**データ戦略はどう進むのか？**

12

[[](https://note.com/hiramoto)](https://note.com/hiramoto)

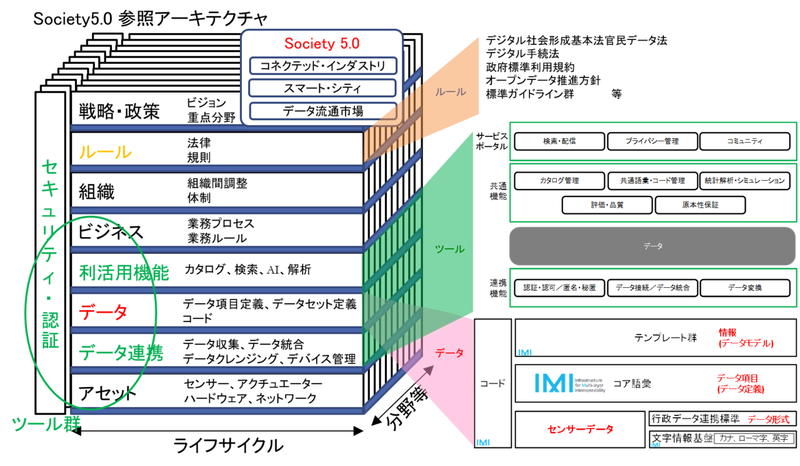
[たまにメモする人](https://note.com/hiramoto)

2021/06/20 01:00

2021年6月18日、IT戦略である [デジタル社会の実現に向けた重点計画](https://www.kantei.go.jp/jp/singi/it2/decision.html)の別紙として[データ戦略](https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20210618/siryou3.pdf)が決定しました。しかし戦略はあくまでも文書であり、どう進めるかが重要になります。これから開始するのかと思うかもしれませんが、実はすでに様々な策が用意、展開されており、着々と外堀が埋まってきているところです。

**アーキテクチャとデータ標準群を中核とした推進**

データ戦略は、論理的にものごとを進め、検討の抜け漏れをなくすためアーキテクチャに基づき検討されています。この[元となるアーキテクチャ](https://www8.cao.go.jp/cstp/tyousakai/juyoukadai/14kai/siryo2-1.pdf" \t "_blank)は[Society5.0](https://www8.cao.go.jp/cstp/society5_0/)の検討の中で作られたものであり、スマートシティにも展開されています。その中でもデータ連携層とデータ層は非常に重要な要素として考えており、さらに細分化して検討されてきました。



特にデータ連携基盤は、[SIP](https://www.nedo.go.jp/activities/ZZJP2_100126.html" \t "_blank)の分野データ連携基盤として検討が進められ、そのノウハウは[DATA-EX](https://data-society-alliance.org/data-ex/" \t "_blank)に引き継がれています。データの標準化は、文字からデータモデルまで必要な領域を着実に進めてきています。

これだけではありません。データを安定的に共有することが重要なことから、マスターデータやコードの設計ガイドを整備し、さらに、データ品質の管理やデータマネジメントの枠組みまで検討しています。

技術的な分野だけでなく制度的な仕組みの検討も進めており、データ契約や取引の在り方もガイドなどにまとめられています。

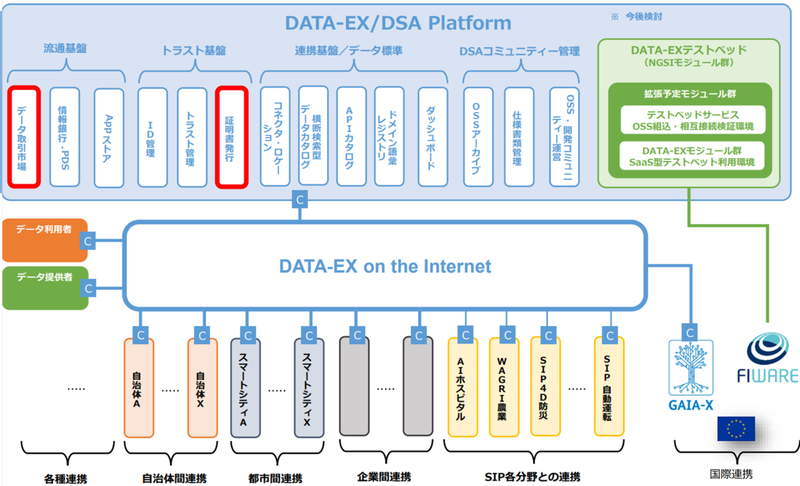
さらに、これを進めるための人材育成の枠組みや経済効果の分析も検討されており、これら一連の流れでデータ戦略実現に向けた外堀は埋まってきているといってよいと思います。

既にこのノウハウは、法人情報を一括して提供する[ｇBizInfo](https://info.gbiz.go.jp/" \t "_blank)や支援制度や事例の情報を提供する[ミラサポplus](https://mirasapo-plus.go.jp/" \t "_blank)に実装もされています。

では、次の見出しから具体的に各項目の取り組みを紹介していきましょう。

**データ連携基盤はどう進んでいるの？**

地道に進んでいます。スマートシティで都市OSという言葉を聞くことが多いと思いますが、これもデータ連携基盤ですし、法人データの連携用に[gbizConnect](https://www.meti.go.jp/policy/digital_transformation/asset/meti-dx/20201001/METI_DX.pdf" \t "_blank)も提供が始まっています。また、[DATA-EX](https://data-society-alliance.org/data-ex/" \t "_blank)のプラットフォームも整備が進んできています。

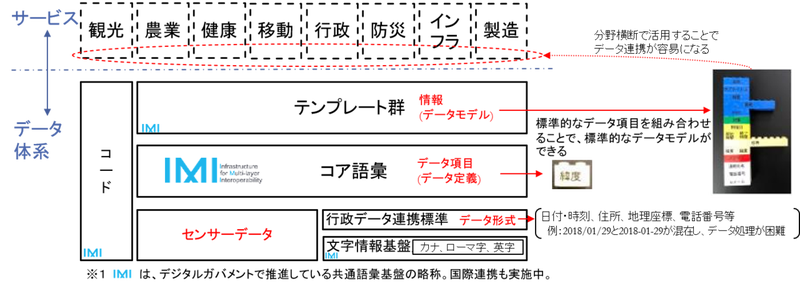
[](https://www.kantei.go.jp/jp/singi/it2/dgov/data_strategy_tf/dai7/siryou7.pdf)

これらの心臓部はデータ連携をつかさどるコネクタやブローカーという仕組みです。エストニア政府の[x-Road](https://e-estonia.com/solutions/interoperability-services/x-road/)も、都市OSなどで有名な[FIWARE](https://www.fiware.org/" \t "_blank)も、欧州のデータ連携基盤として注目を集める[GAIA-X](https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html" \t "_blank)もすべてコネクタで接続する仕組みになっています。

このように世界中でデータ連携基盤の整備は着々と進んでいます。

**データ標準はどうなんだろう**

ここは、現在進行中のものも含めるとかなり進んできています。下図が全体像です。



まず文字の世界です。[漢字の標準化](https://moji.or.jp/mojikiban/" \t "_blank)は完了して、政府から民間にすでに移行していますし、ベース・レジストリに指定しました。一方、一般のコンピュータやスマートフォン等では表示できないいわゆる外字に対して、新しい概念として代替文字がマイナンバー制度と同時に導入されています。いわゆる外字といわれる文字を縮退マップによって、一般の機器でも表示可能な代替文字に置き換える仕組みです。法人名は[法人番号公表サイト](https://www.houjin-bangou.nta.go.jp/)で代替文字になったデータが公表されています。個人氏名については通知カードに記載され、[マイナンバーカードの券面事項入力補助アプリ](https://www.soumu.go.jp/main_content/000393494.pdf)に登録されています。

さらにヨミガナに着手しており、2018年に[法人名のヨミガナ](http://www.moj.go.jp/MINJI/minji06_00109.html" \t "_blank)を登記するようになりました。[氏名のヨミガナ](https://www.kinzai.or.jp/legalization_kana.html)も研究会で検討が進められています。ここまで精緻に検討し、[法務大臣もヨミガナを付けると会見で発言](http://www.moj.go.jp/hisho/kouhou/hisho08_00144.html)しているので、2024年までに解決が図られるはずです。地名のフリガナはベース・レジストリのアドレスDBの検討の中で解決しようと取り組みがスタートしています。onoが小野か大野かといった多くの問題を含むローマ字までの道のりは遠いのですが、その基盤のヨミガナがこの数年で制度化されていくというのは、行政の制度として画期的です。

これらの文字の導入について詳しくは「[文字環境導入実践ガイドブック](https://cio.go.jp/guides)」をご覧ください。

さらに、日付、住所、電話番号等の基本情報の書き方もこれまでバラバラでした。日付の書き方は、西暦和暦、全角半角の混在、／区切り、年月日の漢字区切りなどがあります。こうしたデータをデータ連携するときには2021-06-16で統一する等のルール化しています。同様に住所の表記もバラバラだったものを整理しており、丁目以下を半角数字で記入するとか、英字の住所の書き方を統一表記するなどをルール化しています。このような社会の基本情報のルールを[行政基本情報データ連携モデル](https://cio.go.jp/guides" \t "_blank)で定めています。このような基本データの表記をそろえることで、申請情報とベース・レジストリ情報の自動照合が容易にできるようになります。

基本情報の表記が統一されるだけではなく、データ項目の標準化も進めるとデータ処理が正確にできるようになります。そのために整備されたのが[共通語彙基盤](https://imi.go.jp/goi/)です。サービス時間「10:00～18:00」の書き方は「10時－18時（12時－13時は休憩）」のように様々な表記形式がありますが、これを処理するのは大変です。開始時間「10:00」、終了時間「18:00」というようにデータ項目をきちんと定義することでコンピュータ処理が正確にできるようになります。このように正確にデータ項目を整備できるようにしたものが共通語彙基盤です。

このデータ項目を組み合わせることで、様々なデータのテンプレートというかデータモデルを作ることができます。オープンデータのための[推奨データセットのデータフォーマット](https://cio.go.jp/policy-opendata)は共通語彙基盤のデータ項目の組み合わせでできています。そのため、医療機関やイベント等、データモデルが異なっていても住所や連絡先の記述が共通化できています。

そして6月4日に公開したのが[行政サービス・データ連携モデル（β版）](https://cio.go.jp/guides" \t "_blank)です。申請書や証明書等のデータモデルを公開しています。これに基づき申請システムなどを作ればベース・レジストリと照合しやすくなり、将来、ワンスオンリーや自動審査が容易にできるようになります。β版になっているのは、既存システムへの影響が大きいので慎重に決める必要がある一方で、早く公開してほしいという要望があるので検証のための期間を置いています。

また、コードについては、[コード一覧](https://cio.go.jp/codes)で集約しているものを拡充していきます。これまでクロールで集めていたものを登録型のカタログにすることで、内容も充実していくと思います。

センサーデータはスマートシティでも重要ですし、医療などの準公共データでも重要になります。スマートシティのインタオペラビリティ確保のためにデータ標準の整備をしているところであり、これらの整備により、産業も含めた社会全体のデータ流通が幅広くできるようになると考えられます。

全体図にはないですが地理空間データも重要な要素です。これまで、政府全体で正確に町字情報を系統だって保有しているところがなかったのですが、これを整理するための実証を始めるところです。オープンデータの仕組みを使って、自治体から最新の情報を集める仕組みを作ろうとしています。さらには空家問題や所有者不明土地問題、防災を想定して不動産情報の在り方まで検討は広がりつつあります。

**そして使えるデータにするには管理が重要**

データを連携する基盤を用意して、データ標準を決めればデータ社会が円滑に回るかというとそんなことはありません。データマネジメントが必要になります。どのような体制で管理、確認して、サブスクリプションなどの更新をするかという全体管理が重要になります。既に環境省では[データマネジメントポリシー](http://www.env.go.jp/press/109431.html" \t "_blank)を定めて運用を始めています。このようなマネジメントがしっかりしているからこそ、各種申請サービスはもちろんのことAIを使った業務なども持続的に行えるようになるのです。

データマネジメントの中で、特にデータ品質は重要です。皆さんが買い物に行ったとき何気なく見るものがあります。食品であれば、製造情報や成分、衣服であればタグについた素材や原産国に関する情報です。どのようなところからきて何が入っているのか、実物を見るとともにチェックしますよね。データも同じです、品質情報が付いていないと取引できない時代が来ます。そのため世界各国でデータについて一番のホットトピックといえばデータ品質です。入力時やインポート時にデータを評価するバリデータや、客観的な品質評価フレームワークが求められています。

そのため[データ品質管理ガイドブック（β版）](https://cio.go.jp/guides)を策定しています。標準や解説の寄せ集めではないかという意見もいただいていますが、まずは、データそのものの品質、データを扱うプロセスの品質、そのガバナンス体制の品質の理解を深めることが必要でありそのための検証版としてベータ版で公開しました。そうはいっても机上の空論ではありません。2019年度には、データ品質の[プレ調査](https://www.meti.go.jp/meti_lib/report/2019FY/000589.pdf" \t "_blank)を開始しています。P54以降でISO8000を軸にデータ品質に関する課題などの調査をしています。さらに2020年度は前出のデータ品質管理ガイドブックとその調査票を使って複数企業への[適用評価](https://www.meti.go.jp/meti_lib/report/2020FY/000033.pdf" \t "_blank)をしています。P27以降に調査結果がありますが、調査票以外にもインタビュー調査をしてデータ品質管理の方向性について検討をしています。今後は、これらの結果をデータマネジメントへ組み込んでいくこととなります。

**特に設計を抑え込むことが必要**

データの周りをいくら整備したところでシステムやサービスでうまく使ってもらわないと意味がありません。特に目を付けているのがマスターデータ設計とコード設計です。ここがきれいにできていれば、普通のシステムはきれいに作られます。なぜならマスターデータは一番最初に検討されるから、ここでデータ設計を踏み外さなければ、そのノウハウで他のデータもきちんと設計されるからです。そのため、[マスターデータ等基本データ導入実践ガイドブック](https://cio.go.jp/guides" \t "_blank)を作っています。また、様々なデータの分類コードもシステム全体に影響を与える重要情報ですが、他システムとの関係性など考慮せずに独自に設計してしまうケースが数多く見られます。そのため[コード（分類体系）導入実践ガイドブック](https://cio.go.jp/guides)を整備しています。

これらのガイドを使うとともに、前出のデータ標準類を組み合わせて使い、データ連携基盤も既存のものを活用することで、システム設計の大幅な品質向上と生産性向上が期待されます。

さらに設計では各種モデリング手法を導入して品質と生産性の向上を図っていく必要があります。そのため、データ戦略の検討ではアーキテクチャはarchimateで書き、データはクラス図で記述することを試行しています。このような手法面からの取り組みも進めているところです。

**ルールも重要な要素だよ**

技術はもちろん重要ですけどルールも重要な要素です。そもそも日本の社会はデータというと「目的外利用はしません」というのが定型句のように使われてきましたが、そのような考え方も変えていく必要があります。そのため戦略の中では考え方の変化の重要性にも言及しています。さらに、個人情報保護条例2000個問題のように広域連携しようにもデータ活用ルールが違っていると連携できません。コロナ対策サービスでも多くのところで個人情報の課題にあたっています。このようなところも変えていく必要があります。

データの契約や利活用に関しては経済産業省から多くのドキュメントが提供されています。[AI・データの利用に関する契約ガイドライン 1.1版](https://www.meti.go.jp/press/2019/12/20191209001/20191209001.html)に契約関係のことが整理され、知的財産の観点から[データ利活用のてびき](https://www.meti.go.jp/policy/economy/chizai/chiteki/pdf/datatebiki.pdf)が提供されています。またこの手引きには、より詳細を示す[データ利活用のポイント集](https://www.meti.go.jp/policy/economy/chizai/chiteki/pdf/datapoint.pdf)が提供されています。

**ところで柱にもなっているトラストってなんだっけ**

戦略では、トラストの中でもテクニカルな真正性の証明や非改竄性の証明に大きくフォーカスを当てています。このような基盤的なサービスの利用は、データ流通を行う上では必要です。

一方、データ流通するうえで重要なのは連携先の信頼の面でのトラストです。日本はそのためDFFT（Data Free Flow with Trust）を世界に発信しています。品質について買い物の例でしましたが、再び買い物を例にしてみると、品質表示されていても、得体のしれない製造元の商品は買いたくないってありますよね。世界でMade in Japanが信頼されているように個別の品質だけではなく全体の信頼感の醸成が必要になります。そのためには、データ提供元の透明性なども重要な要素になってきます。

**やっぱり人が欲しいよね**

こうした取り組みを推進するのには体制が必要で、その体制は個々の人材によってなり立ちます。国内のデータ関連人材は、データベース技術者を中心に育成してきており、最近はデータサイエンティストやAI人材の育成が盛んです。一方で、データをゼロから設計できるデータエンジニアがほぼいません。世界では、パッケージソフトを作るためにデータを抽象化、構造化して設計するのが当たり前ですが、日本では御用聞きして、従来の画面や帳票をベースに設計するSEがほとんどであり、大手企業でさえデータ設計をエクセルでやっています。さらにDFDのような古典的技法でデータ設計している人もいます。

国内のデータの参考書を見ても多くがER図であり、世界の主流であるUMLのクラス図の本もほとんど出版されていません。このような教材整備から必要です。

そこで、データエンジニアのロール、スキルを定義し、人材フレームワークとして整理しています。既にそれをもとにした[コースも提供](https://note.com/hiramoto/n/n441578d54b2b" \t "_blank)しているので是非ご覧ください。一方でクラス図やアーキテクチャ等の具体的な講義は、国内のコースがないので[edx](https://www.edx.org/)や[coursera](https://www.coursera.org/" \t "_blank)といった海外の大学のコースの受講を薦めています。

こうした基礎知識とともに、経験も重要であり、今後はOJTも含めてデータエンジニアリングができる人材を育成していくことが急務と考えられます。

**体制はどうあるべきか**

世界では、このような取り組みを進めるためにCDO（Chief Data Officer）を置き、政府横断のガバナンスボードを作ることでデータ戦略を進めています。なぜなら、データの重要性を重視しているとともに、既存制度をデータ社会に合わせて変えていくのに強力な調整力が必要だからです。データ戦略は、優秀な人材を集めるだけでは実現できません。日本でもこのような体制を作れるかどうかがデータ戦略成功のカギになってきます。

また海外の政府では、数十人のデータチームを作るのは当たり前であり、100人規模でもおかしくありません。

リーダーシップ、人材の質、人材の量の3つのポイントをきちんと押さえて推進していく必要があります。

**ところでショーケースってどんな感じかな**

ｇBizInfoやミラサポplusで実装が始めっていると前述しましたが、具体的に紹介していきましょう。

ｇBizInfoでは、サービスの基本設計からデータは共通語彙基盤を使っています。各省から情報を収集し、広くAPIで提供することを目指していたので、データの標準化が重要であり、共通語彙基盤をベースに外部データをクレンジングして蓄積する仕組みにしました。このため、最近運用を開始した法人データ連携基盤のｇBizConnectや後述のミラサポplusなどともデータ連携しやすくなっています。一方で、既存データに依存せざる部分も一部残っています。

ミラサポplusでは、利用者登録のデータをｇBizInfoのデータ項目と合わせて連携しやすくする等、基本部分でデータ連携を考慮しています。さらに、制度や事例のデータ提供に行政サービス・データ連携モデルの標準的なデータ項目を使っており、制度と事例を同じタグで検索出来たり、情報連携させるなどを実現しています。

これらにデータ標準が積極的に導入されているのは、経済産業省のデジタル化チームがデータ標準の重要性を十分に理解して、システム構築時に必ず標準を考慮するようにしているからです。

**ベース・レジストリの指定が行われました**

さてこのようにデータ戦略は、これまでも着実な準備がされてきたのですが、データ戦略の中で重要なことは5月26日に[ベース・レジストリの指定](https://cio.go.jp/node/2764" \t "_blank)が行われたことです。これで、本丸に着手できるようになります。既にベース・レジストリを作るためのデータ標準はできています。今後は、その具であるデータを整備し使えるようにしていくこととなります。

ベースレジストリを指定するということはどういうことかというと、一番大きいのはワンスオンリーの時に使うデータになるということです。ベースレジストリに指定された部門は、ワンスオンリーのために品質の良いデータを提供しなければならなくなります。また、申請や証明の業務システムを持っているところは、システムをワンスオンリーできるようにしていく必要があります。[デジタル手続法](https://www.kantei.go.jp/jp/singi/it2/hourei/digital.html)ができてから2年でやっと実現に向けた方向性ができました。

もう一つの大きな効果である重複データをなくすというところは、ボディーブローのように効いてくる効果です。例えばアドレスデータベースを作るという作業ですが、このベース・レジストリができると、行政だけでなく民間も含め住所処理のコストが抜本的に安くなります。さらに事業所のデータも効果は巨大です。これまでも法人情報はgbizinfoから入手できましたが、事業所のベース・レジストリができることで、エラーのない迅速な手続きや配送などの処理などができるようになります。日本中の組織が顧客データベースを持っていますが、このデータで根本的にメンテナンスコストを下げることができるのではないでしょうか。

既に、アドレスと事業所などを対象としたベース・レジストリのパイロット事業の公募が開始されています。パイロット事業というと、検証して継続できないという事業をよく見ますが、今回の事業は、安定したベース・レジストリを構築して運用することを目的としています。難易度の高いプロジェクトですが、ぜひ100年以上つかえる社会基盤を実現していきたいです。

**ベース・レジストリはコストじゃなくて投資です**

ベース・レジストリをバックオフィスコストとして見ている人も見受けられます。これは明らかに間違いです。高度成長期に鉄道、高速道路や工業地帯に長期的な視点から投資をしてきたのと同様に、デジタル社会の基盤をなす取り組みです。しかも投資対効果はかなり高いと国内外で試算されています。

例えば文字情報基盤の整備ですが、2010年から10年間で約7億円の投資が行われていますが、120億円の効果が試算されています。商業登記の証明書は行政機関（自治体含む）への提出のために年間約1億枚発行されており、１通500円として年間500億円が使われています。現在、府省へのAPI提供の準備が進められており、大幅な投資対効果の実現が期待されています。

このように経済効果を考えながら長期的視野に立った投資が必要になってきます。

**ところで、AIとかデータサイエンスとかキラキラがないよね**

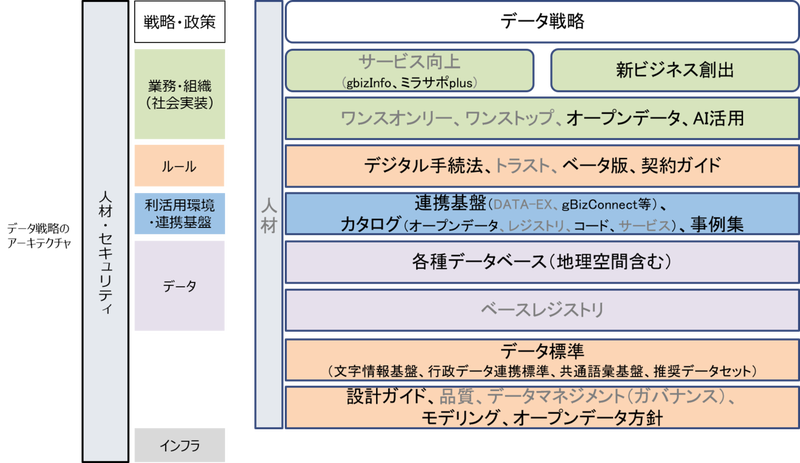
AI戦略は別途策定されていますし、データサイエンスは市場でも盛り上がっているので進んでいくでしょう。そういう点で、データ戦略では、利用領域に正確なデータを供給することにフォーカスしています。

一方で行政でもデータを使った政策立案、いわゆるEBPM（Evidence Based Policy Making）が求められています。データサイエンスの技術を行政機関でも積極的に取り入れていく必要があります。データ整備の投資対効果の分析もそうした流れの一部です。そのような実務的な部分から実績を積んでいくこととなります。

また、ブロックチェーン等の技術にも触れていないという指摘もあります。ブロックチェーン以前の問題として、守るべきデータをまずは揃えなければならないという喫緊の課題があります。今後、ベース・レジストリの構築を進めていく実装の検討で実務的に検討していくことではないでしょうか。

**かなり実務的ですよね**

ここまで書いてきましたように、データ戦略はこれまでの小さな取り組みの積み重ねでできています。でもほぼ体系は整っています。そこに今回のデータ戦略ができました。要するに推進力が上から覆いかぶさってきたわけです。



チームも大きくなってきました。でも、チームが大きくなるスピードより、期待値が膨らむスピードや、要望されている内容が増えるスピードのほうが速い気がします。

これらの戦略を20人程度で推進しているのはかなり大変です。外部との連携も重要と考えていますので、皆さんもこのビッグウェーブに乗ってみませんか。