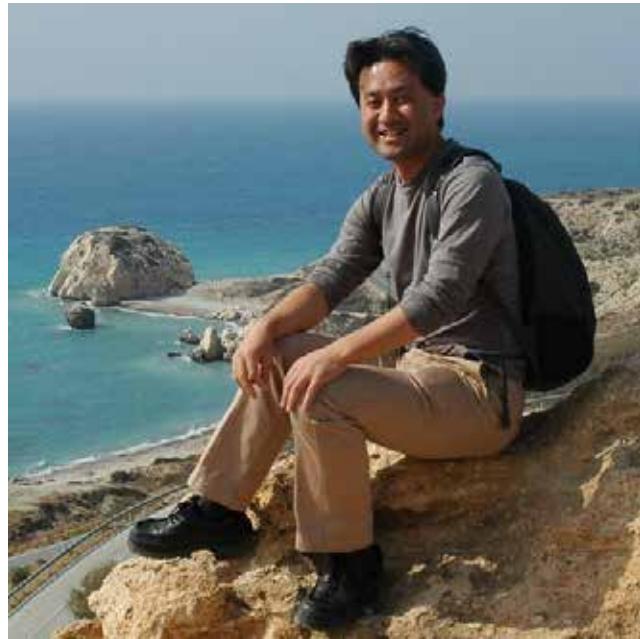
ROIS-DS人文学オープンデータ共同利用センター (CODH)

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

kitamoto@nii.ac.jp

自己紹介

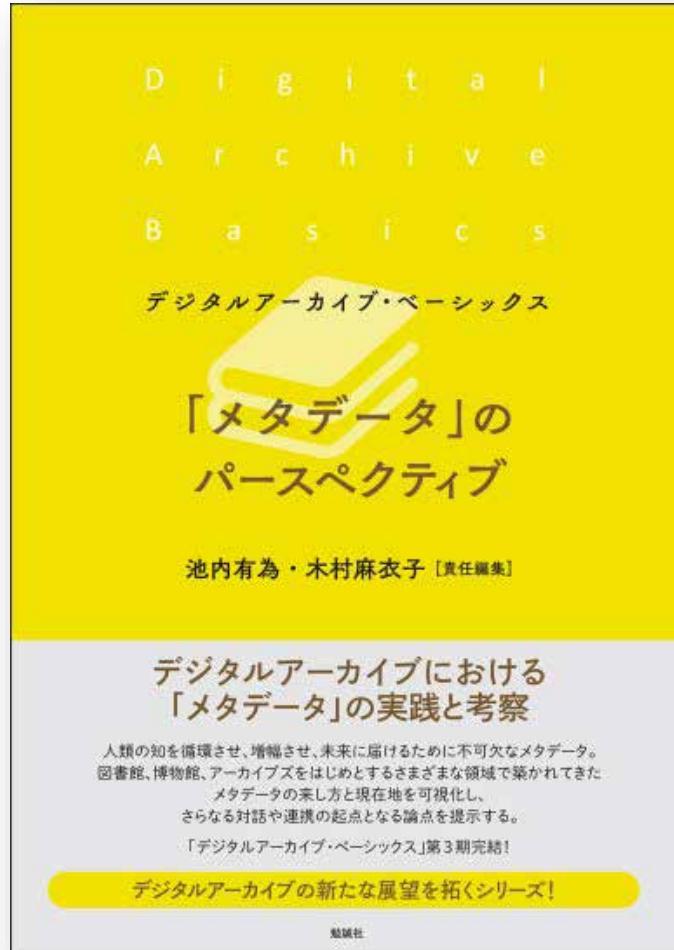
<https://researchmap.jp/kitamoto/>



@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 —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」のパースペクティブ, 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

項目の類型

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的専門性**：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的専門性**：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的専門性**：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？

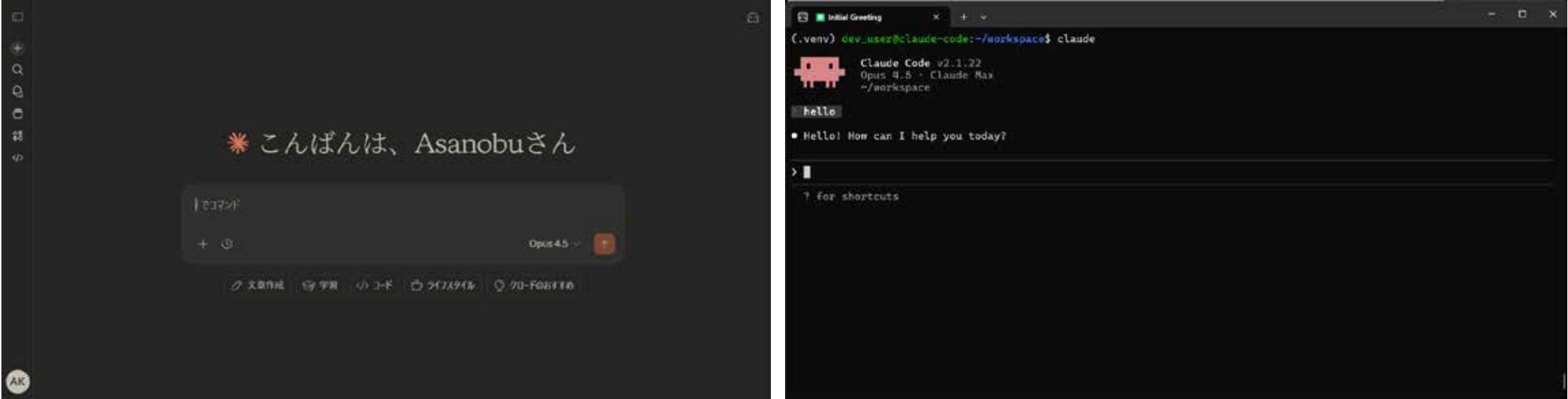
目的

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.nii.ac.jp/humanitiesthroughddps/>

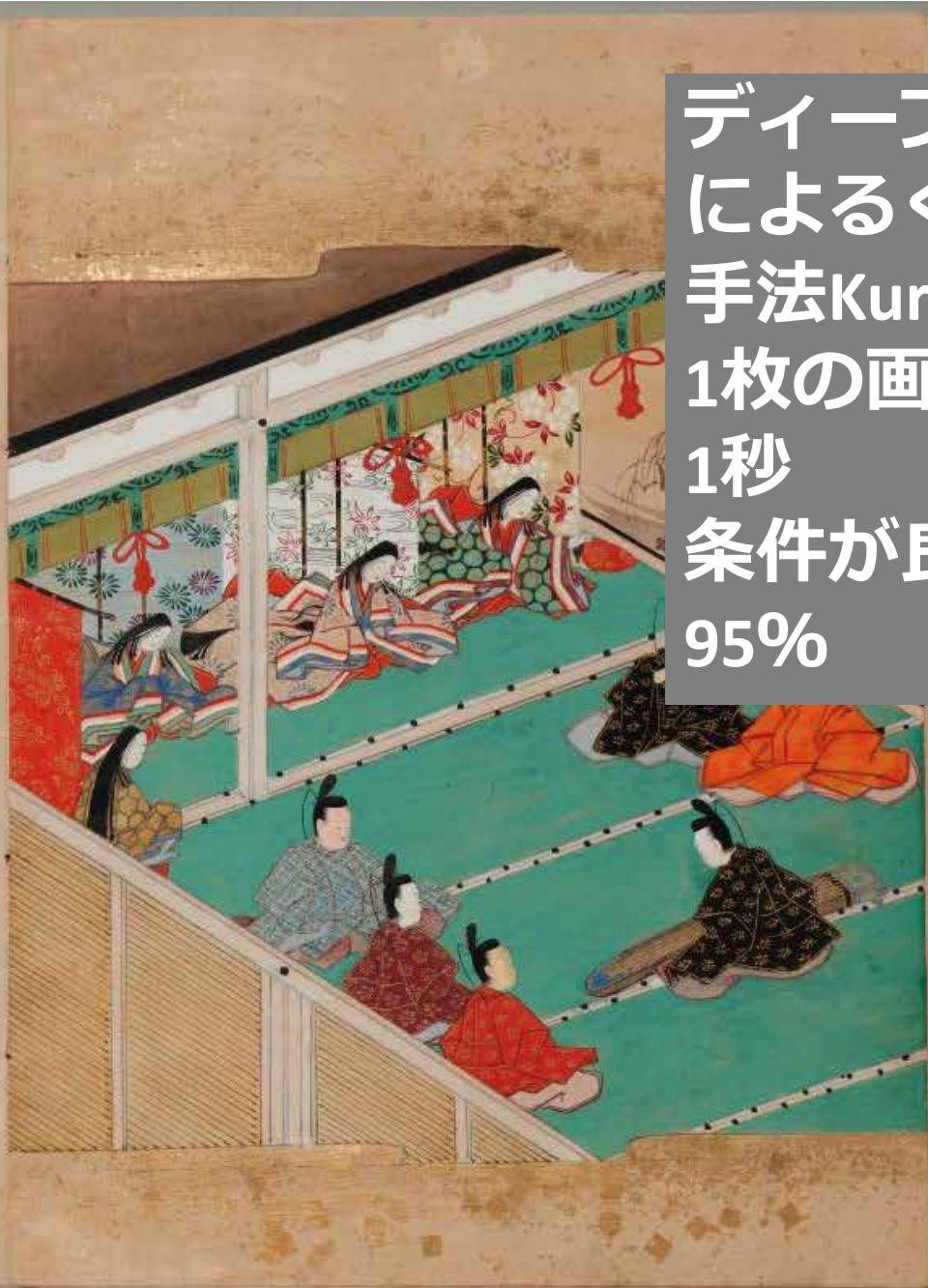
300,000点の日本古典籍
(1868年以前) を、デジ
タル化し、オープンデー
タとして公開



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 ん 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 ど り き ガ 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 さ た む そ め く さ 何 く れ の
と し や う この も の ど も は 一 て う 殿 に も わ か ち
奉 り 給 お は す る 事 は な け れ は 御 か た く
に お ほ し な け き さ ま く に き お と ろ か し



JaLC 「対話・共創の場」 (第12回)

つくし堂

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

2025年12月公開



- 古典籍を親しみやすく提供する「AI生成書店」
- 古典籍の要約をOCRテキストとメタデータから生成
- 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
- 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

日本古典籍データセット
メタデータ（書誌情報）
<https://codh.rois.ac.jp/pmjt/>

日本古典籍データセットの
OCRテキスト（RURI利用）
<https://codh.rois.ac.jp/miwo/>

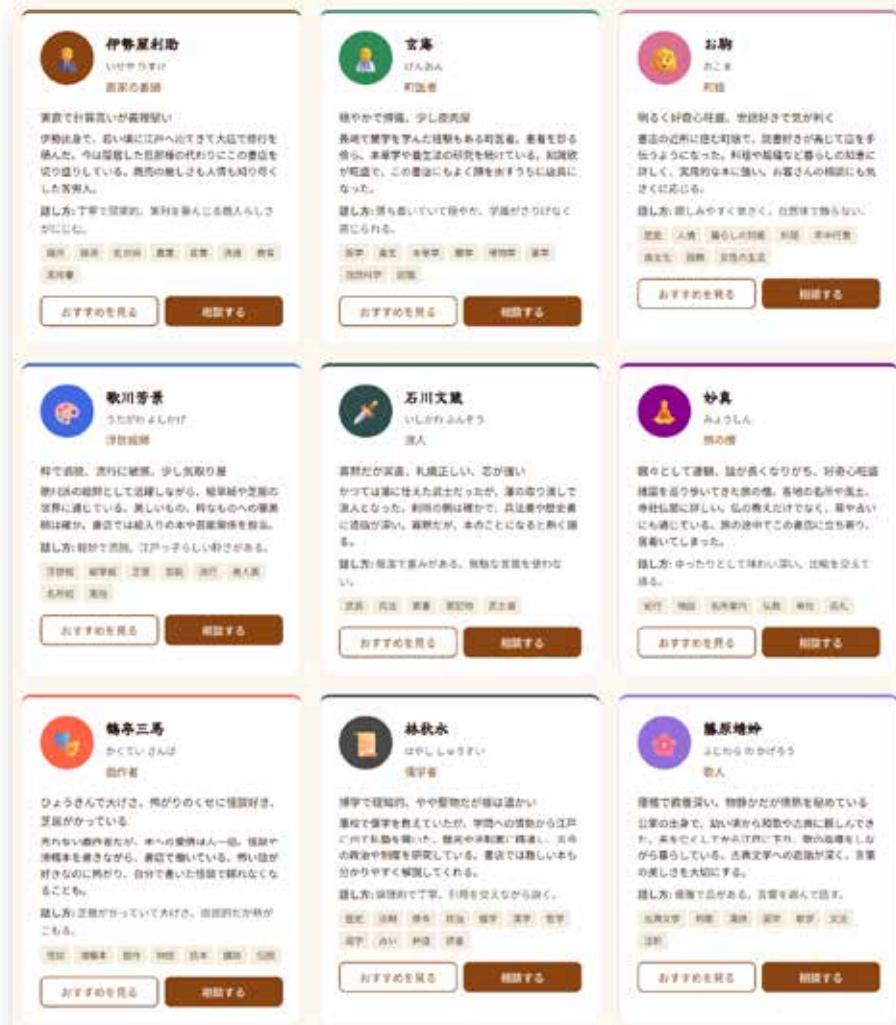
プロンプト：
あなたは日本の書店のプロ
モーション担当者です。
・ ・ ・

メタデータ
OCRテキスト

Anthropic
Claude
Sonnet 4.5

要約・タグ

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

キヤッチコピー生成

自讃歌／苔菴

藤原蜻蛉
歌人

中世の言葉が織りなす、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「苔菴」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・人生の哀歎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

- 要約と店員のペルソナをプロンプトに入れ、キヤッチコピーをAIが生成
- 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)



江戸の古典籍を一般向けに届けるウェブサイトのデザイン

本を並べて検索できるウェブサイトのデザインを検討する。対象は日本の江戸時代の本3000冊。将来的にはもっと増える可能性あり。本ごとに要約やタグがある。要約は本のプロモーションに使いたい。一般になじみがない本を、一般の人に興味を持ってもらうことが目的。そういう場合に、どんなインターフェース、インタラクションが考えられるか？

北本先生、江戸時代の本3000冊を一般の方に届けるウェブサイトですね。とても意義深いプロジェクトだと思います。

なじみのない古典籍を、どうやって「自分ごと」として感じてもらえるか。いくつかの方向性を整理してみました。

入口のデザイン：「検索」より「出会い」

一般の方は検索キーワードを持っていないことが多いので、能動的に探させるより、受動的に出会える仕掛けが有効かもしれません。

ランダム・セレンディピティ型として、ページを開くたびに1冊がハイライトされる「今日の一本」や、要約をカード形式で次々めくれるインターフェースが考えられます。Tinderのスワイプのような気軽さで、興味があれば右、なければ左、という形式です。

パーソナル質問型では、「あなたの興味は？」「好きな時代劇は？」といった簡単な質問から、関連する本をレコメンドする方法があります。「忠臣蔵が好き」と答えたなら武士の廻世物の本を、「料理が好き」なら江戸の料理書を提案するような流れです。

ブラウジングの構造

逆伝...

Opus 4.5

ClaudeはAIだから、誤りを含む可能性があります。誤り内包せよと指示ください。

江戸の古典籍を一般向けに届けるウェブサイトのデザイン

江戸書店プロジェクト仕様書

仕様書を作成しました。これまでの議論を実装可能な形でまとめてあります。

主な内容は以下の通りです：

1. プロジェクト概要：目的、コンセプト、技術スタック
2. データ構造：本、店員、今日のおすすめのJSON形式
3. 店員キャラクター設定：7人分の詳細設定
4. 生成AIによる事前処理：プロンプトテンプレート付き
5. UI設計：画面構成、POPのデザイン要素
6. ファイル構成：ディレクトリ構造
7. 実装の優先順位：Phase 1～5の段階的な進め方
8. 研究としての位置づけ：問い合わせ、評価方法、ログ項目
9. 参考：既存リソース、制約、参考情報

追加や修正したい点はあります？

逆伝...

Opus 4.5

ClaudeはAIだから、誤りを含む可能性があります。誤り内包せよと指示ください。

江戸書店プロジェクト仕様書

1. プロジェクト概要

1.1 目的

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を構築する。

1.2 コンセプト

「江戸の本を売る現代の書店」

- ・ 7人の江戸時代風の架空店員が本を推薦する
- ・ 書店のPOPのような形式で、本の魅力を伝える
- ・ スワイプやガチャで気軽に本と出会える体験

1.3 技術スタック

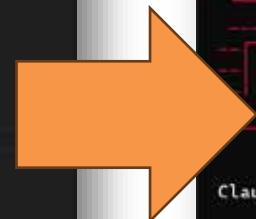
- ・ フロントエンド: HTML + CSS + JavaScript (フレームワークなし)
- ・ データ: JSON形式で静的ファイルとして配置
- ・ 画像: AI生成 (本のイメージ画像、店員のビジュアル)
- ・ 生成AI: 事前処理でコンテンツ生成、リアルタイム処理なし

2. データ構造

2.1 本のデータ (books.json)

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 ### 1.1.1 目的
6 江戸時代の本3000冊を、一般の人興味を持ってもらうためのウェブサイトを構築する。
7
8 ### 1.1.2 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える体験
13
14 ### 1.1.3 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": "繋がりの理由 (発見につながる説明)"
40     }
41   ]
42 }
43
```



```
dev_user@e10680747f46:/var | + - x
Welcome to Claude Code

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*
```

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。数分から數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

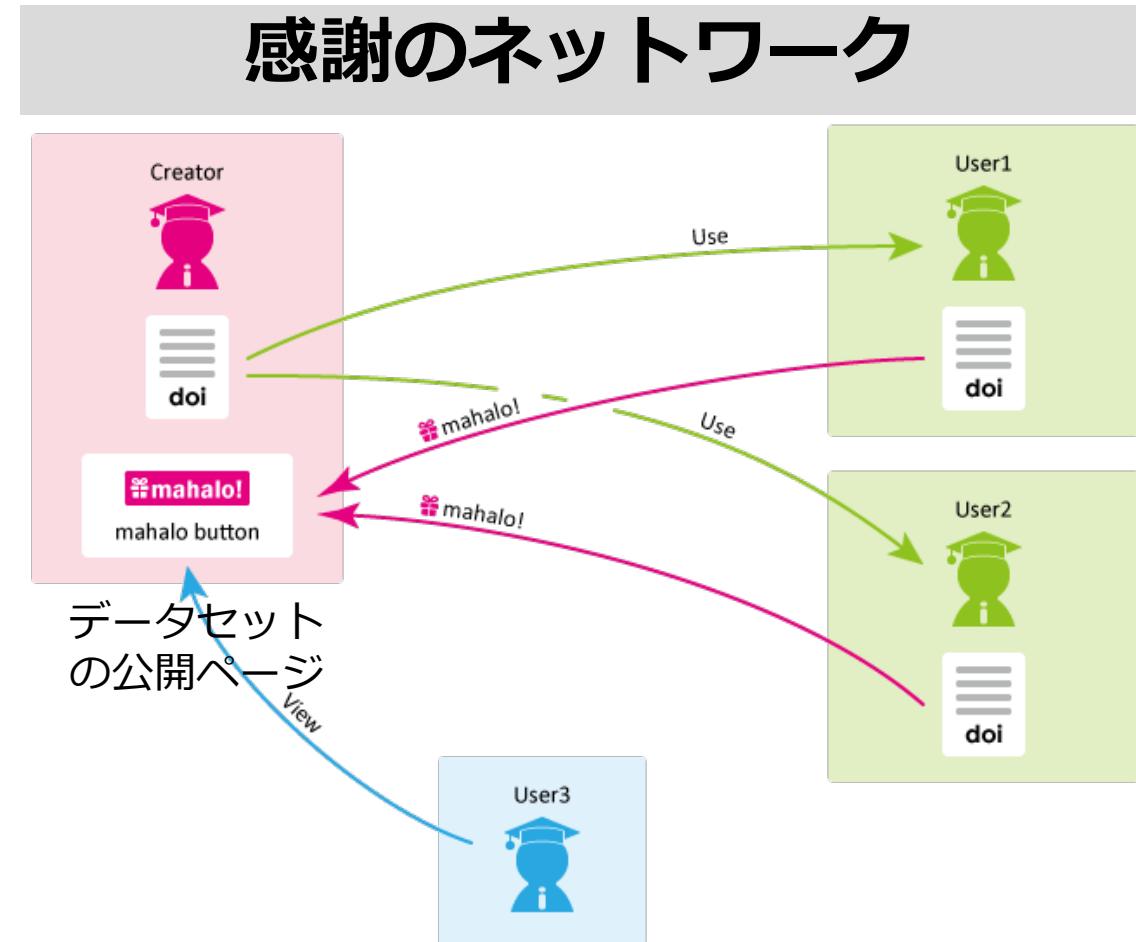
Mahalo Button

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



83

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/>

The screenshot shows the DIAS Dataset Search and Discovery interface. At the top left is the DIAS logo and the text "データ俯瞰・検索システム Dataset Search and Discovery". At the top right are language options ("English") and a search bar. Below the header is a navigation menu with links to "ホーム", "使い方", and "このサイトについて". A large section titled "第3次 全球土壤水分プロジェクト 気象外力 (実験 1)" is displayed, featuring download links for "HTML", "PDF", and "XML". Below this is a section titled "このデータセットの引用文" containing the citation: "金 炯俊. (2017). 第3次 全球土壤水分プロジェクト 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>". A dropdown menu for "引用フォーマット" is set to "APA". At the bottom of the page, there is a red circle highlighting a "mahalo!" button, which is part of the Mahalo Button integration.

2026/2/2

JaLC「対話・共創の場」(第12回)
Connecting Data Creators and Data Users as the "Network of Gratitude"

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

The screenshot shows the Mahalo Button interface. At the top right is a "Show Mahalo Cards" button. Below it is a card for "Global Soil Wetness Project Phase 3 Atmospheric Boundary Conditions (Experiment 1)". The card includes a "mahalo!" button with the number "83", a DOI link (DOI: 10.20783/DIAS.501), and a URL (URL: http://search.dias-project.org/GSWP3_EXP1_Forcing). The card is labeled as "AI-Generated" and has a "Citation" section. A blue arrow points from the "mahalo!" button on the DIAS page to this card. Another card below it is for "Identifying and quantifying the impact of climatic and non-climatic drivers on river discharge in Europe", also with a "mahalo!" button and "AI-Generated" status. A third card for "A synthesis of hydroclimatic, ecological, and socioeconomic data for transdisciplinary research in the Mekong" is partially visible at the bottom.

26

NII

DIAS

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

The screenshot shows the Mahalo Button interface. At the top, there's a header with the Mahalo Button logo, a Dashboard link, and a 'What is Mahalo Button?' link. Below the header is a 'Mahalo Card' section. It contains fields for 'DOI of the citing work:' with the value <https://doi.org/10.48550/arXiv.2411.16421>, a 'Resolve the DOI' button, 'The type of work:' set to 'Paper', and a 'Description (Citation of the dataset)' text area containing a long scientific abstract. A red oval highlights the 'Ask AI to generate summary and tags' button, which is located below the description. Below this button are two input fields: 'Summary' and 'Tags', both with 'clear' buttons. Underneath the 'Tags' field is a checkbox labeled 'AI-Generated'. At the bottom of the card, it says 'Your Name: Kitamoto' with a note to use the top-right button in the header to update it. There are 'Close' and 'Give' buttons at the very bottom.

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

Mahalo Button

Dashboard | What is Mahalo Button? | Hint

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. 雪カ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. 土壤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、土壤水分と炭素吸収の関係性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

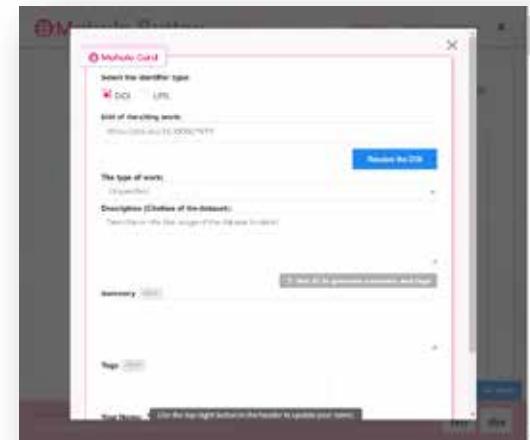
以下は、Global Soil Wetness Project Phase 3 (GSWP3) データセットが気候変動関連の研究でどのように利用されているかを詳述した事例です：

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. 気候変動が土壤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. 土壤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. 土壤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

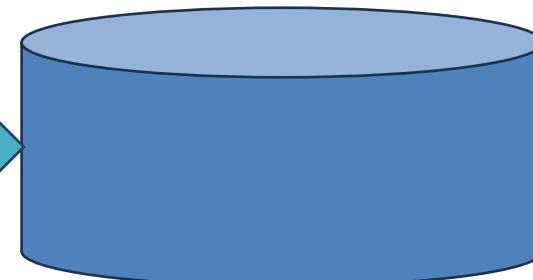
AI Chatの仕組み



Give Card

データ利用事例のテキスト
(資料のTitleと
Description)

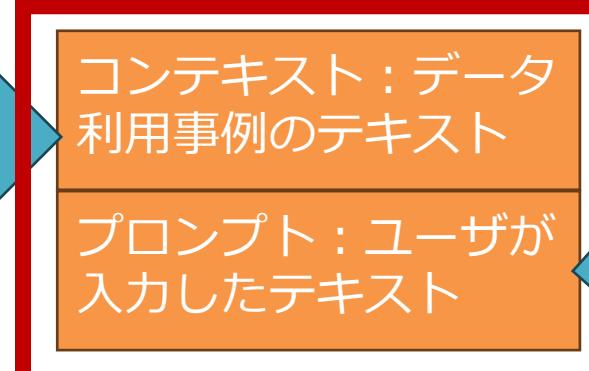
資料のローカル
識別子のリスト



大規模言
語モデル
LLM

プロンプト

このデータセットはどのような目的に使えますか？



NotebookLM

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

令和6年台風第10号気象情報: 日本各地の警戒態勢

この気象情報は、非常に強い台風第10号が日本の広範囲に接近・通過することに伴う防災情報の概況を伝えています。特に九州地方の鹿児島県と宮崎県で、高瀬の特別警報が発令され、土砂災害や河川の氾濫に対する警戒が呼びかけられています。台風の接近が遅いため、線状降水帯の発生により、総雨量が多くなる見込みが強調されています。

台風の進行速度は、北上する傾向があります。また、台風の中心部を含む九州北部から四国、中国、東海、関東、南北地方、暴風、高潮、高瀬といった様々な災害への注意が必要です。

アップロードした資料を対象に、生成AIを活用する

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

台風第10号に関するリスク評価レポート

1.はじめに

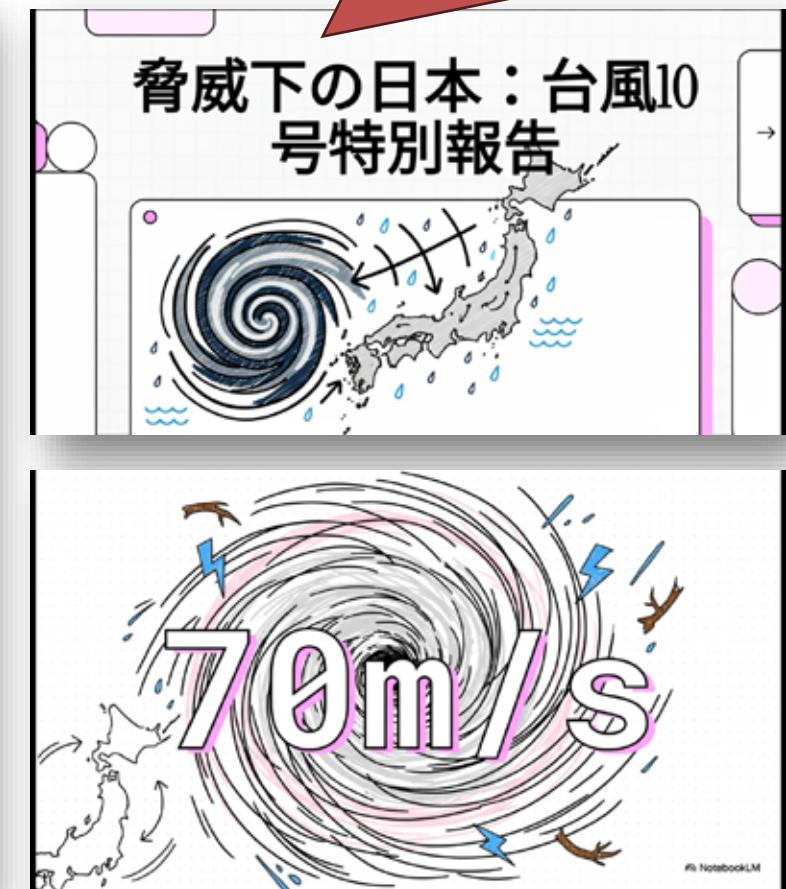
本レポートは、2024年8月29日に発表された各種気象情報に基づき、大型で非常に強い台風第10号が日本各地に及ぼす潜在的リスクを専門的かつ客観的に分析・評価するものです。その目的は、各事業所における事業継続計画(BCP)の策定・見直し、および具体的な防災・減災対策の検討に資する基礎情報を提供することにあります。

台風第10号は、その「非常に強い」勢力に加え、進行速度が遅いという特徴を持っています。これは、特定の地域が暴風、大雨、高潮といった深刻な気象現象に長時間さらされるリスクの増大を意味し、広範囲に甚大な影響を及ぼすことが強く懸念されます。本分析は、これらの気象現象が事業運営に与える具体的な影響を予測し、そのリスクを定量的に評価することを主眼としています。

本稿では、まず台風の基本的な諸元を確認し、その後、特に影響が大きいと予測される九州地方を中心に、各項のリスクを詳細に評価します。

2.台風第10号の概要(2024年8月29日時点)

本セクションでは、後続のリスク分析の基礎となる、特定時点における台風第10号の気象学的特性を概説します。以下のデータは、2024年8月29日午前4時時点の情報に基づいています。



Mahalo Button→画像・音声・動画生成

<https://notebooklm.google.com/notebook/929cdff7-d876-434e-8e6c-86305f00bbd5>



XMLスタイルシート

RSS形式

マークダウ
ン形式

アップロード

NotebookLMへの
アップロードには、
マークダウン形式
が適している=軽
量かつLLMが扱い
慣れている

Google
NotebookLM

GSPW3: 地球システム科学を支える気候データセット

GSPW3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演みます。

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：GSPW3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、GPCC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

土壤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：GSPW3気候データセット

GSPW3データセットの概要

期間: 1901年～2014年
空間解像度: 0.5° × 0.5°
時間解像度: 3時間

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
基礎となる強制データセット。

20世紀再解析データ
観測データ
GPCCやCRU等の観測
データでバイアス補正



GSPW3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）、土壤水分、貯水量の
再現・再構築研究に利用。



GSPW3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（NBP）、植物の
一次生産、炭素循環の変動要因
分析に使用。

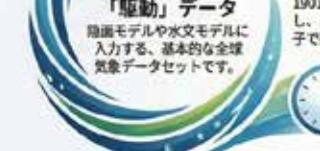
GSPW3の多様な活用事例：気候変動影響評価
CMIP6（統合モデル相互比較プロジェクト）等の国際的な
モデル比較プロジェクトで標準データとして採用。



地球科学研究を支える全球気象データ：GSPW3とは？

GSPW3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。

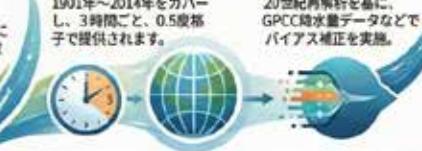


1世紀以上にわたる
高解像度記録

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、
GPCC降水量データなどで
バイアス補正を実施。



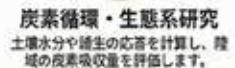
GSPW3の主な活用分野



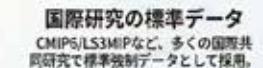
気候変動影響評価
干ばつ、洪水、熱波など、極端気象現象の研究に不可欠です。



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。

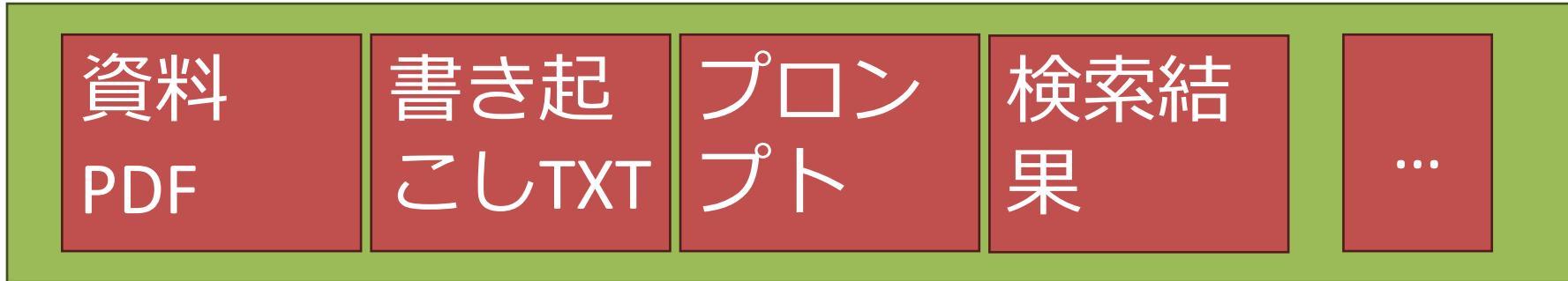


炭素循環・生態系研究
土壤水分や植生の応答を計算し、
陸域の炭素吸収量を評価します。

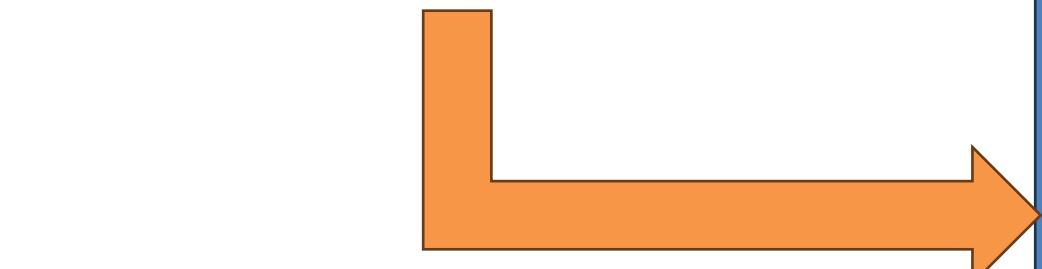


国際研究の標準データ
CMIP6/LS3MIPなど、多くの国際共同研究で標準強制データとして採用。

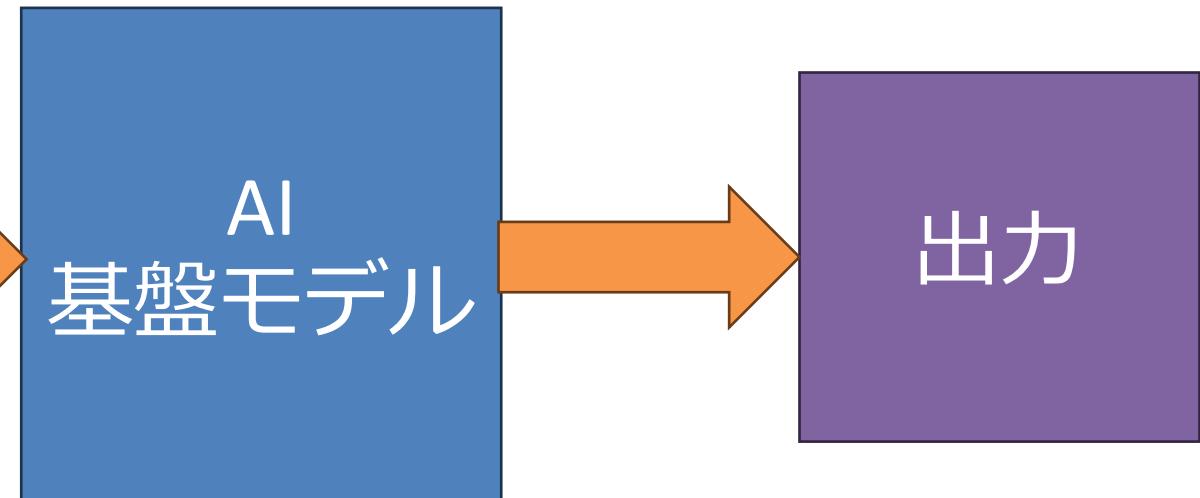
Context Engineering



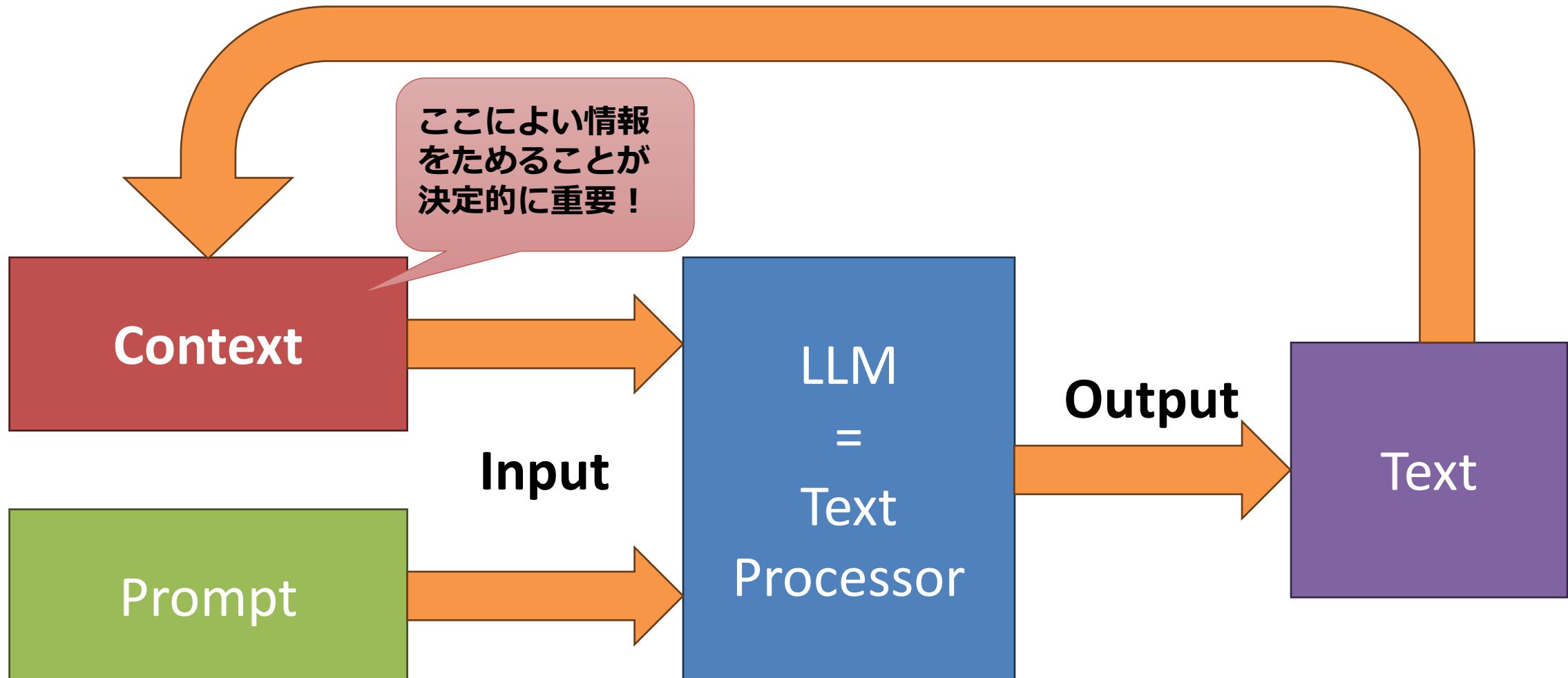
コンテキスト



基盤モデルは、コンテキスト
を入力とし、自身の知識も加
えて、出力に変換する機械



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（ハリシネーション）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

The screenshot shows the Mahalo Research website. At the top right, there are user account details: 'keshikaz@gmail.com' and 'ログアウト'. The main header 'Mahalo Research' includes a gift icon. Below it is the tagline '研究データセット発見システム' and the sub-tagline 'あなたの研究に最適なデータセットを見つめましょう'. A navigation bar at the top has five items: 'データセット要約' (Data Set Summary), 'データセット推薦' (Data Set Recommendation), '関連データセット' (Related Data Set), 'データセット比較' (Data Set Comparison), and 'プロフィール設定' (Profile Setting). A blue info box contains a note about the system being experimental and using DIAS and Mahalo Button data. The main content area features a section titled 'Mahalo Researchへようこそ' (Welcome to Mahalo Research) which describes the system's purpose of helping researchers find suitable datasets. Below this is a '主な機能' (Main Features) section with two cards: 'データセット要約' (Data Set Summary) and 'データセット推薦' (Data Set Recommendation).

データセット要約
利用可能なすべてのデータセットとその要約情報を一覧で確認できます。
[一覧を見る →](#)

データセット推薦
研究計画を入力すると、AIが過去の利用事例を分析し、最適なデータセットを推薦します。
[推薦を受ける →](#)

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 shows the JaLC dataset search results page with the following details:

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

- 63 データセット 527 利用事例 61 メタデータあり
- データセットを検索
- 概要: このデータセットは、1901年から2010年までの110年間にわたり、降水量・気温・放射・風速など9つの大気境界条件変数を3時間ごと・約0.5度解像度で提供する全球陸面モデル用の長期高精度気象強制データです。
- 価値: 20世紀再解析データをスペクトルナッジング技術で力学的ダウンスケーリングすることで、世紀スケールの陸面水文・炭素循環・生態系変動の高精度シミュレーションを可能にし、気候変動の影響評価と帰属分析に必須の基盤データを提供します。
- 応用分野: 陸面モデル評価、水文シミュレーション、炭素循環解析、永久凍土研究、山火事排出量推定、農業干ばつ評価、河川流出予測、生態系モデリングなど、多様な地球システム科学研究に広く活用されています。
- 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

MLIT XRAIN dataset

- 59 利用事例 メタデータあり
- 概要: MLIT XRAINデータセットは、国土交通省が運用するXバンドおよびCバンドマルチパラメータレーダーによる面的降水量などの物理量を、2010年7月から継続的に高時空間分解能（250mメッシュ、1分間隔）で観測したデータです。
- 価値: 局地的豪雨や線状降水帯などの急激な気象現象を詳細に捉えることができ、水災害の防災対策、河川氾濫予測、リアルタイム防災モニタリングに不可欠な高精度降水情報を提供します。
- 応用分野

データセット推薦

- 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
- 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
- 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

The screenshot shows a user profile page with the following sections:

- Top Bar:** Buttons for "プロフィールを更新" (Update Profile), "プロフィールに合うデータセットを表示" (Show datasets matching profile), and "プロフィールを削除" (Delete Profile).
- Profile Summary:** Displays the user's name (Asanobu Kitamoto) and affiliation (National Institute of Informatics (NII)).
- Keywords:** A list of research interests including 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, and Metadata.
- Research Fields:** A list of fields such as Digital Humanities, Machine Learning and AI, Meteorology and Climate Science, Cultural Heritage Informatics, and Geospatial Information Systems.
- Recommended Dataset:** A card for the "Digital Typhoon Dataset".
 - Title:** Digital Typhoon Dataset
 - DOI:** 10.20783/DIAS.664
 - Usage Examples:** 4 pieces | Average Relatedness: 48.3%
 - Relatedness:** 54.9%
- Related Publications:**
 - Title:** Machine Learning for the Digital Typhoon Dataset: Extensions to Multiple Basins and New Developments in Representations and Tasks
 - Abstract:** 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...
 - Authors:** Asanobu Kitamoto, Erwan Delu, Gaspar Faure
 - Year:** 2024
 - Relatedness:** 54.9% - 関連
 - Title:** Digital Typhoon: Long-term Satellite Image Dataset for the Spatio-Temporal Modeling of...
 - Relatedness:** 47.5% - 関連

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 ✘
C:\> home > sync > presentations > jaLC-2026 > library_metadata.yaml
95  file_structure:
96    components:
97      - name: "metadata_json"
98
99
100   files:
101     # --- 画像ファイル (HDF5) ---
102     - name: "image_file"
103       format: "hdfs"
104
105     filename:
106       pattern: "{datetime}-{sequence_id}-{satellite}.h5"
107       example: "2008041300-200801-MTS1-1.h5"
108       regex: "^(\\d{10})-(\\d{6})-(\\w{3}-\\w{2}-\\w{2})\\.h5$"
109
110       components:
111         - name: "datetime"
112           position: "先頭10文字"
113           format: "YYYYMMDDHH"
114
115         - name: "sequence_id"
116           position: "2番目のセグメント"
117           format: "yyymm[年+番号]"
118
119         - name: "satellite"
120           position: "3番目のセグメント"
121
122     hdfs_spec:
123       default_dataset: "Infrared"
124       datasets:
125         - key: "Infrared"
126           dtype: "float64" # 溫度データとして読み込まれる
127           shape: null # 画像サイズは可変 (実データ依存)
128           description: "赤外線画像データ (輝度温度)"
129
130           access_code:
131             import h5py
132             import numpy as np
133             with h5py.File(filepath, 'r') as hsf:
134               image = np.array(hsf.get('Infrared'))
135
136
137     # --- トラックデータCSV ---
138     - name: "track_file"
139       format: "csv"
140       encoding: "utf-8"
```

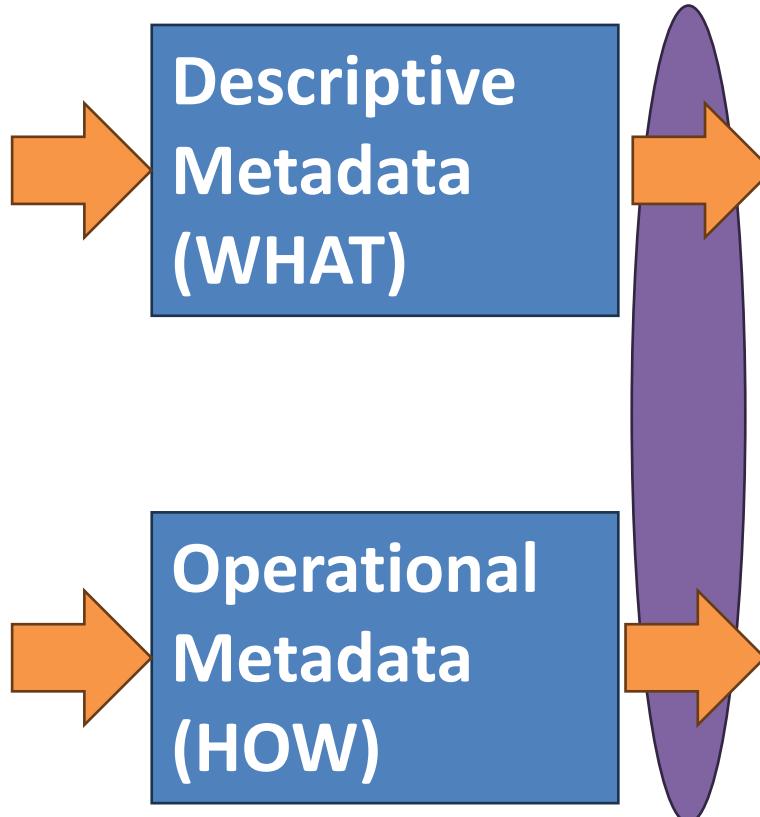
データの処理方法を記述するメタデータを、ソフトウェアから自動的に抽出 (Claude Code Skillsなども活用)

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

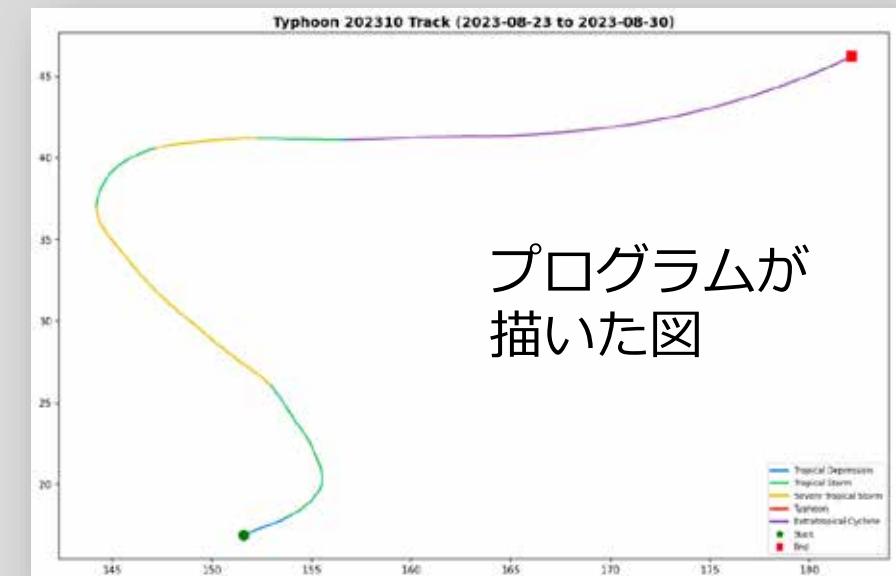
プログラムの自動生成

デジタル台風データセット
<https://doi.org/10.20783/DIAS.664>

Pyphoon2
<https://github.com/kitamoto-lab/pyphoon2>



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



プログラムが
描いた図

AI時代におけるメタデータ再考

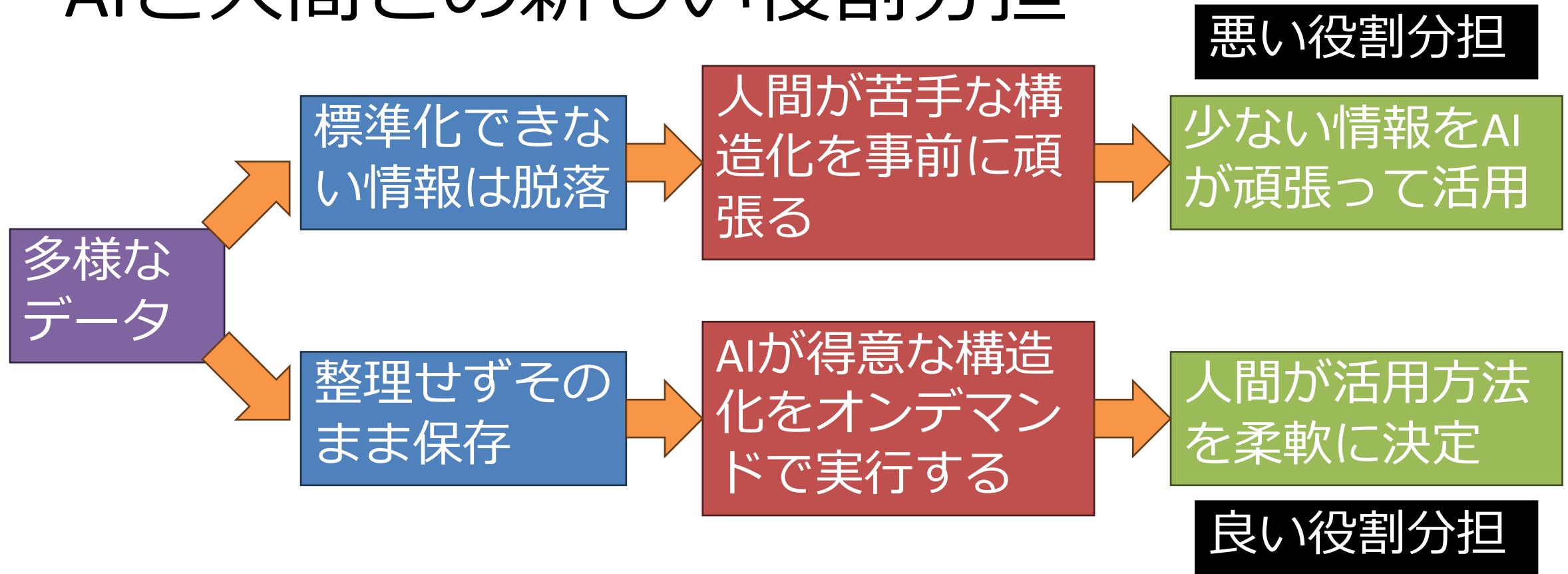
生成AI時代のメタデータ変容

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つた
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