

官参 1 - 45 官相 2 - 1
官総 6 - 4 官番 90
官人 1 - 74 課総 4 - 7
官会 1 - 58 課個 6 - 19
官企 1 - 18 課資 8 - 14
官厚 1 - 102 課法 7 - 11
官際 1 - 31 課酒 6 - 13
官広 1 - 26 課評 1 - 18
官察 14 課鑑 57
官整 3 課消 1 - 24
官督 20 課軽 4 - 2
官協 1 - 26 課審 1 - 13
官公 1 - 64 徴管 4 - 13
官税 1 - 21 徴徴 3 - 6
官情 1 - 11 査調 3 - 23
査察 1 - 20

令和 3 年 7 月 1 日

各 国 税 局 長
沖 縄 国 税 事 務 所 長
税 務 大 学 校 長
国 税 不 服 審 判 所 長
殿

国 税 庁 長 官
(官印省略)

データ活用推進第二次中期計画について（指示）

標題のことについては、別添のとおり「データ活用推進第二次中期計画」を定めたことから、これにより適切に対応されたい。

なお、平成 31 年 3 月 6 日付官企 1 - 8 ほか 20 課共同「データ活用の推進について」（指示）は廃止する。

令和3年7月1日

データ活用推進第二次中期計画

国税庁

1 これまでの取組

近年、ネットワークの高度化等によりデータの流通量は飛躍的に増大しており、また、AIなどの最新技術の進化・普及は著しい。このような中、課題解決や新たな価値創造のため、データをビジネス資源として有効に活用する取組が活発化してきている。

国税庁においても、「税務行政の将来像」（平成29年6月公表）を取りまとめ、「スマート税務行政」の実現に向けて計画的な取組を進めており、その一環として、調査・徴収のパフォーマンス向上や事務運営の改善のため、データを業務に活用する取組（データ活用）の一層の推進に取り組んでいる。

具体的には、平成31年3月6日付官企1-8ほか20課共同「データ活用の推進について」（指示）に基づき、平成30事務年度から令和8事務年度までの9年間で、

- 離陸期：平成30事務年度～令和2事務年度
- 普及期：令和3事務年度～令和5事務年度
- 発展期：令和6事務年度～令和8事務年度

の3期に分け、各期に対応した中期計画を策定し、取組を着実に推進していくこととした。

令和2事務年度までの離陸期においては、

- ① 庁局主務課等が、特に高度な分析・統計スキルを要さず、ITスキルと基本的な分析・統計スキルで実施が可能な、「業務効率化」、「個別事案の分析」、「事務運営の検討」を、おおむね独力で実施できるようになることを目指す
- ② 庁及び先行局のデータ活用チーム・システム部署が独力で、高度な分析・統計スキルを要する、「予測モデルの構築」について、簡単な予測型のデータ分析を実施できるようになることを目指す

として、着実に取り組んできたところである。

2 データ活用推進第二次中期計画の対象年度

データ活用推進第二次中期計画（以下「本計画」という。）は、令和3事務年度から令和5事務年度を対象とする。

なお、本計画の終了に当たっては、令和6事務年度から令和8事務年度を対象として「データ活用推進第三次中期計画」を策定することとする。

3 本計画におけるデータ活用推進の基本的な考え方

データ活用の取組においては、活用の基盤となる「システム」と、活用を担う「人材」が、「自動車」とその「運転手」とも言うべき関係であって、システム整備とともに人材育成に取り組んでいく必要がある。

データ活用推進の「自動車」に当たる「システム」については、令和8年度に向け、現行のKSKシステムの抜本的な刷新（次世代システムの構築）に取り組んでいるところ、次世代システムでは、「活用系システム」として、BIツール、BAツールといったデータ分析ツールを整備することとしており、これらツールの一部は、令和8年度の次世代システム本格稼働の前に、先行導入する予定である。

他方、「運転手」に当たる「人材」については、その育成の更なる充実を図っていく必要があると考えられる。

そこで、本計画では、これまでのデータ活用に関する取組を基礎として、それらの取組をより一層推進し、高度化させるため、取組を体系化するとともに、データ活用を担う人材の育成に焦点を当て、全庁的なデータ活用の推進をより一層図っていくこととしている。

(注) B I ツールとは、「Business Intelligence」ツールの略で、蓄積されたデータを統計的な手法を使って可視化するためのツールをいう。

B A ツールとは、「Business Analytics」ツールの略で、蓄積された大量データから統計分析や機械学習などの高度な分析手法を用いて、法則性を発見して、将来の予測を行うためのツールをいう。

4 本計画における「データ活用」の定義

「データ活用」は、多義的な用語であるが、本計画においては、「収集・加工した上で、データを分析し、利用する」という一連のプロセスとして、その内容を「収集・データ加工」と「データ分析」に分類する。

(1) 収集・データ加工

データ分析を実施する上での必須の前処理。

具体的には、外部データの収集、申告・資料情報等のデータ化、データ加工（データクレンジング、データマッチング等）が当てはまる。

(2) データ分析

データ分析については、多様な分類があるが、本計画では、便宜的に次のとおり、イ又はロの二つの区分に整理する。

なお、実施するデータ分析がいずれの区分に該当するか等について、硬直的に捉えることのないように留意する。

イ 基本的なデータ分析

データの特徴の把握等を目的とした従来の統計分析等をいい、主な例は以下のとおりである。

(イ) 平均値や中央値といった代表値や数字のばらつき（標準偏差）、データ間の関係性（相関）などにより、データの特徴を把握すること

(ロ) データの特徴が一目で分かるよう B I ツールなどを利用し可視化（グラフ化）すること

(注) 母集団の推計を目的とした分析（推測統計）や、結果から原因を推定する分析（ベイズ統計）など、従来の統計分析に該当するものは、「基本的なデータ分析」に分類するが、分析内容に応じて、「高度なデータ分析」に当てはまる場合がある。

ロ 高度なデータ分析

A I を活用したビッグデータ等の分析等をいい、例えば、B A ツール、プログラミング言語による機械学習（参考「高度なデータ分析（機械学習）のイメージ」参照）やテキストマイニング等が当てはまる。

5 データリテラシーレベルの設定

本計画では、データ活用推進の計画終了時点（令和8事務年度末）において職員が備えるべきデータリテラシーを設定する。

（注）本計画におけるデータリテラシーとは、データを理解・解釈し、実際に分析し、日々の業務に利用・活用・応用する能力をいう。

ところで、データ活用の推進は、国税庁・国税局（沖縄国税事務所を含む。以下同じ）・税務署の全ての部署・職員において、その必要性を認識し、取り組むべき全庁的な施策であり、また、一定程度のデータリテラシーは、全ての職員が備えるべき基本的素養と考えられるものの、職務の内容等に応じて、必要とされる知識等は異なるものと考えられる。

そこで、別紙「データリテラシーレベル」のとおり、データ分析のスキル等に応じ、データリテラシーをレベル別に分けることとした。

データリテラシーレベルでは、データ活用の意義を理解する「エントリー」を設定し、全ての職員が備えるべき基本的素養を「ベーシック」とした上で、より高度な知識等として、「スタンダード」、「アドバンスト」、更には、「エキスパート」の各レベルのスキルを定めた。

なお、「高度なデータ分析」の理解に当たっては、必ずしも「基本的なデータ分析」の完全な理解が必要ではなく、例えば、エキスパートレベルに該当する者が、スタンダードレベルの「基本的なデータ分析」能力の一部を有していない場合なども想定される。

このようなことも含め、設定した各レベルやその内容は、現段階における「目安」であって、今後の取組の過程や技術の進展等、必要に応じて、弾力的にその変更等を行っていく。

また、エントリーとベーシックを除き、各レベルのリテラシーを備えるべき職員の範囲等についても、今後の取組の過程で決定する。

6 データリテラシーの向上に関する施策

「研修」、「業務における活用」、「パイロット・プロジェクト」の三本柱の実施により、職員のデータリテラシーの向上を図る。

(1) 研修

データリテラシーレベルに応じて、次のとおり、研修体制を整備する。

イ エントリー

国税組織におけるデータ活用の意義を習得するもの。

通常は、日々の業務により習得するものであるが、新入庁職員に対しては、「普通科（地方研修）」、「専門官基礎研修（本校研修）」及び「社会人基礎研修（本校研修）」により実施する。

ロ ベーシック

「基本的なデータ分析」に関する基礎的な知識、「高度なデータ分析」に関する初歩的な知識を習得するもの。

国税庁から配付した資料に基づき、各国税局において、職場研修を実施する。

なお、研修は、令和3事務年度から実施することとし、実施時期や実施要領等の詳細については別途定めるところによる。

ハ スタANDARD及びアドバンスト

「基本的なデータ分析」に関する一定の専門的な知識、「高度なデータ分析」に関する基礎的な知識を習得するもの。

令和3事務年度中に、国税庁の一部職員に対し、アドバンストレベルの研修を試行的に実施することとし、その結果を踏まえ、各研修の対象範囲や研修内容、実施方法について決定する。

ニ エキスパート

「高度なデータ分析」に関する専門的な知識を習得するもの。

「データ活用研修（本校研修）」により実施する。

(2) 業務における活用

国税庁・国税局・税務署において、データリテラシーレベルに応じた知識・技術等を日々の業務の中で習得することを目的として、「基本的なデータ分析」を中心に業務におけるデータ活用（分析）の取組を推進する。

例えば、会議資料等において、平均値のみならず、中央値や最頻値を用いた計表により各種事績を分析するなどの取組が考えられる。

なお、「高度なデータ分析」の実施を妨げるものではない。

(注) 職員のデータリテラシーの向上、組織におけるデータ分析力の向上に資する施策であれば、幅広く対象として差し支えない。

(3) パイロット・プロジェクト

「高度なデータ分析」に関する取組を一層充実させるため、国税庁と各国税局において、AIを用いた先端的な取組等を「パイロット・プロジェクト」（以下「庁パイロット・プロジェクト」、又は「局パイロット・プロジェクト」という。）として実施する。

局パイロット・プロジェクトの実施の詳細については、別途定めるところによる。

7 体制整備とその役割

全庁的なデータ活用の推進を図るため、以下のとおり、国税庁・国税局・税務署におけるデータ活用の推進体制を整備し、その役割を明確化する。

(1) データ活用チーム（担当）の設置

各国税局の情報システム課に「データ活用チーム（担当）」を設置する。

(注) 各国税局のデータ活用チーム（担当）は、情報システム課職員のみでの構成であっても、各課職員を構成員としても差し支えない。

なお、データ活用チーム（担当）の編成に当たっては、データ活用チーム（担当）と各課との間での積極的な人事交流に配慮する。

(2) データ活用推進に関する各部署の役割

データ活用推進において、各部署の役割は、以下のとおり。

イ 国税庁長官官房参事官

データ活用企画係を中心に次の役割を担う。

- (イ) データ活用の推進に関する全庁的な統括
- (ロ) 庁パイロット・プロジェクトの実施（国税庁各課と共同実施）
- (ハ) 国税庁の各課における「業務における活用」に関する技術的な支援

ロ 国税局情報システム課

データ活用チーム（担当）を中心に次の役割を担う。

- (イ) 国税局におけるデータ活用の推進に関する統括
- (ロ) 局パイロット・プロジェクトの実施（国税局各課と共同実施）
- (ハ) 国税局の各課及び税務署における「業務における活用」に関する技術的な支援

ハ 国税庁各課

- (イ) 担当事務の「業務における活用」に関する統括
- (ロ) 「業務における活用」の実施
- (ハ) 庁パイロット・プロジェクトの実施

ニ 国税局各課

- (イ) 税務署における担当事務の「業務における活用」に関する統括
- (ロ) 「業務における活用」の実施
- (ハ) 局パイロット・プロジェクトの実施
- (ニ) 税務署における「業務における活用」に関する技術的な支援

ホ 税務署

「業務における活用」の実施

（注）国税局各課との協議及び指示に基づき、各税務署の実情に沿った取組を実施する。

8 報告

(1) 「業務における活用」の報告

各国税局は、「業務における活用」について、その取組状況を定期的（半年に1回程度）に庁へ報告する。

なお、報告の詳細については別途定めるところによる。

(2) 局パイロット・プロジェクトの報告

各国税局は、局パイロット・プロジェクトについても、その取組内容を定期的に庁へ報告する。

なお、報告の詳細については別途定めるところによる。

9 本計画の見直し

本計画については、データ活用の取組状況、研修の実施状況、職員のデータリテラシーの達成度合い等を適切に見極めた上で、必要に応じて、見直しを実施する。

データリテラシーレベル

レベル	データ分析に関する知識の「理解」	ツール等を駆使し結果を導くデータ分析の「実施」
エントリー	○ データ活用（分析）の意義について理解している。	
ベーシック	○ 基礎的な統計知識を有している。 【代表値、分散、標準偏差、相関など】 ○ データ可視化の基礎知識(グラフの見方)を有している。 【棒グラフ、ヒストグラム、散布図など】 ◎ 予測分析など、高度なデータ分析の種類等について理解している。 【予測分析の種類など】	○ データ加工（マッチングや集計など）を行い、可視化をすることができる。 【Excel】
スタンダード	○ 統計について一定程度の専門的知識を有している。 【回帰分析など】 ◎ 予測分析など、高度なデータ分析手法等について基礎的な知識を有している。 【予測分析結果の理解など】	○ 自らの業務に係るデータについて、業務課題の発見のため、基礎的な統計知識を活用し、適切に可視化することができる。 【Excel、BI ツール】
アドバンスト	◎ 予測分析など、高度なデータ分析の手法等について理解し、高度なデータ分析の対象とすべき業務課題の抽出ができる。また、適切な分析アプローチかどうかについて判断できる。 【予測モデルの構築プロセス、分析結果の活用方法など】	○ 高度なデータ加工（データクレンジング、外れ値の処理、標準化など）を行うことができる。 【Excel、Access】 ○ 回帰分析等を行うことができる。 【Excel の分析ツール機能】
エキスパート	◎ 統計学や機械学習等について専門的知識を有している。 【時系列分析、多変量解析、決定木、ランダムフォレスト、勾配ブースティング決定木、ニューラルネットワーク（ディープラーニング）、クラスター分析、アソシエーション分析など】	◎ 大量のデータを扱いながら、高度なデータ加工、可視化、統計解析を行うとともに、与えられた課題に対応する適切なアルゴリズムを選択し、予測モデル等の構築を行うことができる。 【SQL、BA ツールやPython、R といったプログラミング言語】

※1 「○」は基本的なデータ分析に関する能力、「◎」は高度なデータ分析に関する能力とする。

※2 原則として、上位レベルの者は、下位レベルのスキルを習得しているものとする。

高度なデータ分析（機械学習）のイメージ

A I (Artificial Intelligence・人工知能)

機械学習

機械学習(教師あり)

正解のラベルがついた学習用データから一般化した法則を導き出し、将来の出来事が起こる確率等を統計的に予測する手法

- **線形回帰**(単回帰、重回帰)
データの分布ルールを直線として仮定し、その直線がどのようなものかを学習し数値を予測する手法
- **ロジスティック回帰**
0~1の発生確率、いわゆる予測の対象が2値のデータを予測(判別)する手法
- **サポートベクターマシーン(SVM)**
パターン認識モデルの一種で、データを高次元の特徴空間にマッピングし、2つのグループ間の最も距離を離れた箇所を見つけ出し、データを区分し予測する手法
- **決定木**(C5.0、CHAID、CART、QUEST)
木構造を用いて、影響度の高い情報で、データを分割しそのルールを算出することで判別や回帰を行う予測の手法
- **ランダムフォレスト**
決定木のアンサンブル学習のうち、バギングを応用した手法(過学習を解消し未知のデータに強い汎用的な性能が期待できる)
- **勾配ブースティング決定木**(Xgboost、LightGBM)
決定木のアンサンブル学習のうち、ブースティングを応用した手法(ベースとなるモデルを学習させ、そのうち誤判定となった部分を加味して次のモデルを学習させる手法で、精度が高いモデルが期待できる)
- **ニューラルネットワーク**
人間の脳の仕組み(情報処理)に着想を得たもので、脳の特性をコンピューター上で表現(ニューロンを模して「入力層」、「隠れ層(中間層)」、「出力層」の3つから構成)し、予測する手法

機械学習(教師なし)

事前に正解が与えられていない学習用データから一般化した法則を導き出すために利用される手法

- **クラスター分析**…入力されたデータの類似度に基づきグループ化(分類)する手法
- **アソシエーション分析**…事象間の関連性から、条件を満たすルールを探索する手法
- **主成分分析**…量的な説明変数を、情報の損失を最小限にしつつ、少ない変数に縮約する手法
- **新規性(異常)検知**…新しい事例(データ)が与えられたとき、それが正常であるか異常であるかを判定する手法

強化学習

試行錯誤を通じて、「価値を最大化するような行動」を学習・予測する手法(目的とするスコアを最大化するためにはどのような行動をとればよいかを導き出す)

ディープラーニング(深層学習)

ニューラルネットワーク(NN)を発展・応用させたものであって、画像、音声、動画、自然言語などのデータを用いて分析を行うことができる手法