修士論文

二次的生成データに対するメタデータ自動生成に関する研究

荒川　彰太郎

主指導教員 丹　康雄

北陸先端科学技術大学院大学
先端科学技術研究科
（情報科学）

令和５年３月

Abstract

In recent years, as exemplified by the government-led data utilization in Society 5.0, it is expected that the sophistication and efficiency of production activities and the convenience of services will be improved by linking and utilizing data dispersed and held by the national government, local governments, private companies, and other entities.

However, there have been issues in terms of interoperability, such as the lack of uniformity in the notation of information in data catalogs in the linkage of data from various fields in industry, government, and academia. In response to this situation, the Data Society Promotion Council is promoting guidelines for data catalog creation in order to utilize sensing data and device data. This will enable standardization of description methods when explaining metadata, and will make it possible to search for data using metadata and link data between fields in the future.

In data utilization, there is an increasing trend toward the use of not only primary data, which is raw data directly collected from IoT devices, etc., but also secondary data generated by processing multiple pieces of primary data, etc. The number of combinations of original data and processed data is increasing. As the number of types of secondary data increases with the number of combinations of original data and processing, it may become difficult to manually assign metadata to the secondary data. In addition, how to assign metadata is becoming a problem.

In addition, in the case of complex structures that have been processed multiple times, it is impossible to obtain information on the original data and processing from the last generated data, and it is difficult to know how the secondary data was generated.

In this study, we proposed a system to solve these problems. As a proposed system, for the problem of secondary data that increases with the number of combinations, the processor creates a program for metadata generation when the primary data is processed and secondary data is generated. This will allow metadata to be assigned at a stage before the number of types increases, thus solving the problem of secondary data that continues to increase with the number of combinations. To address the issue of the features of the generated secondary data, metadata of the primary data used in the processing process can be added to the metadata of the secondary data as a history to explain the processing route, etc. of the generated secondary data. We believe that this will allow secondary data to be registered in data distribution.

By utilizing the proposed system, we believe that users will be better served by being able to provide secondary data. We also believe that the data distribution market can be expected to be activated as these data become even more available.

目 次

第 1 章　はじめに **1**
1.1 研究背景 1
1.2 研究目的 2
1.3 本文構成 3

第 2 章　関連研究 **4**
2.1 メタデータ 4
2.2 データ流通エコシステム 5

第 3 章　提案手法 **7**
3.1 システムの概要 7

第 4 章　記述例 **10**
4.1 単体の一次データから処理 10
4.1.1 入力メタデータの生成 10
4.1.2 メタデータ記述例 11
4.1.3 出力メタデータ生成 13
4.2 時間方向に処理 14
4.2.1 メタデータ記述 14
4.2.2 出力メタデータ生成 16
4.3 一次データと二次データで単位が変わる処理 17
4.3.1 メタデータ記述 17
4.3.2 出力メタデータ生成 19
4.4 外れ値の処理 20
4.4.1 メタデータ記述 20
4.4.2 出力メタデータ生成 22

第 5 章　評価 **23**
5.1 組み合わせで増える二次データの問題 23
5.2 生成された二次データの由来の問題 24
5.3 提案手法の課題 25
5.3.1 メタデータ生成プログラムの作成の検討 25
5.3.2 メタデータ記述量の検討 29

第 6 章　考察 **31**
6.1　提案システムの有用性 31

第 7 章　おわりに **32**
7.1　まとめ 32
7.2　今後の課題 32

図 目 次

1.1 元データと加工処理の組み合わせ 2
1.2 データの加工経路 2

2.1 メタデータのイメージ [5] 4
2.2 IoT エコシステム 6

3.1 提案システムの概略 8
3.2 データ流通システムの利用方法 9

4.1 単体の一次データから処理 10
4.2 入力メタデータ 11
4.3 入力メタデータ生成 11
4.4 入力メタデータ 12
4.5 出力メタデータ 13
4.6 時間方向に処理 14
4.7 入力メタデータ 15
4.8 出力メタデータ 16
4.9 一次データと二次データで単位が変わる処理 17
4.10 入力メタデータ 18
4.11 出力メタデータ 19
4.12 外れ値の処理 20
4.13 入力メタデータ 21
4.14 出力メタデータ 22

5.1 提案前 23
5.2 提案後 23
5.3 提案前 24
5.4 提案後 24
5.5 単体の一次データから処理（例１）の分類 26
5.6 時間方向に処理（例２）の分類 27
5.7 一次データと二次データで単位が変わる処理（例３）の分類 28
5.8 出力メタデータ分類表 28
5.9 複数回加工処理した出力メタデータ 30

表 目 次

2.1　メタデータとして必要な情報の例 [6]　．．．．．．．．．．．．．．．．．．．　5

第1章　はじめに

本章では，研究背景と研究目的，本論文の構成を示す．

1.1　研究背景

近年，政府が主導するSociety5.0におけるデータ利活用に代表されるように，国，地方公共団体，民間企業などで分散して保有するデータを連携して，活用することで生産活動の高度化・効率化，サービスの利便性の向上等が実現すると期待されている．さまざまな場所に存在するIoTデバイスがネットワークに接続され，そこで生成されたセンシングデータを中心とするデータがクラウド上にビックデータとして収集・蓄積されつつある[1]．こうしたセンシングデータ等の多様な組み合わせによって，画期的なサービスが数多く生まれる期待がある．そのためには，センシングデータやデバイスデータを使ったサービスを開発・提供するデータ利用者が，安全かつ容易にデータを入手できる流通市場の重要性が高まっている．

また，データ利用者が必要とするデータを専門的な知識がなくとも容易に判断してデータを収集し，サービス等に利用できる環境の整備すること等を目的として，データにメタデータを付与する動きが高まっている[2]．データ流通市場においては，IoTデバイスから随時，取得される一次データ等のデータとや，国や自治体などが公開するオープンデータなどの多種多様なデータについて，メタデータを整理・管理することを目的としたデータカタログを整備することが必要である．データカタログを利用することによって，データ利用者側は必要とするデータをデータカタログから検索することができるようになる[3]．

しかし，産官学の多種多様な分野のデータ連携において，データカタログに記載される記述方法等が統一されていなかったため，相互運用をしていくうえで問題があった．こうした状況を受け，センシングデータやデバイスデータを活用するために，データカタログ作成のガイドラインがデータ社会推進協議会[4]によって推進されている．これにより，メタデータを説明する際の記載方法の標準化が可能になり，メタデータを用いたデータの検索や分野間のデータ連携を行うことが今後可能になると考えられる．

データ利活用の場面においてIoTデバイスなどから直接収集した生データである一次データのみならず，複数の一次データ等を加工して生成された二次データの利用が増加傾向にある．図1.1のように元データと加工処理の組み合わせで種類

が増える二次データに対し，人手によるメタデータ付与が困難になることが考えられる．また，メタデータをどのように付与していくかが問題になりつつある．

図 1.2 のように複数回，加工処理を行なった複雑な構造の場合，最後に生成されたデータ E から，元となったデータの情報や加工処理の情報を得ることができず，データ E という二次データがどのようにして生成されたのかがわからなくなることが挙げられる．

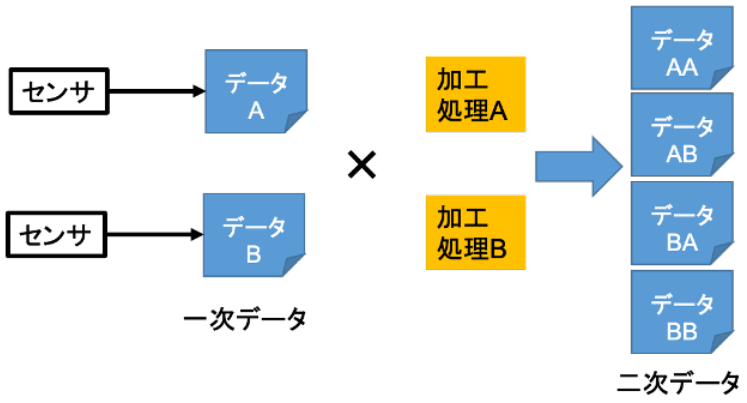

図 1.1: 元データと加工処理の組み合わせ

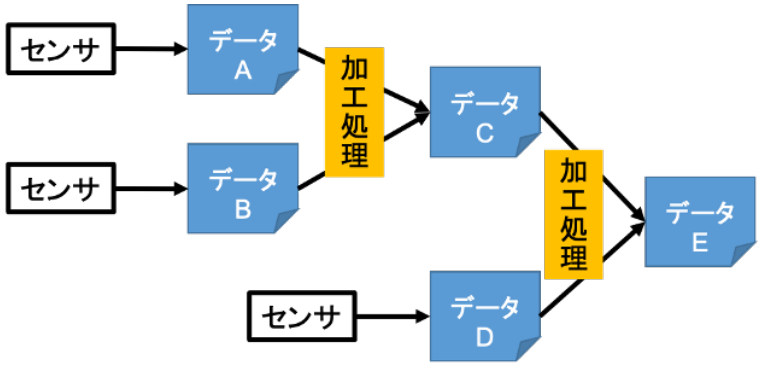

図 1.2: データの加工経路

1.2 研究目的

本研究では，現状を踏まえ，元データと加工処理の組み合わせで種類が増える二次データあるいは高次データに対しては，データを加工処理する際に，加工者側がメタデータ生成のプログラムを作成する．これにより，種類が増える前の段階でメタデータ付与が行えるため，組み合わせで種類が増える二次データに対し，人手によるメタデータ付与が困難になる問題は解決できるのではないかと考える．また，複数回の加工処理を経たデータに対しては，そのデータが生成されるまで

の履歴というものをメタデータに記載する．こうすることで，そのデータが直接，物理的なセンサから取得したデータなのか，二次的に他の物理センサから計算して生成されたデータなのか区別することができる．また，メタデータの中身によっては生成されたデータがどれくらいの精度があるのか，どのくらい信用できるものなかが分かるようになると考える．

1.3　本文構成

本論文では，本章を含めて５章で構成される．１章では研究背景と目的について述べる．２章では関連研究について述べる．３章では提案手法について述べる．４章では提案したメタデータ記述例について述べる．５章では提案したメタデータ記述について評価を行う．６章では考察を行う．７章では本論文のまとめを行う．

第2章　関連研究

2.1　メタデータ

データ利活用において，メタデータはデータを管理するために使われるデータであり，データ利用者にデータの特徴や利用方法などについて理解してもらうための情報である．メタデータの例としては，文書ファイルを作成した日付けや作成者の名前，ファイルデータのサイズ，規格などがある．目的に応じたデータの分類等を検索するためのメタデータをデータの種類ごとにまとめたものをカタログという．メタデータでは，カタログ情報も含め，そこにどのような情報が含まれるのかを階層的に管理するものであり，以下の構造で表される．表2.1[6] に，データ利用プロセスにおいて，データ利用者が必要とするであろう情報の例を示した．このように，メタデータとして必要な情報を付与したメタデータを利用することで，様々な分野で蓄積された膨大なデータに対し，データの効率的な管理，処理，検索などが可能になる．

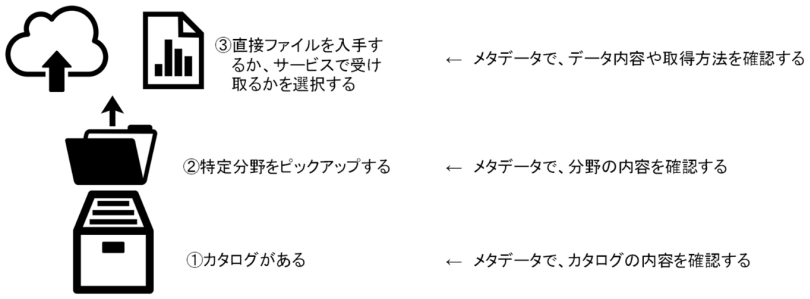

図 2.1: メタデータのイメージ [5]

表 2.1: メタデータとして必要な情報の例 [6]

プロセス	必要な情報の例
検索	適切なキーワード
有用性の判断	データの内容，品質についての説明
ダウンロード	データファイルを入手するための手順
解析	データファイルの構造，数値の単位，精度
結果の解釈	データに影響を与えうる要因（測定機器の修理履歴など）
成果の公表	利用条件（引用や謝辞の書式など）

2.2 データ流通エコシステム

データ流通エコシステムとは，データを収集，保存，処理，共有する方法や手段，およびその過程で関わるものの集合体を指す．データ市場やデータ流通・取引は AI や IoT 技術の発展と実世界から取得したセンシングデータなどのデータ利活用を企業やメーカー間でデータの流通・取引とともに成長した．[7] 今までは，IoT デバイスなどから収集された多種多様なデータをそれぞれの企業などの中で分析，利用していたが，多対多でのデータの交換・取引を目的としたデータ市場ビジネスの社会実装が進んできている．[8]

エコシステムの例を図 2.2 に示す．システムの流れは，まずデータ提供者が IoT デバイスから取得した一次データおよび一次データのメタデータをクラウド等に蓄積する．データ加工者とデータ利用者はクラウドのデータ市場・データ流通システムなどから，必要なデータを検索し取得する．データ加工者は，取得したデータを基に加工処理を行い二次データを生成する．そして，生成した二次データをクラウドに登録する．データ利用者は，取得したデータを使い App やサービス等の開発を行う．

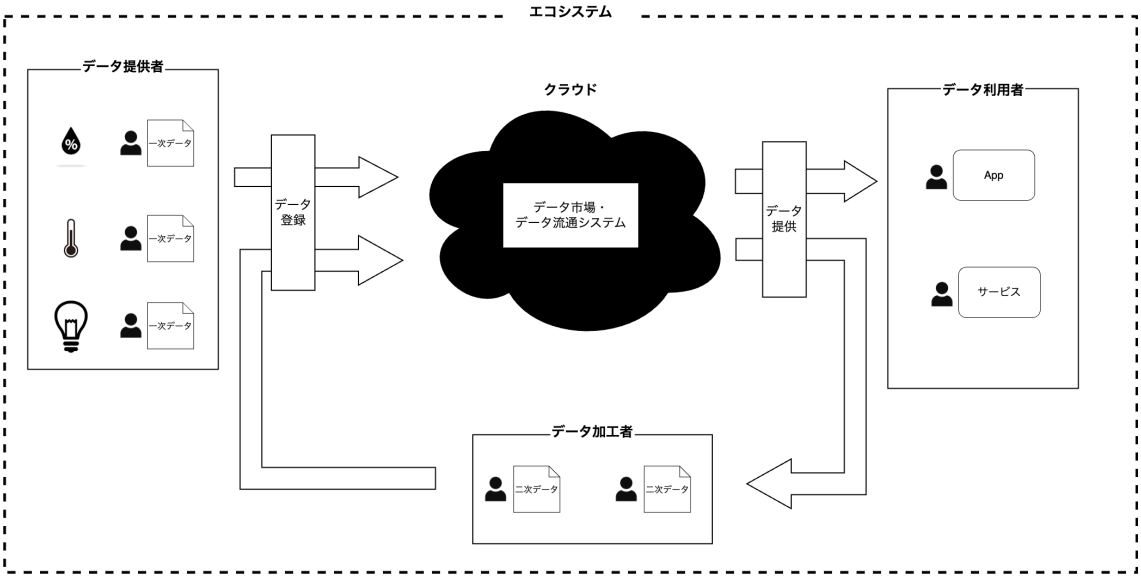

図 2.2: IoT エコシステム

第3章　提案手法

本章では，メタデータ活用のためのデータ流通システムを提案し，その詳細について述べる．

3.1　システムの概要

本研究では，メタデータ活用のためのデータ流通システムを提案する（図 3.1）．データ流通では，以下の三種類のユーザの利用を想定する．

データ提供者

自身の保有する一次データと一次データのメタデータをシステムに登録する．

データ利用者

システムに登録されたデータの中から，特定のデータを発見・取得し，アプリケーションやサービスに活用する．

データ加工者

システムに提供されたデータの中から，特定のデータを発見・取得した後，必要に応じたデータの加工処理を行うことで二次データを生成し，再度システムにメタデータを提供する．また，データを加工処理する際に，加工者がメタデータ生成のプログラムまで作成する．これにより，種類が増える前の段階でメタデータ付与が行えるため，組み合わせで種類が増える二次データに対し，人手によるメタデータ付与が困難になる問題を解決することができる．

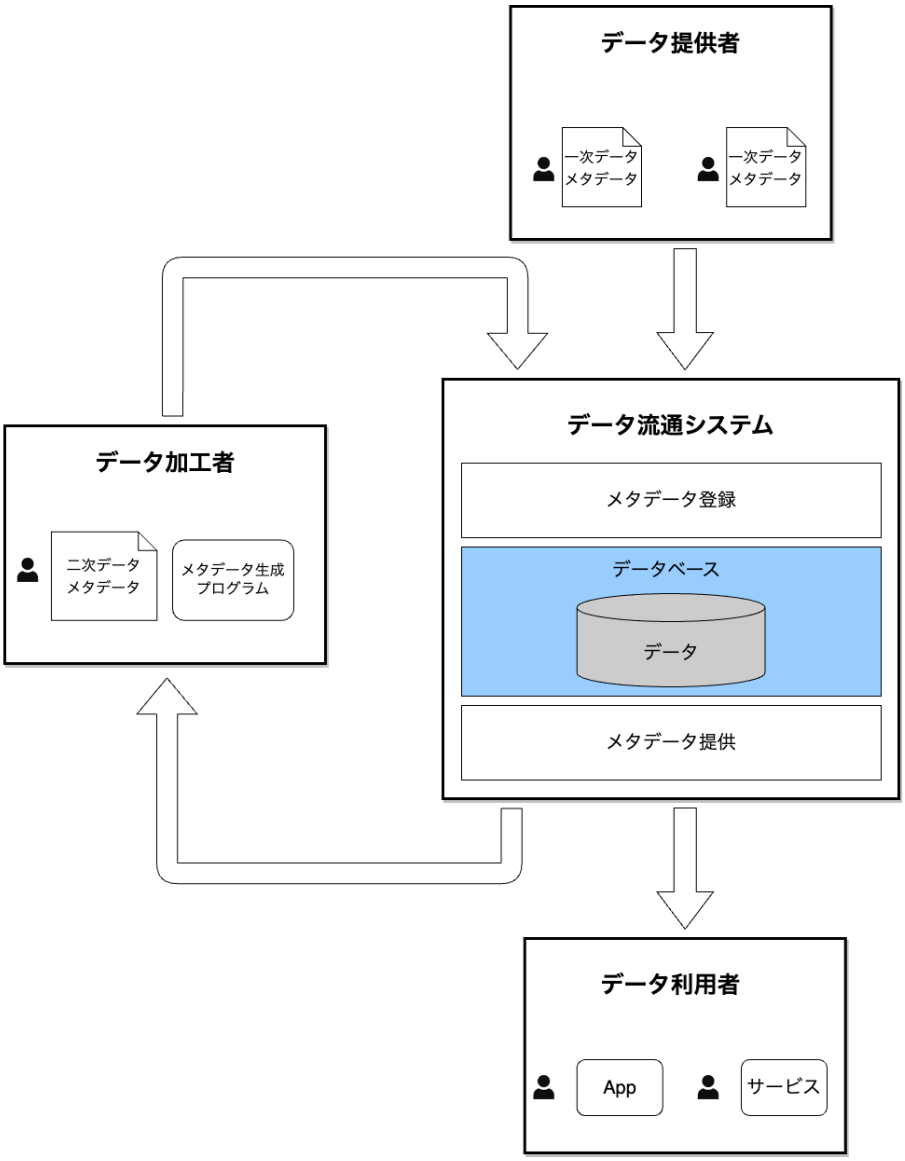

図 3.1: 提案システムの概略

データ流通システムの利用方法

図 3.2 は，前述したデータ提供者，データ利用者，データ加工者を踏まえて，データ流通システムの利用方法を示したものである．データ利用者およびデータ加工者が「メタデータ検索」を行う場合は，データの種類や取得データ，デバイス情報で条件を指定する．データの種類や取得データ，デバイス情報の条件については，後述する実例のメタデータ記述にて詳細を説明する．データ加工者は「メタデータ取得」で得たデータを基に加工処理を行う．その際に，メタデータ生成のプログラムを作成し，二次データのメタデータを「メタデータ登録」する．これによ

り，メタデータの検索から利用までの一連の流れが機能し，提案システムの活用が可能となる．

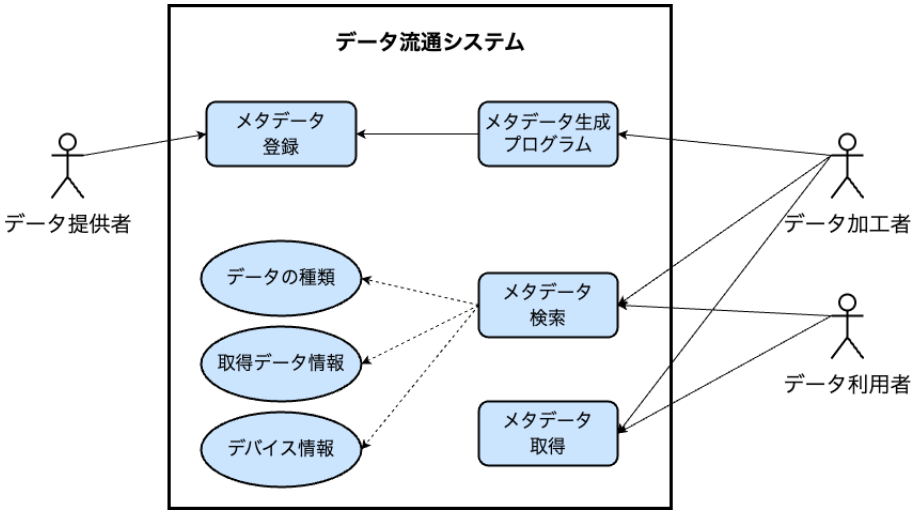

図 3.2: データ流通システムの利用方法

第4章　記述例

４章では，加工処理を行い生成される二次データの種類の特徴，違いを検討する．本来であればあらゆる種類の二次データ処理を網羅した検討をせねばならないところ，以下の特徴を持つこれらの例について検討することにした．
4.1 では単体の一次データを処理する特徴を持つ例を挙げる．
4.2 では時系列に対して処理する特徴を持つ例を挙げる．
4.3 では処理することで単位が変わる特徴を持つ例を挙げる．
4.4 ではセンシングやセンサのノイズを対処する特徴を持つ例を挙げる．

4.1　単体の一次データから処理

単体の一次データから処理は，一つの一次データを加工処理して二次データを生成する．データを加工処理する中で最も単純かつ標準的な例となる．例として，温度センサからセンシングされた一次データを加工処理し，二次データとして飽和蒸気圧を生成する．

図 4.1: 単体の一次データから処理

4.1.1　入力メタデータの生成

入力メタデータでは，ユーザが求めるデータを検索する際に必要となるメタデータを記載している．
今回は, 温度のデータを検索するために測定単位と位置情報, デバイス名を記述する．また, 記述しない項目に関しては「ー」と記述される．記述した例を図 3.4 に示す．入力メタデータの生成の流れは, 必要な項目を記述した入力メタデータをデー

タ連携基盤で検索する．そして，一致するデータがあれば，提供されたデータを基に記述されていない項目の中身が図 3.5 の赤字のように記載され，入力メタデータが生成される．

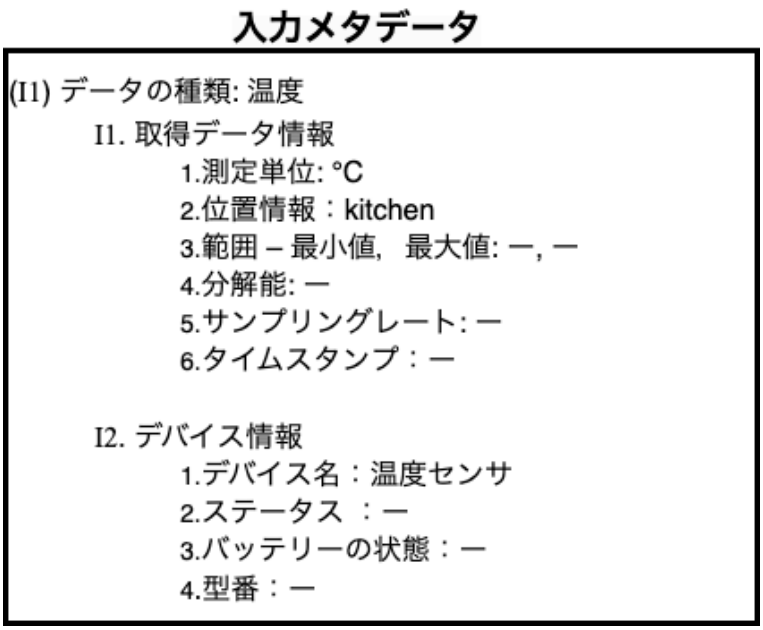

図 4.2: 入力メタデータ

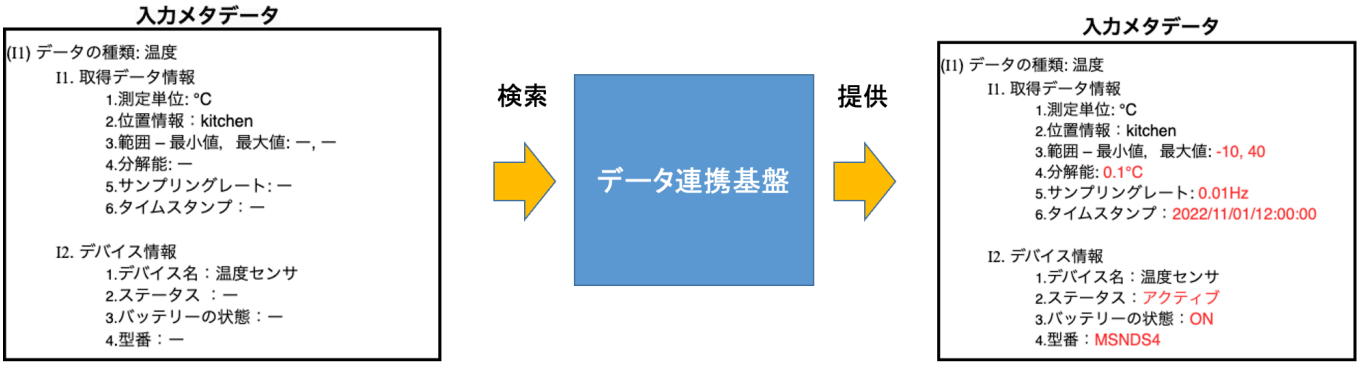

図 4.3: 入力メタデータ生成

4.1.2 メタデータ記述例

（I1）のデータの種類はセンシングされた一次データの種類を示す．
I1. 取得データ情報では，測定単位，取得する温度の範囲，分解能，サンプリングレート，データを取得した時間と位置を示す．
I2. デバイス情報では使用したセンサについての情報を示す．

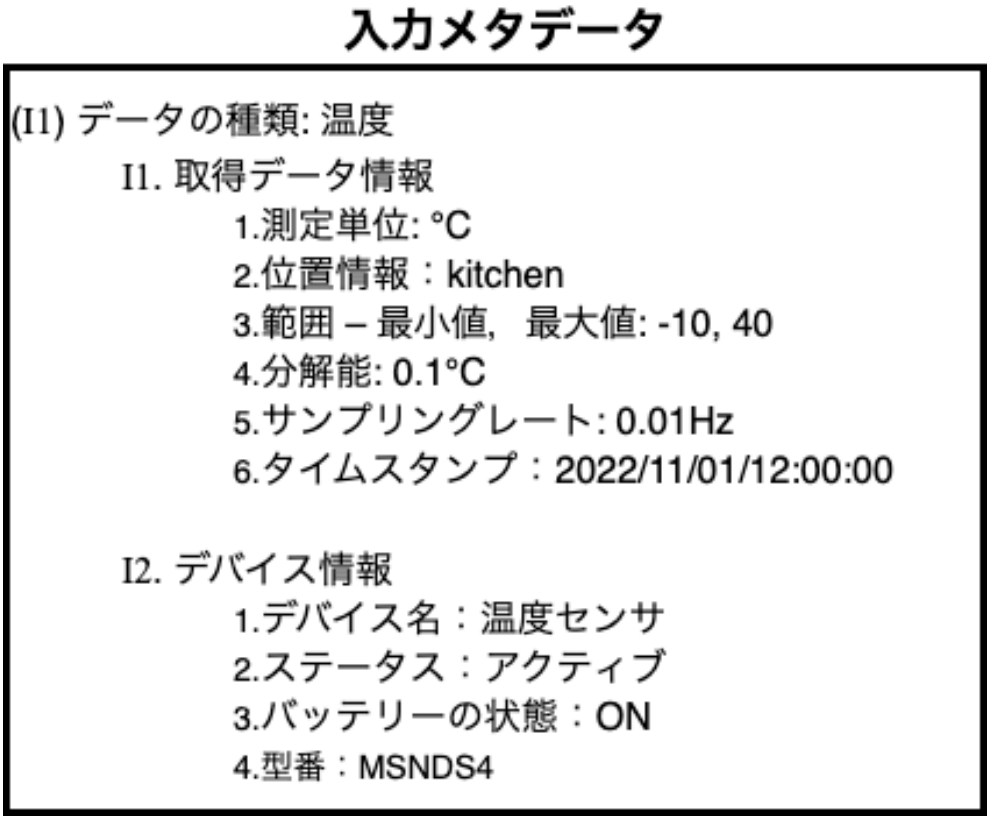

図 4.4: 入力メタデータ

出力メタデータは，入力メタデータを基に加工処理して生成された二次データに対してのメタデータを記載している．

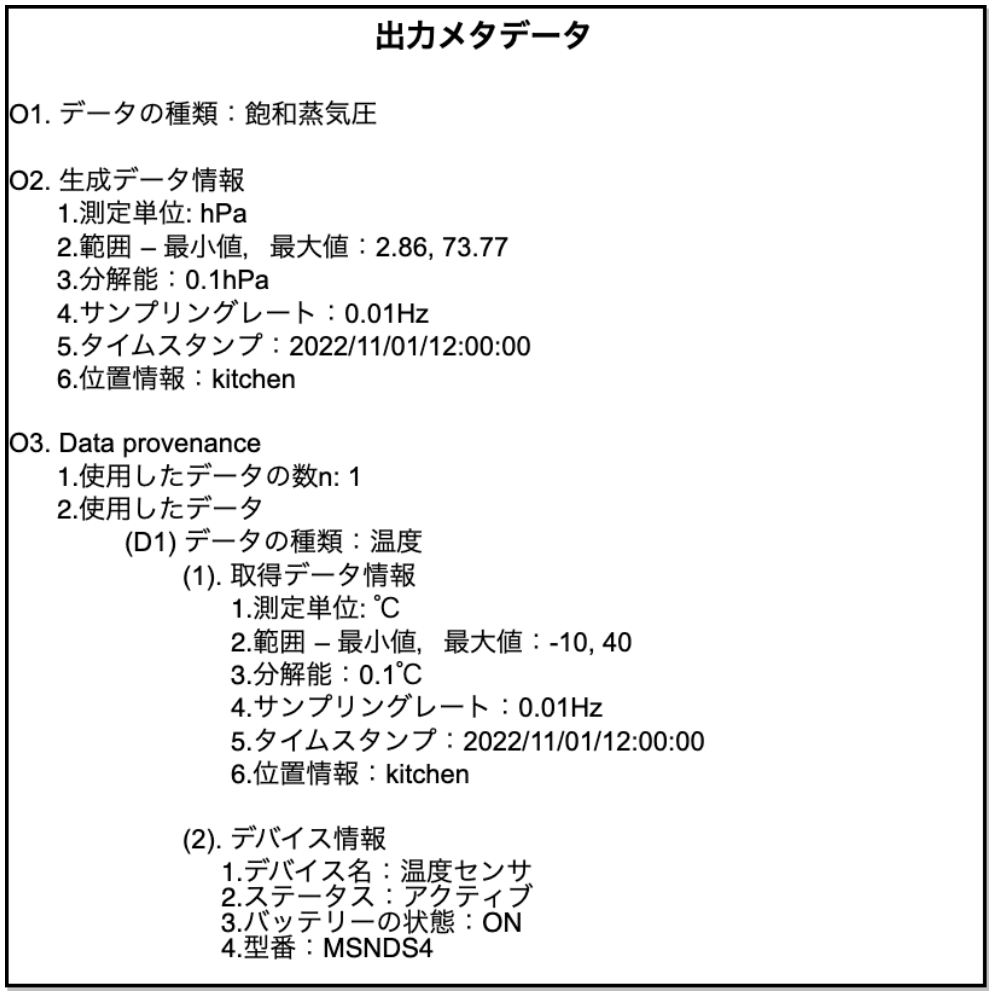

図 4.5: 出力メタデータ

4.1.3 出力メタデータ生成

以下に出力メタデータの生成の手順を示す．

O1. 飽和蒸気圧と記入する．

O2. 1 の測定単位は，飽和蒸気圧のため hPa で表す．2 の範囲は使用したデータ（D1）の最小値，最大値と Tetens の式 $E(t) = 6.11 \times 10^{(}7.5t/(t + 237.3)$ により指定した温度 t ℃における飽和水蒸気圧 E(t) hPa から生成する．3〜6 については使用したデータ（D1）（1）3〜6 に記載されている内容をコピーする．

O3. 1 は 使用したデータの数は 1 つのため，使用したデータの数 n: 1 となる．2. データ 1 には履歴として，使用した一次データのメタデータの内容をコピーする．

4.2　時間方向に処理

時間方向に処理では，複数の同じ種類の一次データを加工処理して二次データを生成する．実例として，一日を通して相対湿度を観測したときの最も小さい値（最小湿度）や当日だけでなく過去の湿度も考慮して算出する実効湿度，最高・最低・平均温度などが該当する．メタデータの記述例として，1 日の最小湿度を示す．湿度センサからセンシングされた複数の一次データを加工処理し，二次データとして最小湿度を生成する．

図 4.6: 時間方向に処理

4.2.1　メタデータ記述

入力メタデータではユーザが求めるデータを検索する際に必要となるメタデータを記載している．
（I1）のデータの種類はセンシングされた一次データの種類を示す．
I1. 取得データ情報では，測定単位，取得する湿度の範囲，分解能，サンプリングレート，データを取得した時間と位置を示す．
I2. デバイス情報では使用したセンサについての情報を示す．

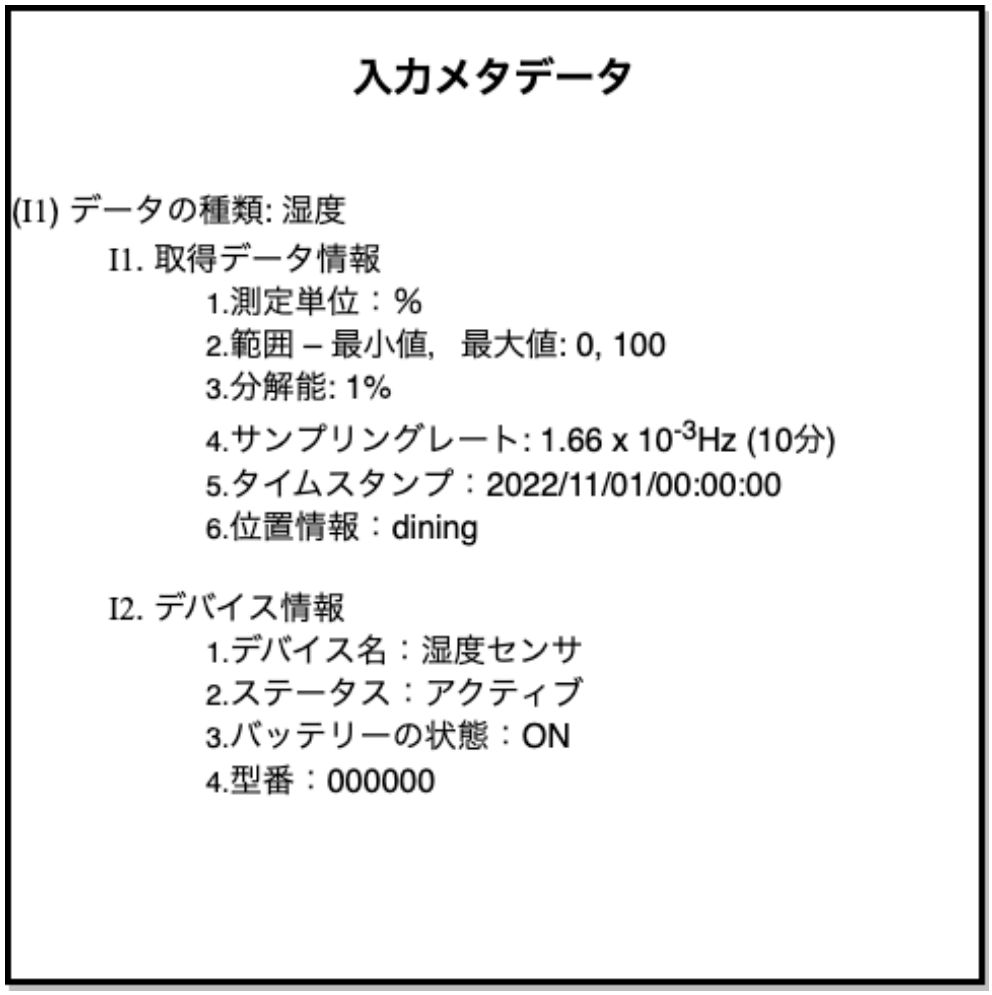

図 4.7: 入力メタデータ

出力メタデータは，入力メタデータを基に加工処理して生成された二次データに対してのメタデータを記載している．

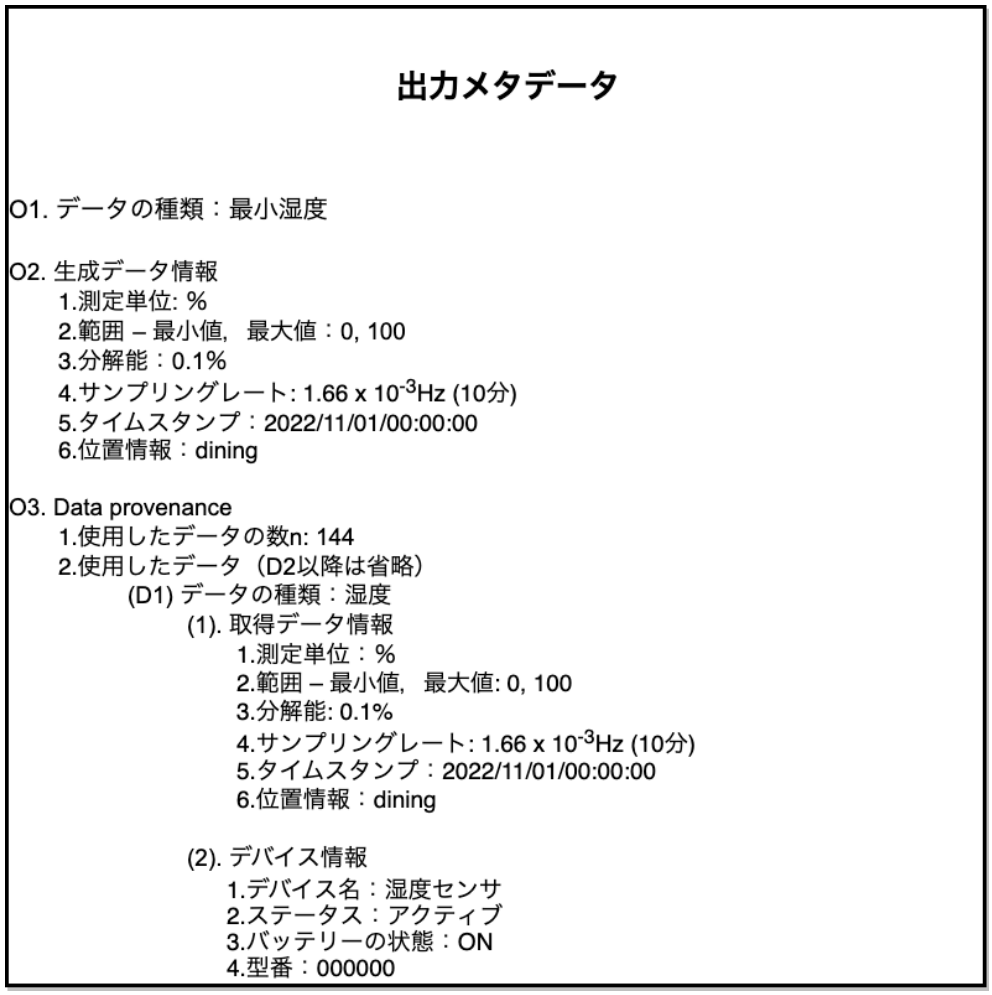

図 4.8: 出力メタデータ

4.2.2 出力メタデータ生成

以下に出力メタデータの生成の手順を示す.

O1. データの種類には最小湿度と記入する.

O2. 1〜5 は使用したデータ（D1）（1）1〜5 に記載されている内容をコピーする．6 は使用したデータの位置情報が同一の場合のみ使用する．7 では蓄積されたデータから最小の湿度を出力する.

O3. 1 は使用したデータの数は 144 のため，使用したデータの数 n: 144 となる．2 は履歴として，使用した一次データのメタデータの内容をコピーする．これを使用したデータの数だけ行う.

4.3　一次データと二次データで単位が変わる処理

一次データと二次データで単位が変わる処理では，データの種類が違う単体・複数の一次データを加工処理して二次データを生成する．実例として，温度 (℃) から湿度 (%）から不快指数（単位なし），絶対湿度（体積絶対湿度 g/m^3）求める手法や黒球温度 (℃) と湿球温度 (℃) から屋内の暑さ指数 WBGT（単位なし）などが挙げられる．

例として，温度センサと湿度センサからセンシングされた一次データを加工処理し，二次データとして不快指数を生成する．

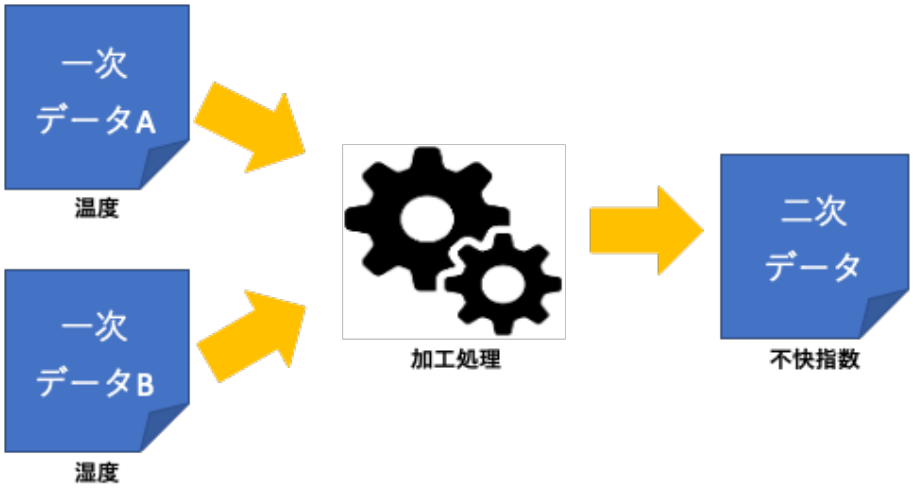

図 4.9: 一次データと二次データで単位が変わる処理

4.3.1　メタデータ記述

入力メタデータではユーザが求めるデータを検索する際に必要となるメタデータを記載している．

（I1）と（I2) のデータの種類はセンシングされた一次データの種類を示す．

I1. 取得データ情報では，それぞれの測定単位，取得する湿度の範囲，分解能，サンプリングレート，データを取得した時間と位置を示す．

I2. デバイス情報では，それぞれで使用したセンサについての情報を示す．

入力メタデータ

(I1) データの種類: 温度

I1. 取得データ情報

1.測定単位: °C
2.範囲 – 最小値，最大値: -10, 40
3.分解能: 0.1°C
4.サンプリングレート: 0.01Hz
5.タイムスタンプ：2022/11/01/12:00:00
6.位置情報：kitchen

I2. デバイス情報

1.デバイス名：温度センサ
2.ステータス：アクティブ
3.バッテリーの状態：ON
4.型番：MSNDS4

(I2) データの種類: 湿度

I1. 取得データ情報

1.測定単位：%
2.範囲 – 最小値，最大値: 0, 100
3.分解能: 1%
4.サンプリングレート: 0.01Hz
5.タイムスタンプ：2022/11/01/00:00:00
6.位置情報：kitchen

I2. デバイス情報

1.デバイス名：湿度センサ
2.ステータス：アクティブ
3.バッテリーの状態：ON
4.型番：000000

図 4.10: 入力メタデータ

出力メタデータは，入力メタデータを基に加工処理して生成された二次データに対してのメタデータを記載している．

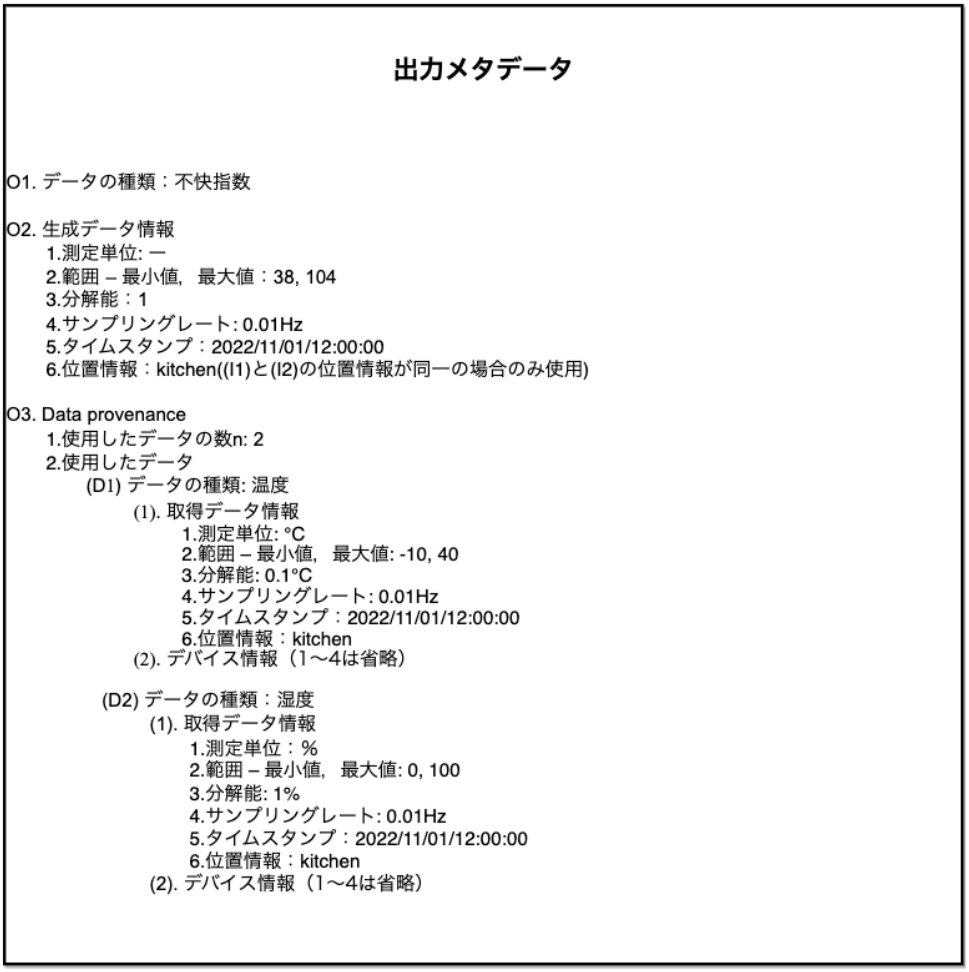

図 4.11: 出力メタデータ

4.3.2 出力メタデータ生成

O1. データの種類には不快指数と記入する．

O2. 1. 測定単位は，不快指数には単位がないため ー と表記する．2. 最小値・最大値は，使用したデータ (D1) の（1).2 と不快指数 DI（T は気温℃，H は湿度%）$DI = 0.81T + 0.01H \times (0.99T - 14.3) + 46.3$ をもとに生成する．3. 分解能は，不快指数は自然数で表すため，1 とする． 4〜5 は使用したデータ（D1）(1) 4〜5 と（D2）(1) 4〜5 に記載されている内容をコピーする． 6. 位置情報は，使用したデータの位置情報が同一の場合のみ使用する．

O3. 1. 使用したデータの数は 2 つのため，使用したデータの数 n: 2 となる．使用したデータには履歴として，使用した温度と湿度のデータのメタデータをコピーする．

4.4　外れ値の処理

外れ値の処理の位置づけとして，3.3 のような時系列の平均をとったものではなく，センシングやセンサのノイズの対処することを目的として処理する．実例として，部屋の四隅に温度センサを設置し，あるセンサの値が他のセンサから推測される値から大きく離れている場合の処理を示す．

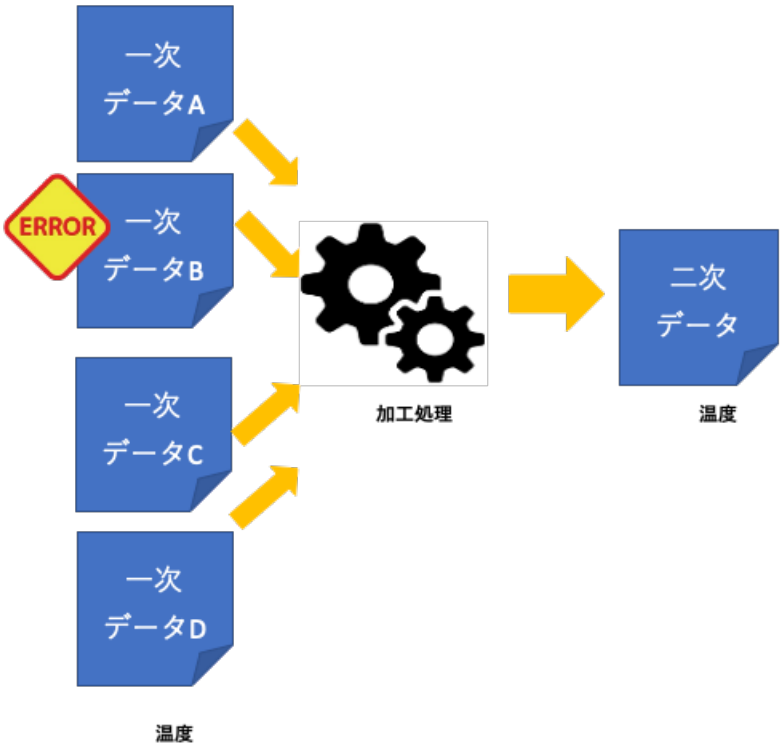

図 4.12: 外れ値の処理

4.4.1　メタデータ記述

入力メタデータではユーザが求めるデータを検索する際に必要となるメタデータを記載している．それぞれを I1（Input1）などで分類する．
（I1）のデータの種類はセンシングされた一次データの種類を示す．
I1. 取得データ情報では，測定単位，取得する湿度の範囲，分解能，サンプリングレート，データを取得した時間と位置を示す．四隅に設置された温度センサが天井か床に設置されているか判断するために設置高度を記述する．
I2. デバイス情報では使用したセンサについての情報を示す．

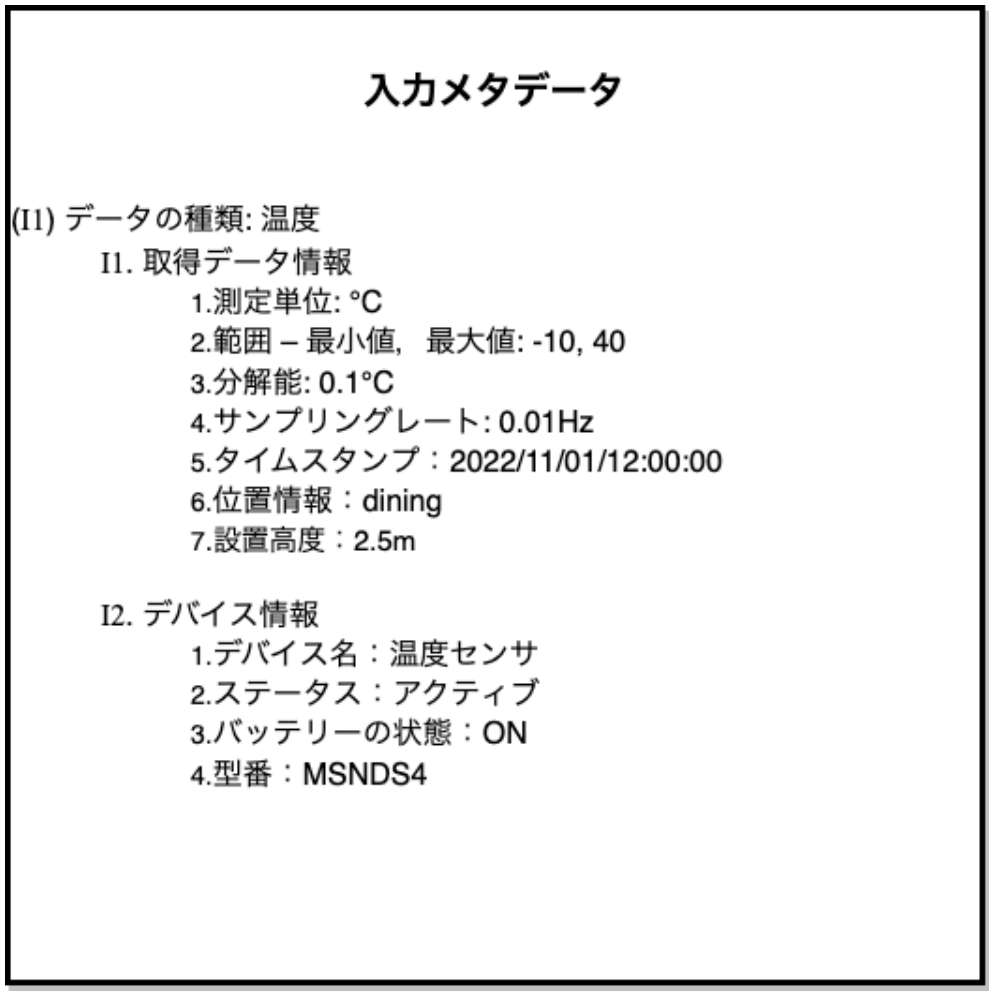

図 4.13: 入力メタデータ

出力メタデータは，入力メタデータを基に加工処理して生成された二次データに対してのメタデータを記載している．それぞれを O1（Output1）などで分類する．

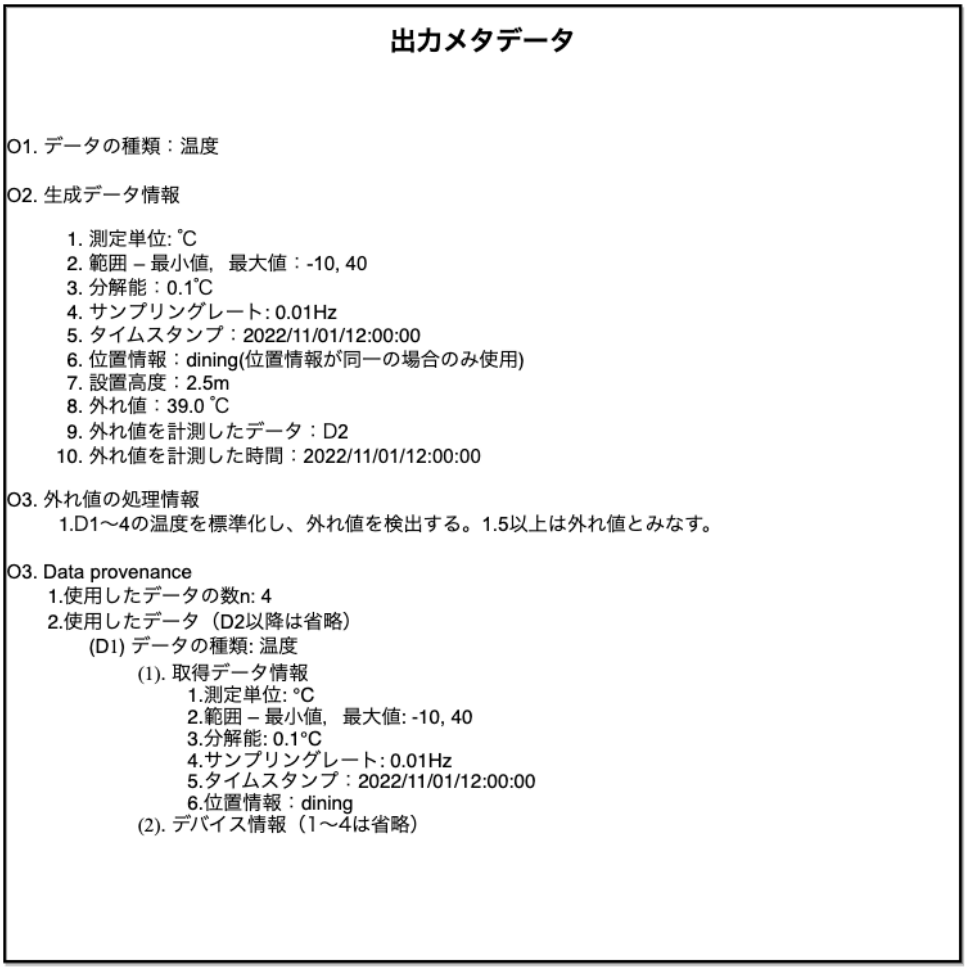

図 4.14: 出力メタデータ

4.4.2 出力メタデータ生成

O1. データの種類には温度と記入する．

O2. 1〜5 は使用したデータ (D1)(1)1〜5 に記載されている内容をコピーする．6. 位置情報は，使用したデータの位置情報が同一の場合のみ使用する．7. 設置高度は，センサの設置が床か天井かを判断するのに必要となる．8. 外れ値では，使用したデータ（D1 4）の温度を標準化し，外れ値を検出する．今回は 1.5 以上を外れ値とみなす．9. 外れ値を計測したデータでは，使用した 4 つのデータのうちどれが外れ値を計測したかを記述する．10. 外れ値を計測した時間では，外れ値を検出したと時間を記述する．

O3. 1. 使用したデータの数は 4 つのため，入力の数 n: 4 となる．2. 使用したデータには履歴として，使用したデータのメタデータの内容をコピーする．これを使用したデータの数だけ行う．

第5章　評価

5.1　組み合わせで増える二次データの問題

図 5.1 で示すように，元データと加工処理の組み合わせで種類が増える二次データに対し，人手によるメタデータの付与が困難になる問題があった．これに対して，データを加工処理する際に，加工者側がメタデータ生成のプログラムを作成する．これにより，図 5.2 のように種類が増える前の段階でメタデータ付与が自動的に行えるため，生成された二次データに対して，一つ一つ人手でメタデータ付与を行う必要がなくなると考える．

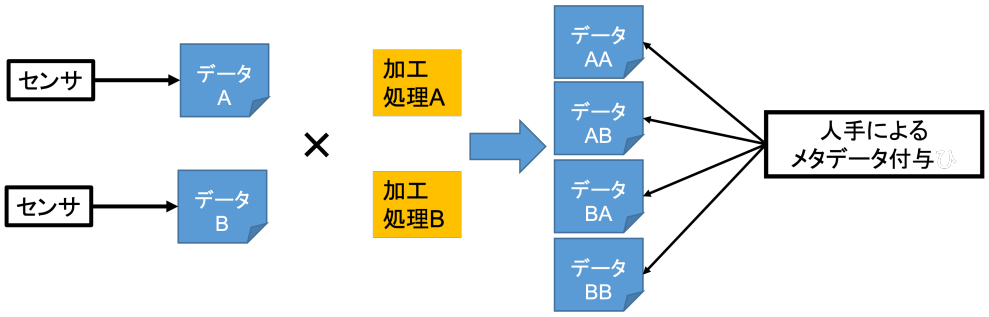

図 5.1: 提案前

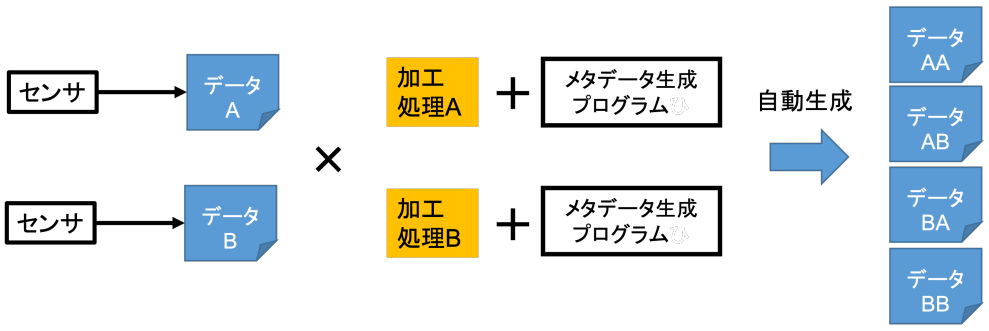

図 5.2: 提案後

5.2　生成された二次データの由来の問題

図 5.2 で示すように，提案前までは, 生成された二次データの由来や加工経路を説明することができなかったため, データ利用者は目的に合ったデータか判断できず, データの信頼性が担保されないということが問題となっていた．これに対して, 生成された二次データのメタデータに履歴となる部分をつけることで, 二次データの由来や加工経路を説明することができ, データ利用者が求める信頼できるデータかどうか判断することができるようになる．

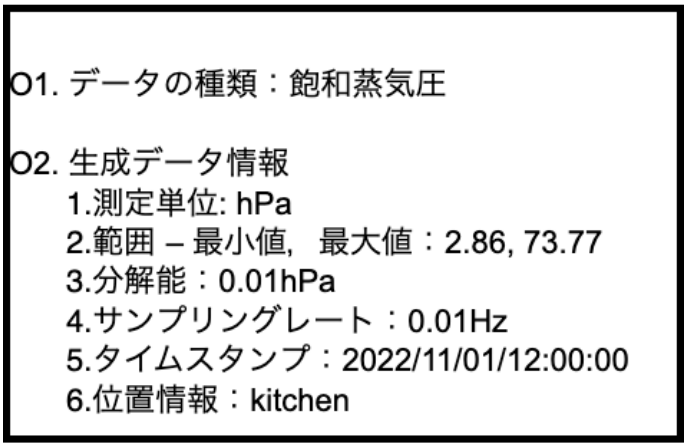

図 5.3: 提案前

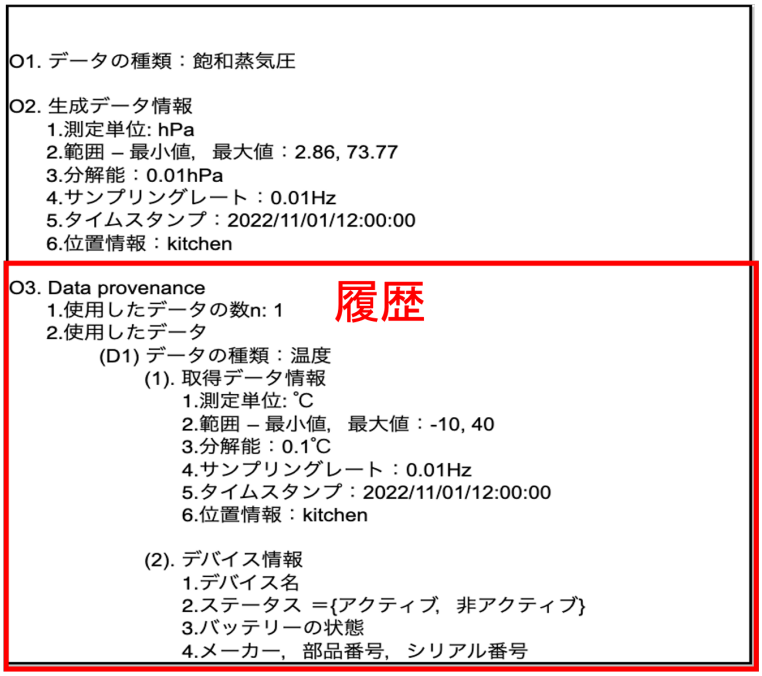

図 5.4: 提案後

5.3 提案手法の課題

提案をする中で 2 点の課題があり検討した. 一つは, 組み合わせで増える二次データの問題に対し, 加工者がメタデータ生成プログラムを作成することを提案したが, メタデータ生成プログラムの作成はどの程度で作成できるのかという点. もう一つは, 生成された二次データの由来の問題に対し, 生成された二次データに履歴として使用したメタデータを記載することを提案した. しかし, 複数回加工処理した場合, メタデータの記述量というのはどれくらい増加するのかという課題がある.

5.3.1 メタデータ生成プログラムの作成の検討

メタデータ生成プログラムの作成の難易度を検討するために, メタデータの生成プログラムの流れについて 4.1 で用いた単体の一次データから処理（例１）を例に挙げて説明する.

入力メタデータを基に作成した出力メタデータの項目を生成方法ごとに３つに分類した. 分類したものを図 5.5 に示す. 固定の項目は, 加工処理の内容に応じて, 最初から固定的に決まるもののことであり, データの種類と測定単位, 使用したデータの数がそれに該当する. そのため, 今回であれば, 飽和蒸気圧に加工するプログラムを書く以上, 機械的にメタデータを記述することができる. 使用したデータをコピーは, 入力メタデータで得られたメタデータをコピーすることで記述できるもののことである. 処理部分は, 固有の値で加工処理を行いメタデータを生成しないといけないものに限られるため, 本来のデータの加工処理のプログラムと関係のある処理を行う. そのため, メタデータ生成のプログラムは加工処理のプログラムと付随して作成が可能ではないかと考える.

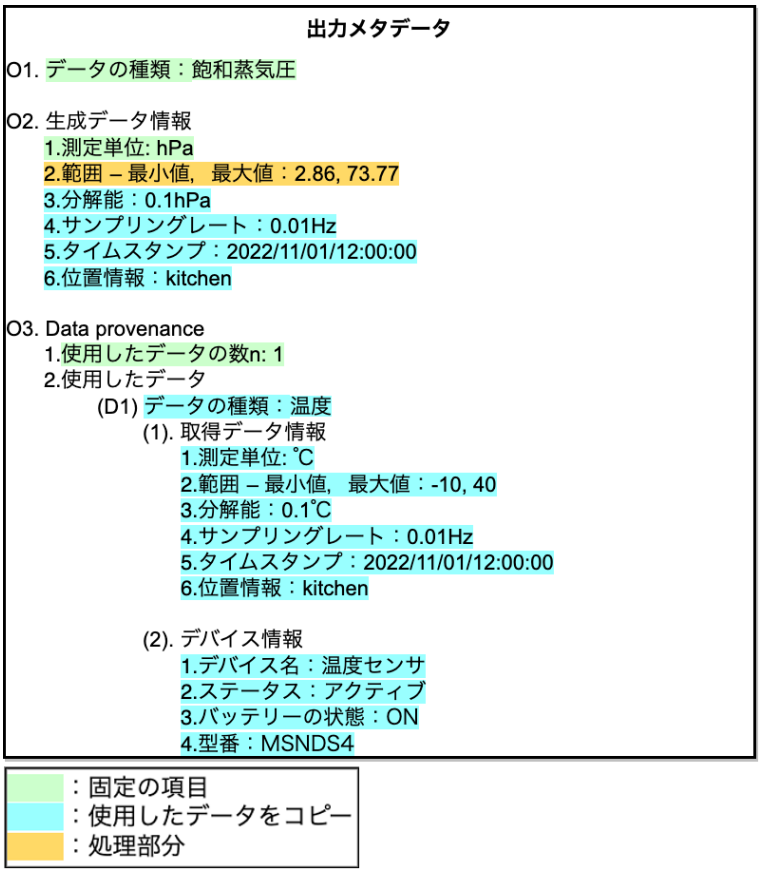

図 5.5: 単体の一次データから処理（例１）の分類

実例の時間方向に処理（例２）と一次データと二次データで単位が変わる処理（例３）の出力メタデータに対しても分類を行った. その結果を図 5.6, 図 5.7 に示す. 時間方向に処理は, 固定の項目とコピーして記述するもののみで構成されているため機械的にコピーして作成が可能であり, 一次データと二次データで単位が変わる処理は, 処理部分は一つのみで大部分は, 固定の項目とコピーして記述するもので構成されているため, 作成が可能だと考える.

実例のメタデータの項目を分類した結果, どの実例も全体の８割以上が使用した一次データのメタデータをコピーしたものとなったため, ほとんどがメタデータを機械的にコピーして記述し作成することが可能だと考える. 固定の項目部分も加工処理をする時点で決まる項目のため, メタデータ生成プログラムは比較的容易に作成することができると考える. 処理部分については, 処理内容に応じて計算し記述する必要があるが, メタデータ生成プログラムは加工処理のプログラムと並行して作業が可能である.

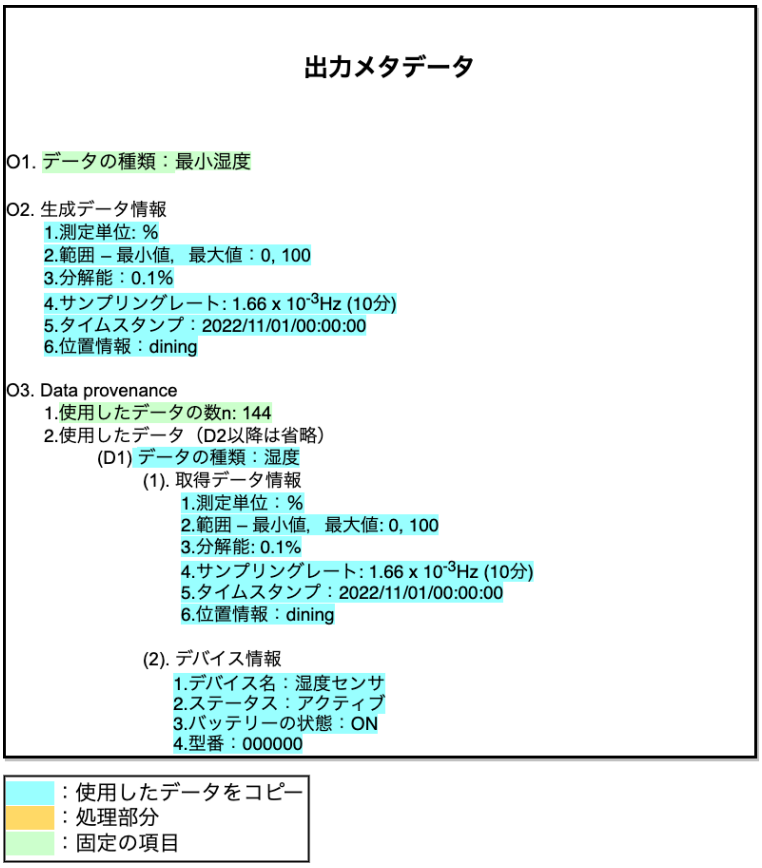

図 5.6: 時間方向に処理（例２）の分類

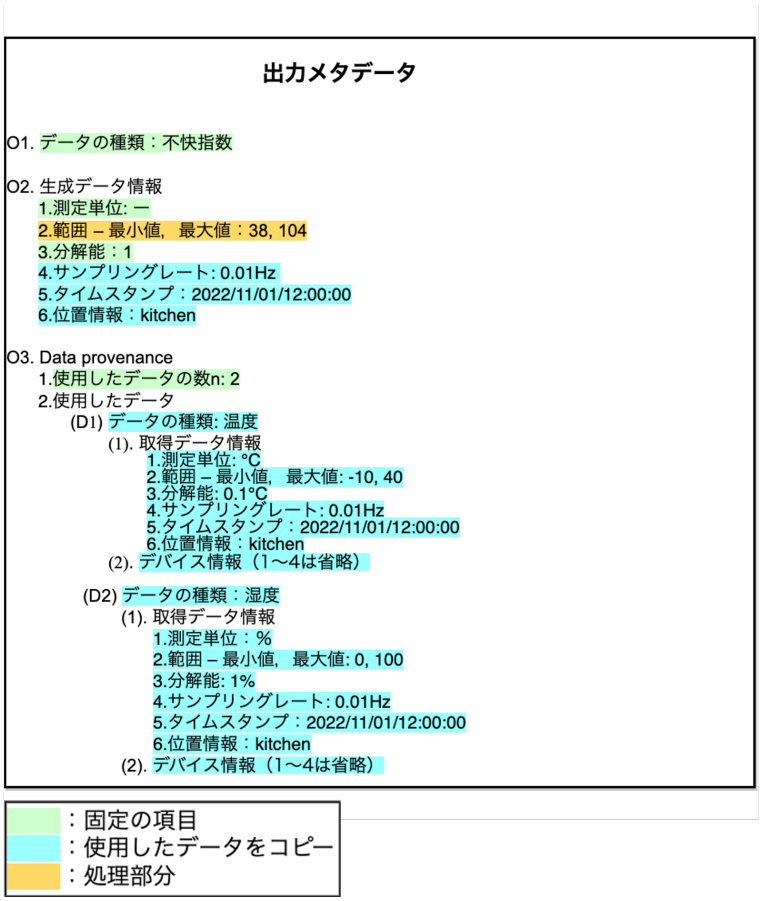

図 5.7: 一次データと二次データで単位が変わる処理（例３）の分類

項目の分類	例１	例２	例３
固定の項目	3	2	4
使用したデータをコピー	15	1590	25
処理部分	1	0	1

図 5.8: 出力メタデータ分類表

5.3.2 メタデータ記述量の検討

一次データと二次データの二つに対して, 加工処理を行い生成される高次データのメタデータの記述の例を図 5.9 に示す. 例から, 複数回加工処理を行ったメタデータの記述量は, 使用したメタデータ数の足し算で増えるため, 複数回加工処理をすることで記述量が膨大になることはないと考える.

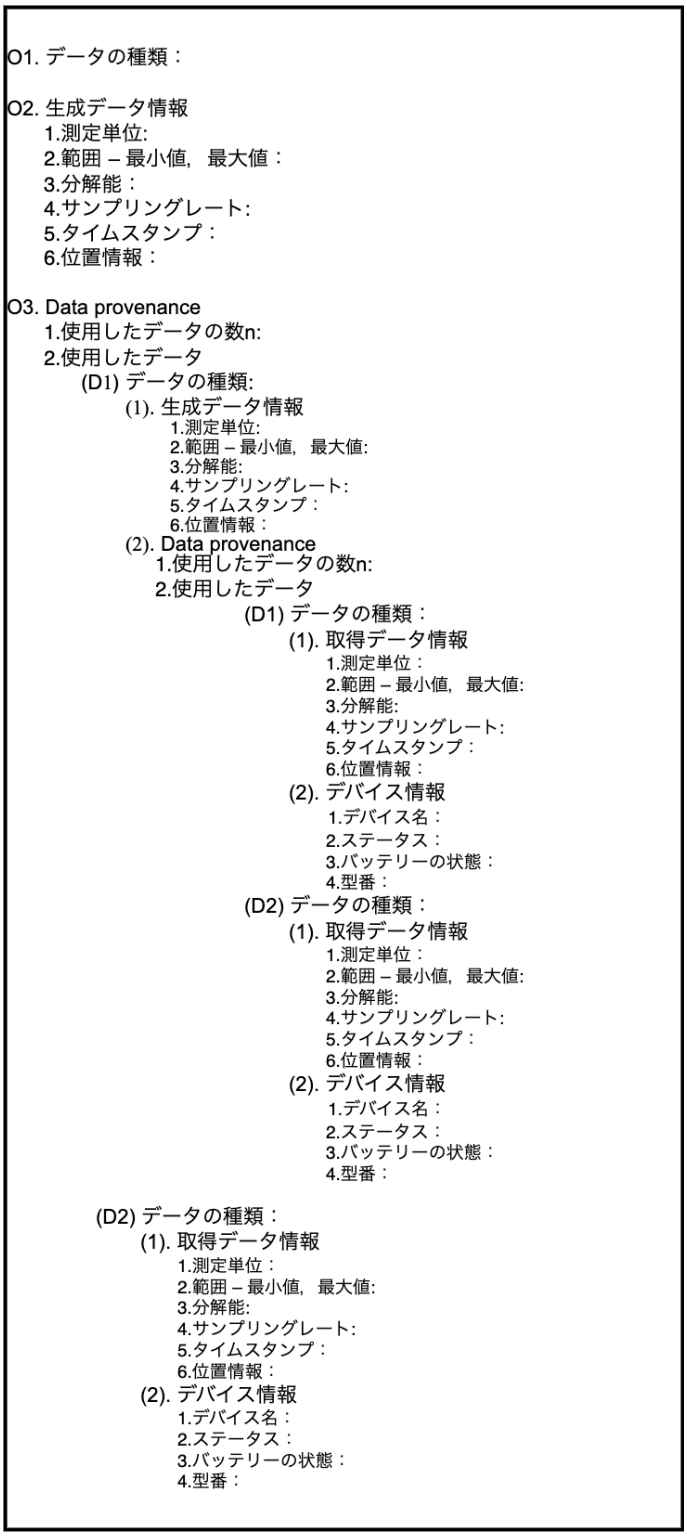

図 5.9: 複数回加工処理した出力メタデータ

第6章　考察

6.1　提案システムの有用性

本研究は，元データと加工処理の組み合わせで種類が増える二次データに対し，人手によるメタデータの付与が困難になる問題と生成された二次データから，元となった情報を得ることができず二次データがどのようにして生成されたのかわからなくなる問題に対しての検討を行った．

増加する二次データに対して人手によるメタデータ付与が困難になる問題に対しては，提案システムで示したようにデータを加工処理する際に，加工者がメタデータ生成のプログラムまで作成する．これにより種類が増える前の段階でメタデータ付与が行えるため，組み合わせで増加し続ける二次データの問題を解決することができると考える．

生成された二次データの素性に対する問題に関しては，二次データのメタデータに履歴として加工処理に使用した一次データのメタデータを付与することで，二次データの説明をすることができる．

しかし, 提案する中で 2 点の課題があり, それについて検討した．メタデータ生成プログラムの作成の難易度については, 多くのメタデータ項目は機械的にコピーして記述し, また固定の項目もあるため, 作成は比較的容易だと考える．複数回加工処理した場合のメタデータの記述量が増加することについては, 記述量が膨大になるということはないため, 実用上問題にはならないと考える．

これにより，二次データをデータ流通に登録することができる．また，データ利用者は，メタデータ検索から今まで以上に広範な選択肢からデータを選択し，より目的に合致したデータが利用可能となり，実現されるサービスの質の向上が期待される．

第7章　おわりに

7.1　まとめ

本研究は，元データと加工処理の組み合わせで種類が増える二次データに対し，メタデータをどのように付与していくかという問題と生成された二次データから，元となった情報を得ることができず二次データの素性がわからなくなる問題に対して解決するためのシステムを提案した．

提案システムとして，組み合わせの数で増加する二次データの問題に対しては，一次データを加工処理し二次データを生成する際に，加工者がメタデータ生成のプログラムを作成する．これにより，種類が増える前の段階でメタデータ付与が行えるため，組み合わせで増加し続ける二次データの問題を解決することができると考える．

生成された二次データの素性に対する問題に対しては，二次データのメタデータに履歴として加工処理に使用した一次データのメタデータを付与することで，生成された二次データの加工経路などの説明をすることができる．

提案する中で, メタデータ生成プログラムの作成の難易度と複数回加工処理したメタデータの記述量の増加について課題があり検討した. 結果として, どちらの課題についても実用上問題にならないことがわかった.

提案したシステムの活用することで，二次データを提供することができるようになることで，ユーザーがより良いサービスが実現できると考えられる．また，これらのデータを更に利用できるようになり，データ流通市場が活性化することも期待できると考える．

7.2　今後の課題

本研究では，提案システムを実装し，性能の評価をすることができなかった．そのため，システムの実装に向けた調査や設計開発を行う．また，サンプルデータ及び処理手法をもとに，生成されたメタデータが適切に記述されているか評価を行う必要がある．実装したシステムと従来のシステムを比較して，メタデータの生成速度など，問題点が改善したかを比較検討する必要がある．

メタデータの生成プログラムは, 多くのメタデータ項目は機械的にコピーして記述し作成でき, 処理内容に応じて計算し記述する必要があるが, 加工処理のプログ

ラムと並行して作業が可能である. そのため, 機械的にメタデータを自動生成する手法について検討したいと考える.

謝辞

本研究を行うにあたり，終始ご指導ご鞭撻を賜りました丹康雄教授に深く感謝致します．審査員をお引き受け頂いた本学リム勇仁准教授，BEURAN Razvan 准教授，谷川忍教授には，新たな視点から多大なご助言を頂きました，深く感謝致します．本研究室の PHAM Van Cu 特任助教，博士後期課程 XIN Tao 氏には，研究に関して活発な議論，ご指導を賜りました．心から感謝致します．また，丹・リム研究室の皆様には，論文や発表資料の添削などの研究活動ばかりでなく日常生活においてもご協力いただきました．厚く感謝申し上げます．

最後に，学生生活を支えて頂いた家族に感謝いたします．ありがとうございました．

参考文献

[1] 小田利彦，今井紘，内藤丈嗣 and 竹林一．センシングデータ流通市場におけるメタデータの定義・生成・活用の一方向，人工知能学会，pp.1–2（2018）

[2] 始動するＩｏＴデータ流通と EverySense ジャパンの取り組み URL:https://www.jftc.go.jp/cprc/katsudo/bbl_files/221th_bbl.pdf

[3] 総務省・経済産業省, データ流通プラットフォーム間の連携を実現するための基本的事項, 総務省 経済産業省, 4 平成 29.

[4] 一般社団法人データ社会推進協議会データカタログ作成ガイドライン【公式公開資料】URL:https://www.jeita.or.jp/japanese/pickup/category/190314.html

[5] メタデータルールと利用イメージの検討，政府 CIO 補佐官等ディスカッションペーパー（2021）

[6] 真板英一. データペーパー投稿者のためのメタデータ作成ガイド, 日本生態学会誌, vol.63, no.2, pp275–281,(2013)

[7] 早矢仕晃章 and 大澤幸生. データ市場ビジネスの動向調査によるデータ流通エコシステムの成長に関する一考察, 人工知能学会全国大会論文集 第 34 回, pp.1F5OS404–1F5OS404,(2020)

[8] 早矢仕晃章 and 大澤幸生．データ流通エコシステムのデザインと実践的課題, 人工知能学会, pp.605–608, (2021)