

統計マイクロデータ利用の安全性確保に関する調査研究

統計データの二次的利用促進に関する研究会 資料

2017年3月17日

本資料の構成

1. 本調査研究の位置づけ
2. 各国／機関のオンサイト利用の実態
3. 我が国オンサイト利用への示唆・今後の課題抽出

1. 本調査研究の位置づけ

本調査研究の背景と目的・方法

【背景】

- 総務省では、「公的統計の整備に関する基本的な計画」(平成26年3月25日閣議決定)に基づき、オンサイト利用の実用化に向けた検討を進めている。
- オンサイト利用で実績のあるイギリスの取組み等も参照し、我が国では、利用者に対する信頼を重視した「人間モデル」を想定し、分析結果等をチェックする基準が検討されている。
- 基準等の検討においては、特に
 - ①. 調査票情報の安全確保のアプローチの仕方
 - ②. それに基づく利用者への教育プログラムの実施
 - ③. 利用者の作成した分析結果等のチェックの仕方(技術・プロセス等)に関し、オンサイト利用で先行する諸外国／機関の取組みを踏まえ、我が国のオンサイト利用の取組みへの示唆を得ることを考えている。

【目的・方法】

- 上記背景に鑑み、本調査研究は、後述する5つの国／機関を対象に、オンサイト利用に係る法制度全般、教育プログラム、分析結果等のチェックの仕方を中心とした実態把握を実施し、我が国オンサイト利用への示唆・課題抽出を検討することを目的とする。
- 実態把握においては、公開されている情報の分析とそれに基づく論点の整理をした上で、各国／機関への実地調査(平成29年2月12日～18日)を実施した。

本調査研究の対象

■ EUを中心に、以下5つの国／機関のオンサイト利用にかかる実態を調査研究した。

対象国	組織※	サービス名	調査対象とした理由(初期仮説)
EU	Eurostat	safe centre	<ul style="list-style-type: none"> ESS Handbook等を手掛け、EU各国の取り組みを踏まえたオンサイト利用を実施していることが期待される
イギリス	イギリス国家統計局 (ONS)	VML (Virtual Microdata Laboratory) ※VMLの内、オンサイト利用のみ を対象とする	<ul style="list-style-type: none"> 5 safe model等マイクロデータの利活用に係る考え方を提示し、それを実際の運用に落とし込んでいる
ドイツ	ドイツ連邦統計局 (FDZ-Bund)	safe centre	<ul style="list-style-type: none"> 2002年よりオンサイト利用を開始した 連邦統計法の厳しい制約下の中で、サービス提供を継続している
	ドイツ連邦雇用庁 (FDZ-IAB)	on-site use	<ul style="list-style-type: none"> 連邦統計局とは異なり、海外にもオンサイト利用の施設を展開し、リモートアクセスも提供する等、取り組むが先進的であるように推察される
オランダ	オランダ統計局 (CBS)	on-site	<ul style="list-style-type: none"> 秘匿処理に係るtau-Argus等のツールを保有し、高度な技術とツールを持って、秘匿処理を実施しているものと推察される

※) 順序は、仕様書記載の通り

2. 実地調査を踏まえた各機関のオンサイト利用の実態報告

本調査研究の主要調査事項

- 諸外国／機関の外形的なサービスプロセスを把握するだけでなく、プロセス構成の前提となる考え方や、定量的な側面等に関しても配慮をし、実態調査を実施した。

調査項目の大分類

定性的な調査項目

定量的な調査項目

調査項目の大分類	定性的な調査項目	定量的な調査項目
概要	<ul style="list-style-type: none"> ・オンサイト利用におけるオンサイト利用の位置づけ(リモートアクセス等との差異等) ・オンサイト利用で利用可能なデータ等 	<ul style="list-style-type: none"> ・利用者数 ・プロジェクト数 及びその推移
法制度	<ul style="list-style-type: none"> ・根拠となる法制度とそのポイント ・規約等の設定状況 	<ul style="list-style-type: none"> ・過去の罰則発生状況と制裁の適用状況
技術	<ul style="list-style-type: none"> ・サーバ・ネットワーク等の構成概要 ・物理面／IT面でのセキュリティ担保の手法 	<ul style="list-style-type: none"> ・施設数 ・利用可能なソフトウェアのライセンス数
運用	<ul style="list-style-type: none"> ・利用申請～利用後のプロセスの構成と各プロセスの特徴 	<ul style="list-style-type: none"> ・申請件数、審査通過率 ・各プロセスにかかる期間
	<ul style="list-style-type: none"> ・経験則、原則ルールそれぞれの適用状況 ・審査に利用しているツールの有無 ・技術的に高度な問題(占有度ルール等)への対応 	<ul style="list-style-type: none"> ・審査件数 ・審査人員 ・審査期間 ・審査通過率
	<ul style="list-style-type: none"> ・教育プログラムの有無とその理由 ・教育プログラムの内容・修了要件 ・講師の資質 	<ul style="list-style-type: none"> ・受講者数と受講者の種類 ・受講期間 ・講師人数

※) 濃い青色が、本調査研究の仕様の要求事項である。

※) 全て調査項目を個々に把握しきることは目的としていない。これらの視点に従い、各国／機関のオンサイト利用の特徴を把握することを目的としている。

参考 | 重要な言葉の定義

- 後述で頻繁に使用する、オンサイト利用、リモートアクセス、リモートエグゼキューション等の違い、また、秘匿審査における経験則と原則ルールの違いを、以下に説明する。

テーマ	用語	定義
マイクロデータ利用に係る代表的なサービス	オンサイト利用	<ul style="list-style-type: none"> ● 情報セキュリティ等が確保された、サービス実施者によって認定された施設内に限り、公的統計のマイクロデータを利用できるサービスで、直接利用型(オンサイト利用型)の一種
	リモートアクセス	<ul style="list-style-type: none"> ● インターネット経由でサービス実施者が提供しているシステムにアクセスし、利用者自らのコンピュータ上でマイクロデータを参照し分析することができるサービスで、リモートアクセスプログラム送付型(参照可能型)の一種
	リモートエグゼキューション	<ul style="list-style-type: none"> ● インターネット経由でサービス実施者が提供しているシステムにアクセスし、利用者自らのコンピュータ上でダミーデータを参照し分析することができるサービスで、リモートアクセスプログラム送付型(参照不可能型)の一種
秘匿審査に係る代表的な用語	経験則 (Rule of thumb)	<ul style="list-style-type: none"> ● 明示的でかつ厳格な安全性を要求するルール ● 統計的開示抑制技術に精通していない審査者も機械的に運用できるが、厳しすぎる安全性を求める可能性がある
	原則ルール (Principle-based rule)	<ul style="list-style-type: none"> ● 申請の状況を考慮し柔軟な審査を行うルール ● 研究者の目的を踏まえ安全性を審査するため、統計的開示抑制に精通した審査者による審査が必要となる

※) サービスの定義は、三菱総合研究所『統計データ・アーカイブの整備に関する調査研究報告書』(平成23年度)を参照

参考 | 秘匿審査における技術的な問題

- 我が国では、Eurostatのハンドブック等を参照し、オンサイト利用に係る秘匿審査基準を検討してきた。ただし、技術の確立や運用に落とし込むには、不明瞭さが残る技術的テーマが存在した。
- 本調査研究では、特に下記の技術的な論点に対し、実地調査での聴取を実施した。
 - 各国／機関の取組み状況を踏まえて聴取を行い、全ての国／機関で、下記項目を網羅的に把握した訳ではない。

占有性ルール	二次秘匿	差分攻撃	最小度数の緩和	準識別子の決定
<ul style="list-style-type: none"> ● 対象とする統計は？ 必要な統計は事業系のみと考える ● 役割分担は？ 占有性の確認には、元の個票まで確認する必要がある どこまで利用者を信頼するかがポイントとなる ● 採択したルールは？ (n,k)ルールやp%ルール等のルールが存在する 	<ul style="list-style-type: none"> ● 秘匿インターバルの閾値は？ 二次秘匿結果を安全にするには、秘匿セルが取りうる値のインターバルを大きくする必要がある 	<ul style="list-style-type: none"> ● 差分攻撃への対策実施の有無、及び対処の方法は？ どこまでの公表物を審査の対象に入れているか等、運用の仕方がポイントとなる 	<ul style="list-style-type: none"> ● 標本調査やサンプリングした母集団に対する最小度数の緩和の有無は？ 	<ul style="list-style-type: none"> ● 何を個人情報との再識別に繋がる情報と重視し、それ以外の変数のチェックを緩くするか？

各国／機関の実態調査結果 | まとめ

（オンサイト利用の位置づけ・考え方）

- ミクロデータにアクセスさせるサービスの中で、オンサイト利用は特に、**利用者が利用中に、機関職員と密なコミュニケーションを通じ、研究活動と結果として得る生成物を、効率よく作り込む点**が強調され、これは他のサービスとの違いになっている。
- **オンサイト利用とリモートアクセスの両方を提供することで、利用者にとって利便の高いサービス提供が可能となると考えている。**

（オンサイト利用で利用可能な統計・素材）

- **オンサイト利用では、利用できる素材の秘匿度が低かったり、リンケージ研究ができたりと、他のミクロデータへのアクセスと比較した違いが明確に作られている。**

（設備環境とプロセス）

- オンサイト利用の考え方にもとづき、サテライト利用だけでなく、**機関本体内にオンサイト利用施設を設置**、さらにそこで、**データに詳しい機関職員(研究者等)が、利用者と一緒に、秘匿処理も含めて生成物を作りこむプロセス**を採用している。
- ソフトウェアは、Rに限らず、各種分析ツールの利用を可能としている。

（秘匿審査）

- **経験則と原則ベースの二段階審査を実施している訳ではなく、一体として審査を実施している。**
- 非常に多くの知見と経験を積んだ審査者に、**「水際」で審査をさせるのではなく、利用者が利用する段階から、生成物を作りこんでいくことを重視**、結果、**秘匿に係る高度な技術的問題も、一定程度は「利用者の信頼」の元、検証を省いている。**
- 利用者との信頼関係構築のもと、必要最小限の生成物を作成する雰囲気[※]が形成されているため、**秘匿処理をするためのツールは利用せず、マニュアルで審査や処理を行っている。**

（教育プログラム）

- 利用者[※]と機関職員は、利用時のコミュニケーションを密に取ることで、信頼関係を構築し、生成物を作りこむため、**教育プログラムによって、利用者の秘匿処理等に係るスキルや知識を強化させようとする考え方を持った機関はいない。**

※）各国／機関に共通し、我が国のオンサイト利用に参考となる事実のみを取り上げている。

各国／機関のオンサイト利用の概要

■ オンサイト利用の利用者数は、各国／機関で様々。拠点配置の特徴として、どの国／機関も、サテライトオフィスだけでなく、機関内に施設を設置している点が挙げられる。

	Eurostat	イギリス国家統計局	ドイツ連邦統計局	ドイツ連邦雇用庁	オランダ統計局
利用者	<ul style="list-style-type: none"> ● 利用者は20名弱 (CISでの実績) ● 民間企業も利用可能 ● 海外の利用者も存在 	<ul style="list-style-type: none"> ● 登録された利用者は500名 ● 現状230件のプロジェクトが稼働、近い将来1,000件に達すると想定 ● 民間企業も利用可能 ● 海外の利用者も存在 	<ul style="list-style-type: none"> ● 受理された申請件数(≒プロジェクト数)は毎年130件程度で安定 ● 民間企業は利用不可 ● 海外の利用者も存在 	<ul style="list-style-type: none"> ● 累積利用者は1,110名超 ● 毎年のフローは100名程度 ● 稼働中のプロジェクトは500件超 ● 商用利用は不可 ● 海外の利用者も存在 	<ul style="list-style-type: none"> ● - ※ リモートアクセスを含め、利用者は100名、プロジェクト数は200～300件／年 ● 民間企業も利用可能 ● 海外の利用者も存在
拠点	<ul style="list-style-type: none"> ● Eurostat施設内に1サイト設置 	<ul style="list-style-type: none"> ● イギリス国内に5サイト設置 (統計局支局内) ● ティッチフィールド支局の席数は4席 	<ul style="list-style-type: none"> ● 統計局を含め、国内に18サイト設置、手薄な地域へのサイト展開を計画中 ● 統計局内のサイトの席数は3席 	<ul style="list-style-type: none"> ● 雇用庁を含め、国内に7サイト、他、海外に8サイト設置(海外では秘匿度の高いデータのみ、利用可能) ● 雇用庁内サイトの席数は3席 	<ul style="list-style-type: none"> ● 統計局支局内に2サイト設置(サイト数を縮小中)

※) CIS=Community innovation survey(科学技術研究調査)

参考 | オンサイト利用施設の室内

- 撮影を許諾頂いたドイツ連邦統計局(左)とドイツ連邦雇用庁(右)の施設は、それぞれ以下の通り。広めのスペースを三席が利用可能となっていた。
- 一方、撮影できなかったイギリス国家統計局の施設は、窓もほとんどない狭い部屋に、三席が用意されていた。
 - 簡易な間仕切りはあったが、監視カメラなどの存在は確認されなかった。



各国／機関のオンサイト利用の概要

■ ミクロデータの利用に関し、各国、各機関は、複数のサービスを提供。その中で、オンサイト利用の特徴は、いずれも、利用者と職員が密に連携し、効率的に生成物を作りこんでいる点にある。

	Eurostat	イギリス国家統計局	ドイツ連邦統計局	ドイツ連邦雇用庁	オランダ統計局
ミクロデータへのアクセス手法	<ul style="list-style-type: none"> ● 匿名データ提供とオンサイト利用の2手法を提供 ● 匿名データ提供に係るリスクを懸念 ● リモートアクセスは、予算上、未実施 	<ul style="list-style-type: none"> ● オンサイト利用とリモートアクセスを提供 ● リモートアクセスは政府機関及びイングランド銀行のみ <p>※ 上述は統計局のサービスのみ、エセックス大学が提供するUK Data Service等は除く</p>	<ul style="list-style-type: none"> ● SUF、リモートエグゼキューション、オンサイト利用の3手法を提供 ● SUFの準備は、非常に手間と認識 ● リモートアクセスは、法制度上、実施不可 	<ul style="list-style-type: none"> ● オンサイト利用とリモートアクセスの2つを提供 ● リモートアクセスは、分析を試行錯誤するためのツールとして使い、本番をオンサイト利用で実施するケースも多い 	<ul style="list-style-type: none"> ● PUF、オンサイト利用、リモートアクセスの3手法を提供 ● 現状主流になりつつあるのは、リモートアクセス
オンサイト利用の位置づけ	<ul style="list-style-type: none"> ● 秘匿度の弱いミクロデータにアクセスする唯一の手段 ● 密なコミュニケーションを通して、利用者・職員が共同で生成物を作成 	<ul style="list-style-type: none"> ● 信頼される利用者と共に、生成物を作成 ● コミュニケーションを非常に重視し、利用中も、利用者と職員は密に連携 	<ul style="list-style-type: none"> ● 利用者を信じ、生成物を共に作成 ● 統計局職員が、分析のコーディネーグレベルから助言をし、効率的な生成物作成を支援 	<ul style="list-style-type: none"> ● 秘匿度の弱いミクロデータにアクセスする唯一の手段と位置づけ ● 雇用庁職員の助言を受け、成果の質を向上させている 	<ul style="list-style-type: none"> ● リモートアクセスシステムダウン時のバックアップとしての位置づけ ● ビギナーは、職員の助言を受け、効率的に研究を進めることが可能

※) PUF = Public use file、SUF = Scientific use file

各国／機関のオンサイト利用の概要

- 国／機関によっては、オンサイト利用でのみ扱えるデータセットを定義したり、リンケージ研究等を許している。利用可能な素材は、法的制約のあるドイツ連邦統計局以外は、仮名化データである。

	Eurostat	イギリス国家統計局	ドイツ連邦統計局	ドイツ連邦雇用庁	オランダ統計局
利用可能な統計の種類	<ul style="list-style-type: none"> ● オンサイト利用でのみ利用可能な統計は3種類 	<ul style="list-style-type: none"> ● オンサイト利用で利用可能な統計は59種類 ● 人口センサスは当該サービスのみ利用可能 ● リンケージされたデータセットも一部利用可能 	<ul style="list-style-type: none"> ● オンサイト利用で利用可能なデータセットは140種類程度 ※公開データは全てドイツ語 	<ul style="list-style-type: none"> ● オンサイト利用で利用可能な統計は10種類 リンケージデータは6種類を用意 ● 雇用庁の実施した調査や、行政記録などのリンケージデータの利用も可能 	<ul style="list-style-type: none"> ● オンサイト利用で利用可能な統計は21種類以上 ※英語での案内は断片的 ● リンケージ研究のための素材結合は、利用者自ら実施可能
利用可能な素材	<ul style="list-style-type: none"> ● 上記マイクロデータは、仮名化データレベルで利用可能 	<ul style="list-style-type: none"> ● 上記マイクロデータは、仮名化データレベルで利用可能 	<ul style="list-style-type: none"> ● 統計法上、弱い匿名化データは参照不可 ● 各種秘匿処理を入れた、de facto anonymized(事実上匿名化した)な素材を使用可能 	<ul style="list-style-type: none"> ● 上記マイクロデータは、仮名化データレベルで利用可能 	<ul style="list-style-type: none"> ● 上記マイクロデータは、仮名化データレベルで利用可能

各国／機関のオンサイト利用の法制度

- 法制度の中で、利用者・利用目的が特定され、それに則った規約や申請書／宣誓書類等が規定されている。以降で示すプロセス等も、この考え方に沿って形成されている。

	Eurostat	イギリス国家統計局	ドイツ連邦統計局	ドイツ連邦雇用庁	オランダ統計局
準拠する法制度	<ul style="list-style-type: none"> ● regulation 223 /2009 23条 ● regulation 557 /2013 8条2項 	<ul style="list-style-type: none"> ● Statistics and Registration Service Act 2007 39条 ● Freedom of Information Act 	<ul style="list-style-type: none"> ● BstatG 16条 6項、7項、8項 	<ul style="list-style-type: none"> ● SGB X 75条 ● SGB III 282条 	<ul style="list-style-type: none"> ● Statistics Netherlands Act 37条、41条、42条
法制度の中核	<ul style="list-style-type: none"> ● 223/2009 23条で、科学的な目的のみに限定 ● 557/2009 8条2項で、オンサイト利用の申請～公表等のプロセスとセキュリティ対策を規定 	<ul style="list-style-type: none"> ● 承認された研究者のみに利用主体を限定 ● 生成物は公共に属し、誰でも見られることを規定 	<ul style="list-style-type: none"> ● 高等教育、独立した研究機関の科学的なプロジェクトに目的を限定 ● 事実上匿名化されたデータのみ提供可能 ● 研究後のデータ抹消を要請 	<ul style="list-style-type: none"> ● SGB X 75条で、公共の利益、労働市場研究の目的での利用を許可 ● SGB III 282条で、海外オンサイト利用施設を規定 	<ul style="list-style-type: none"> ● 利用目的を統計的な目的に限定 ● 利用可能な主体を限定
規約・宣誓書等	<ul style="list-style-type: none"> ● Application form for research entities ● Research proposal application form 	<ul style="list-style-type: none"> ● HOW TO BECOME AN APPROVED RESEARCHER 	<ul style="list-style-type: none"> ● Application for access to microdata for scientific research project 	<ul style="list-style-type: none"> ● Request for Access to Social Data or Data 	<ul style="list-style-type: none"> ● Application for access to microdata – form for organisations ● Application form research project

各国／機関のオンサイト利用の法制度

- 利用者からの課金で、オンサイト利用の体制を賄っているケースはほとんどない。なおイギリスは、オンサイト利用の必要性を政府に粘り強く説明し続け、現在の体制を維持することに成功している。

	Eurostat	イギリス国家統計局	ドイツ連邦統計局	ドイツ連邦雇用庁	オランダ統計局
課金体系	無料	無料	有料	無料	有料
課金に係る考え方	<ul style="list-style-type: none"> ● 以前は有料であった ● EU内の足並みが異なることから、Eurostatとして、無料に舵を切った ● 運用予算は、European commissionの予算で賄っている 	<ul style="list-style-type: none"> ● 利用者には課金はしていない ● 必要な経費は、数百万€に上るが、これは政府予算で賄う ● 政府には継続的な説明を実施、分析結果を持ち込み、オンサイト利用の必要性等を説得し、予算を確保している 	<ul style="list-style-type: none"> ● 以前は無料段階的に課金する体系にシフト変更は、KVI※からのレコメンドによるところが大きい模様 ● ただし、利用者からの収入で賄っているのは、全体の5～10%程度 	<ul style="list-style-type: none"> ● - 	<ul style="list-style-type: none"> ● データではなく、サービス「行為」ごとに、細かに課金体系を構築している ● 利用者の幅広いニーズに対しても、「予算があれば、幅広に対応する」スタンスを採用している

※) German Commission on Improving the Information Infrastructure between Science and Statistics

各国／機関のオンサイト利用の技術

- 各国／機関は、プロジェクト単位でアクセス制御を実施している点が特徴。なおイギリスは、利用者を信頼する一方で、操作ログを監視する等、彼ら／彼女らが信頼に値するか、トレースできるようにしている。

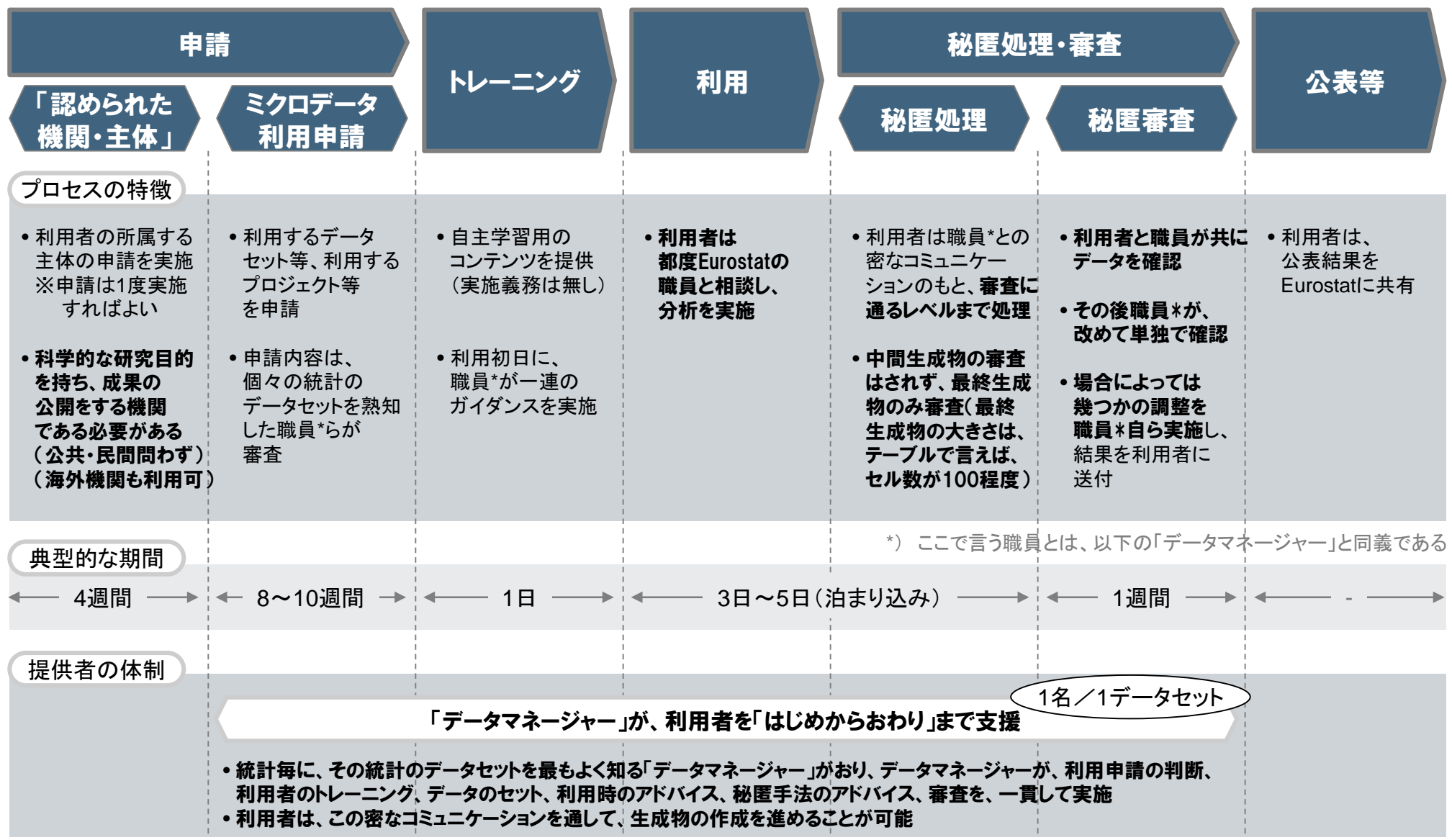
	Eurostat	イギリス国家統計局	ドイツ連邦統計局	ドイツ連邦雇用庁	オランダ統計局
DBの構成	<ul style="list-style-type: none"> ● 利用可能な統計は3種類のみよってDBは構築せず、ファイルをフォルダで管理する程度 	<ul style="list-style-type: none"> ● 国内民間企業に委託し、クラウド化 ● NWに専用線を用いる計画はなし 	<ul style="list-style-type: none"> ● SINAワークステーションによるシンクライアント化 ● NWは行政専用切り替え予定 	<ul style="list-style-type: none"> ● Citrixを用いてシンクライアント化 	<ul style="list-style-type: none"> ● ローカルでのアクセス
セキュリティ対策 (物理面)	<ul style="list-style-type: none"> ● IDカードによる、建屋、施設への入退館管理の実施 	<ul style="list-style-type: none"> ● IDカードによる、建屋敷地内、建屋内への入退館管理の実施 ● 移動には職員の同行が必要 	<ul style="list-style-type: none"> ● IDカードによる、建屋、施設への入退館管理の実施 	<ul style="list-style-type: none"> ● IDカードによる、建屋、施設への入退館管理の実施 	<ul style="list-style-type: none"> ● IDカードによる建屋、施設への入退館管理の実施 ● 移動には職員の同行が必要
セキュリティ対策 (システム面)	<ul style="list-style-type: none"> ● ID、パスワード認証を実施 	<ul style="list-style-type: none"> ● ID、パスワード認証を実施 ● IDはプロジェクト単位で発行 ● システム操作ログを監視するソフトを導入 	<ul style="list-style-type: none"> ● ID、パスワード認証を実施 	<ul style="list-style-type: none"> ● ID、パスワード認証を実施 ● スタッフと共同でなければシステムにログインができない認証方法を採用 	<ul style="list-style-type: none"> ● ID、パスワード認証を実施

各国／機関のオンサイト利用の技術

- 各国／機関、高額なソフトウェアも合わせて、オンサイト利用で利用可能としている。一部ソフトは、ベンダと有効な関係を構築した上で、優位な価格で契約している模様である。

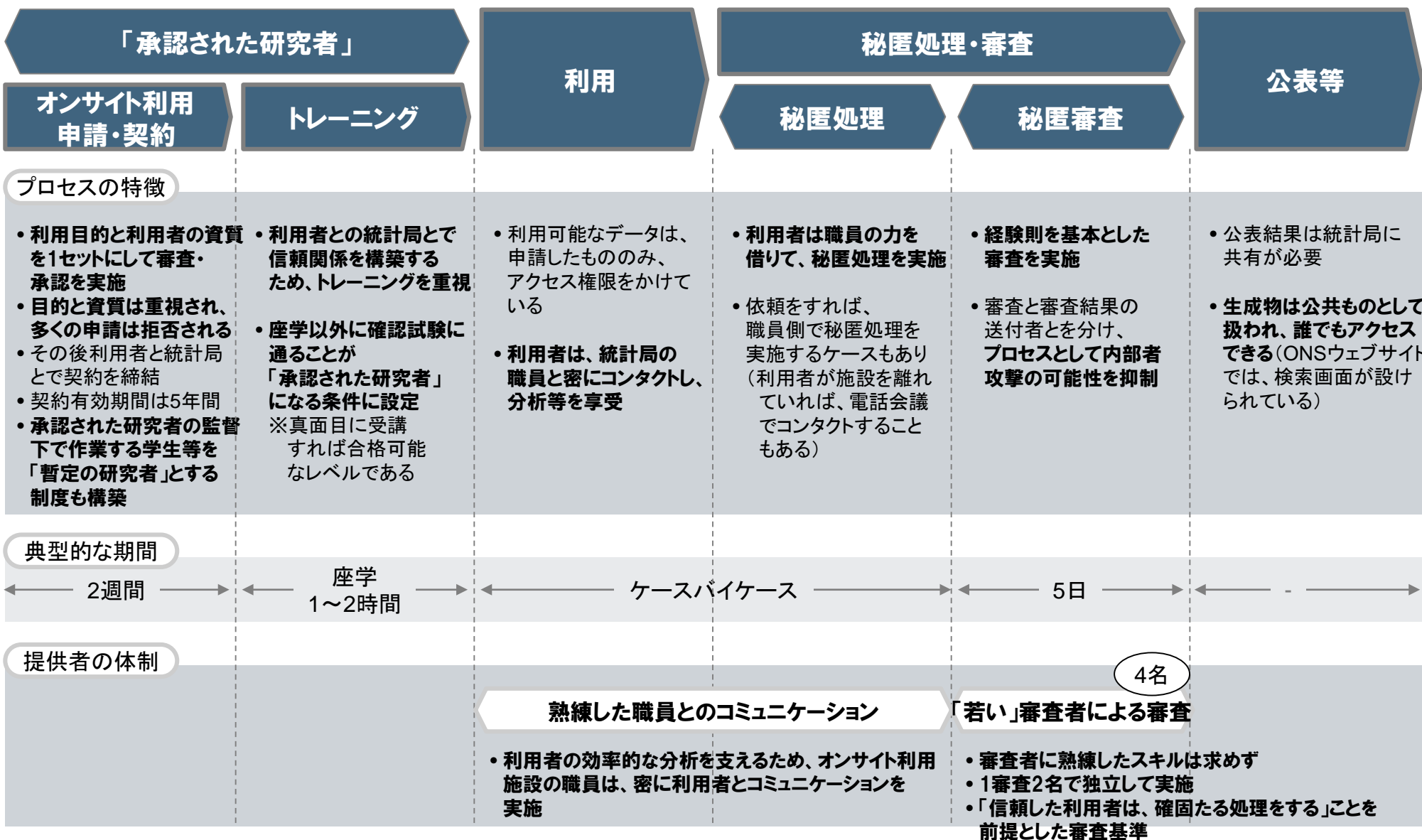
	Eurostat	イギリス国家統計局	ドイツ連邦統計局	ドイツ連邦雇用庁	オランダ統計局
利用可能なソフトウェア例	<ul style="list-style-type: none"> ● Microsoft office ● Matlab ● SAS 9.2 ● STATA 13 ● R 等 ※STATAでプログラムを作成し、利用者に提供 ※結果的にSTATAの利用率が高い	<ul style="list-style-type: none"> ● Microsoft office (xls, doc, ppt) ● STATA 9 ● SPSS 12 ● SAS ● R ● Python ● ArcMap 等 ※STATAで記載したプログラムを保有	<ul style="list-style-type: none"> ● R ● STATA ● SPSS 等 	<ul style="list-style-type: none"> ● STATA ● Matlab 等 	<ul style="list-style-type: none"> ● Microsoft office ● SPSS ● Stata ● R ● Stat Transfer 等
ライセンス契約の考え方	<ul style="list-style-type: none"> ● スタンドアロン端末ごとにライセンス契約を実施 	<ul style="list-style-type: none"> ● ー(不明) ※ニーズに応じて、ソフトウェアの導入も考えるが、費用の観点で、拒否される可能性もある	<ul style="list-style-type: none"> ● STATAは、ベンダよりディスカウントを受けている 	<ul style="list-style-type: none"> ● STATAは雇用庁自身が導入しており、ベンダと良好な関係を構築 	<ul style="list-style-type: none"> ● ー(不明) ※ニーズに応じて、ソフトウェアの導入も考える ※ソフトウェアの持込可能、ただしライセンスは利用者負担

各国／機関のオンサイト利用の運用(プロセス) | Eurostat

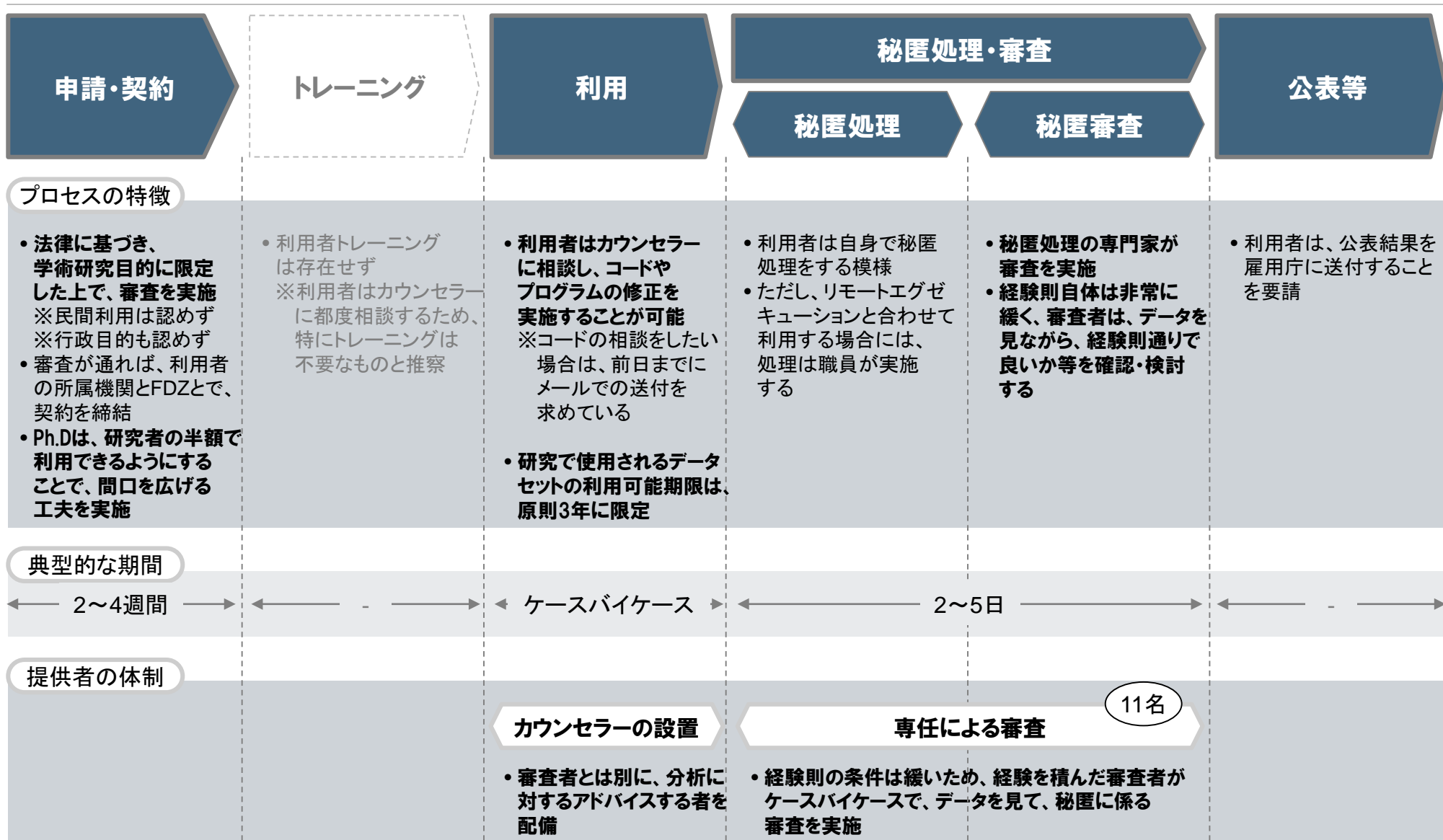


*) ここで言う職員とは、以下の「データマネージャー」と同義である

各国／機関のオンサイト利用の運用(プロセス) | イギリス国家統計局

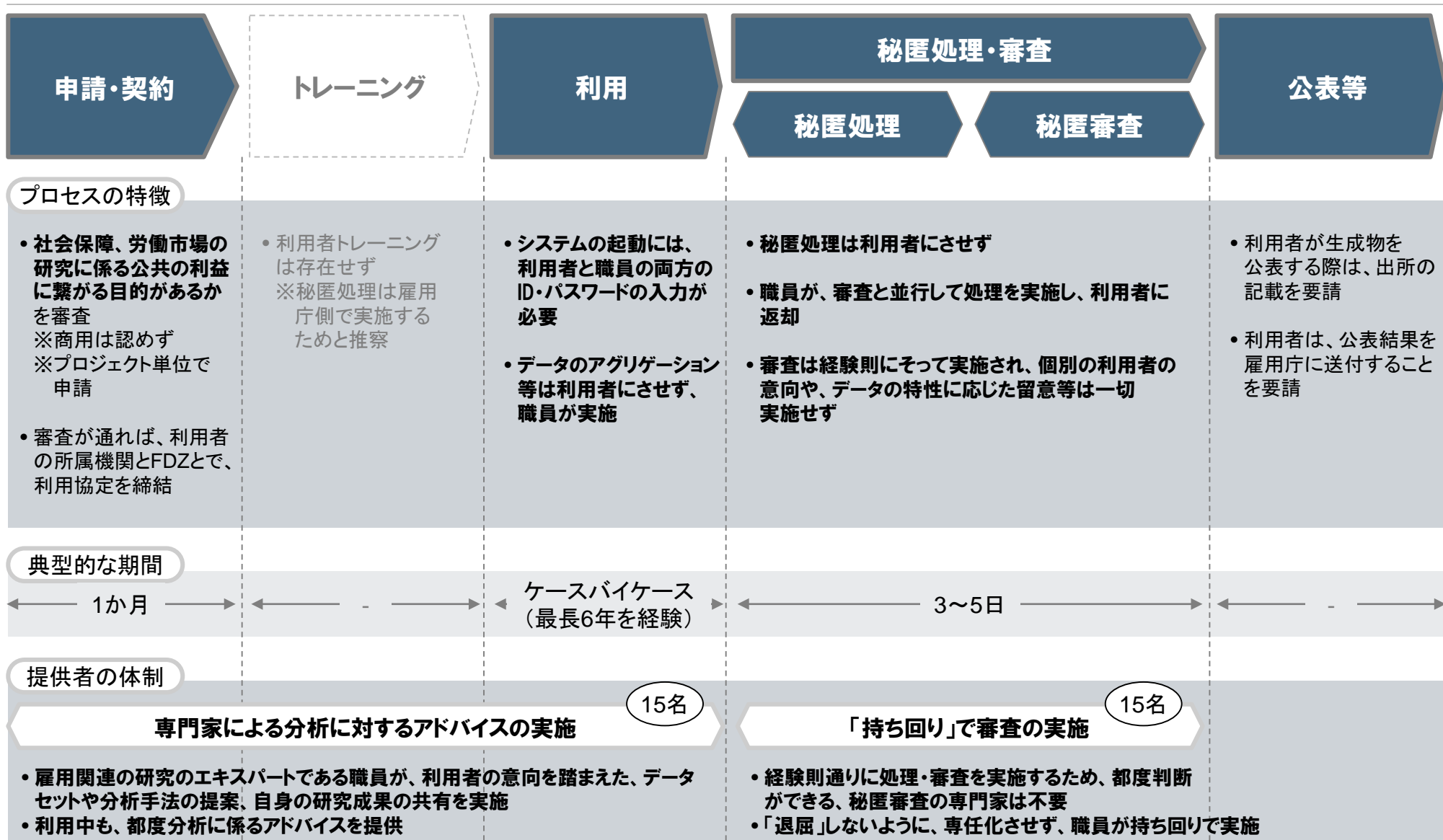


各国／機関のオンサイト利用の運用(プロセス) | ドイツ連邦統計局



※) ここで言うFDZは、統計局のオンサイト利用施設のことである

各国／機関のオンサイト利用の運用(プロセス) | ドイツ連邦雇用庁



※) ここで言うFDZは、雇用庁のオンサイト利用施設のことである

各国／機関のオンサイト利用の運用(プロセス) | オランダ統計局



各国／機関のオンサイト利用の運用(秘匿審査)

- 経験則、原則ルールを双方を扱うも、プロセスは渾然一体としており、かつ、経験則を主体的に使用。「水際」で専門家により秘匿開示を防ぐのではなく、利用者と職員の密な連携の上での審査を前提としている。

	Eurostat	イギリス国家統計局	ドイツ連邦統計局	ドイツ連邦雇用庁	オランダ統計局
審査の仕方	<ul style="list-style-type: none"> ● 経験則と原則ルールを活用 ● 経験則での対応がメインだが、プロセスは明確には、分けず 	<ul style="list-style-type: none"> ● 経験則と原則ルールを活用 ● 経験則での対応がメインだが、プロセスは明確には、分けず 	<ul style="list-style-type: none"> ● 経験則と原則ルールを活用 ● 経験則は緩く、必要に応じ、原則ルールで確認 	<ul style="list-style-type: none"> ● 経験則のみ ● ガイドラインで提示した処理を「粛々と」実施し、例外は認めず 	<ul style="list-style-type: none"> ● 経験則と原則ルールを活用 ● 経験則での対応がメインだが、プロセスは明確には、分けず
審査体制	<ul style="list-style-type: none"> ● データマネージャが、利用～審査を一貫して実施 ● 利用者と一緒に秘匿状況を確認し、更に、自身で審査を実施 ● 必要に応じて、結果の調整も実施 	<ul style="list-style-type: none"> ● 審査担当者を専任で設置(高いスキルは不要) ● 1つの審査に対し、2名の担当者が独立で審査 ● 審査者と結果の利用者への送付者を分け、内部者攻撃への対処を実施 	<ul style="list-style-type: none"> ● 審査担当者を専任で設置 	<ul style="list-style-type: none"> ● 秘匿の処理と審査の両方を実施を、持ち回りで実施 	<ul style="list-style-type: none"> ● 専任の秘匿処理専門家と、兼務のデータ専門家で実施 ● 1つの審査に対し、上述の担当者が独立で審査
責任	<ul style="list-style-type: none"> ● 利用者、機関の両方 	<ul style="list-style-type: none"> ● 利用者、機関の両方 	<ul style="list-style-type: none"> ● 利用者、機関の両方 	<ul style="list-style-type: none"> ● ー 	<ul style="list-style-type: none"> ● 利用者、機関の両方

各国／機関のオンサイト利用の運用(秘匿審査)

- 審査者の育成は、OJTをベースとする国／機関が多い。なお、オランダは自身の経験をライブラリとしてまとめ、暗黙知を形式知化することで、職員のレベルの底上げを実施している。

	Eurostat	イギリス国家統計局	ドイツ連邦統計局	ドイツ連邦雇用庁	オランダ統計局
審査者の素養	<ul style="list-style-type: none"> ● 経済学出身で、統計を使用してきた経験を保有 	<ul style="list-style-type: none"> ● 数学、統計学、経済学、社会科学等を大学、大学院で学んできた人材が多い 	<ul style="list-style-type: none"> ● 社会科学出身の人材が多い 	<ul style="list-style-type: none"> ● 経済学出身の人材をはじめ、出自は様々 ● そのまま研究を継続し、FDZで自身で労働市場等の研究を実施 	<ul style="list-style-type: none"> ● CBS自体が国内で情報の開示に最も挑戦的な存在であり、職員は其中でスキルを研鑽
スキル底上げの手法	<ul style="list-style-type: none"> ● 審査に閉じたスキル底上げは実施していない ● もともと、データセットを管理し、そのデータに最も精通している担当者が、審査を兼務 	<ul style="list-style-type: none"> ● OJTが基本 ● 新人はスーパーバイザの指導のもとにOJTで育成され、1年も経たない内に、独自に審査可能と認定 	<ul style="list-style-type: none"> ● 審査者にはOJTの他、内部に保有する文書類を提供(Off-JT)し、知見を拡充 	<p>—</p> <p>審査だけに閉じたスキル底上げは実施していない</p>	<ul style="list-style-type: none"> ● 審査結果は、ライブラリー化し、ノウハウを共有
<ul style="list-style-type: none"> ● EurostatはEU向けに審査者育成のための教育を実施している 					

各国／機関のオンサイト利用の運用(秘匿審査)

- 閾値、占有性ルール、一次秘匿には、具体的な数字を定めていることが多い。多くの機関が閾値を10とする中、ドイツは雇用庁が20、統計局が3と、極端な取組みをしていた。

	Eurostat※	イギリス国家統計局	ドイツ連邦統計局	ドイツ連邦雇用庁	オランダ統計局
経験則の 体系	<p>【閾値】</p> <ul style="list-style-type: none"> ● 度数表の構成要素は10以上 <p>【占有性ルール】</p> <ul style="list-style-type: none"> ● 1要素の占める割合は70%以下 ● 大きい2つの要素の占める割合は85%以下 <p>【統計記述量の扱い】</p> <ul style="list-style-type: none"> ● 最大、最小、中間値の秘匿 <p>【二次秘匿】</p> <ul style="list-style-type: none"> ● 他のデータとの照合も考慮されなければならないと規定 	<p>【閾値】</p> <ul style="list-style-type: none"> ● 度数表の構成要素数は10以上 <p>【占有性ルール】</p> <ul style="list-style-type: none"> ● 1要素の占める割合は43%未満 ● 大きい2つの要素を除いた合計は、最大の要素の10%以上 <p>【統計記述量の扱い】</p> <ul style="list-style-type: none"> ● 最大、最小値の秘匿 <p>【二次秘匿】</p> <ul style="list-style-type: none"> ● 適切さを利用者が確認 	<p>【閾値】</p> <ul style="list-style-type: none"> ● 度数表の構成要素数は3以上 <p>【統計記述量の扱い】</p> <ul style="list-style-type: none"> ● 最大、最小値の秘匿 ただし、パーセンテージ表記の場合、構成要素数が300以上あれば、開示可能 <p>【二次秘匿】</p> <ul style="list-style-type: none"> ● 他のデータとの照合を避けることを要請 	<p>【閾値】</p> <ul style="list-style-type: none"> ● 度数表の構成要素数は20以上パーセンテージで表示する際も、その構成要素が20以上あること(1%を開示する場合、元のデータは2,000は必要) <p>【最小度数】</p> <ul style="list-style-type: none"> ● 重み付けをしたデータを開示する場合、重み付けをしていない生成物で、条件を満たすかを確認 	<p>【閾値】</p> <ul style="list-style-type: none"> ● 度数表の構成要素数は10以上 ※重み付けがない状態 <p>【占有性ルール】</p> <ul style="list-style-type: none"> ● 度数表で1要素の占める割合は50%未満 <p>【統計記述量の扱い】</p> <ul style="list-style-type: none"> ● 最大、最小値の秘匿 <p>【グループ開示】</p> <ul style="list-style-type: none"> ● 数量表のセルが行・列の総和の90%未満

※) Eurostatの経験則は、1つの統計(CIS)で用いられているものより転記している。

※) ドイツ連邦雇用庁のデータセットは規模が大きく、1つのデータセットで、8,000万人分のデータセットが存在するものもある。仮に数%のサンプルであっても、万単位のデータセットが存在し、その中で、閾値を20にセットしている点では、一概に「厳しい」条件をセットしているとも言い難い。

各国／機関のオンサイト利用の運用(秘匿審査)

- 秘匿審査のベースの考え方は経験則、またそもそもの利用者との信頼関係にあり、高度な問題に対して、ツールや厳密な統計的手法を用いての解決を試みる姿勢は、確認されなかった。

	Eurostat	イギリス国家統計局	ドイツ連邦統計局	ドイツ連邦雇用庁	オランダ統計局
占有性ルール	— (経験則以上のことはせず)	● 時折審査をしている模様(具体的な取組みは不明)	● 時折原則ベース(p%ルール)で審査を実施	— (経験則以上のことはせず)	● (n,k)ルールの採用が基本 ※なお、統計局自身はp%ルールを採用
二次秘匿	● 生成物のサイズは小さいため、特段検証はせず	● たいていの場合、リコーディングを推奨(問題の質を簡易化)	● マニュアルで確認を実施	— (経験則以上のことはせず)	● 研究者に基本的に委ね、また、審査者の「経験」に委ねる状態
差分攻撃	● 低リスクと判断し、特段対処は実施しておらず	● 低リスクと判断し、原則ベースの審査はせず	● レビューは困難、利用者を信頼しているのが現状	— (経験則以上のことはせず)	● 特にリモートを注力しているため、問題意識はあるが、対処はせず
最小度数の緩和	● 全て同じレベルで対処	● 全て同じレベルで対処	● 全て同じレベルで対処、経験則の基準は元々緩め	— (経験則以上のことはせず)	● 配慮をしている模様(具体的な取組みは不明)
準識別子の決定	● 全て同じレベルで対処	● ミクロデータとしては全て同じで、差異は設けず	● 全て同じレベルで対処、経験則の基準は元々緩め	— (経験則以上のことはせず)	● 配慮をしている模様(具体的な取組みは不明)

各国／機関のオンサイト利用の運用(教育プログラム)

- 教育プログラムの実施は、国／機関によって分かれた。ただし、実施している国／機関も、ガイダンスレベルのコンテンツで教育し、具体的なポイントは利用中にコミュニケーションを通して伝えるスタンスであった。

	Eurostat	イギリス国家統計局	ドイツ連邦統計局	ドイツ連邦雇用庁	オランダ統計局
プログラムの有無	あり	あり	なし	なし	なし
プログラム概要	<ul style="list-style-type: none"> ● 事前学習資料・理解度チェック資料の提供 ● 利用初日の講義 ● 年1回のオープンな公開講座 	<ul style="list-style-type: none"> ● 「承認された研究者」になるために教育プログラムの受講と試験合格が必要 ● コンテンツはEurostatに類似 	<ul style="list-style-type: none"> ● 利用者が利用中に、機関職員と密にコンタクトを取る形でカバーしている 	<ul style="list-style-type: none"> ● そもそも秘匿処理を利用者にさせていない 	<ul style="list-style-type: none"> ● ブローシャを渡す程度 ● 利用者が利用中に、機関職員と密にコンタクトを取る形でカバーしている
プログラムの特徴	<ul style="list-style-type: none"> ● 利用初日にデータマネージャが、概要を説明 ● 事前学習の徹底は求めておらず、利用者の理解度も最初はバラバラ 	<ul style="list-style-type: none"> ● 試験を実施し、二度落ちれば、永久的に利用を認めない ● 試験自体は、真面目に講習を受ければパスできる程度 	<ul style="list-style-type: none"> ● — 	<ul style="list-style-type: none"> ● — 	<ul style="list-style-type: none"> ● —

参考 | Eurostatの利用初日講義資料(抜粋)

- ユーザーガイドに掲載されているレベルの内容が、パワーポイントにまとめられた程度のものである。

オンサイト利用の位置づけ

1. CIS: Overview and data structure



1.1. Overview

- CIS: Survey on innovation activity in enterprises

Information on:

Different types of innovation: Product/ Process/ Organisation/ Marketing

Development of an innovation: Objectives/ Sources of information/Public funding/ Expenditures, etc.

- **Business statistics**

4

オンサイト利用での規則

CIS Confidentiality rules




- Eurostat Safe centre = secure room. It's **NOT** possible to
 - print documents;
 - copy data to CD-ROM, DVD, Zip drives or USB;
 - copy data to the local hard disk;
 - connect recording devices to the serial, parallel and USB ports;
 - use and connect a laptop;
 - use email and make Internet connections.

14

秘匿審査に係る規則


CIS Confidentiality rules



- **Primary confidentiality**: tabular cell data, whose dissemination would permit attribute disclosure.
 - ➔ Too few enterprises in a cell;
 - ➔ Dominance of one or two enterprises in a cell.
- **Secondary confidentiality**: data which is not primary disclosure, but whose dissemination, when combined with other data permits the identification of an enterprise or the disclosure of an attribute of the enterprise

17

CIS Confidentiality rules



Statistics not allowed to be published:

- (1) if they consist of less than 10 enterprises;
- (2) where one enterprise represents more than 70% of the total sub-population expenditures, employment or turnover;
- (3) where two enterprises represent more than 85% of the total sub-population expenditures, employment or turnover

18

3. 我が国オンサイト利用への示唆・今後の課題抽出

諸外国の事例に基づく示唆・今後の課題抽出の考え方

- 我が国は既にオンサイト利用のトライアルを開始し、連続的に本格展開に移行することとなることに鑑み、短中期、中長期の以下に関する観点で、諸外国事例を踏まえた示唆を検討する。

	短中期的に見たポイント	中長期的に見たポイント
オンサイト利用の位置づけ	オンサイト利用をどのようなサービスと位置付けるか？	中長期的に見たマイクロデータ利用のサービスポートフォリオは？
オンサイト利用の利用者	利用者の確保・増加に資する有用な施策は？	(同左)
オンサイト利用の設備環境	セキュリティ・アクセスコントロールをはじめどのように行うべきか？	利用者が増えた中で、どのような拠点展開をしているべきか？
サービスプロセス	オンサイト利用の肝を何に設定するか？	(同左)
秘匿審査	短中期的に見てどのような審査が有用か？ 審査者育成で有用な施策は？	中長期的に見てどのような審査が有用か？
教育プログラム	利用者にどのような教育を実施するか？	(同左)

諸外国の事例から見た我が国オンサイト利用への示唆・今後の課題

■ 諸外国の取組みのポイントを踏まえ、今後我が国オンサイト利用において検討が必要と考えられる5つのテーマを抽出する(詳細は、別紙A3資料を参照)。以下、①～④に関して、詳細を述べる。

オンサイト利用の位置づけ

- 利用者とサービス実施者が、利用中に密なコミュニケーションを取るプロセス設計、規約、体制等の設計を検討する。
- オンサイト利用とリモートアクセスの双方の提供に向けた検討をする(中長期視点)。

1

オンサイト利用の利用者

- 海外の利用に関する方策を検討する。
- 学術研究者以外に、学生等の利用に関する方策の検討する。

2

オンサイト利用の設備環境

- アクセスコントロールの付与を含め、セキュリティ対策のあり方を検討する。
- 統計局内も含めた拠点展開のあり方を検討する(中長期視点)。

3

サービスプロセス

- 利用者とサービス実施者が、利用中に密なコミュニケーションを取るプロセス設計、規約、体制等の設計を検討する(上述と同様)。

秘匿審査

- 経験則、原則ルールの2段階の審査を行うプロセス、審査基準と体制の再考する。
- 高度な技術的問題に対する対処方法の位置づけ、それに伴うツール導入等を再考する。

4

教育プログラム

- オンサイト利用のサービスプロセスの再考に伴う、教育プログラムの再考を行う。
- 利用者増加、および、オンサイト利用の認知度を向上させるための、公開講座などの設置を検討する。

5

※) 既に我が国において課題意識が醸成されていると考えられるテーマは、示唆としてここでは抽出していない。

示唆・課題① | オンサイト利用の位置づけ（1）

- 諸外国／機関の実態調査より、オンサイト利用の価値は、利用者とサービス提供者が信頼関係のもとで、密にコミュニケーションを取り、生成物を作成することにあることが分かった。
- この示唆を我が国に取り入れるとした場合、プロセス設計およびプロセスを支える体制に関する検討が、今後の課題になると考える。

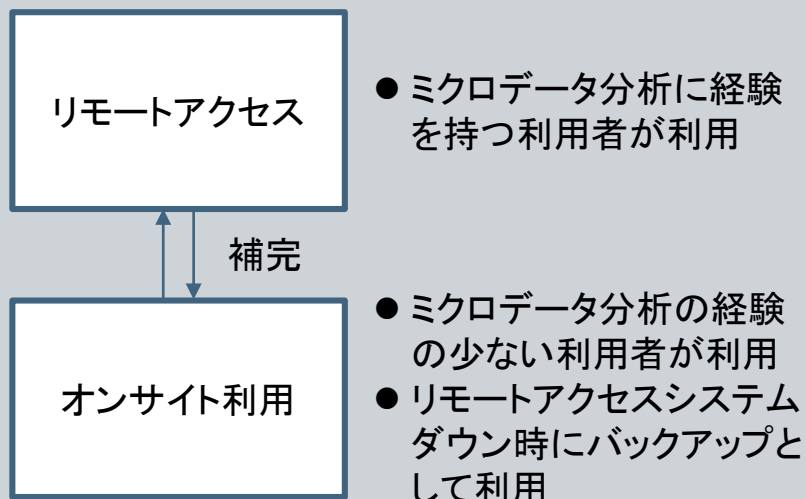
項目	説明			
望ましいプロセス	<div style="border: 1px solid black; padding: 5px; text-align: center;">申請</div> <ul style="list-style-type: none"> ● 利用者の素養や目的を厳しく審査 ● 信頼に足る対象を早期に絞る 	<div style="border: 1px solid black; padding: 5px; text-align: center;">利用</div> <ul style="list-style-type: none"> ● 利用中は密にコミュニケーション ● 利用者と共に生成物を作り上げる 	<div style="border: 1px solid black; padding: 5px; text-align: center;">秘匿審査</div> <ul style="list-style-type: none"> ● 利用時に生成物を作りこむことで、効率的な審査を実施 	<div style="border: 1px solid black; padding: 5px; text-align: center;">公表等</div> <ul style="list-style-type: none"> ● 公表結果はサービス提供者側にも共有
プロセスを支えるために必要な体制等の課題	<p>＜人的資源＞</p> <ul style="list-style-type: none"> ● オンサイト利用時に利用者の相談に乗るミッションを持った組織・人材の設置の検討が必要。 <ul style="list-style-type: none"> ● Eurostatのようなデータマネージャーを設定するか、ドイツ雇用庁のような研究者による対応とするか等を検討していくことが必要。 <p>＜物的資源＞</p> <ul style="list-style-type: none"> ● 短期的には、我が国ではサテライトオフィスでのオンサイト利用が中心となる。利用時のコミュニケーションを増やすための、システム面での工夫(例えば、リモートで同一画面を利用者・サービス提供者が共有できる 等)の検討が必要。 			

示唆・課題① | オンサイト利用の位置づけ（2）

- 諸外国／機関の実態調査より、秘匿度の弱いマイクロデータの利活用においては、オンサイト利用と共にリモートアクセスを提供し、利用者の利便性を向上させることが有効であることが分かった。
- この示唆を我が国に取り入れるとした場合、リモートアクセスが我が国において実施可能であることを確認した上で、双方の役割分担を明確にした上でのサービス提供の検討が、今後の課題になると考える。

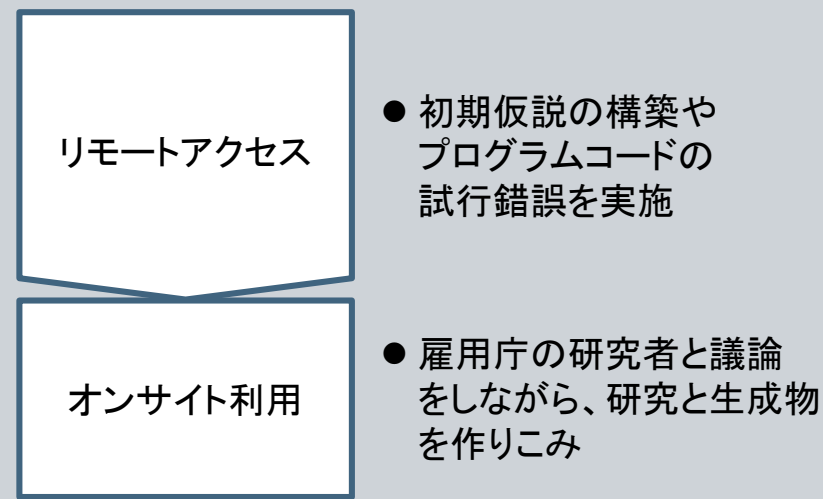
オンサイト利用とリモートアクセスの棲み分け例①

- オランダ統計局の考え方を採用
- ターゲットとする利用者、およびシステム面でオンサイト利用とリモートアクセスの間に補完関係を構築する



オンサイト利用とリモートアクセスの棲み分け例②

- ドイツ連邦雇用庁の考え方を採用
- リモートアクセスを、初期仮説構築のためのツールとして提供し、オンサイト利用で本格的な研究を行う前の準備をするツールとして提供する※



※) ドイツ連邦雇用庁のリモートアクセスの使い方は、上述以外にも存在するが、上述の利用傾向は、実績として高い

示唆・課題② | オンサイト利用の利用者

- 諸外国／機関の実態調査より、オンサイト利用の認知向上や、利用者の確保・増加に向けて、1)海外の利用者を認めたり、2)研究者の「徒弟」に対する利用を認めたりしていることが明らかとなった。
- この示唆を我が国に取り入れるとした場合、どのように利用規約や申請書、申請プロセスを策定していくかを検討することが、今後の課題になると考える。

項目	説明
海外からの利用の許可	<ul style="list-style-type: none">● 利用規約や申請書類の英語版の用意、海外向けウェブサイトの用意等の検討が必要 (諸外国からのコメント)<ul style="list-style-type: none">● 『EurostatのCISの利用者は、EUに留まらない。オーストラリアや韓国の研究者も、「認められた主体」に登録し、利用している。』 (By Eurostat)● 『ドイツ連邦統計局は、利用規約やウェブサイトは英語も用意しているが、肝心の契約書はドイツ語のみしかない。自分が海外の人間だったら、まず利用しようとは思わない。』 (By ドイツ連邦統計局)
研究者以外の利用者 (例えば「徒弟」) の利用許可	<ul style="list-style-type: none">● 研究者以外の利用の範囲等に関するルール等を利用規約で整備・明確化する検討が必要。 (諸外国からのコメント)<ul style="list-style-type: none">● 『承認された研究者の枠組みだけでは、オンサイト利用を活用できる機会が少ない。よって、これを緩和する措置として、暫定の研究者という枠組みを用意している。』 (By イギリス国家統計局)● 『Ph.Dには、通常の半額でオンサイト利用を利用させている。彼らにオンサイト利用のコミュニティに入ってもらい、将来、本格的に利用してもらうことを重視している。』 (By ドイツ連邦統計局)

示唆・課題③ | オンサイト利用の設備環境

- 諸外国／機関の実態調査より、1)どの国も機関の本拠にオンサイト利用施設を設置していること、2)そこで利用者が職員と密に会話をしていること、3)研究者の利便性に鑑み、各種ツールを用意していること、4)国／機関によっては、セキュリティ対策を厳格に実施していることが明らかとなった。
- 中長期的には、本拠へのオンサイト利用の設置も考慮するとしたうえで、当面、サテライトオフィスでの運用が想定される我が国において、この示唆を取り入れるとした場合、ツールやプロセスの設定において、より深い検討することが、今後の課題になると考える。

項目	説明
利用者とのコミュニケーションの促進	<ul style="list-style-type: none">● 利用時のコミュニケーションを増やすための、システム面での工夫(例えば、リモートで同一画面を利用者・サービス提供者が共有できる 等)の検討。 ※ イギリス国家統計局は、利用者が施設利用後も電話等でコミュニケーションを取ることがあるとのことであった。
利用環境の向上	<ul style="list-style-type: none">● ソフトウェアベンダとの端末ベースでのライセンス契約の獲得による、多種多様なソフトウェア環境の構築の検討。 ※ Eurostatは、このスキームで、STATA、SPSS、SASなどのツールを導入しているとのことであった。
厳格なセキュリティ	<ul style="list-style-type: none">● ID、パスワードによる入退館管理、システム認証は勿論のこと、ファイルのどの部分までアクセスを可能とするか等のアクセス権限、システム作動のログ取得による各種攻撃の回避の検討。 ※ イギリス国家統計局は、上述の方策を既に運用に落としこんでいるとのことであった。

示唆・課題④ | 秘匿審査

- 諸外国／機関の実態調査より、生成物の秘匿審査においては、1)多くのケースで中間生成物の持ち出しは認めていないこと、2)経験則と原則ルールを切り分けた審査プロセスを敷いてはいないこと、3)技術的に高度な問題はある程度リスクの存在を認め、審査をしていること、が分かった。
- この示唆を我が国に取り入れるとした場合、秘匿審査に係るプロセス設計およびプロセスを支える体制・ツールに関して、下の図表に挙げるような再検討を行うことが、今後の課題になると考える。

項目	現状想定しているモデル	Eurostatモデル	ドイツ連邦雇用庁モデル
概要	<ul style="list-style-type: none"> ● 経験則と原則ルールの二段階審査 ● 利用者の資質や技術的な問題に対し、ツールの支援を受けながら、積極的に原則ルールを適用 	<ul style="list-style-type: none"> ● 経験則と原則ルールを同時に実施する一段階審査 ● 公表に資する最終生成物を利用者とサービス提供者が一緒に生成、利用段階で審査にほぼ耐えられる生成物を作りこむ 	<ul style="list-style-type: none"> ● 経験則のみでの審査 ● 審査は公表するに資する最終生成物に対して、厳格なルールを適用、ルールが厳格なため、特にツールは必要とせず
適用シーン仮説	<ul style="list-style-type: none"> ● ー(諸外国／機関では該当する事例は確認されず) 	<ul style="list-style-type: none"> ● 利用者数が少ない段階で質の高いサービスを提供する際に有効か 	<ul style="list-style-type: none"> ● 利用者数が多い場合または、審査の経験が少ない場合に有効か
適用に向けた課題	<ul style="list-style-type: none"> ● 諸外国で実現してはいないモデルを遂行する、実現可能性の担保 ● 技能やツールを含めた技術に係る実現可能性の担保 	<ul style="list-style-type: none"> ● 利用～審査までを一貫して対処するサービス実施者の体制・組織の確保 	<ul style="list-style-type: none"> ● 利用者の満足度や利用に資する生成物の作成を実現するための、サービス実施者の体制・組織の確保および、経験則の作りこみ

※) 上述の選択肢はあくまで例であり、今回調査した諸外国／機関以外の国／機関の取組みも、有用であれば取り入れていくべきである。

— 終 —