

IoT データ収集システムのデータアーキテクチャ

松波 成行^{1,a)} 松田 朝彦¹ 知京 豊裕¹ 原田 善之¹ 吉川 英樹¹

受付日 2020年8月17日, 再受付日 2020年11月5日,

採録日 2020年12月21日

概要: データ駆動型研究の進展に伴い, 効率的かつ実用的なデータ収集の仕組みが求められている。物質・材料研究機構では, 主に計測・プロセスデータを対象としたデータ収集システムを構築し運用を行っている。本システムでは非ネットワーク環境に置かれている実験装置の制御 PC 等に対して通信セキュリティを施したデバイスを装着させて, 制御 PC を IoT 化させている。この PC からアップロードされたファイルは, XML スキーマに基づいて計測パラメータのメタデータが抽出され, データサーバで自動的にデータベース化される。本稿では主に材料科学分野での IoT 技術を活用したデータ集積の設計指針 (データアーキテクチャー) について論じる。

キーワード: データアーキテクチャ, IoT, データ収集システム, メタデータ, XML スキーマ

Data Architecture for IoT Data Collection System

SHIGEYUKI MATSUNAMI^{1,a)} ASAHIKO MATSUDA¹ TOYOHIRO CHIKYOW¹ YOSHITOMO HARADA¹ HIDEKI YOSHIKAWA¹

Received: August 17, 2020, Revised: November 5, 2020,

Accepted: December 21, 2020

Abstract: With the development of data-driven research, an efficient and practical data collection system is required. National Institute for Materials Science (NIMS) has been constructing and operating a data collection system mainly for measurement and process data. In this system, even if a control PC of a laboratory equipment are placed in a non-network environment, the PC is able to converted to IoT by attaching with a communication security device. The files uploaded from the PC are automatically converted into a database via extracting the metadata of the measurement parameters based on the XML schema. This paper discusses the “data architecture” and data collecting system mainly used in the field of materials science by applying the IoT technology.

Keywords: data architecture, IoT, data collection system, meta data, XML schema

1. はじめに

近年, 実験データやシミュレーションによる計算値に機械学習や深層学習によるインフォマティクスの技術を用いて新規材料を開発するマテリアルズインフォマティクス (MI: Materials Informatics) の研究が進展している [1]。また, 計測分野ではスパースモデリングによるスペクトル解析 [2] や解析条件のバイズ最適化 [3] など人為的な主観に頼らず最適解を導く計測インフォマティクスの取り組みもデータ駆動型研究分野で注目されている。実

験手法では学術研究分野においてもコンビナトリアル手法によるマルチスクリーニングやハイスループット実験の評価係が取り入れられつつある [4]。これにより, 研究開発ではこれまで以上に扱うデータ量は肥大化・多様化し, データの探索空間は大きく広がっている。このような背景から, エキスパートの高度な判断や経験に基づく識別・分類といった作業を AI や IoT, ロボット技術等に任せ, 研究者がより付加価値の高い仕事に注力することで研究開発の高度化・効率化・高速化を実現するデジタルトランスフォーメーション (DX: Digital Transformation) が期待されている。

この次世代の研究環境の実現, すなわち DX によるデータ駆動型研究の推進を実現するためには, データのライフサイクルで初端となる計測装置等からのデータをできるだ

¹ 国立研究開発法人 物質・材料研究機構
National Institute for Materials Science, Tsukuba, Ibaraki 305-0044, Japan

^{a)} MATSUNAMI.Shigeyuki@nims.go.jp

け人手を介さず、自動的に取得および集積される仕組みが必要である。しかし、現状は依然としてデータの収集段階が最も障壁となっている。その理由の多くは、以下の事由の組み合わせによっている。

- ①データそのものが測定装置本体や個人のローカル PC に保管されており、データの所在が分散している。そのため、データの所在を人に聞いて回らないと集められないこと。
- ②測定装置の多くはウイルスによる誤動作や OS のアップデートによる測定の中断などのトラブルを回避するため、ネットワーク環境から切り離れたスタンドアロンで運用されている。そのため、データの取り出しや受け渡しは USB メモリなどの外部記録媒体で人手が介在していること。
- ③測定データの素性が不明であること。たとえば、新規材料の開発においては試料情報が略語表記やコードネーム化されている。また、計測や分析における前提条件に関する付帯情報がないために、測定当事者以外の第三者による再利用・再解析が難しいこと。
- ④計測データが出力状態のまま (raw データ) となっていること。特に出力がメーカー固有のバイナリー形式の場合には可読化のためのデータ変換に難航することのほか、記載内容もメーカー固有の計測パラメータの語彙の意味が解読できず、データ駆動にかけられるためのデータ構造化が難しいこと。

物質・材料研究機構 (NIMS) では上記のようなデータ収集の課題に対応するため、IoT 技術を活用したデータ収集基盤のデータ設計 (データアーキテクチャ) を行い、実践的な運用を行っている。第 2 章では①および②の課題に対して、IoT 技術を導入したシステム (IoT システム) を構築しており、その設計要件ならびに技術について述べる。また、③の課題に対しては、ローカル環境にある PC で動作する HTML ファイルによる実験記録簿 (ここでは Local HTML と呼ぶ) を開発し、IoT システムと一体となった運用を進めている。第 3 章ではこの Local HTML で記録された実験記録をデータファイルと一緒にデータサーバへアップロードすることによりデータサーバ側で HTML パーサーによる読み取りが行われ、データベース (DB) にメタデータとして管理する事例について述べる。既報の④の変換ツール・可読化方法 [5] をあわせることで、計測条件のメタデータも一体となってデータと紐づけられ、DB へは特定の開発テーマや意味付けをもたせた「データセット」としてデータの収集が行える。

第 4 章は総括であり、このようなハードウェア面およびソフトウェア面の両面における整備を進めるうえでのデータアーキテクチャの課題と今後の展望について述べる。

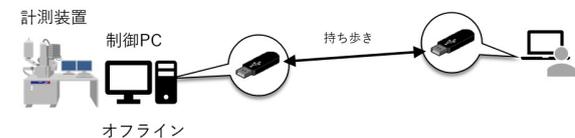
2. IoT システム設計

2.1 IoT 技術を用いる背景

ビッグデータ解析などを活用した材料研究開発のためには、各種実験装置から効率的に計測データを収集して大量に蓄積するシステムを構築することが求められる。しかし科学技術分野では、装置に接続されてその制御やデータ記録を行う PC は、セキュリティや稼働安定性の観点から、ネットワークには接続されずにスタンドアロンで運用される場合が多い。そのためデータの取り出しは USB メモリなどを用いて人の手で行われていることが一般的であり、データの大量収集と系統的な蓄積には適さない (図 1(a))。

一方、本 IoT システムの特徴は、計測装置やプロセス装置のパソコン (制御 PC) に、通信セキュリティを施したデバイス (以後、セキュリティデバイス) を装着することでネットワーク接続していない PC を IoT 化させ、raw データを安全にデータサーバへ転送させることができ、計測装置やプロセス整備から送られてくる大量のデータを効率よく集積することを可能にする点にある。セキュリティデバイスとしてはデータ容量や装置の利用形態を勘案し、Wi-Fi による無線方式と Ethernet に直結する有線方式の二つの方式を併用する。図 1(b) は無線方式によるデータ転送の例を示している [6]。USB メモリの代わりにセキュリティデバイス (ここでは Wi-Fi 機能をもつ SD カード) にデータをコピーすると、認証サーバを経てデータ蓄積サーバにアップロードされ、続くデータ収集サーバでは raw データが構造化されて DB へ蓄積される。データは NIMS のネットワークにつながる環境にあれば、ウェブブラウザを通じてどの端末からでもダウンロードできる。現行、NIMS では約 70 台の計測装置についてセキュリティ

(a) 従来方法



(b) IoTシステム

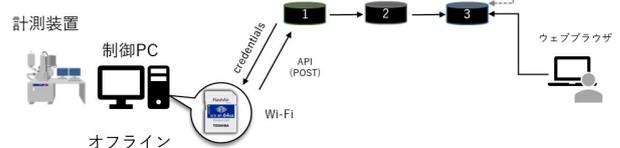


図 1 装置から取り出される計測データの流れ：(a) 従来方法と (b) 本 IoT システムの概念図

Fig. 1 Flow of measurement data retrieved from the device: (a) conventional method and (b) conceptual diagram of this IoT system.

デバイスが導入済みであり、月単位で約 10 GB のデータが流通している。なお、本システムは組織内運用における観点からオンプレミス開発を進めているが、システム構成はクラウドでの運用も可能である。

2.2 IoT システムの設計指針

IoT を利用したデータ収集の導入事例では、インターネット化する“モノ”の対象としてエネルギーハーベスティング等の自立電源で駆動するセンサーを対象とすることが多い。この場合は限られた自立電源の出力でセンサー信号を送信する LPWA (Low Power Wide Area) の無線方式を前提とした設計となっており、消費電力の兼ね合いから扱うデータ容量は比較的に小さい観測データである。

対して本 IoT システムで対象とするネットワークへ接続する“モノ”は、科学技術分野における計測装置や先端プロセス装置、より正確には装置を制御し、装置からの信号をデータとして取り込む制御 PC である。ネットワークにつながらない制御 PC は“モノ”化してしまっているが、セキュリティデバイスを装着させることで IoT 化され、ネットワークを通じて外部のサーバと連携ができるようになる。セキュリティデバイス自体も IoT と見なすことができるが、本稿ではこのセキュリティデバイスを介して IoT 化された制御 PC を「IoT 化 PC」と呼び、このような IoT 化 PC からデータを収集するシステム全体を「IoT システム」と称して述べる。

データの形態は、一回の測定で一つのスペクトルデータを取得するといった一般的な計測のデータ様式のみならず、近年の先端計測分野においてはハイパースペクトルに代表される多次元の面データや三次元 X 線 CT のような三次元データの普及も目覚ましい。そのため、データ容量も MB (メガバイト) から GB (ギガバイト) を扱うケースが増え、データ形式も画像形式、動画形式などバラエティに富んできている。加えて、“モノ”が多数の利用環境にある共用設備や共有装置である場合には、“モノ”からのデータ転送に対しては人と結びついた識別とデータ格納においては適切な分類や秘匿性が求められる。そのうえで、情報セキュリティの三大要素である機密性 (Confidentiality)、完全性 (Integrity)、可用性 (Availability) を担保する必要がある。運用においては環境における様々な制約をクリアしながらセキュリティ機能を保持させる必要がある。このように、IoT 化 PC が接続されても、システムとして高度なセキュリティを保持しつつ、かつ研究開発用途に特化できるように様々な工夫を織り込ませることが、研究開発分野における IoT システムのデータアーキテクチャの要諦である。

2.3 セキュリティデバイス

研究開発分野における多様なデータ形式に対応するた

め、セキュリティデバイスとしては無線 LAN 接続機能を備える SD カード (キオクシア製 FlashAir™ W-04) と有線方式では暗号・認証技術を基盤としたセキュリティプロキシデバイス (東芝インフラシステムズ製 CYTHEMIS™) とを併用して運用している。双方の通信にかかる仕様と運用におけるスループットならびに NIMS で取り決めた一回あたりのアップロード許容量を表 1 に示す。

FlashAir は市販の SD カードであるが、そのカードには無線機器が内蔵されていることに特徴がある。無線通信にかかる周波数帯は 2.4 MHz で IEEE 802.11b/g/n の三つの通信規格を有する。組織内では Wi-Fi のアクセスポイント (AP) 側の接続条件や通信安定性の観点から IEEE 802.11g にて運用している。制御 PC と接続させるインターフェースは USB 2.0 の SD カードリーダー経由でつなげる。ユーザはその制御 PC のファイルシステムを通じて FlashAir にアクセスできる。なお、SD カードリーダーのインターフェースが USB 3.0 の場合には、USB 3.0 の端子から発生するノイズが 2.4 GHz 帯と干渉し、通信の不安定化や高い割合で送信エラーが生じる現象が確認された [7]。また、高圧電源などを使用する特殊装置では、それから発生する干渉ノイズで送信エラーが生じるケースもみられた。このように無線方式においては設置環境や対象機器も十分に考慮してセットアップすることに留意を必要とする実践上の重要な指針を獲得した。スループットは AP の設置位置や設置数、もしくは時間帯にも依存するが、平均して 5 Mbps の伝送速度である。一回あたりのアップロードできる許容量はこの伝送速度も考慮して 100 MB としている。小規模・軽量のデータを主体とする計測装置はこの FlashAir で対応するようにしている。実際、セキュリティデバイスを装着した測定機器のうち八割が FlashAir でまかなっている状況にある。

一方、二割の測定装置については、一回あたりの測定で

表 1 無線/有線方式のセキュリティデバイスの諸元
Table 1 Specifications of wireless and wired security devices.

	無線方式	有線方式
セキュリティデバイス	FlashAir	CYTHEMIS
通信規格	IEEE 802.11g (2.4 GHz)	IEEE 802.21
インターフェイス	USB 2.0 (SD カードリーダー)	RJ-45
スループット (実測値)	約 5 Mbps	約 70 Mbps
一回あたりのアップロード許容量	100 MB	10 GB

100 MB 以上のデータが出力される機器、もしくは FlashAir の設置が困難であった環境にある測定機器である。これらは産業向けのセキュリティデバイスとして開発された有線方式の CYTHEMIS で対応している。CYTHEMIS は装置の通信を暗号・認証処理でセキュアにする外付けのデバイス群と、それらのデバイスを管理する PC と一体となって運用する。インターフェースは一般的な RJ-45 であるため、古い型でサポート切れの OS の制御 PC であっても“モノ”側に特別の設定や改変を施すことなく接続させることができる特徴をもつ。スループットは平均して 70 Mbps の伝送速度であり、構内におけるネットワークの生活線での影響を見定めたくて一回あたりのアップロードできる許容量は 10 GB と定めて運用している。

2.4 セキュリティ

セキュリティデバイスの選定とそれを用いたシステム開発は、以下の三点を特に重要な要件とした。

- ①アップロードの送信のみに特化させて、ダウンロードは許可しない制限をもうけること。
- ②外部からのマルウェアやウイルスの侵入を防ぐのみならず、Windows Update など外部トリガーとなるタスクスケジューラの実行を遮断すること。
- ③セキュリティデバイスのためのドライバなどを制御 PC にインストールしないこと、もしくは環境設定などの設定変更を行わないこと。

この三要件は、長時間計測では失敗の許されない計測装置や OS の更新によっては動作保証の確約のない特殊な制御プログラムで動かしている先端計測や精密加工プロセスならではの事項である。FlashAir および CYTHEMIS ともにこの要件を満たし、非ネットワーク環境にある制御 PC であったとしても、効率的かつセキュアにデータ転送ができています。

なお、無線方式による高速な伝送速度を有するデバイスとしては Raspberry Pi 等のシングルボード PC も検討の俎上に載せていた。シングルボード PC は廉価でありながら、高いパフォーマンスと開発の自由度がある利点は魅力であったが、組織におけるセキュリティポリシーをクリアするには至らず採用を見合わせた。

(1) 無線方式のセキュリティ

FlashAir は一眼レフカメラなどデジタル撮像機器に装着し、その撮像を SD カード内に記録するとともに、FlashAir 自身も Wi-Fi 経由で画像や動画をクラウドに自動的にアップロードする用途で用いられている。我々は、そのシンプルな一方通行性の送信方式に着目した。装着対象を PC とした場合、FlashAir のストレージはカード内のマイコンからのアクセスおよび PC へのマウントが可能である一方で、カードからは PC のストレージへのアク

セスはできない。これにより、データは構造上一方向にしか流れず、PC 自体をネットワークに接続することなくデータのみアップロードによる収集が可能となる。

また、FlashAir は SD カード内で動作する軽量言語の Lua スクリプトを実行できる機能を有している。本 IoT システムでは、FlashAir にファイルが追加されるとそれを検出し、指定したサーバへ送信する Lua スクリプトをもたせている。FlashAir は実験室内の Wi-Fi に接続し、計測日時や実験者を示すメタデータとともにデータ蓄積サーバへ送信される。データ受信認証は FlashAir からアップロードされたデータ蓄積サーバ側で認証され、サーバに登録された FlashAir 以外からのアップロードが行えないようにしている（詳細は 3.3 節で述べる）。

(2) 有線方式のセキュリティ

有線方式の CYTHEMIS は、そのデバイス本体に高い耐タンパー性を備えた IC チップをもち、認証方式によるセキュリティ機能を有する [8]。さらには管理システムと一体となって、アクセス制限やフィルタリング機能、不正侵入防止機能を持ち、許可されていない通信や接続（双方向の送受信）は遮断する。本システムにおいては IoT 化 PC がデータ蓄積サーバへファイル送信するとき、SMB /FTP/SCP 等のプロトコル・通信方向だけを許容し、それ以外の通信を遮断する設定としている。加えて管理システムは、接続されている当該デバイスとその情報がホワイトリスト等で管理されており、条件が満たされた CYTHEMIS からの指定のプロトコルの通信のみを許可する。これにより、IoT 化 PC とデータ蓄積サーバ間のみの通信とし、CYTHEMIS が装着されていない PC からの通信は制限をかけている。

ホワイトリストの情報設定やアクセス制限の運用は、ネットワーク上の管理システムを通じて一元的に管理できるようにし、増設対応においても柔軟性をもたせている。CYTHEMIS は IoT 化 PC の OS の種類やバージョン等に依存しないため、ほぼ機器を選ぶことはない。また汎用的な通信プロトコルによる方法で接続を可能にしていることや有線 LAN (RJ45 Cat6 準拠) に接続できるため、組織内ネットワークに仮想専用線 (VPN: Virtual Private Network) で通せる環境にあれば、遠隔の計測機器に対しても CYTHEMIS を装着させることで当該装置のデータをデータ蓄積サーバにセキュアに集積することもできる。たとえば、NIMS は茨城県のつくば地区 (千現地区、並木地区、桜地区) のほかに兵庫県の大型放射光施設 (SPring-8) に分室をもつ (西播磨地区)。西播磨地区との接続は機構専用の VPN でつくば地区と結んでいるが、CYTHEMIS では VPN に対してもセキュリティを確保して西播磨地区からつくば地区へのデータのアップロードを可能としている (図 2)。

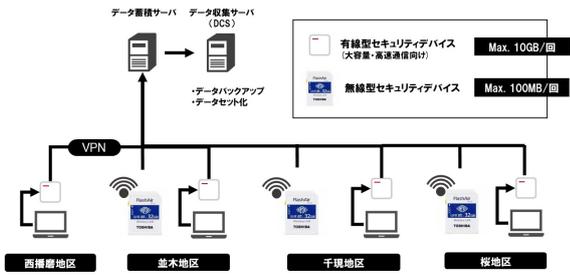


図 2 無線/有線方式を併用したデータ収集の概念図

Fig. 2 Conceptual diagram of data collection using both wireless and wired methods.

3. IoT システムによるデータ集積化

3.1 データ集積にかかる先行技術

IoT 化 PC から取り出されたデータを有効に活用させるためには、データ収集、蓄積、そしてサービスまでにまたがるインフラストラクチャの構築が不可欠である。科学データの集積にかかるシステムや DB については数多くの取り組みがなされてきた。先行技術としては、1990 年代後半から化学分析分野で普及がはじまった研究ラボのデータを一元的に管理運用するラボ情報管理システム (LIMS: Laboratory Information Management System) の取り組みが挙げられる。主には低分子創薬系の分析データの収集で導入が図られ、サンプル管理から各種の化学分析のトラッキングを含め、測定データを同一 DB に一元的に集積するシステムが開発された。異種メーカーの機器であっても同一のシステム内で接続ができ、かつデータ流通が図れるように、データフォーマットも互換性をもつ AnDI [9], JCAMP-DX [10] といった標準フォーマットも整備された。とはいえ、イントラネット内での計測機器のつなぎこみはドライバが必要となるケースやデータフォーマットも DB 側であらかじめ受け入れる様式を厳密に定めておかなければならない。すなわちデータ収集の考え方は中央制御方式であり、規格に対応しない計測機器や各種センサー、データロガーの接続への拡張性には制限があった。

米国の国立再生可能エネルギー研究所 (NREL) では太陽電池分野で得られるデータ収集および管理について LIMS の概念導入の試みを行っている [11]。市販の LIMS アプリケーションについて上述と同様の拡張性の不備を認識しており、時間や費用の観点から市販品をカスタマイズよりも自製による LIMS 構築の判断をとっている。データ登録ではメタデータの記述方法について XML スキーマを定義し、計測器からの自動的に収集されたデータは、そのメタデータとともにデータウェアハウス (DWH) に蓄積するシステムを構築している。旧来型の LIMS とは大きく異なる点は、NREL では「データ利活用」として公開基盤を意識した設計をとっていることにある。各研究分野に特化したデータセットについては extract-transform-load

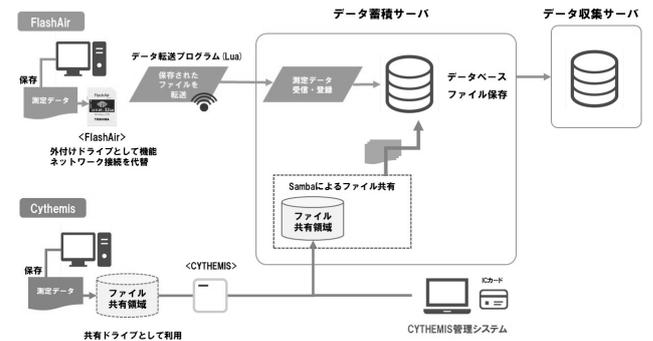


図 3 セキュリティデバイスの併用におけるファイル転送とデータ蓄積サーバの関係

Fig. 3 Relationship between file transfer and data storage server in combination with security devices.

(ETL) のプロセスで構築できるようにしており、実際、薄膜半導体材料については NREL のウェブサイトを通じてデータ公開を行っている [12]。米国の国立標準技術研究所 (NIST) では 2011 年に始動した Materials Genome Initiative (MGI) の取り組みの中でデータ集積にかかるインフラストラクチャの整備を担当し、計測機器データや外部からのデータ受け入れも想定した「Material Data Curation System (MDCS)」を開発した。ここでも特徴とするのは計測データについては装置ごとに XML スキーマを定義し、データの互換性を持たせる設計思想をとっていることにある [13]。

このように NREL, NIST のデータ収集方式は中央制御方式ではなく、計測機器ごとに XML スキーマを定義したエッジ制御方式を念頭に置き、多様な計測データを受け入れることを可能にしている。NIMS におけるデータ収集においても、数多くの計測機器・プロセス機器からのデータ種を受け入れるために、エッジ制御方式がとれる XML スキーマ方式を採用し、システム設計および DB 設計を行っている。

3.2 ファイル蓄積の概要

NIMS のシステムでは、制御 PC から無線方式もしくは有線方式のセキュリティデバイスを経てアップロードされたデータは、一旦、受け口なる「データ蓄積サーバ」に記録される (図 3)。ただし、そのファイル転送と受け入れ方法は両者のセキュリティデバイスの仕様の違いにより次のような相違がある。無線方式の FlashAir では、セキュリティデバイスは IoT 化 PC から外付けドライブとして認識される。ユーザが IoT 化 PC から FlashAir へファイルを配置すると FlashAir の Lua スクリプトが動作し、ファイルはデータ蓄積サーバへ転送される。一方、有線方式の CYTHEMIS は、そのデバイスの中にはストレージやファイル送信機能を有してはいない。ここではデータ蓄積サーバに共有フォルダ領域を設け、IoT 化 PC からはその

共有フォルダを介しファイルの転送を行う形態としている。具体的にはデータ蓄積サーバ (Cent OS) に Windows ネットワークを実装しファイルサーバ化ができる samba を導入した。IoT 化 PC の共有フォルダにファイルが配置されると同時にデータ蓄積サーバに共有がかり、それをトリガーとして DB (PostgreSQL) へファイルが取り込まれる。

DB への書き込みは、次節で述べる送信時で使うファイルシステムの階層構造情報を利用し、計測機器やデータ送信者を識別し蓄積される。また、データ蓄積サーバへは LDAP 方式による機構内認証サーバによりアクセスが認められたユーザについてはウェブブラウザからのアクセスを可能としている。データ蓄積サーバの段階ではメタデータは付与されておらず raw データのみが取得できるが、ネットワークに接続できる環境であれば機構内外からも raw データをダウンロードから入手ができ、実験の速報を知りたい場合などに活用されている。

IoT 化 PC からのデータは、データ蓄積サーバに蓄積されるものの、データ蓄積サーバの着信がトリガーとなり、「データ収集サーバ」へコピーされる。データ収集サーバは広義にとらえれば DWH の役割を担うが、本システムでは DB 兼バックアップサーバとして機能する。ここで XML スキーマが定義できている計測機器については、データ収集サーバで raw データの内部に記載されている計測パラメータの抽出が行われメタデータを生成する。raw データがメーカー固有のバイナリー形式であるような場合、メーカーからの協力の元、テキスト形式への変換を行っている。また、第三者による再利用性を高められるようにするために計測パラメータ部と数値行列部とを分け、後者については csv ファイルとして出力する。これらの構造化されたデータは抽出された計測パラメータのメタデータとともにデータ収集サーバで一体化されて DB に書き込まれる。このようにして、IoT 化 PC より送信されたファイルデータは、できる限り人の手を介さずにコンテンツに富んだメタ情報を付与させて DB へ蓄積し、かつ再利用ができるようにしている。

3.3 データの識別・分類と秘匿性

ここでは IoT 化 PC からのアップロードされるデータの識別・分類とその秘匿性の制御について述べる。装置識別は、セキュリティデバイスを設置の段階に運用側で装置 ID を付番する。汎用性の高い機器である場合には、組織をまたいで同一メーカーで同一機種モデルを購入しているケースも想定されるため、異なる装置 ID を付番することで区分できるようにしている。また FlashAir と CYTHEMIS はいずれも MAC アドレスを持つ。この MAC アドレスを利用し、一台の装置 (制御 PC) につき一つのセキュリティデバイスを装着することで、セキュリ

ティデバイスと装置 ID とを 1:1 で紐付けするようにしている。仮にユーザが意図せずセキュリティデバイスを取り外し、別の装置の制御 PC に装着した場合には、そのセキュリティデバイスは IoT システムの中では認識されない。別の装置からのデータ転送ができないようにするため、FlashAir においては、クレデンシャルファイルによる制御、CYTHEMIS においてはホワイトリストの情報における管理システムの設定で行っている。

無線方式および有線方式ともに IoT 化 PC ではファイルシステムに基づくフォルダ階層で表示させている (図 4)。第一階層は装置 ID 名のフォルダ、第二階層はユーザの職員番号にも紐付いている LDAP ID 名のフォルダとして定め、セキュリティデバイスの設置時に運用管理者側でセットアップを行う。この両階層のファイル名はユーザには改変をしないようにルール化している。第三階層はユーザが実験テーマなどで自由に設定できるフォルダである。

ユーザは自分の職員 ID の配下の第三階層にデータファイルをドラッグ&ドロップすると、それがトリガーとなってデータ蓄積サーバへアップロードされる。上記で述べたようにファイル転送方法は FlashAir と CYTHEMIS とでは原理的に異なるが、ユーザはファイルシステム上ではセキュリティデバイスの違いによる操作を変える必要はない。ここは無線方式と有線方式の併用において最もこだわった点である。

データ蓄積サーバ内では送られてくるファイルのほか、ファイルシステムの第一階層/第二階層/第三階層の各階層情報から「装置」と「人」、そして「テーマ」が識別されて DB に分類される。また、データ送信側の情報としてデータが生成または更新された日時も記録されることから、「人 - 装置 - データ - 時刻」の識別ができるようになっている。複数名の利用による共用装置や共有設備であったとしても、第二階層にユーザの ID フォルダを設けるだけでよい。各ユーザは、自分のフォルダにデータを移すだけの操作でアップロードは完了となる。

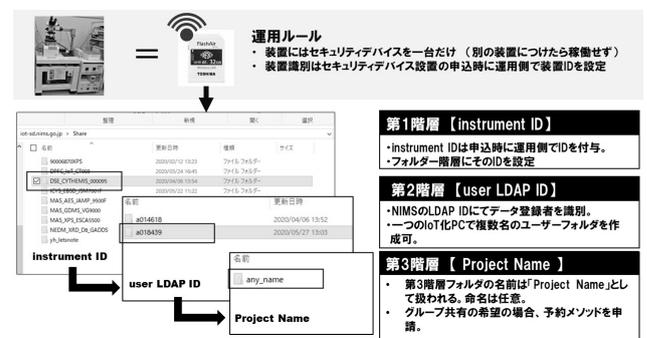


図 4 ファイルシステムに基づく階層構造による装置および人の識別の方法

Fig. 4 Method of instrument and person identification by hierarchical structure based on file system.

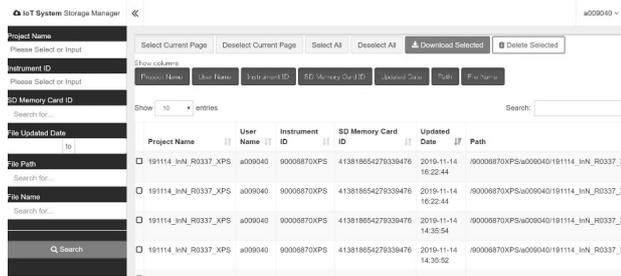


図5 データ蓄積サーバのウェブサイトにおける表示画面（ダウンロード画面）

Fig. 5 Display screen on the data storage server website (download screen).

データのダウンロードは、ユーザが自分のLDAP IDでデータ蓄積サーバへログインすると、図5に示される画面に遷移する。この画面で、ダウンロードのファイルを選択することで取得することができる。この画面にはユーザが登録したデータのみが表示され、別ユーザのデータは表示されない。これは、郵便における私書箱のようなイメージであり、我々はこれをIoTシステムにおける「私書箱機能」と呼んでいる。

なお、第三階層目のフォルダは各ユーザが任意に作成できるフォルダと上述したが、フォルダ名の最初の文字がアンダーバーから始まる「_フォルダ名」の場合にはあらかじめ申請のあったグループメンバーでデータ共有をすることができる予約メソッドを設けている。これにより、「個人」もしくは「グループ」のどちらかでの秘匿範囲の選択をもたせ、限られた範囲内でデータ共有を行うことができるようにしている。この方法の場合、図5の画面においては、グループやチームのメンバーから予約メソッドでアップロードされた共有データも表示される。測定の実操レーターとデータ解析を担う研究者との役割分担が明確なる研究テーマでは、この予約メソッド形態が多用されている。

ファイルの共有方法の観点でみると、フォルダ名の先頭文字の命名規則による予約メソッドの方法は決して最善な方法ではないことに留意は必要である。システムの設計段階ではデータは測定した個人のみにも属する秘匿性と要件定義して開発を行っていた。ところが、実際にサービスを組織で供用したところ、ユーザからはチーム単位や特定のテーマにかかわるグループでのデータ共有の要望が多くあげられた。予約メソッドはその要望を受けて、システム改修上の工夫によるアジリティ的な施策である。

3.4 計測メタデータ

メタデータのXMLスキーマが定義できている計測機器については、計測ファイルに記載されている計測パラメータからメタデータを抽出し、機械可読性の高いXMLファイルへと変換する（図6）。このようにして抽出された

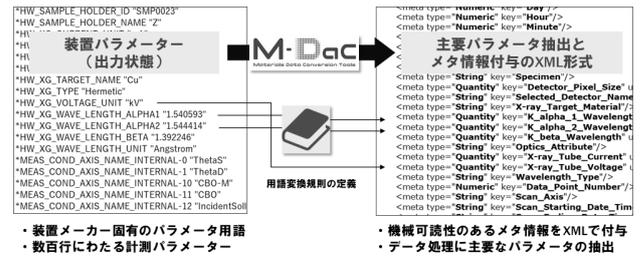


図6 XMLスキーマによる計測ファイルからのメタデータの抽出の例 [14]

Fig. 6 Example of metadata extraction from a measurement file using XML Schema [14].

「計測メタデータ」はXML形式でデータ収集サーバのDBに書き込まれるようにしている。メタデータによって、DB内での高速検索を可能にするとともに、異なる形式とデータを比較する際においても、メタ情報の語彙を統制することで膨大な数の計測データ群からの特定データファイルの検索を容易にすることに利用している。

また、機械学習やデータ駆動型解析で利用しやすいデータの創出・蓄積を効率的に行えるようにするために、バイナリデータである場合にはテキスト変換を通じてcsvの数値データ行列化を行うとともに、データセットの内容の判別を容易にするためにスペクトル等へのグラフ化（視覚化）変換を同時に行ってDBのサムネイルに供している（図7）。なお、X線回折（XRD: X-ray diffraction）、X線光電子分光（XPS: X-ray Photoelectron Spectroscopy）、およびオージェ電子分光（AES: Auger electron spectroscopy）については、装置メーカーの協力のもと、バイナリデータの変換プログラムの提供や、データファイルに記述された測定にかかるパラメータの語彙の意味について定められている。これらはMaterials Data Conversion Tools (M-DaC) と命名したツール群としてGitHubで公開し、第三者利用を可能にしている [14]。

3.5 Local HTMLによるメタデータ記録

メタデータの中には計測ファイルからは取得できない、または記述できない計測や実験の付帯情報がある。たとえば、自作で組み上げた計測系では、制御信号や調整条件が計測ファイルには記録されないことは往々にして起こる。また、真空装置であればその到達真空度といった日々変動する測定環境の情報も、第三者がデータを利用するうえでは重要な情報になる。それらの付帯情報は主には測定担当者の記録メモや実験ノートに記載される事項であるが、それをデジタル化し、測定ファイルに紐づきデータが管理されることは稀である。

検討初期には、手書きに代わる電子記録の手段として市販の電子ラボノートの導入調査を行ったが、対象は化学向け・バイオ向けの研究が主たるもので、このような簡便な計測記録や機器管理向けに合致するアプリケーションは見

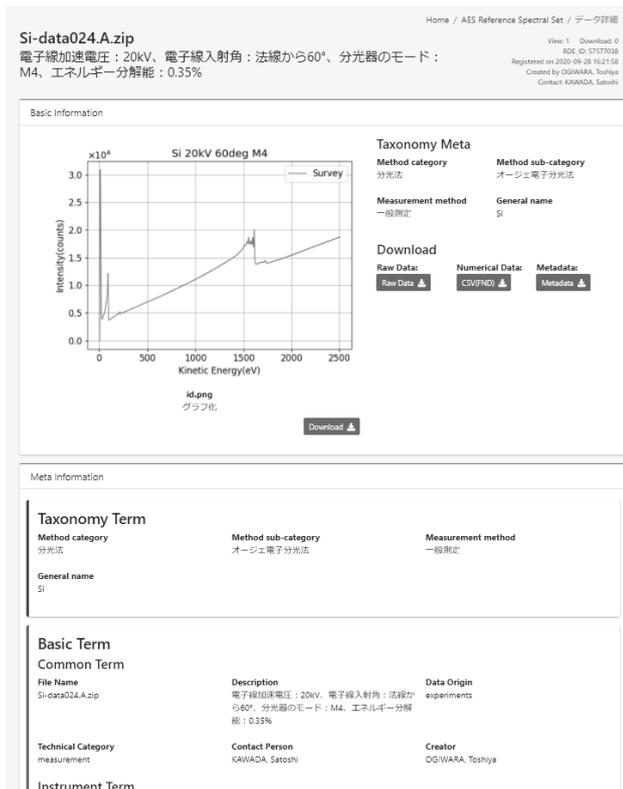


図 7 IoT から集積された AES 計測データのデータセットの閲覧画面。メタデータを使った計測情報の表示のほか、計測の可視化情報をサムネイルとして表示させている。

Fig. 7 Dataset view of AES measurement data. In addition to displaying the measurement information using metadata, the measurement visualization information is displayed as thumbnails.

い出すことができなかつた。また医療向けの電子カルテのようなタブレットを利用するシステムも調査検討を行った。しかし、所内のヒアリングでは数行の計測記録のために、機器利用者全員にタブレットを購入することには、各部署の予算上では捻出が厳しいとの声がよせられた。また、共用設備のような不特定多数の利用者が扱う装置ではタブレットの紛失なども想定した管理は負荷が高いとされた。

導入コストを抑え、かつ現場の作業者が日々、確実に記録を続けることができる検討の中で見いだされたのは、計測の制御 PC を入力端末として利用することであった。ただし、2.4 節でも述べたように、一般に先端計測や精密加工プロセスは、ウイルスによる制御の誤動作や制御遅延を招く恐れがあるため、プレインストールされているソフトウェア以外のインストールを禁止して運用していることが一般的である。

この制約をクリアする方法として、Windows であればプレインストールされている「メモ帳」のような軽量エディタで記録し、そのテキストファイルをデータとともに紐付けて送信するアイデアに至った。しかしながら、ガ

イド機能がない軽量エディタでは作業者にミスのない記録を促すことができない。標準作業化するためには、計測機器の利用形態に応じた記入すべき項目がまとめられたテンプレートが必要である。Microsoft Word や Excel のような高機能アプリケーションであればテンプレートの作成は容易であるが、プレインストールされている軽量エディタではそのような要求を満たすことは難しい。

我々は、このようなローカル環境の制御 PC でも高機能エディタ並の要求に応えられる仕組み作りのツールとして HyperText Markup Language (HTML) に着目した。一つにはテンプレートで必要となる記入項目のガイドやデザインはカスケーディング・スタイルシート (CSS) 機能を使うことで解消できる。また、JavaScript (js) を HTML 内のコードに埋め込めば、たとえば、入力内容が要件を満たしているかを判定するバリデーションも可能である。この HTML, js, CSS の構成であれば、制約となっているインストーラの形を取らずに動かすことができる。もしくは、外部のウェブサーバに接続をさせて CGI (Common Gateway Interface) のような処理プログラムを呼び出すことをせずとも、ネットワークにつながっていないローカル環境において、高機能エディタで作成されたフォーム並の高度な処理を成し得る。このデータの付帯情報を HTML 形式のファイルとして作成し、実験・計測情報の入力を支援する方法は、試行段階で協力を得た利用者からも支持が認められ、手書きに代わる計測条件の記録方法として、またデータの再利用にかかる手法としては有効な方法であると考えている。(このようなローカル環境にある PC での HTML ファイルによる実験記録簿を、次節以後は「Local HTML」として表記する。)

3.6 Local HTML の様式

Local HTML による入力フォームの書式の事例を図 8 に示す。記入する項目は、大別すると「基本情報」、「サンプル情報」、そして実験由来の「固有情報」の三つの象限を設け、A4 版で 1 枚程度に収まる内容としている。上記のような計測ファイルに書かれない項目は「固有情報」に欄を設け HTML エディタで書き換えるだけで簡単に編集できることに特徴があり、主要な装置についてのテンプレートは運用側で作成している。HTML 構造も各計測装置でのカスタマイズや拡張性をもたせ、細かい修正はユーザでも対応できるようにしている。

「データ投入者」といった装置利用でほぼメンバーが固定化している記入項目の場合には、プルダウンで選択させることができる。作業項目が「有/無」のような選択性の場合にはラジオボタンを設けることで記録情報をとることもできる。また、測定ファイルと Local HTML との紐づけは、「ファイル選択ボックス」を設けて、対象となるファイルを選択できるようにしている。

図8 Local HTMLによる入力フォームの書式の例

Fig. 8 Example of input form formatting with Local HTML.

入力内容は「submit ボタン」を押すと、Local HTMLの内部に埋め込まれたjs ファイルが呼び出され、メタデータの key-value 構造がとられるほか、測定データ(ファイル)との紐づけはハッシュ値を取得し、入力フォームとは別のHTML ファイルがデスクトップ上に出力される。

ユーザは測定ファイルと一緒にこの出力されたHTML ファイルをIoT化PCからアップロードすれば、データ収集サーバ側では受け取ったHTML ファイルをパーサーで読み取り、key-value に基づいて自動的にデータ収集サーバのDBに書き込みが行われる。またファイルについてはハッシュ値を頼りにデータファイルを認識し、そのファイルからデータ変換やパラメータ抽出をかけ、描画に必要な数値データなどを読み出す。これら一連の処理はその装置に特化したコードをpythonで作成している。そのような一連の処理を終えて、上記の計測メタデータとともに書き込まれて図7の画面に反映される。なお、本稿では詳細は示さないが、Local HTMLでは事前にデータ登録者が特定のテーマのデータ群として分類・収集をする「データセット」単位での仕分けを可能にしている。この「データセット」はデータカタログが付与され、将来、REST APIを通じてこのデータセットの単位でデータを一括してダウンロードを可能とする予定である。

4. 今後の展望

2020年1月のWindows 7のサポート終了により、古いOSであってもセキュリティを保証してネットワーク接続にも対応できる「レトロフィット」の利点も見いだされ、IoTの利用が増えてきている状況にある。また、既存設備で高速通信が可能となるWi-Fi 6や5G(第5世代移動通信システム)の組織内運用が可能な「ローカル5G」の次世代通信技術への適用も見込み、大容量・低遅延化検討も着手し始めているところである。

このようにデータ収集基盤のデータ設計の指針(データアーキテクチャ)や安定運用のためのノウハウは整ってきたところであるが、利活用をより高めてゆくためには、ま

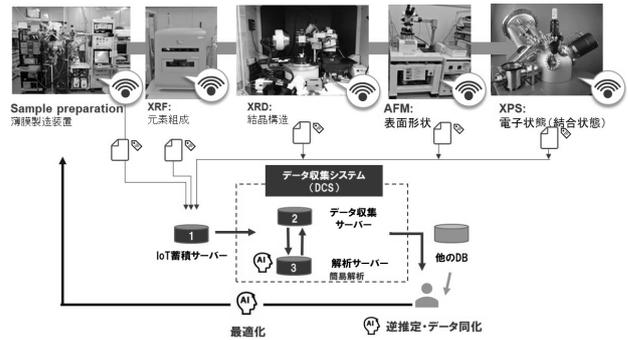


図9 IoTシステムによる多様な計測データの統合による高付加価値化

Fig. 9 Adding value by integrating various measurement data with IoT system.

だ依然として改善項目は多い。インフォーマティクスの本丸である機械学習や深層学習などにかかるまでにデータの前処理や定形処理に膨大な作業工数がかかっている実情もある。そのため、rawデータを再利用のしやすいデータ構造へといかに整えるかがポイントにもなる。そのような観点から、データ解析ツールや前処理スクリプトなどによる自動処理化の付加価値を備えることを進めている。図9に示されるようなデータ収集サーバと解析サーバと連携・同期させる試みである。この取組は計測装置や学術分野によって個別最適化する要素もあるが、繰り返しの単調なる人手による作業を緩和させることで利便性を訴求することが期待できる。

一方で、運用側において過度の負荷とならないようにするために、ルーチン化できるデータ構造化などはできるだけメニュー化する必要もある。また、データ構造化したデータを自動的に「データセット化」するだけでなく、二次利用性を高めるためのユースケースやショーケースも整備してゆかなければならない。

今後、IoTシステムのスケラビリティにおける課題として、数百人または数百台規模でGB単位のデータがローカル5GやWi-Fi 6に大量に流通した場合における分割処理技術や分散システムなどの導入が不可欠となることが予想される。また、最新のデータ技術にも長けたデータ設計家(データアーキテクト)の養成も急務である。そのようなシステムの拡張性におけるIoTシステムの運用や人材育成のあり方については今後検討を進めてゆくものである。

謝辞 Local HTMLのアイデアと助言をいただいた柳生進二郎博士(物質・材料研究機構)およびCYTHEMISのセキュリティデバイスに関する技術的な助言をいただいた友枝裕樹氏(東芝インフラシステムズ株式会社)に感謝を申し上げます。また、本システムの開発に携わったNIMSの諸氏に深い感謝の念を表します。

参考文献

- [1] 知京豊裕：マテリアルインフォマティクスの現状と課題～海外の動向と日本の挑戦，情報知識学会誌，Vol.27, No.4, pp.297-304 (2017).
- [2] Murakami, R., Kageyama, H., Nakamura, K., Tanaka, H., Shinotsuka, H., Yoshikawa, H. and Yoshihara, K.: Background estimation in X-ray photoelectron spectroscopy data using an active Shirley method with automated selection of the analytical range, e-Journal of Surface Science and Nanotechnology, Vol.17, pp.61-68 (2019).
- [3] Ozaki, Y., Suzuki, Y., Hawaii, T., Saito, K., Onishi, M. and Ono, K.: Automated crystal structure analysis based on blackbox optimisation, npj Computational Materials, Vol.6, No.1, pp.1-7 (2020).
- [4] Kusne, A. G., Gao, T., Mehta, A., Ke, L., Nguyen, M. C., Ho, K. M., Antropov, V., Wang, C. Z., Kramer, M. J., Long, C. and Takeuchi, I.: On-the-fly machine-learning for high-throughput experiments: Search for rare-earth-free permanent magnets, Scientific Reports, Vol.4, pp.1-7 (2014).
- [5] Suzuki, M., Nagao, H., Harada, Y., Shinotsuka, H., Watanabe, K., Sasaki, A., Matsuda, A., Kimoto, K. and Yoshikawa, H.: Raw-to-repository characterization data conversion for repeatable, replicable, and reproducible measurements, Journal of Vacuum Science & Technology A, Vol.38, No.2, p.023204 (2020).
- [6] 松田朝彦, 吉川英樹, 知京豊裕：無線LAN搭載SDカードをIoTデバイスとして利用した実験データ収集システム，第79回応用物理学会秋季学術講演会講演予稿集, p.18a-231A-8 (2018).
- [7] Intel Corporation. USB 3.0 Radio Frequency Interference Impact on 2.4 GHz Wireless Devices, 22p (2012)
- [8] 友枝裕樹, 福岡寛規, 杉湖 慶：ICカード技術を応用してセキュアなネットワークを実現するIoTセキュリティソリューション，東芝レビュー，Vol.1, No.1, pp.59-62 (2018).
- [9] ASTM International: Standard Specification for Analytical Data Interchange Protocol for Chromatographic Data, ASTM E1947 - 98. (<https://www.astm.org/Standards/E1947.htm>).
- [10] Baumbach, J. I., Lampen, P., Davies, A. N. IUPAC/JCAMP-DX: An International Standard for the Exchange of Ion Mobility Spectrometry Data. International Journal for Ion Mobility Spectrometry, Vol.1, No.1, pp.64-67 (1998).
- [11] White, R. R. and Munch, K.: Handling Large and Complex Data in a Photovoltaic Research Institution Using a Custom Laboratory Information Management System, MRS Proceedings, Vol.1654, p.mrsf13-1654-nn11-04 (2014).
- [12] Zakutayev, A., Wunder, N., Schwarting, M., Perkins, J. D., White, R., Munch, K., Tumas, W. and Phillips, C.: An open experimental database for exploring inorganic materials, Scientific Data, Vol.5, No.1, p.180053 (2018).
- [13] Dima, A., Bhaskarla, S., Becker, C., Brady, M., Campbell, C., Dessauw, P., Hanisch, R., Kattner, U., Kroenlein, K., Newrock, M., Peskin, A., Plante, R., Li, S. Y., Rigodiat, P. F., Amaral, G. S., Trautt, Z., Schmitt, X., Warren, J. and Youssef, S.: Informatics Infrastructure for the Materials Genome Initiative. Jom, Vol.68, No.8, p.2053-2064 (2016).
- [14] NIMS Materials Data Conversion Tools (M-DaC), (<https://dice.nims.go.jp/services/M-DaC/>).



松波 成行 (非会員)

1998年北海道大学大学院地球環境科学研究科博士課程修了。博士(地球環境科学)。2015年より物質・材料研究機構調査分析室長，2018年より同統合型材料開発・情報基盤部門参事役。



松田 朝彦 (非会員)

2013年京都大学大学院工学研究科博士課程修了。博士(工学)。2013年NIST客員研究員。2015年物質・材料研究機構ポスドク研究員。2018年より同統合型材料開発・情報基盤部門エンジニア。



知京 豊裕 (非会員)

1987年早稲田大学理工学研究科電気工学専攻後期課程中退，1989年工学博士。2020年より物質・材料研究機構統合型材料開発・情報基盤部門特命研究員。



原田 善之 (非会員)

2001年岩手大学大学院工学研究科博士課程修了。博士(工学)。2017年より物質・材料研究機構 NIMS エンジニア。



吉川 英樹 (非会員)

1992年大阪大学大学院工学研究科博士課程修了。博士(工学)。1995年物質・材料研究機構の前身の無機材質研究所に入所。2017年より同機構の材料データプラットフォームセンター副センター長。