Advanced Analytics for Better Tax Administration

PUTTING DATA TO WORK

OECD

Advanced Analytics for Better Tax Administration

PUTTING DATA TO WORK

BETTER POLICIES FOR BETTER LIVES

Please cite this publication as:
OECD (2016), *Advanced Analytics for Better Tax Administration: Putting Data to Work*, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264256453-en

ISBN 978-92-64-25644-6 (print)
ISBN 978-92-64-25645-3 (PDF)

Photo credits: Cover © a-image | Shutterstock.com

Corrigenda to OECD publications may be found on line at:
www.oecd.org/about/publishing/corrigenda.htm.

Preface

The last ten years have seen marked progress in our ability to capture, organise, store, and manipulate data. Operations that were once prohibitively time-consuming can now be completed in seconds. This transformation in the technical possibilities has created a whole range of opportunities for organisations with the wherewithal to understand, analyse, and act on the information at their disposal.

Many tax administrations, including Ireland, are already using the wealth of information now available to them in order to better understand taxpayers and improve operational performance. As Commissioner, I am particularly conscious of our responsibility to fully utilise the external data sources available to us now, and those we will acquire in future, through third-party returns, real-time systems, supply chain data and international exchanges, especially given the administrative burden that these returns impose on individual taxpayers and businesses.

However, the task of using this information effectively is not straightforward – “Big Data” does not automatically translate into “Big Improvements”. To capitalise on the opportunities available, administrations must solve a range of statistical, organisational, and technical problems.

This report is about how tax administrations are extracting value from data using “advanced analytics”, a set of statistical techniques and practices that can help distil insight and clarity from masses of information. By applying advanced analytics techniques, tax administrations can begin putting their data to work to identify compliance and other risks, to tailor customer service, and to design more effective treatment and intervention programmes. The report is intended as a practical resource for managers and senior leaders looking to establish or further develop an analytics function within their administration.

It begins by outlining how administrations are currently applying advanced analytics techniques, describing the operational problems being solved and discussing the analytic principles and procedures being applied. It goes on to discuss a number of organisational and technical considerations in the application of advanced analytics: how to integrate an analytics function into the wider organisation; how to prioritise, manage, and evaluate advanced analytics projects; how to resource projects with the right tools, technology, and – crucially – data. The report provides a range of practical case studies and examples of how advanced analytics has been deployed for better tax administration.

I hope you will find the report prompts you to ask the right questions about how to use advanced analytics effectively, and points you toward the right answers for your administration.

I believe that there is significant scope for further collaboration on these topics. The FTA provides a unique opportunity to share best practice and to draw on our combined resources to address issues of common interest. Tax administrations could seek – on a bi-lateral and multi-lateral basis – to develop closer project work, to second staff, and to collaborate on the development of capabilities in order to maximise the benefit of our extensive data resources.

I also believe that there is scope for developing countries to learn from the experiences of developed countries in order to realise the potential of advanced analytics to bring about significant advances in modernising their tax systems.

Finally, I would like to thank everyone who has been involved in producing this report: the teams in Ireland and the UK that led the work, the OECD Secretariat who supported and contributed valuable insights, and all the FTA member administrations who contributed their time and expertise to help make this an engaging and valuable document.

Niall Cody

Chairman, Office of the Revenue Commissioners
Ireland.

Foreword

Tax Administration of the future series

This report is one of three produced in 2016 by the FTA with a particular focus on how technological and business developments can be leveraged by tax administrations to help realise operational and programme efficiencies and improve their effectiveness.

While the three reports have consistent themes around the use of data, changing customer expectations and the role of emerging technologies, they take different perspectives.

- *Advanced Analytics for Tax Administration: Putting data to work* provides practical guidance on how tax administrations are using analytics to support compliance and service delivery.
- *Rethinking tax services: the changing role of tax service providers in SME tax compliance* looks at developments in the domain of tax service providers and explores how tax administrations can better co-operate with them to improve outcomes for SME taxpayers.
- *Technologies for better tax administration: a Practical Guide for Revenue Bodies* explores how tax administrations can utilise emerging technologies to further enhance their electronic services. It also offers a framework for administrations to assess the maturity level of these services.

Caveat

Tax administrations operate in varied environments, and the way in which they each administer their taxation system differs in respect to their policy and legislative environment and their administrative practice and culture. As such, a standard approach to tax administration may be neither practical nor desirable in a particular instance. Therefore, this document and the observations it makes need to be interpreted with this in mind. Care should be taken when considering a country's practices to fully appreciate the complex factors that have shaped a particular approach. Similarly, regard needs to be had to the distinct challenges and priorities each administration is managing.

Table of contents

Tables

Boxes

Abbreviations and acronyms

ATO	Australian Tax Office
BI	Business Intelligence
BT	Business Tax
CGE	Computable General Equilibrium
CRA	Canada Revenue Agency
CRS	Common Reporting Standard
CRISP-DM	Cross-Industry Standard Process for Data Mining
D&A	Data and Analytics
FTA	Forum on Tax Administration
HMRC	Her Majesty's Revenue and Customs (UK)
IR	Inland Revenue (New Zealand)
IRAS	Inland Revenue Authority of Singapore
RAG	Revenue Analytics Group
SAFe	Scaled Agile Framework
SNA	Social Network Analysis
SQL	Structured Query Language

Executive summary

This report describes the findings of two work streams commissioned by the FTA following the Advanced Analytics Conference held in Dublin in March 2015. The work streams cover the areas of *Advanced Analytics Delivery* and *Emerging Tools and Technologies* for advanced analytics; the report is based on the information and insight gathered through a survey of 18 FTA member administrations.

"Advanced Analytics" is the practice of using statistical techniques to make predictions and draw inferences about cause and effect. Both prediction and inference are everyday tasks within a tax administration. Whether selecting cases for audit, determining the next steps for debt management, or in designing taxpayer communications, tax officials are constantly making predictions and drawing conclusions about the likely impact of their actions. In this regard, advanced analytics does not aim to achieve anything fundamentally new: rather it simply seeks to carry out these tasks and make judgements with more reliance on data.

Advanced analytics is proving an extremely valuable tool in improving tax administration effectiveness. From its initial use in the selection of cases for audit, the scope of advanced analytics applications has broadened, to the extent that analytic techniques are now used to optimise debt-management processes, secure filing and payment compliance, improve taxpayer service, and understand the impact of policy changes. As a result, advanced analytics is becoming a cornerstone capability for operational and strategic decision-making in tax administration.

Alongside this expansion in application, there has been a broadening in the range of analytics techniques being used. Administrations that began with general, high-level models to recognise the patterns in yielding audits are now using advanced analytics teams to build distinct models for each risk type, mine taxpayer correspondence for insight, use unsupervised learning to identify new risks, and integrate predictive modelling with experimental design to create targeted and evidence-based treatment programmes.

This expansion in activity has brought with it a range of organisational and technical challenges. Establishing an effective advanced analytics function requires administrations both to create a dedicated, cohesive team (in order to ensure quality control and build capabilities) and to integrate analytics into the wider organisation (in order to establish an effective working relationship between analytics and operational teams). Leaders of advanced analytics functions therefore need to strike a balance between centralisation, which supports internal cohesion, and de-centralisation, which supports integration into the wider organisation. Survey responses suggest that in the early phases of development centralisation may be more appropriate, with activity becoming increasingly de-centralised as the analytics function matures.

Once the organisational structure and culture is established, administrations must determine how to manage the complexity and uncertainty associated with advanced analytics projects. Analytics initiatives are similar in nature to R&D projects, where the outcome is arrived at through an exploratory process. For this reason, most administrations are taking an iterative, "test-and-learn" approach, using methodologies such as Agile or CRISP-DM to deliver incremental improvements by adapting regularly in response to new data and feedback. The novelty of the typical project also places a burden on advanced analytics departments to follow through fully on their work: careful evaluation and active change management are essential parts of any advanced analytics initiative.

Beyond the organisational and project-management challenges, administrations also face complex decisions relating to analytics software and programming. Although commercial analytics software still pre-dominates, many of the more mature analytics functions are experimenting with open-source tools such as the programming language "R". Open-source technology offers significant cost advantages over commercial options, and generally enables administrations to deploy cutting-edge techniques before they are integrated into the major commercial packages. However, the technology also imposes new and stringent capability requirements: administrations must therefore take care that they do not solve an IT problem by creating a fresh set of skills and capability challenges.

Finally, administrations must actively manage their data to ensure it is suitable for analytic purposes. Administrations that wish to exploit fully the opportunities presented by advanced analytics must do more than record and store large volumes of information. They must re-assess the way they collect, evaluate and manage that information. Rather than seeing data as the residue of operational activities, administrations need to ensure it is viewed as a key input into the analytic process, and therefore an asset that requires careful and active management. This means that IT divisions have a key role in laying the foundations for advanced analytics. More specifically, it means trying to build representative (as opposed to just large) datasets, and actively searching for and integrating new data sources – whether internal or external, structured or unstructured – that might offer an analytic model valuable signals about taxpayer preferences, behaviour, and risk. This will also be the case in the context of expanding international information exchange, such as Country-by-Country Reporting (Action 13 of the BEPS Action plan) and the Standard for the automatic exchange of financial account information under the Common Reporting Standard (CRS).

To build on the findings of this report, FTA member countries are encouraged to consider the key organisational and technical opportunities raised, and to continue to share their experiences and learning in relation to advanced analytics. In particular, it is recommended that tax administrations collaborate to share knowledge in relation to:

- Specific applications where advanced analytics techniques have been found to be effective;
- Best practices in securing analytical capabilities, with particular reference to the skills needed to deploy open-source technologies;
- Sources and types of data that have proved most productive for analytical purposes.

Chapter 1

Introduction to analytics use in tax administrations

This chapter outlines the report's background, key objectives, and methodology. It also defines the term "advanced analytics", introduces the distinction between predictive and prescriptive analytics, and explains the basic logic of each approach.

Background

While tax administrations have been using analytics for more than 25 years, much of this use has focused narrowly on case selection. The emergence of advanced analytics, with its ability to examine data or content using sophisticated approaches such as pattern recognition, outlier detection, cluster analysis, experimental design, network analysis, and text mining, has opened up new opportunities for the use of intelligence across all aspects of tax administration.

While FTA member countries have been quick to exploit the opportunities these new approaches provide, it was not until 2011, when Ireland hosted the first FTA conference on advanced analytics, that FTA countries formally started to share their experiences in the use of advanced analytics, and signalled an interest in establishing an on-going forum for sharing their learning on the topic. Four years later, FTA tax administrations again gathered in Dublin to report on their work in this area and to identify areas of mutual interest where countries could collaborate or share knowledge more systematically.

Following the March 2015 Dublin conference, the FTA commissioned Ireland as sponsor of the Advanced Analytics Programme to undertake work in two areas: *Advanced Analytics Delivery* and *Emerging Technologies and Tools*. Twenty-one FTA member tax administrations agreed to participate in this work. It was agreed that Ireland would lead the programme on *Advanced Analytics Delivery*, while the United Kingdom would lead the work on *Emerging Technology and Tools*. What follows is a combined report on the findings of these work programmes.

Report objectives

The objectives of this report are twofold. Firstly, it highlights the key opportunities and challenges in establishing, operating, or improving advanced analytics functions in tax administrations. The report provides practical examples of how administrations are currently using advanced analytics, and discusses the topics of organisational structure, governance, project management, data, and software, among others. As such, it is intended as a practical resource to prompt senior leaders and analytics professionals to ask the right questions of their analytics functions, and to highlight where to begin looking for answers.

Secondly, the report identifies a number of areas that may be suitable for future research and collaboration between FTA member countries.

The report is intended for a wide audience of senior leaders and managers, whether they consume or produce analytics outputs, and whether their administration is highly experienced with advanced analytics or wholly new to the field. Administrations may wish to use the report as a basis for discussion between senior management and analytics professionals, to help highlight the operational, technical, and organisational challenges and opportunities that advanced analytics presents.

Methodology

The report is based on the results of a survey carried out in late 2015, together with information gathered through follow-on correspondence and meetings.

In addition, the report contains seven case studies contributed directly by participating administrations. These case studies are based on items raised in the initial survey that were identified as being of particular significance, and therefore worth elaborating on in the report.

Box 1.1. **What is "Advanced Analytics"?**

"Advanced Analytics" is the process of applying statistical and machine-learning techniques to uncover insight from data, and ultimately to make better decisions about how to deploy resources to the best possible effect. Most advanced analytics projects fall into one of two categories:

I. ***Predictive analytics*** aims simply to anticipate likely problems – for instance with the accuracy of a tax return or the timeliness of a payment – so that tax administrations can consider which actions should be taken and when;

II. ***Prescriptive analytics*** aims to help tax administrations understand the impact of their actions on taxpayer behaviour, so that they can select the right course of action for any chosen taxpayer or group of taxpayers.

The nature of the analytic task is quite different in each category.

Predictive analytics is ultimately about pattern recognition: we need to recognise the fact of relationships in the data, but we do not necessarily need to understand the nature of those relationships. To understand the impact of our actions, by contrast, it is essential to understand the nature of any relationships we identify. ***Prescriptive analytics*** is about causal reasoning: the aim is to determine whether our action caused, rather than just coincided with, a change in taxpayer behaviour.

Box 1.1. **What is "Advanced Analytics"?** *(continued)*

As the tasks are different, so are the statistical techniques required. Predictive modelling uses a wide range of techniques which, through a process of pattern "fitting" and systematic trial-and-error, aim to discover regularities in historical data. These regularities are then used as a basis for prediction. Common techniques used include multiple regression, logistic regression, decision trees, nearest neighbours and neural networks. These methods can be used to predict outcomes of interest directly, or to highlight anomalies (cases that deviate substantially from predicted values).

By contrast, prescriptive analytics uses a set of techniques that aim to address the problem of confounding variables: the fact that there are typically many influences on taxpayer behaviour, making it difficult to distinguish the impact of our actions from the effects of other relevant factors. These techniques, which include randomised controlled trials, regression discontinuity designs, and instrumental variable analysis, all aim to estimate the "counter-factual" – what would have happened if no action, or a different action, had been taken? The difference between the actual outcome and this counterfactual estimate gives a measure of the effect of our intervention.

In certain cases, predictive and prescriptive techniques can be combined in order to anticipate how a particular taxpayer is likely to respond to a given intervention. For instance, a predictive model may be fitted to experimental data to identify exactly which types of debtor will respond to a particular debt-management intervention. This approach can also be used in conjunction with cluster analysis to identify taxpayer segments that require different treatment.

It is worth noting that the tasks of prediction and causal inference are ordinary, everyday tasks within a tax administration. In selecting cases for audit, in determining next steps in debt management, in designing taxpayer communications, officials are constantly making predictions and coming to conclusions about the likely impact of their actions. Advanced analytics does not aim to achieve anything fundamentally new: it simply seeks to carry out these tasks with more reliance on data, and less on human judgement.

Source: FTA Advanced Analytics Project team.

Chapter 2

Advanced analytics activities

This chapter describes how administrations are focussing their advanced analytics efforts across the range of operational activities, including audit case selection, filing and payment compliance, debt management, taxpayer service, policy evaluation and taxpayer segmentation. It also provides a high-level description and discussion of the main analytic approaches and techniques used in each area. It does not attempt to identify best practices (since these will vary widely depending on context), but it does highlight a number of significant applications and opportunities that administrations may wish to consider.

Advanced analytics activities by country

Table 2.1 provides an overview of where survey respondents that are actively using advanced analytics have allocated their efforts. The table makes no assessment of the relevant capability of administrations working in these areas; it seeks only to identify where work is being carried out.

Table 2.1. **Summary of activities by country**

	Audit case selection	Filing & payment compliance	Taxpayer Service	Debt Management	Policy
Australia	■	■	■	■	■
Canada	■	■	■	■	
China		■			■
Finland	■	■		■	■
France	■				
Ireland	■	■	■	■	■
Malaysia	■			■	
Mexico	■	■			■
Netherlands	■			■	
New Zealand	■	■	■	■	■
Norway	■	■	■	■	
Singapore	■	■	■	■	■
Sweden	■			■	■
Switzerland	■				
United Kingdom	■	■	■	■	■
United States	■	■	■	■	■

Source: FTA Advanced Analytics survey, 2015.

Advanced analytics for audit case selection

Survey responses indicate, perhaps unsurprisingly, that audit case selection is the principal application of advanced analytics techniques. Of the 16 administrations that responded to the survey and are actively utilising advanced analytics, 15 indicated that they had deployed analytics to prioritise cases for investigation, audit, or other compliance intervention.

The survey highlighted a number of issues and practices for consideration in case selection:

a) Value of large datasets

Combatting VAT fraud or error, both in payments and re-payments, emerged as a particular area of focus in case selection, with France, Ireland, Mexico, the Netherlands, Norway, Sweden, Switzerland, and the United Kingdom all making reference to analytical models built to identify claims which might be fraudulent or otherwise non-compliant.

Many administrations choose to begin their analytics efforts in VAT non-compliance, attracted by the plentiful data arising from the high volume of payment and re-payment claims. Data-rich environments such as this favour analytics for two reasons:

- Firstly, because "noise" in the data (movements that are due to chance rather than any underlying cause of interest) tends to cancel itself out in larger samples, leaving a clear signal which can be learned by a model;
- Secondly, because the volume of returns is such that direct examination by experts is typically not feasible.

b) Importance of assessing against next-best alternative

The latter raises an important point about predictive models: a key part of the decision to build a model is an assessment of the next best alternative. Where this alternative is under-developed, advanced analytics can add significant value without difficulty; where this alternative is mature, and particularly where it uses information that cannot be made available to a model (for instance, in the form of "local knowledge"), it can be very challenging to add value through predictive analytics. This should be borne in mind when selecting advanced analytics projects.

c) The role of social network analysis

In addition to building statistical models to predict VAT fraud or error, several administrations (including Ireland, Malaysia the Netherlands, New Zealand and Singapore) are carrying out social network analysis (SNA) to help detect VAT carousel fraud (a VAT carousel is a complex form of missing-trader fraud which exploits the VAT-free treatment of cross-jurisdictional sales) and other group-level risks.

SNA helps administrations to identify risky groups in situations where individual-level assessments may fail to detect anything of concern. It identifies links between individuals (for instance, through company directorships, joint bank accounts, or shared telephone numbers), and assembles connected individuals into easily visualised networks. Caseworkers can then browse these networks to profile individual risks. Equally, the networks can be scored for risk using either a rules-based assessment or a statistical model trained on historical data.

Perhaps because SNA is still a relatively new approach, our survey found that rules-based assessment is most common. As the practice matures and historical data accumulates, it is likely that more administrations will build analytic models to determine network risk.

d) Power of fine-grained models

It is significant that at least two administrations deploy Social Network Analysis alongside predictive models for the assessment of VAT risks. This illustrates an important limitation of individual analytical models: it can be very difficult for one model to predict a range of different risks with accuracy.

This is because the underlying relationships in the data are likely to be different depending on the nature of the risk in question: for example, the patterns that predict an error on a VAT return are likely to be very different to the patterns that predict a VAT carousel fraud. When a single model is asked to recognise both patterns, its performance tends to suffer. For this reason it may be advisable – resources permitting – to build a different model for each distinct risk type. This is acknowledged in Norway, where the model that predicts fraud and error in VAT declarations is not used to detect carousel fraud.

Clearly this raises a question as to how best to define each risk type: for analytic purposes, the key question is whether the underlying patterns and relationships are likely to be substantially different for the various risk types under consideration. It is likely that this will be primarily a question for expert judgment.

A variety of administrations, including Ireland, Norway and Sweden, have developed models to assess income or payroll tax risk, including identifying the likelihood of fraud or error on relevant deductions. Interestingly, the Netherlands has developed a suite of models to address the issue. One model predicts the likelihood that a case will yield; a series of models beneath this predict the likelihood that individual elements of the return are fraudulent or otherwise incorrect. Another model predicts the likelihood that the case will yield over a pre-specified amount.

Although more time-consuming and resource-intensive to implement, this method can offer several advantages over the usual single-model approach: Firstly, it helps caseworkers to direct their efforts more effectively at the riskiest aspects of the case, thereby facilitating more efficient case-working and encouraging adoption. Secondly, if the underlying patterns that link input variables and risk differ across risk types, this approach will generate better predictive performance.

e) Role of unsupervised learning methods

The applications described above are generally examples of *supervised models*: these are models that seek to learn from historical data where the outcome of interest (e.g. whether or not a case was non-compliant) is known. Clearly these models must be built using data collected from previous audits or other interventions; the exercise then is to "train" a model to recognise the patterns that best predict non-compliance.

The upshot is that models of this type will typically refine and automate our existing understanding of risk. In general, the main contribution of a supervised model will be to reduce the number of cases wrongly flagged for intervention. This will save caseworkers' time and lessen the burden on compliant taxpayers, but in general will not help to identify new or previously unknown types of risk.

To achieve this, administrations typically need to create *unsupervised models* – that is, models that seek to identify interesting or anomalous patterns in the data, rather than trying to learn from the outcomes of specific cases. Survey responses provide two good examples of such models: Australia's *nearest neighbours* model, which is designed to identify incorrect income-tax deductions, and Ireland's *income-consumption* model, which aims to identify under-declaration of income. Although the two models use different statistical techniques (*k* in the case of Australia's *nearest neighbours* and multiple regression for Ireland's *income consumption*), both operate on the same intuition: by comparing a taxpayer's return to those of his or her peers, it is possible to identify outliers for further investigation, and also to identify cases which, though they may appear unusual on initial inspection, are in fact normal when compared to other, similar cases.

Which approach is more effective depends on the specific problem at hand: where the administration's main aim is to reduce nil-yielding interventions (eliminate false positives), *supervised models* are generally most appropriate; where the aim is to identify previously undetected types of non-compliance (eliminate false negatives), *unsupervised models* are typically more suitable.

f) Other case-selection projects of interest:

Other initiatives undertaken include:

- *Structured Income flows:* A model that links analysis of related entities to uncover misreporting at the entity-level and non-compliance associated with the structure of income flows (United States).
- *Tax Agent risk:* A predictive model to assess risk at the level of the tax agent, rather than just the individual taxpayer (Australia).

- *Unreported Income:* A predictive model to specifically identify unreported income, as distinct from over-claiming of deductions (Sweden).

Advanced analytics for filing & payment compliance

The objective in filing and payment compliance initiatives is either to secure an outstanding payment or return, or preferably to prevent the problem from occurring in the first place: in either case, the operational aim is to change taxpayer behaviour. To achieve this outcome administrations are applying both *prescriptive* and *predictive* techniques. Survey respondents indicate that *predictive* techniques are used to identify taxpayers who are likely to fail to meet their obligations, while *prescriptive* techniques are used to determine how to communicate most effectively with this group.

Survey responses showed that advanced analytics techniques are being successfully applied to improve both the timeliness and extent of filing and payment compliance, including:

Use of experimental designs

Experimental design is a prescriptive-analytics technique in which treatment and control groups are partitioned and observed in order to isolate the effects of specific actions, interventions, or treatments. Direct taxpayer communications are particularly well suited to this type of work, since the number of cases tends to be large and the cost of creating variations and partitioning treatment and control groups is minimal. Many survey respondents mentioned this type of work: the Norwegian administration, for example, has engaged with a behavioural economics researcher to test a variety of communications intended to improve compliance on declarations of foreign income.

Blending predictive modelling and experimentation

Tax authorities including Australia, Canada, Norway and the United Kingdom have implemented programmes of risk modelling and controlled experimentation that identify which cases are likely to fail to meet payment or filing obligations, and which interventions are likely to remedy the problem. In initiatives such as these, analytic outputs are used both to prioritise cases and to determine treatment paths.

The United Kingdom has commenced building models that assess taxpayer risk prior to filing. The models predict which taxpayers are most likely to miss filing deadlines, in order to target interventions to encourage compliance. These interventions are based on insights gathered from the

United Kingdom's Behavioural Insights Team, which was established to apply *nudge theory* (a mix of behavioural economics and social psychology) to try to improve government policy and services. A typical outcome of such upstream models may be to support a decision to communicate intensively (e.g. by phone) with a small set of taxpayers believed to present high risk, as opposed to using blanket communications which may be expensive and inefficient.

In general, filing and payment compliance initiatives offer interesting examples of the use of analytics to support business processes from end-to-end. Rather than using statistical techniques only to make predictions, or only to identify effective interventions, it is possible to combine the two tasks to design and evaluate intervention programmes that might otherwise be thought to be prohibitively expensive.

Box 2.1. **Canada's use of data mining models for non-filer programmes**

The Canadian tax system is based on voluntary compliance. In Canada more than 25 million individuals pay and file their tax returns without intervention. The Canada Revenue Agency (CRA) manages its programmes using a risk-based approach, to direct resources to cases with the highest risk of failing to file on time.

The CRA has developed and continues to refine several predictive models to assist in the delivery of its non-filer programmes. The models support improved workload selection and prioritisation for the programmes, and also supply estimates for cases that have not filed returns. In its first year in production, one non-filer model resulted in a total of CAD127.6 million in additional positive assessments. The CRA is now moving away from a pure predicted value to a relative ranking indicator, dynamically scoring accounts on an ongoing basis. The CRA has also developed several other models to improve programme effectiveness and enhance taxpayer services by predicting self-resolution and responsiveness to a specific compliance action.

In addition to predictive techniques, CRA applies prescriptive analytics to support improved strategic and operational programme delivery. Prescriptive analytics is used to enrich the CRA's understanding of the non-filer population, optimise operational processes, and direct the application of compliance activities, allowing for more fact-based decisions. Complementing the use of predictive models, the non-filer programme is expanding its use of behavioural economics through nudge experiments to influence taxpayer compliance behaviours.

Source: Canada Revenue Agency (CRA).

Advanced analytics for debt management

Survey responses indicated that most administration's approach debt-management analytics in the same way that they approach analytics for filing and payment compliance: applying a mix of predictive modelling and experimental techniques to identify which cases should be subject to intervention, and which specific interventions should be carried out.

A number of administrations commenced their work in debt-management analytics by modelling the risk that an individual or company will fail to pay: Finland, Ireland, Singapore and Sweden have all built models that attempt to assess the likelihood of insolvency or other payment problems. In a similar vein, Australia and Norway have built real-time debt management systems that put in place different payment arrangements depending on a taxpayer's predicted propensity and capacity to pay. The Australian Tax Office (ATO) also uses predictive analytics to send SMS messages to individuals found to be a payment risk. In relation to interventions, the Canada Revenue Agency has run controlled experiments to determine the impact of automated work processes and different taxpayer communications.

Traditional predictive modelling in debt management helps tax administrations identify groups of potentially high-risk debtors. While this approach helps focus resources on targeting those cases of highest risk, a number of administrations consider it possible to take the application of advanced analytics one step further by using a combination of experimentation and modelling to identify which cases are most likely to respond to a debt-management intervention. These may or may not be the highest-risk cases as identified by traditional predictive modelling – models may highlight risky cases that are not amenable to intervention or cases where debts would be re-paid even in the absence of intervention.

To help administrations identify the course of action that will yield maximum incremental return, a technique known as *uplift modelling* is required. This approach starts by running a controlled experiment to determine the incremental impact of a particular intervention. It then applies predictive modelling techniques to identify which types of taxpayer show the greatest response. This model can then be used as a basis for targeting future interventions. This approach has become increasingly common in the private sector in recent years. In the field of tax administration, it is most likely to be used in debt management, taxpayer service and programmes encouraging voluntary compliance. Survey responses indicated that, to date, minimal use has been made of these techniques in tax administration outside of a few isolated projects. Administrations should seek to explore this area over the coming years.

Advanced analytics for taxpayer service

To date, tax administrations have tended to use advanced analytics mainly to inform case selection, compliance activities, and debt management. However, as discussed in the OECD publication *Increasing Taxpayers' use of self-service channels* (OECD, 2014), many administrations have commenced using analytics in support of taxpayer service. The use of analytics to assist in developing views on taxpayer channel use, inform design decisions and to identify opportunities to offer self-services are assisting tax administrations improve outcomes. The use of pro-active messaging, calling, and other interventions, especially in the case of possible non-compliance has encouraged administrations to look more closely at how advanced analytics can more broadly improve service delivery to taxpayers. Survey responses highlighted a number of initiatives in this area, with more planned for 2016.

One particularly innovative service analytics project was highlighted by Singapore, where the text of incoming customer emails is mined in order to classify, analyse, and gain insight into the content of taxpayer inquiries. This insight is then use to devise and prioritise initiatives to improve service delivery. This initiative is the subject of a more detailed case study below.

New Zealand applies similar tools for sentiment analysis and question extraction, and has also worked to achieve a customer-centric view of its data, integrating customer complaints, survey results, and risk management data to offer a more rounded picture of each taxpayer. This provides a platform for new operational practices – for example, fully customer-centric data makes it possible to treat complaints from compliant taxpayers differently to complaints from their non-compliant counterparts.

A range of administrations – including Canada, Ireland, Norway and the United Kingdom – are using a mix of predictive and prescriptive analytics techniques to manage which channels taxpayers use for inbound communications. In general, analytics is used to encourage greater adoption of digital channels. This in turn is expected to open up new opportunities for analytics, since tracking and experimentation are simpler and less costly in the digital environment.

Box 2.2. **Text mining of inbound emails in Singapore**

In 2014, the Inland Revenue Authority of Singapore (IRAS) began using text-mining techniques to analyse the content of emails received from taxpayers. The findings from this analysis complemented existing analyses of structured data and helped the IRAS to pre-empt or reduce contacts and improve service delivery for taxpayers.

Box 2.2. **Text mining of inbound emails in Singapore** *(continued)*

The objectives of the project were to identify the nature of taxpayer inquiries and highlight important changes and trends that might require response. Text data from taxpayer correspondence was extracted, cleansed, and structured to derive patterns and insights. The project was an iterative process that required close collaboration between the analysts and the business users to contextualise the findings and improve the text mining process.

As a result of the project, the IRAS was able to uncover insights, otherwise locked in textual data, on issues pertinent to taxpayers. In one project, text-mining helped to identify the common queries taxpayers had after an existing tax policy was changed. Based on this analysis, the IRAS was able to push out appropriate campaigns in a timely manner, to provide more guidance on the IRAS website, and to proactively initiate updates to taxpayers, thereby reducing the need for taxpayers to contact the IRAS.

Ongoing tracking of the nature of email enquiries over time has also enabled the IRAS to identify trends in certain topics and respond accordingly. Text mining has now replaced the manual tracking of email enquiries, which has saved time and improved staff productivity. It has also enabled the IRAS to track the nature of enquiries more objectively, avoiding the inconsistencies of interpretation typical of manual tracking.

Source: Inland Revenue Authority of Singapore (IRAS).

Advanced analytics for policy evaluation

Although most analytics work is carried out to support operational decision-making, survey responses highlighted that analytics is also being used for decision-making in relation to strategy and policy. The most common analytic applications in this field are tax gap measurement, and assessing or forecasting the impact of changes in tax policy.

China, Finland, the United Kingdom and the United States all use analytic techniques to carry out tax gap analysis. Here, the focus tends to be on using yield data from random audit programmes, since these typically give an accurate representation of the wider taxpayer base.

Singapore has deployed visual analytics and simulation methods to explore the likely impact of proposed policy changes. The use of data visualisation has also enabled policymakers to quickly identify patterns, trends, and anomalies, improving the efficiency of policy review and decision-making. China's analytics function has carried out assessments of the impact of major tax reform initiatives based on simulation modelling, as discussed in the case study below.

It is worth noting that a simulation model is quite different to the type of predictive model discussed above. Where predictive models fit patterns to historical data, simulation models tend to draw more heavily on modeller input allied to economic theory (specifically, mathematical representations of the macro-economy). These models can help policymakers to understand and explore complex relationships, but their predictions or other findings will only be as accurate as the inputs supplied by the modeller.

Box 2.3. **China's assessment of the impact of value-added tax reform**

In 2012, the Chinese government implemented a value-added tax reform pilot programme, which replaced the previous business tax (BT) with a value-added tax in selected sectors. In order to analyse the overall effect of the policy reform, including effects on the wider economy, on tax revenue, on industry structure, on social welfare, and on a wide variety of economic indicators, China's analytics department built a Taxation Computable General Equilibrium (CGE) model.

The Taxation CGE Model consists of six parts: the production sector, private sector, government sector, tax sector, trade sector, and macro close condition. The social accounting matrix table was established on the basis of the Chinese input-output table, national economic statistics, and tax statistics.

Under the changes made in 2012, the manufacturing, wholesale, and retail industries were made subject to VAT while other industries remained subject to the existing business tax. Where the business tax was replaced by VAT, the output price would vary because VAT and BT applied different administration and calculation principles. This price change is the source of the effects modelled. Through the CGE's price system, the model makes it possible to estimate the different effects of VAT and BT on output price. The model even mimics the invoice deduction method of VAT.

Through the Taxation CGE model, the analytics team was able to estimate the economic and social consequences of the VAT reform. Their report was approved by the Premier of the State Council, and played a key role in the policy reform process.

Source: State Administration of Taxation (SAT), People's Republic of China.

Advanced analytics for taxpayer segmentation

Many survey respondents reported an interest in using analytical techniques to segment their taxpayer base more effectively, though few dedicated projects have to date been carried out in this area by tax administrations.

A number of administrations (including Ireland and the Netherlands) have experimented with unsupervised segmentation techniques. These techniques, which fall under the broad heading of "cluster analysis", seek to identify groups of taxpayers who are similar to each other in some significant respects, and dissimilar to the other groups identified. These projects have often provided interesting general insight into the taxpayer population, but have typically not shown a strong practical impact as the segments identified have not had obvious business applications.

An alternative approach to segmentation, currently being pursued in Ireland, looks to group taxpayers based largely on their predicted response-to-intervention. If all taxpayers respond in the same way to a given intervention, then there is little practical value in segmentation; where there are large and consistent differences in response-to-intervention, then segmentation is worthwhile, and should follow the observed differences in response. This approach, which uses the *uplift modelling* techniques described above, is likely to create multiple segmentations – ultimately, each type of intervention may require a different segmentation of the taxpayer base. With tax administrations increasingly looking to personalise service and develop appropriate and timely interventions, further work in this area will be of prime importance.

Chapter 3

Advanced analytics in the wider organisation

This chapter discusses the challenge of fitting an advanced analytics function into the structure and culture of the wider tax administration. It highlights the importance of building an effective working relationship between operational departments and the analytics function, while maintaining high technical standards and strict quality control. Finally, the chapter outlines the main approaches taken by survey respondents to achieve these ends.

A major issue identified by survey respondents was how to fit analytics into the wider organisation, both structurally and culturally, in order to develop a technically strong analytics function that also enjoys an effective working relationship with the operational teams it is intended to serve. If the analytics function is not sufficiently distinct and cohesive, it can be difficult to maintain quality and develop technical capabilities; if the relationship with the business units is not sufficiently close, there is a risk that operational teams will not understand or engage with analytics, and that analytics teams will not appreciate and work within operational priorities and practical constraints.

Structural integration

Achieving the proper structural integration of analytics and operational functions was highlighted as being particularly challenging: the main barriers cited by respondents were the natural scepticism of operational staff toward a new and unfamiliar approach, and the wide gaps in mind-set, expertise, and even terminology between analytical and operational specialists.

Administrations have taken varying approaches to the organisational positioning of analytics functions; further differences emerge when reporting lines and working practices are taken into consideration. However, the various approaches can be broadly characterised as either *centralised* or *de-centralised*.

a) The centralised approach

Ireland, Mexico and Singapore each offer examples of the centralised approach, where consolidated analytics departments sit in "head office" divisions. This encourages collaboration within the department and makes it easier to manage the technical development of analytics staff. It also enables more experienced team members to supervise analytics work closely, ensuring close quality control. The approach is therefore well suited to the early phases of analytics development, where building capabilities and producing reliable outputs are likely to be the main priorities.

However, there is a risk that, as the analytics function matures and seeks to widen its influence across the organisation; the centralised approach may act as a hindrance rather than as a support. It is notable that the three administrations mentioned above have all deemed it necessary to adopt special initiatives to promote collaborative working practices and overcome the possible "silo" effects of centralisation. These initiatives include exchanges of staff between analytics and operational teams, the transfer of governance of certain analytics projects into operational departments, and the joint conduct of project evaluations.

A highly distinctive approach has been taken in the Netherlands, where a consolidated analytics department operates quite separately from the rest of the organisation. To date, analytics projects in the Netherlands typically have not had a direct business sponsor. This has allowed the analytics function to prioritise and explore projects free from operational requirements, and to develop and innovate quickly without having to adjust to the pace of other departments. Several major projects have been delivered successfully using this approach, and the analytics section has expanded rapidly over the last three to four years.

Potential drawbacks of the centralised approach are:

- In instances where there is a dependency on other departments (e.g. for data exchange or infrastructural developments), it may be difficult to have analytics initiatives prioritised;
- The separation between analytics and the rest of the administration means that the analytics department may lack process and tax knowledge; and
- There is a risk that the analytics department may not address the most pressing business needs, and operational staff may not adopt the analytics solutions developed.

Nevertheless, the structure has worked well in the Netherlands as a "proof-of-concept", allowing the analytics department to develop and demonstrate the value of statistical techniques more quickly than might otherwise have been possible. It is anticipated that, as the analytics department matures, it will form closer connections to the relevant operational teams.

b) The de-centralised approach

Australia, Canada, New Zealand and the United States all follow forms of a de-centralised approach, in which analytics and operational teams are integrated or co-located in order to encourage more collaborative working practices. Administrations following this approach generally reported good relations with the relevant operational teams, but were often required to set up special arrangements in order to ensure sufficient quality control, and to maintain a repository of experience and technical expertise. The de-centralised approach seems to work best where analytics capabilities are already well developed.

The "hub-and-spoke" systems developed in Canada and New Zealand are typical of the de-centralised approach. In this structure, a central "hub" is responsible for spreading good practice and providing quality control, while the "spoke" applies analytics across the different levels of the organisation. Typically the hub takes responsibility for more complex projects, while the spoke tends to focus on more straightforward work.

The United States also uses a blended approach, with an analytics team reporting directly to the agency head, and further research and analytics units embedded in each operating area. Model development happens throughout the organisation, with co-ordination provided by regular meetings of the research directors and senior leadership. This group also ensures that analytics projects are aligned with the organisation's strategic priorities.

Australia recently launched its *Smarter Data Program* to pull together expertise across the administration into a "virtual" analytics department. Analytics staff continue to co-locate with business operations, but now report through a single analytics business line. This enables analysts to retain a close connection with the business, while also offering the advantages of consolidation.

Although there is a general tendency to de-centralise as an analytics function matures, it is striking that most administrations are actively seeking a balance between external integration and internal consolidation.

The major potential drawbacks of a de-centralised approach are:

- While being located with business partners may improve business connectedness it can increase the challenge of establishing a strong analytics culture among decentralised analysts.
- Control over the quality of analytics work requires more active management, to ensure standards are adhered, learning and analytical methods are shared across the analytics community and staff capability and learning remains a focus.

Cultural integration

Many survey respondents highlighted the challenge of achieving a cultural fit between analytics and the wider organisation. Switzerland observed that cultural barriers were often a greater obstacle to progress than any of the technical or statistical challenges that might arise. Survey respondents and interviewees highlighted the major differences in perspective that exist between analytics departments and operational teams, with several observing that analytics can be perceived as a threat by staff who are used to making decisions according to experience and instinct.

While many of these difficulties can be addressed through specific change management initiatives (discussed in Chapter 4), respondents also reported a number of more general approaches:

- *Integration with strategic objectives:* As well as pursuing a mixed organisational model, the United Kingdom has achieved close collaboration between operations and analytics thanks to an organisation-wide commitment to digital development and to evidence-based decision making. This puts an onus on operational managers to establish that they have used the best available information and evidence in prioritising cases, determining interventions, etc. This in turn creates a natural customer base for analytics outputs.

- *Education and training:* Ireland has commenced a "bottom-up" approach to generating demand for analytics: an "*Introduction to Analytics*" training course has been rolled out to promote better understanding of the nature and potential of analytics. In addition, an analytics module has been added to the university degree course in Applied Taxation that is pursued by many auditors.
- *Operational focus:* In order to establish strong working relationships, managers in Canada have worked to build trust amongst operational staff to highlight that analysts are focused on practical business problems, and are using analytics to identify opportunities, address risks and respond to pressing business needs. The aim has been to take incremental steps, and to ensure that the first priority of every advanced analytics project is to meet core business objectives.
- *Active communications:* In both Canada and Ireland, there is an emphasis on actively communicating about analytics using plain terms, attempting to explain the "common-sense" logic behind the techniques used. In Ireland, there is a particular emphasis on conveying the limitations of analytics as well as selling the benefits. These initiatives are intended to remove the mystery often associated with the topic, and to encourage staff to see analytics as a useful practical tool to help achieve operational goals, rather than as an intellectual exercise or a threat to traditional roles.

Chapter 4

Managing advanced analytics projects effectively

This chapter describes the main practical challenges associated with the governance, prioritisation, management, and evaluation of advanced analytics projects, and outlines a number of initiatives that have enabled survey respondents to overcome these challenges. It focuses particularly on how administrations can manage the uncertainty associated with the typical advanced analytics initiative. The chapter also addresses the issue of "change management" and identifies a range of approaches deployed by administrations to ensure that operational staff act on analytic outputs.

Unlike traditional projects, where it is typically possible to scope and deliver new tools or functionality to meet user requirements within a specified timeframe, analytics projects often bring with them significant uncertainty as to outcome and timing. Survey responses highlight that, in many cases, it is not possible to tell how well an outcome can be predicted until the underlying relationships have been modelled, or how effective an intervention will be until the relevant experiment has been conducted. In this sense advanced analytics projects are analogous to R&D investments. As such, the approach to project governance, prioritisation, management, and evaluation requires careful consideration.

Project governance

Analytics governance requires a strong focus on integrating the business, IT, and analytics perspectives. Each department has a distinctive contribution to make:

- Business representatives provide an understanding of operational priorities and constraints;
- Analytics staff can determine where statistical techniques can and cannot be effective;
- IT specialists add understanding of data sources and technical constraints.

Many respondents indicated that they had established integrated governance bodies to prioritise, resource, and oversee analytics projects. A wide variety of approaches was evident:

- A number of administrations had centralised the governance of analytics projects under a single dedicated body. For instance, Ireland has recently established a senior management group comprised of IT, business, and analytics representatives to oversee analytics work across the organisation. This is discussed in more detail in the case study (see Box 4.1).
- Singapore pursues a mixed model. A permanent, centralised Steering Committee (made up of the senior management of the business and analytics departments) prioritises projects across the business units and governs the overall direction of analytics in the organisation. Specific projects are then governed by dedicated committees (made up of the management of the departments that are involved) established specifically to oversee the initiative at hand.
- Canada provides a final variation: advanced analytics activities are governed through a Steering Committee whose remit covers ordinary business intelligence (BI) as well as predictive modelling and other

advanced techniques. This data-oriented approach may help to co-ordinate BI and analytics activities, and ensure that each project deploys a level of sophistication appropriate to the problem at hand.

While it is not yet possible to draw firm conclusions as to specific "best practices", most administrations indicated a desire to consolidate analytics governance in a permanent body, which could then build expertise and experience across multiple projects.

Box 4.1. **Centralised governance of advanced analytics in Ireland**

The Irish tax administration uses advanced analytics to identify cases for intervention, to forecast measures such as debt available for collection, and to evaluate and enhance the impact of its actions and interventions. Advanced analytics methods have been used in Revenue since 2011, but 2015 saw a re-doubling of efforts in this area.

Prior to 2015, governance was organised on a project-by-project basis. A number of highly effective models – notably in the areas of VAT and payroll taxes – were introduced under this system, but the lack of a centralised, permanent governance structure made it difficult to build organisational momentum behind analytics, and to maintain and upgrade the predictive models that had been built. Each project group developed significant expertise over the course of its work, but was then disbanded as the individuals involved began to focus on other work.

In 2015, a decision was made to establish a senior management group – the Revenue Analytics Group (RAG) – to prioritise and oversee all analytics work in Revenue Ireland. The group is led by the Chairman of the Revenue Commissioners, and consists of representatives from the business, analytics, and IT functions. The RAG also has direct links into the key operational and IT governance bodies, the latter of which oversees the advanced analytics budget. Business Intelligence initiatives are governed through a separate but linked structure.

Beneath the RAG is a steering group comprising the relevant IT and analytics managers, and a liaison group which includes the same managers and a wide range of business representatives. The steering group meets fortnightly to manage the details of ongoing projects; the liaison group meets quarterly to gather feedback and project ideas from across the operational departments. These ideas are then developed and prioritised by the RAG.

This structure provides cohesive governance of Revenue's analytics work: it aligns analytics projects to organisational priorities; it ensures the analytics function works within the appropriate technical infrastructure; and it monitors the progress and value delivered by analytics projects.

Source: Revenue Commissioners, Ireland.

Project prioritisation

Perhaps not surprisingly, project prioritisation has emerged as a significant challenge in administrations with more mature analytics functions. Owing to early successes, many administrations have seen the demand for analytics work outstrip the supply, making prioritisation essential. Creating an effective prioritisation process – particularly one that can handle the uncertainty associated with analytics projects – has proved to be a challenge for many administrations.

In general, administrations have taken an exploratory approach to prioritisation: rather than attempting to design and execute detailed long-term project plans, analytics teams tend to begin work on a wide range of areas, narrowing their focus only as it becomes clear from experience that a particular project is likely to yield results. For example, the Netherlands has explicitly adopted a "funnel" approach, where work begins on a wide variety of projects, and formal checkpoints are used to assess and compare progress. Over time, under-performing projects are identified and stopped, and a small group of "winners" emerge. Clearly this process entails a certain amount of wastage, but it offers the compensating advantage of facilitating evidence-based prioritisation decisions.

Australia has pursued a mixed approach to project prioritisation, blending the "exploratory" process described above with a formal scoring process, as outlined in Box 4.2.

Box 4.2. **Australia: Use of "weighted shortest job first" for project prioritisation**

Australia's *Smarter Data* organisational unit is implementing the use of the Scaled Agile Framework (SAFe) methodology to deliver value to internal and external customers more rapidly utilising agile practices. The principles of the Agile framework include: taking an economic view, applying system thinking, and building incrementally with fast, integrated learning cycles and decentralised decision making.

The SAFe methodology makes use of the "weighted shortest job first" prioritisation model, which considers the elements of business or user value, time criticality, risk reduction or opportunity enablement and job size. Typically, business owners will submit a wide variety of ideas for analytics projects. An advisory group consisting of business, IT, and analytics representatives will meet to evaluate these ideas, assigning explicit scores to each proposal based on the elements described above. After debate and discussion, the group typically reaches consensus on the appropriate scoring.

Box 4.2. **Australia: Use of "weighted shortest job first" for project prioritisation** *(continued)*

This approach brings together business owners along with data, analytics, and IT capabilities to make a comparative assessment of each component to determine project priorities.

The benefits of this approach have been:

- All parties are learning to consider the minimal viable product that could be delivered to realise value, shifting away from defining in detail a final "gold-plated" solution which may not be relevant once delivered.
- Involving the right stakeholders in the prioritisation process has meant they feel more empowered in the delivery process and have more confidence in making decisions at a local level to ensure the right outcomes are met.

Source: Australian Tax Office (ATO).

Project management

The empirical, "test-and-learn" approach is also visible in the project management techniques favoured by analytics departments, with iterative approaches such as Agile and CRISP-DM generally preferred to sequential models such as the "waterfall".

"Agile" is a project-management approach developed in the software industry which focusses on making progress through rapid cycles of testing and adaptation, as opposed to executing a minutely detailed plan. As such, it is a useful response to the challenge of uncertainty in analytics projects: rather than trying to plan every step in advance, projects are adjusted quickly in response to regular feedback, whether that comes from users, validation tests, or pilots. The CRISP-DM approach, deployed in Canada and New Zealand, has a similar focus on iteration, and on making adjustments in response to real-world data.

Beyond the difficulties of managing in uncertain environments, many survey respondents highlighted the challenge of drawing together the wide range of skills necessary in an advanced analytics project: IT knowledge, statistical understanding, communications and change management, operational and tax expertise.

To address the challenge of project complexity, most administrations assemble multi-disciplinary project teams, consisting of a mix of IT, analytical, and business experts at a minimum. In the United States, this is often supplemented by a qualitative researcher; in the Netherlands a dedicated change management team supports the implementation of any completed projects (this is the subject of a case study in Box 4.3). Several administrations second staff into and out of analytics teams; others embed tax experts within the analytics function, and vice-versa.

The United Kingdom takes this emphasis on multi-disciplinary skills one step further, aiming to train individual analysts to perform a wide variety of tasks. The analyst role is seen as having three main elements: a "consulting" role to bridge the business and statistical understanding; a "data engineering" role to extract and organise the relevant data; and an "analytic" role to create effective statistical models. In practice, more experienced analysts tend to take the lead in the "consultancy" roles, and provide quality assurance for the other activities. Nonetheless, there is an expectation and ambition that each analyst should aim to develop all three capabilities.

Project evaluation

Advanced analytics remains a novel and – to many – unfamiliar approach, making the task of measuring outcomes and demonstrating value especially important. However, it can be very difficult to isolate and measure the contribution made by a new analytics model, particularly as they tend to be deployed in contexts with many moving parts. For instance, if there are changes in the economic environment, in case-working practices, or in conventional case-selection methods, it can be problematic to assess the effect of a new analytics model using the traditional before-and-after comparison. This can make it hard to determine whether analytics is providing value for money.

Survey responses indicated that analytics departments invest substantial time and effort into impact evaluation, deploying a range of approaches tailored to suit the practical constraints and the exact evaluation question to be answered.

In general, respondents indicated that they first make a quantitative estimate of the impact of a new model on audit yield, caseworker productivity, or other metrics. In addition to this quantitative assessment, both Singapore and the United States indicated that they often add a qualitative element to project evaluations. In Singapore, for example, experts make an assessment of the type of cases highlighted by a new model, in an effort to understand whether the model is finding new varieties of non-compliance, or whether it can add to the wider understanding of taxpayer behaviour.

In making a quantitative estimate of a model's impact, the most common approach cited is a before-and-after study: a "baseline" performance measure is taken prior to the introduction of a new model; the same measure is taken after a model has bedded in, and any difference between the two is assumed to indicate the model's impact. This approach has the virtue of simplicity but does not account for the possible influence of confounding factors. Where the new model offers a very large improvement over existing methods, this is typically not a problem. However, a more sophisticated approach may be needed where we are trying to identify smaller effects.

In Canada and a number of other administrations, systematic randomised testing has been used in order to adjust for confounding factors and make a more accurate assessment. This can take two forms. In the more basic, a control group of randomly selected cases is used to give a measure of the effectiveness of the model compared to no case selection method at all. While this is an unbiased measure, it may not be especially useful for operational purposes: typically, the question of interest is whether a model outperforms some pre-existing selection approach. To address this issue, many administrations are using a "champion vs. challenger" approach, running different case selection methods in parallel to determine which is more effective.

Change management

The process of embedding analytic outputs into everyday operational practices was cited as a significant challenge by almost all survey respondents. Given the inherent complexity of analytics, the discipline can be seen by experienced operational staff as unreliable or somehow threatening. Moreover, new models often entail changes to longstanding work practices that tend to provoke a natural resistance. This can make it extremely difficult to implement analytics solutions even after the challenges of data preparation, modelling, and evaluation have been overcome.

It is clear from survey responses that several administrations begin considering the change-management aspects of an initiative as early as the design stage of an analytics project. For instance, as well as ranking cases according to risk, Mexico has built an analytics and BI tool to supply the caseworker with all necessary profiling information. This approach is distinctive in that it combines analytic modelling with an effort to place more data in the hands of the caseworker, and may improve caseworker buy-in and trust in the analytic outputs.

Taking a similar "early-intervention" approach, Norway has a mechanism in place to screen out at source any initiatives that are unlikely to secure business buy-in: projects will not be prioritised unless the relevant business unit commits its own resources. In general, business units provide a project manager who will support implementation and take responsibility for change management; this has the additional benefit of freeing analysts to focus on the more technical aspects of the work.

Australia and Canada highlighted the importance of individual advocates – referred to as *evangelists* or *power users* – in catalysing the success of their advanced analytics function. These individuals are typically experienced and trusted operational managers who have enjoyed some success with analytics projects, and so can act as credible advocates for the approach across the wider organisation.

In a similar vein, Singapore aims to forge strong partnerships with business users and domain experts in order to tailor analytics solutions to operational needs. This is achieved by actively involving business stakeholders throughout analytics projects, from the agreement of objectives and scope all the way to project evaluation. Together with the analytics team, these partners will actively share success stories in order to garner buy-in and promote the use of analytics.

Of all the tax administrations surveyed, the Netherlands has taken the most systematic approach to change management, establishing a dedicated team to support all major analytics projects. The set-up and activities of this team are described in the case study (see Box 4.3).

Box 4.3. **The Netherlands: Advanced analytics change management**

The primary goal of the Dutch Tax Administration's data and analytics (D&A) team is to improve existing systems and ways of working in the organisation by using data and analytics. The analytics products created by the team have a huge impact on the way tax employees carry out their day-to-day jobs. Where employees have been operating the same processes for many years, they may find it difficult to adapt to the new D&A standards. In order to make sure that the new products are used – and used properly – by frontline caseworkers, the data and analytics department has created a dedicated change management team.

This team is staffed mainly by psychologists and other experts in organisational behaviour, and forms a part of the wider analytics department. The team bridges analytical and organisational understanding, and is therefore well placed to help embed new practices based on analytical recommendations and products.

Many of the changes proposed as a result of D&A projects require considerable standardisation of processes. When this happens employees may feel less free to make their own choices or do things their own way. Often this creates resistance, with a resulting need for a professional change management approach. The approach taken in the Netherlands is to implement the new system by also changing the context of work and changing the whole working process. To assess and improve the current processes, a lean approach is used: What is the caseworkers' current working process? What steps in the process add value and which steps can be improved to be more time-efficient? Crucially,

Box 4.3. **The Netherlands: Advanced analytics change management** *(continued)*

input is sought from staff on how they want the new process to work, and what they need in order to work effectively and efficiently. This approach has helped to shape the work of the main analytics team, ensuring a focus on building user-friendly caseworker dashboards and other outputs. Its main benefit has been in facilitating the transition from old working practices, ensuring that predictive modelling and other analytics initiatives achieve a practical impact on caseworker behaviour and taxpayer outcomes.

Source: Netherlands Tax and Customs Administration.

Chapter 5

Resourcing advanced analytics projects: Tools & data

This chapter discusses two key resources for advanced analytics projects: software and data. It outlines the range of software tools used by survey respondents, and describes each administration's self-reported level of familiarity with each. It also reports on the (mainly positive) experiences of administrations that have begun using open-source tools. Finally, the chapter outlines the main challenges administrations have encountered in sourcing data for analytics projects, and the steps being taken to address these challenges.

Advanced analytics tools – commercial and open-source options

The last ten years have seen a wide range of analytics packages released onto the market, each with its own features and associated marketing pitch. The sheer volume and range of packages and features makes choosing the right tools and technology for a tax administration a major investment decision. In an effort to assist administrations in this regard, this section of the report examines the different tools and technologies that tax administrations are currently using, highlighting the benefits and drawbacks of the various options, with a particular focus on the differences between commercial and open-source packages.

Satisfaction with current tools

As indicated in Table 5.1, survey respondents generally reported themselves satisfied with the tools they currently have in place.

Table 5.1. **Reported satisfaction with analytics tools along 4 key dimensions**

	Very satisfied	Satisfied	Somewhat satisfied	Somewhat dissatisfied	Dissatisfied	Very dissatisfied
Ease of use						
Richness of functionality						
Scalability						
Performance						

Key: *Number of countries selecting each option*

10-12 7-9 4-6 1-3

Source: FTA Advanced Analytics survey, 2015.

Overview of tools used

It is clear that commercial software packages, and especially SAS, continue to dominate over open-source tools such as the statistical programming language "R". Table 5.2 gives an overview of the main analytics software tools used, together with each administration's assessment of its own capabilities with each tool. Survey responses indicate that the main advantage of commercial packages is usability: with off-the-shelf software, analysts can manage data and build working models without extensive training in statistical programming.

Table 5.2. **Usage of and familiarity with analytics software and programming languages (self-reported)**

	SAS E Guide	SAS E Miner	SAS Other	SPSS	IBM Modeller	IBM Other	SQL	Oracle Data Miner	Stata	R
Australia		High	High			Medium	Very high		Low	High
Canada	High		High	High	High		High			
China									Medium	Medium
Finland				Very high			Medium		Very high	Very high
France	High									
Ireland	Very high	Very high							Very high	Very low
Malaysia	Very high	Very high		Medium		Medium				
Mexico	High						Medium	Medium		
Netherlands		High	High							Low
New Zealand								High		High
Singapore	Very high	Very high	High				Medium			Very low
Sweden	High	High						Low		Low
Switzerland							Very high		High	High
United Kingdom	Very high		High	High	High		High			Low
United States	Very high	Very high					Very high	High		Medium

Key: *Level of familiarity*

Very high High Medium Low Very low

Source: FTA Advanced Analytics survey, 2015.

Use of open-source tools

Although most administrations report a low level of maturity with R, it is striking that 11 administrations have begun to use or at least experiment with this language. The primary reasons for respondents' interest in R are its cost advantages over commercial packages and the flexibility and broad functionality that the language provides. Furthermore, new algorithms are typically made available in R years in advance of their integration into commercial packages. Australia cited this speed of innovation as a particular advantage when trying to combat aggressive fraud: with fraudsters deliberately probing the limitations of predictive algorithms, it is vital that tax administrations stay ahead by using the latest available technology.

In general, respondents reported positive experiences in their experiments with R to date. A few minor concerns were raised about security, deployment, the reliability of external libraries, and the ease of handling large datasets, but in general respondents did not consider these to be a major barrier to adoption.

Typically, analysts' own lack of familiarity with the language has proved to be the major obstacle; conversely, the greatest strength of the commercial packages is their usability and customer support. It is notable that some of the administrations with the greatest maturity in R are those who recruit most heavily from academia, where R tends to be very widely used. No respondents referred to large-scale efforts to train existing staff to code using R. This indicates the risk associated with switching to open-source analytics tools: for some administrations, there is a risk of solving an IT problem simply by creating a fresh set of personnel and skills challenges.

Given the large numbers of tax administrations using SAS and R, and the differing levels of familiarity in the latter especially, there are clear opportunities for sharing experience and expertise with these tools.

Box 5.1. **New Zealand: Open source and innovation**

New Zealand's Inland Revenue (IR) utilises open-source tools for identity resolution: the linking of references (registrations, refunds, etc.) to a real taxpayer identity. IR uses a combination of open-source tools to develop new algorithms to identify those who intentionally misuse identities, ranging from simple misappropriation of family members' identities to complex organised crime rings such as carousel fraud.

The innovative algorithms developed reproduce and optimise state-of-the-art mathematical and machine-learning models which help IR to predict linkages between references and identities, compute social network metrics, and traverse relationships with several degrees of separation. As a result, IR is able to identify networks of unusual behaviours which would not be easy to identify using proprietary technologies and tools. The main open source tools used are Neo4J, Gephi, R, and Java (for algorithm development).

Open-source tools are essential for IR's analytical strategy. They enable the rapid adoption of state-of-the-art technologies which would be virtually impossible to adopt in a proprietary technology stack. They allow the analytics department to utilise the knowledge produced and shared by a wider user community, whether in academic or industrial fields. Finally, they enable the analytics department to work more flexibly, creatively, and rapidly, encouraging collaboration and the sharing of expertise within the team.

Source: New Zealand Inland Revenue.

Access to the right data for advanced analytics

Survey respondents cited a range of difficulties securing timely access to the right data for analytics projects. While general questions of collecting, storing, and exposing data are beyond the scope of this report, a number of analytics-specific issues were raised, and are discussed below.

Access to data

Even more than other data analysis activities, advanced analytics requires integration of many data sources. However, many administrations reported that their information was "siloed" across multiple IT domains, slowing the process of data gathering and preparation; several highlighted that delays in access to data frequently made it impossible for the analytics function to contribute to projects with tight deadlines.

To address the issue of data access, Ireland has created a dedicated Analytics Technical Services team within its IT division. As well as managing analytics software, this team is responsible for making the relevant data available for analysis in a timely manner. The Technical Services team also creates and maintains metadata to help analysts to find, understand, and utilise the available information.

Singapore has custom-built a centralised database for analytical purposes, running on a "massively parallel processing platform". This database consolidates data from different sources and stores it in a format that is suitable for predictive modelling and other advanced analytics uses. Both Ireland and the United Kingdom are in the process of establishing similar analytics environments running on the Hadoop framework. These initiatives are intended to facilitate data access, allowing advanced analytics teams to respond more flexibly to organisational priorities.

Data representativeness

Beyond the issue of access, several respondents cited challenges relating to the *nature* of the data collected. In particular, China, Ireland and Sweden raised concerns about the "representativeness" of the available data. Often the only data available to train predictive models comes from highly biased samples (e.g. cases previously selected for intervention). This means that models can only learn about a small segment of the population, reducing their effectiveness and applicability. A related issue was highlighted by Finland, who observed that the use of "naturally" occurring data was often problematic in projects designed to identify causal effects. This is because selection effects may introduce a bias in the data that skew any conclusions drawn.

The United States addresses the problem of data "representativeness" by making extensive use of data from randomly selected audits (which are unbiased representations of the population) in order to develop its case-selection methodologies. The same approach is under consideration by a number of other administrations, though concern has been raised about the small sample sizes available – although the models created would be unbiased, they would suffer from higher variance than may be desirable. To address the impact of selection effects on our ability to make causal inferences, both the United Kingdom and Finnish analytics departments seek to engage operational teams as early as possible in any project, with a view to building some degree of randomisation into the intervention programme from the beginning.

Advanced analytics and unstructured data

As illustrated by examples provided in Chapter 2, advanced analytics techniques create the potential for systematic, comprehensive analysis of unstructured data such as that contained in taxpayer emails and other inbound communications. Almost all respondents expressed the view that this type of data contains potentially valuable information. Notwithstanding this, only five administrations (Australia, Ireland, New Zealand, Singapore and the United States – see case study in Box 2.2) reported that they are currently using this type of data for analytics. It is likely that the next five years will see a significant increase in tax administrations' use of unstructured data for analytical purposes.

Data comprehensiveness

Mexico, Singapore, and others suggested a need for a change in philosophy when it comes to data collection. Typically, analytics projects work from data that is collected and stored primarily for operational purposes. While valuable results can be produced from such data, significant opportunities may be missed: data points that are not relevant for operational purposes may be hugely valuable for analytics. For instance, case management systems may not record the type of non-compliance identified in a successful audit. While this information may not be required for operational purposes, it could prove valuable for analytics, since it would make it possible to build distinct models for each risk type, potentially improving both accuracy and usability.

Furthermore, it is difficult for analytics to outperform human judgement in cases where there is significant "local knowledge" that is not encoded in data. As far as possible, administrations should encourage operational staff to input this type of knowledge – whether in structured or unstructured form – so that it is available for analytic purposes.

To ensure it is considering data needs from an analytics as well as an operational perspective, Singapore is initiating a holistic review of current and potential data sources and requirements. Similar reviews are proposed in a range of other administrations, including Ireland that has made its IT division a core part of all its analytics efforts. Such initiatives are in part recognition of the need to re-think the approach to data and information in light of the opportunities created by analytics.

While the "big data" revolution is widely reflected on and discussed, the items above suggest that perhaps "good data" should be an equally important topic of conversation: Are tax administrations collecting the right data from the right sources, whether internal or external? Are they ensuring that caseworker knowledge is encoded as structured data? Are they paying sufficient attention to possible biases in any data collected? Given that so many of these challenges are common across tax administrations, there is a strong case for further collaboration on these topics. These developments are highlighted in the report on *Technologies for better tax administration: A Practical Guide for Revenue Bodies* (OECD, 2016).

Chapter 6

Findings and recommendations for better use of advanced analytics

This chapter summarises the findings reported in previous chapters, and sets out a number of recommendations for future collaboration and research.

This report has documented the range of areas in which different administrations are applying advanced analytics techniques. Advanced analytics approaches, while still playing an important role in case selection, are now also used to target and optimise non-filer interventions, to determine debt management strategies, to improve taxpayer service, and to evaluate the impact of policy changes.

The techniques used are evolving, with high-level supervised models giving way to more granular models, unsupervised learning techniques, and the integration of predictive and prescriptive methods. These developments allow analytics to support operational processes from end-to-end, finding known risks more efficiently, identifying previously unknown risks, and equipping managers with an evidence-based understanding of how their actions influence taxpayer behaviour.

Findings

This report has identified a number of organisational and technical challenges relating to analytics that were common across survey respondents:

- Leaders of analytics functions need to balance the need for centralised management (which supports quality control and staff development) against the need to engage and integrate with the wider business (which facilitates change management and the exchange of expertise). Survey responses suggest that centralisation may be appropriate in the early phases of development, with functions becoming increasingly de-centralised as they mature.
- Analytics departments must also find ways to manage the uncertainty and complexity associated with analytics projects, and to ensure that finished models achieve a tangible impact on operational performance. Survey responses show that most administrations handle uncertainty by using iterative, "test-and-learn" approaches in order to gather regular feedback and deliver incremental improvements. To ensure that completed models deliver tangible results, administrations are deploying a range of methods, including establishing specialist change-management units dedicated to analytics implementation.
- Beyond these organisational issues, administrations are faced with complex choices about analytics software. Although commercial software packages still pre-dominate, survey responses indicate that many administrations are beginning to experiment with open-source tools such as R. Open-source languages offer significant advantages in terms of cost, flexibility, and access to the latest algorithms and techniques, but are significantly harder to use than the tools offered

by commercial providers. To make the switch to open-source with confidence, administrations must find a way to recruit or develop staff with advanced statistical programming skills.

- Finally, the survey suggests that tax administrations should think differently about their data if they wish to realise the full potential offered by advanced analytics. Instead of seeing data as the residue of operational processes, administrations must treat it as an asset to be managed and developed actively. If analytics is to fulfil its promise to help tax administrations make better predictions and draw more robust inferences, it needs a foundation of accurate, representative datasets that capture the full facts of taxpayer characteristics and behaviour.

Recommendations

In the course of the research, a number of opportunities for further collaboration have become apparent. FTA member countries are encouraged to consider the key organisational and technical opportunities raised in the report, and to continue to share their experiences and learning in relation to advanced analytics.

In particular, it is recommended that tax administrations collaborate to share knowledge in relation to:

- Specific applications where advanced analytics techniques have been found to be effective;
- Best practices in securing analytical capabilities, with particular reference to the skills needed to deploy open-source technologies;
- Sources and types of data that have proved most productive for analytical purposes.

Bibliography

OECD (2014), *Increasing Taxpayers' Use of Self-service Channels*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264223288-en.

OECD (2016), *Technologies for better tax administration: A Practical Guide for Revenue Bodies,* OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264256422-en.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where governments work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The European Union takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

OECD PUBLISHING, 2, rue André-Pascal, 75775 PARIS CEDEX 16
(23 2016 18 1 P) ISBN 978-92-64-25644-6 – 2016

Advanced Analytics for Better Tax Administration

PUTTING DATA TO WORK

Contents

Consult this publication on line at ***http://dx.doi.org/10.1787/9789264256453-en.***

This work is published on the OECD iLibrary, which gathers all OECD books, periodicals and statistical databases.
Visit ***www.oecd-ilibrary.org*** for more information.

OECD publishing
www.oecd.org/publishing

ISBN 978-92-64-25644-6
23 2016 18 1 P

9 789264 256446