

公的統計分野の データサイエンス力 強化支援の取組み

独立行政法人統計センター
椿 広計

2018/09/12 統計関連学会連合大会
特別セッション「データサイエンスへの展望」

内容：総務省統計コミュニティの取組み

- はじめに：問題意識
- データ提供の取組み
 - オープンデータの取組み：マクロデータのユーザー利活用環境の整備
 - 統計教育標準データ提供とコンペティション
 - データサイエンス駆動型研究のためのマイクロデータと分析環境提供
- 教育環境提供の取組み
 - 一般市民のデータサイエンス力育成
 - 行政データサイエンティストの育成
- データサイエンスの公的統計実務研究と実務展開
- おわりに

はじめに：問題意識と公的統計改革

- コトづくり限界費用 $\cong 0$ の時代
 - データとアナリティックスの価格崩壊
 - コトづくり経済価値の増大
- コトづくりのもたらず産業と社会の革命
 - 経済価値・知識価値・感情価値の高いプログラム創生
 - コトづくり生産手段の社会共有
 - コトとモノの共有
- コト生産技術としてのデータサイエンス定義の再確認
 - 情報社会設計科学
 - ⇒ 情報に基づくオペレーション(シグナル性プログラム)の設計科学
- 「公的統計改革」⇒ 統計法・統計センター法改正
- 公的統計分野の支援：教育・研究用データ・環境＋教材

統計改革推進会議最終取りまとめの 全体構成（イメージ） [内閣官房2017年5月25日]

以下2スライド：内閣
官房最終とりまとめ
参考資料より抜粋

[http://www.kantei.g
o.jp/jp/singi/toukeik
aikaku/pdf/saishu_s
ankou.pdf](http://www.kantei.go.jp/jp/singi/toukeik aikaku/pdf/saishu_s ankou.pdf)

政策・統計の改善

EBPMプロセスを通じた
経済統計の改善

1. EBPM（証拠に基づく政策立案） 推進体制の構築

- (1) 基本的な考え方
- (2) 推進の要の整備
- (3) 政策、施策、事務事業の各段階における取組



2. GDP統計を軸にした経済統計の 改善

- (1) GDP統計の体系的整備の全体像
- (2) より正確な景気判断に資する基礎統計改善、GDP統計の加工・推計手法改善に向けた取組
- (3) 生産面を中心に見直したGDP統計への整備

経済構造の正確な把握
によるEBPMの促進

利活用促進

環境・基盤の整備

3. ユーザーの視点に立った統計シス テムの再構築と利活用促進

- (1) 各種データを用いた統計的分析の推進
- (2) 社会全体における統計等データの利活用の促進

リソース確保

4. 報告者負担の軽減と統計業務・ 統計行政体制の見直し・業務効 率化、基盤強化

- (1) 報告者負担の軽減
 - (2) 統計業務の見直し・業務効率化及び各種統計の改善
 - (3) 統計行政体制の見直し
- 統計(4) 統計改革の推進の基盤強化

ユーザーの視点に立った統計システムの再構築と利活用促進

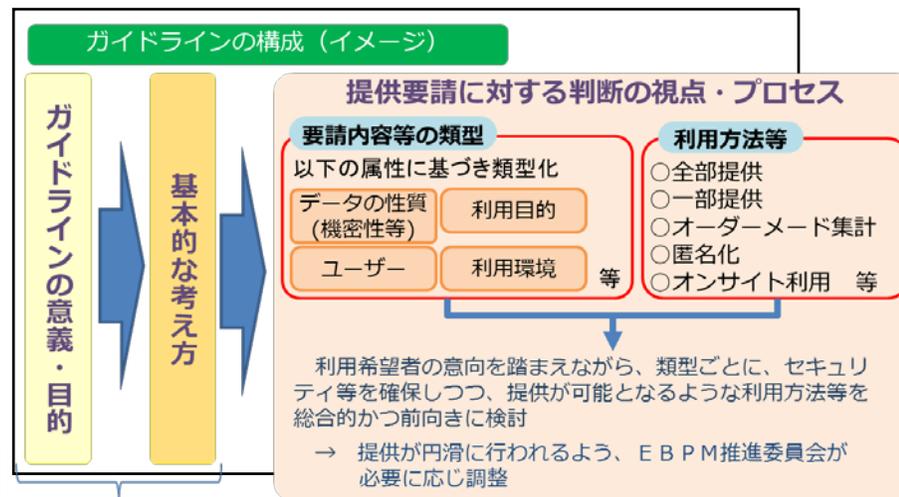
内閣官房
2017/05/25

各種データの利活用推進のための統計関係法制等の見直し・整備

- 利活用が限定的な統計の個票データや、利活用規定の不十分な行政記録情報や地方自治体・民間が保有する各種データについて、**セキュリティを確保しつつ、利活用を促進**するための**統計関係法制等の見直し・整備**

【具体的な取組（例）】

- 地方自治体・民間が保有する各種データへの各府省による**提供要請**や提供された**データの保護**、統計委員会を通じた各府省と地方自治体・民間の**あっせん**等の仕組の整備
- 各府省が、未提供のデータ等の提供要請を受けた場合に、データの性質や利用目的等の類型に応じ、**提供の判断を適切に行うためのガイドライン**の策定



最終取りまとめで骨格を明示

→ EBPM推進委員会において、統計委員会の意見を聴取しつつ全体を具体化

社会全体における統計等データの利活用の促進

- EBPM推進委員会が、統計等データの**ユーザーからの提案募集**を実施 → EBPM推進委員会において、統計委員会の意見を聴取しつつ全体を具体化
- 各府省のEBPM推進統括官の下、外部からの統計等データの**問合せや要望への対応**のための体制を整備
- 統計等データの利活用のための**基盤の整備**

【具体的な取組（例）】

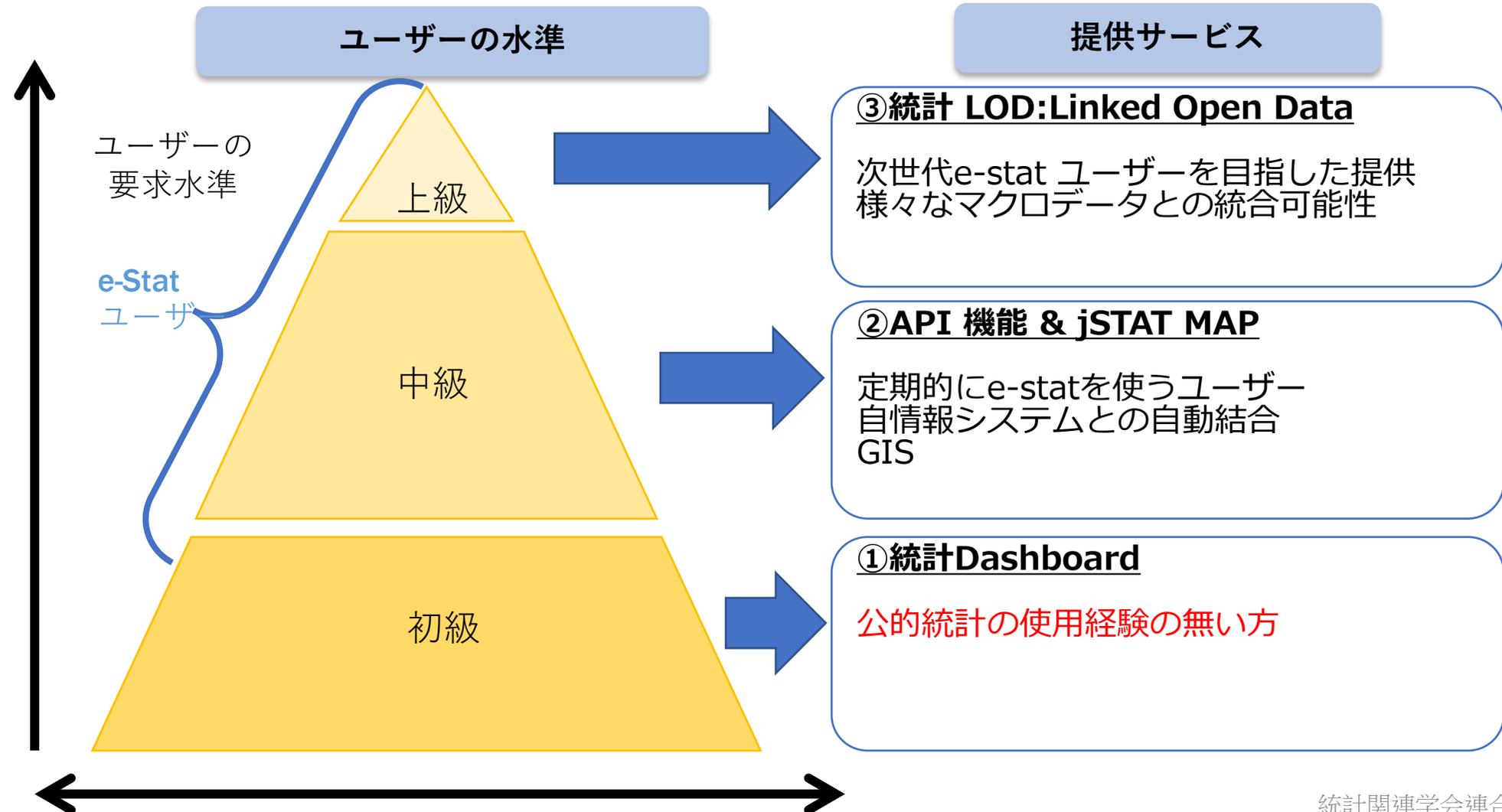
- **e-Stat（政府統計の総合窓口）**への行政記録情報の検索機能追加など、抜本的な機能強化
- **オンサイト施設**（p.12参照）の整備の推進
- 一般の人が利用できる**匿名データ（匿名化した個票データ）**の提供
- **行政記録情報**の標準化・電子化

統計関連学会連合大会

データ提供の取組み

オープンデータの取組み：マクロデータのユーザー利活用環境の整備
統計教育標準データ提供とコンペティション
データサイエンス駆動型研究のためのデータと分析環境提供

マクロデータのユーザー利活用環境の整備



統計 Dashboard : 統計データの見える化



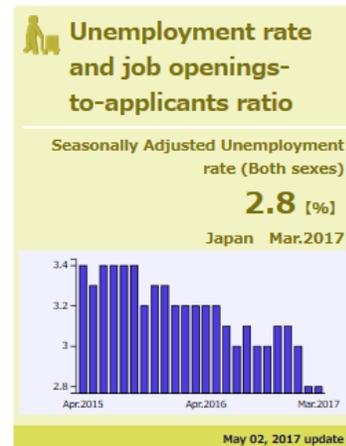
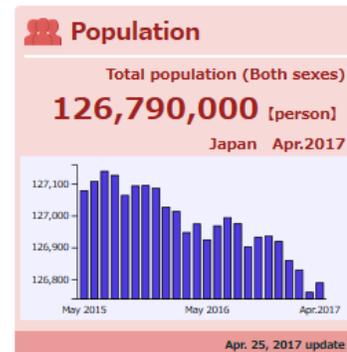
統計 Dashboard

多様な統計情報を初心者にもわかりやすいように図表化して要約提供するシステム (2017~)

Example:
2017/03の
英文トップページ



キッズ・スタット
小中学生向け
統計データ検索サイトも
整備(2018~)



統計関連学会連合大会

統計におけるオープンデータの高度化

独立行政法人
統計センター

統計を
つくる・
活かす・
支える

政府中央統計機関
の一翼を担う
ビッグデータ時代の
支援活動
「統計を活かす」
マイクロデータの
研究利活用

オンサイト拠点
試行事業開始
2017/01

統計データ利活用
センター2018/04
和歌山市に設置

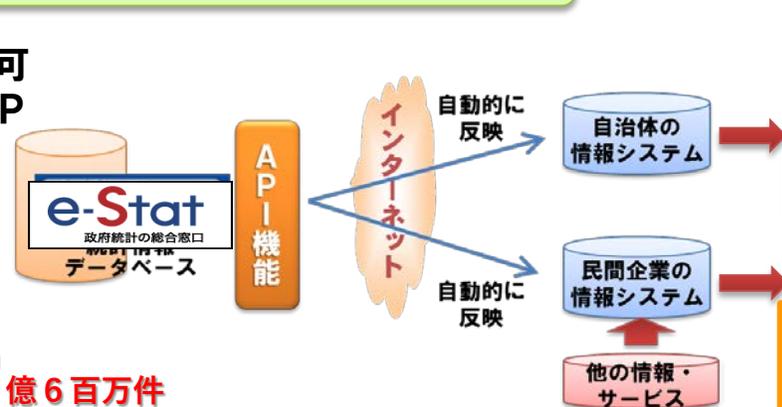
2018/9/11

統計データの提供方法を高度化し、新たな付加価値を創造するサービスや革新的な事業の創出などを支援する取り組みを、総務省統計局と連携し行っており、政府が取り組んでいるオープンデータの推進を先導。

API機能による統計データの提供

2014.10.31から運用開始

統計データを機械判読可能な形式で提供するAPI機能（Application Programming Interface）を提供中



利用登録者数約1万3千名
APIリクエスト件数：約1億6百万件
2018年3月31日現在

活用例1：利用者の情報システムにe-Statのデータを自動的に反映

活用例2：ユーザー保有やインターネット上のデータ等と連動させた高度な統計データ分析



地図で見る統計 (jSTAT MAP)

2015.1.20から運用開始

任意に指定したエリアによる集計や利用者が保有するデータの取り込み集計する機能などを提供



利用登録者数 約4万5千名
ログイン件数：約76万件
2018年3月31日現在

活用例1：任意に指定したエリアによる集計や、利用者が保有するデータと統計データを組み合わせ、集計結果を地図上で視覚的に把握可能



活用例2：選択したエリアの年齢構成等の基本的な分析結果のレポート作成

統計関連学会連合大会

LOD - TOP Levels of Open Data -

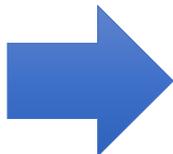
統計Linked Open Data : リンクされた項目数世界1

From link to files to link to data

Link to files

Link to data

Addresses が各ファイルに付与
(<http://www.e-stat.go.jp/xls/0001.xls>)



	Total 総数 (男女別)		Male 男		Female 女	
	44歳【人】	45歳【人】	44歳【人】	45歳【人】	44歳【人】	45歳【人】
さいたま市	16,130	19,245	8,293	9,938		
川口市	6,582	8,022	3,526	4,289		

Addressが各データ項目に付与
(<http://data.e-stat.go.jp/lod/.../obs00001>)



2018/06/27 統計教育用標準データの提供とコンペ Standardized Statistical Data Set for Education

<https://www.nstac.go.jp/SSDSE/>

• 都道府県・市町村のすがた（社会・人口統計体系）

- 全市区町村（東京23区を含む）：1741自治体
 - 791市, 744町, 183村, 23区
 - 生徒誰もがどこかに所属。有名な市町村も多い。
- 111変量抽出：完備データセットに加工
 - 人口・世帯, 自然環境, 経済基盤, 行政基盤, 教育
 - 文化・スポーツ, 居住, 健康・医療, 福祉・社会保障
- 高校**探究科目**, 大学データ分析教育, **Good Practice 共有化**への活用
 - 文科省の日本食品標準成分データベース（2198×63）

• 教育用標準データの量的階層性

- 都道府県データ・教室内データ取得（50件規模）：小中学校
- 市町村データ, 上場企業データ（1500～2500件規模）：高校・大学
- 定型的ビッグデータ（数万件規模）：匿名マイクロデータ：大学院
- 画像・音声などのビッグデータ

標準データセットによる

Good Modelling Practicesの知の全国共有

• 統計データ分析コンペティション

- 総務省，独立行政法人統計センター，一般社団法人日本統計協会共催
- 後援（予定）：国立研究開発法人科学技術振興機構，
一般社団法人日本統計学会，全国統計教育研究協議会
- 高校生の部（高等学校，高等専門学校1-3年次）
- 大学並びに一般の部
 - 個人またはグループ，指導教員の指導可
- エントリー期間：8月10日締め切り：合計100件
- 論文提出締め切り：9月18日
 - 論文要旨＋論文（8ページ以内）を所定の様式で提出
- 受賞論文の決定・発表：10月18日（統計の日）
- 1月全国統計大会で表彰：総務大臣賞，優秀賞，日本統計協会賞
 - 優秀論文を雑誌「統計」に掲載することを検討中

和歌山県和歌山市2018/04/01

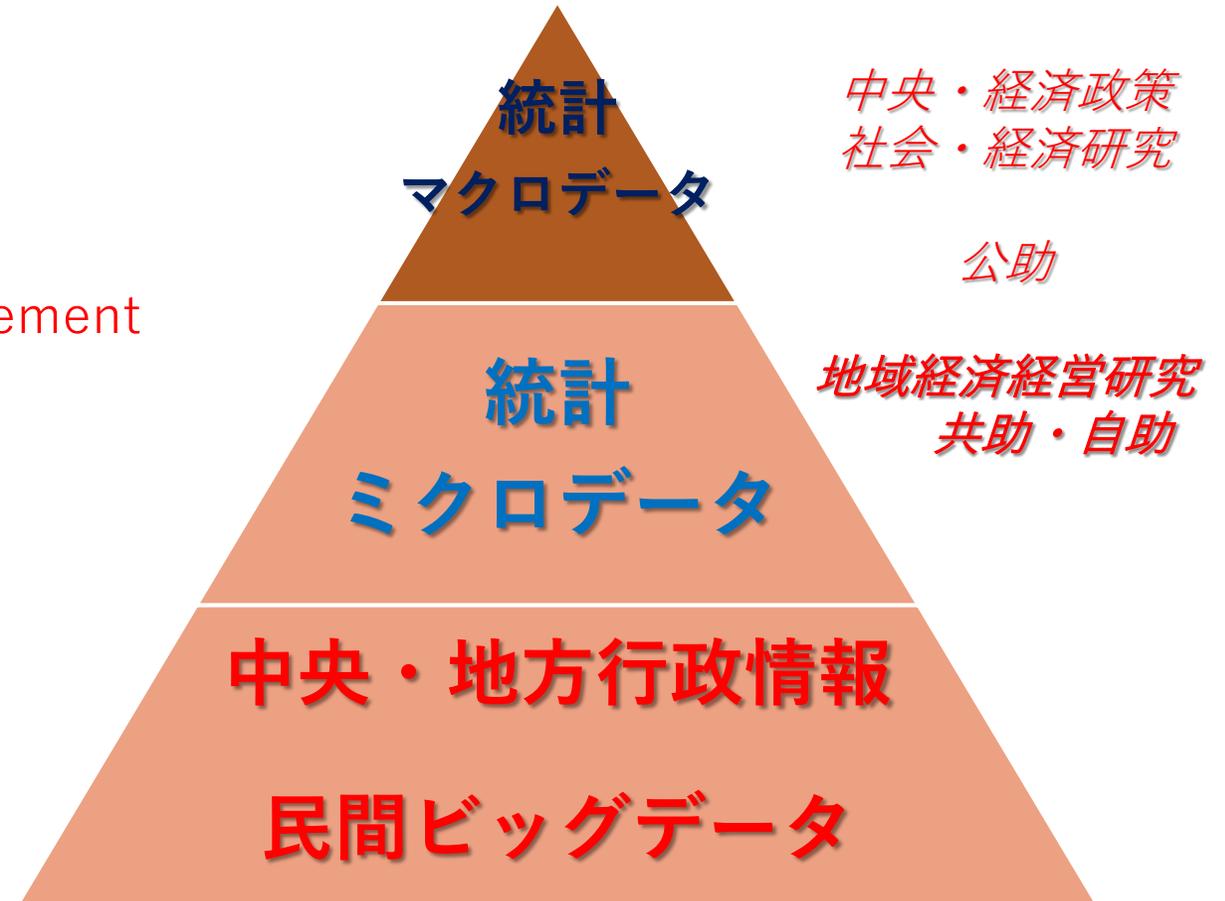
地方創生Good Practiceをデータで：和歌山モデル From Wakayama to the World

- データ利活用推進拠点の設置
 - 総務省統計局・統計センター「統計データ利活用センター」
 - 和歌山県「データ利活用推進センター」
 - 地域創生EBPM研究活動：和歌山大，和歌山医大，同志社大，東大などの協力
- 啓発活動：第2回 和歌山県データ利活用コンペティション参加者募集について
 - 和歌山県では全国の高校生、大学生等を対象に、様々な行政課題に対するデータを活用した解決アイデアを募集して、優秀な提案を表彰する「第2回和歌山県データ利活用コンペティション」を開催します。詳しくは、HPをご覧ください。
 - **【募集期間】** 2018年9月3日(月)～10月16日(火)
 - **【対 象】**
 - 高校生の部：高等学校、高等専門学校1年から3年まで、専修学校高等課程、特別支援学校高等部の在学学生
 - 大学生の部：大学、大学院、短期大学、高等専門学校4年・5年、専修学校専門課程・一般課程の在学学生
 - **【詳細】** 以下ウェブサイトをご覧ください。
<https://www.pref.wakayama.lg.jp/prefg/020100/compe.html>

データサイエンス駆動型研究のための マイクロデータと分析環境提供

• 相互に関連する目標

- 効果的・効率的な証拠に基づく
政策マネジメント(EBPM)の実現
 - Evidence Based Macro Policy Making
⇒ Evidence Based Micro Policy Management
 - トップダウン+ボトムアップ
- データ中心社会設計科学の確立
 - ミクロ社会の経営研究
- ビッグデータ時代の情報基盤確立
 - **統計データとのリンケージ**
 - 民間・社会ビッグデータの価値向上
 - 囲い込まれているデータの偏り補正
 - 質の高いデータ編成(Editing)技術の向上



公的統計マイクロデータ利活用 日本が実現したこと + 目指していること

提供データの分類軸

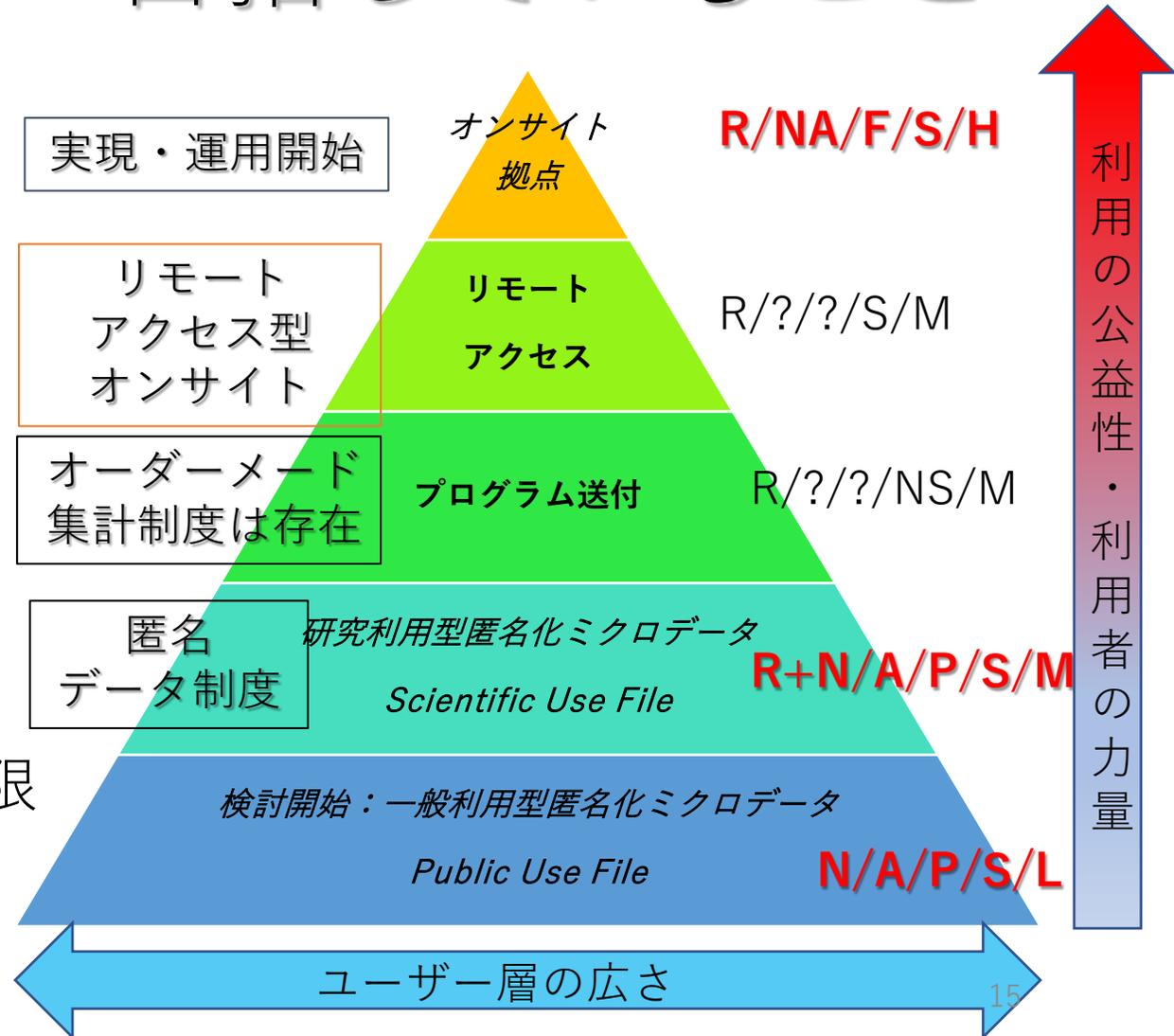
- R:実データ vs N:ノイズ付加データ
- A:匿名データ vs NA:非匿名データ
- F:完全データ vs P:一部データ

データを直接眺められるか？

- S, NS

提供サービスのAccessibility

- L:常時・誰でも・何処でも利用可能
- M:利用者・利用期間の制限
- H:利用者・利用期間・利用場所の制限
 - 分析結果の持ち出し審査



オンサイト施設におけるマイクロデータの活用

オンサイト施設

監視カメラ

シンクライアント端末

すべての調査項目を利用した探索的、
創造的な分析・研究を行うことが可能。

一橋大学	神戸大学	滋賀大学	多摩大学
群馬大学	新潟大学	統計センター	統計データ活用センター(和歌山)

(8月1日現在)

- ・シンクライアント端末
ユーザーが使用する端末の機能は必要最小限にとどめ、サーバー側で処理を行う仕組み

中央データ管理施設

仮想PC

仮想PCサーバ

調査票情報



SINETを活用した
専用線による接続

[インターネットに接続できない]

専用線による接続

統計データ活用センター(和歌山)

審査

登録

管理

等運用管理業務

霞が関オンサイト
設置に向けて

- ・SINET
全国の大学、研究機関等を含む
学術情報ネットワーク
(国立情報学研究所が構築、運用)

オンライン施設で利用可能な統計調査一覧 (平成30年8月1日現在)

調査名	年次, 月次	調査名	年次, 月次
国勢調査	平成22年	社会生活基本調査	昭和56年, 平成23, 28年
住宅・土地統計調査	平成25年	経済センサス-基礎調査	平成26年
家計調査	平成26年1月～平成27年12月分	経済センサス-活動調査 (総務省統計局所管分)	平成24年
労働力調査	平成26年1月～平成27年12月分	経済センサス-活動調査 (経済産業省所管分)	平成24, 28年
労働力調査特別調査	平成26年1月～平成27年12月分	工業統計調査	平成26年
就業構造基本調査	平成24年	商業統計調査	平成26年
個人企業経済調査(動向編)	平成26年1月～平成27年12月分	特定サービス産業実態調査	平成27年
個人企業経済調査(構造編)	平成26年1月～平成27年12月分	経済産業省生産動態統計調査	平成28年
サービス産業動向調査(拡大)	平成25年	経済産業省企業活動基本調査	平成28年1月～平成29年12月分
サービス産業動向調査(月次)	平成25年1月～平成27年12月分	商業動態統計調査	平成28年1月～平成29年12月分
家計消費状況調査	平成27年1月～平成27年12月分	経済産業省特定業種石油等消費統計調査	平成28年1月～平成29年12月分
科学技術研究調査	平成24, 27年	石油製品需給動態統計調査	平成28年1月～平成29年12月分
全国消費実態調査	平成26年	ガス事業生産動態統計調査	平成28年1月～平成29年12月分

黒字
 総務省
 統計調査
赤字
 経済産業省
 統計調査

統計関連学会連合大会

総務省の教育支援の取組み

一般市民のデータサイエンス力育成

- 統計局の取組み
 - 統計指導者研修事業(2002~)
 - 統計を学ぶサイト
 - なるほど統計学園(2010~)
 - なるほど統計学園高等部(2013~)
 - 統計力向上サイト
 - データサイエンス・スクール
 - オンライン講座提供(2014~)
 - 社会人のためのデータサイエンス入門
 - 誰でも使える統計オープンデータ
- 統括官室監修教材提供(日本統計協会)
 - 生徒のための統計活用～基礎編～
 - 同指導者用 (H28)
 - 高校からの統計・データサイエンス活用～上級編～
 - 同指導者用 (H29)

行政データサイエンティスト

- 総務省統計研究研修所
 - 公務員の統計教育がミッション
 - 1921：統計職員養成所として設立
 - 主として公的統計に関する研修
 - 統計入門課程：4課程（オンライン2課程）
 - 統計基本課程：7課程（本科研修を含む）
 - 統計思考力・データ分析力
 - 統計専門課程：9課程：統計実務活用力
 - 2018年度からの新たな取組み
 - 官庁データサイエンティスト育成
 - データサイエンス入門
 - ビッグデータ利活用
 - EBPM関連
 - 政策立案と統計，政策評価と統計
 - 2019年度概算要求による事業拡大？

データサイエンスの 公的統計実務研究と実務展開

標本理論に基づく集計

→モデルの活用 Since 2018/01

- 総務省統計局消費統計化課
- 消費動向指数開発・実装
 - CTIミクロ：家計消費指数
 - プロペンシティブスコア
 - 非標本抽出データとの融合
 - CTIマクロ：総消費動向指数
 - 状態空間モデル
 - 速報性のある指標
 - 2017/07 消費動向指数研究協議会

家計統計Codingのエキスパート投入

→自動化

- Supervised multiclass classifier for auto-coding based on partition coefficient
 - 統計センター統計技術研究課
 - 床上級研究員・佐藤研究担当理事
 - **KES-IDT-18:Best Paper Award**
 - Knowledge Based and Intelligent Engineering Systems, Intelligent Decision System

おわりに

統計関連学会への期待

データ利活用のGood Practice地域創生研究などへのご協力

オンサイト拠点の各県設置への立候補と積極利用

科研費縛りは無くなると予想，利用料問題の発生

統計データ標準データセットを活用した教材開発へのご支援

匿名化マイクロデータを活用した教育実践と教材開発へのご支援

関連コンペへの研究室などの学生の参加省令

統計研究研修事業の高度化への支援