

A large, stylized graphic on the left side of the cover, resembling a pie chart or a fan. It consists of several overlapping, semi-transparent shapes in shades of yellow, orange, and teal, all pointing towards the center. The shapes are outlined in a light teal color.

**African Tax
Administration Paper 27**

**Data Management at the
Senegalese Tax Authority:
Insights from a Long-term
Research Collaboration**

Leo Czajka, Florence Kondylis
Bassirou Sarr & Mattea Stein

November 2022

African Tax Administration Paper 27

Data Management at the Senegalese Tax Authority: Insights from a Long-term Research Collaboration

Léo Czajka, Florence Kondylis, Bassirou Sarr and Mattea Stein

November 2022

Data Management at the Senegalese Tax Authority: Insights from a Long-term Research Collaboration

Léo Czajka, Florence Kondylis, Bassirou Sarr and Mattea Stein

ICTD African Tax Administration Paper 27

First published by the Institute of Development Studies in November 2022

© Institute of Development Studies 2022

ISBN: 978-1-80470-060-0

DOI: [10.19088/ICTD.2022.020](https://doi.org/10.19088/ICTD.2022.020)



This is an Open Access paper distributed under the terms of the Creative Commons Attribution 4.0 International license (CC BY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited and any modifications or adaptations are indicated. <http://creativecommons.org/licenses/by/4.0/legalcode>

Available from:

The International Centre for Tax and Development at the Institute of Development Studies, Brighton BN1 9RE, UK

Tel: +44 (0) 1273 606261 Fax: +44 (0) 1273 621202

E-mail: info@ictd.ac.uk

Web: ictd.ac/publication

Twitter: twitter.com/ICTDTax

Facebook: facebook.com/ICTDTax

LinkedIn: linkedin.com/company/ICTDTax

IDS is a charitable company limited by guarantee and registered in England

Charity Registration Number 306371

Charitable Company Number 877338

Data Management at the Senegalese Tax Authority: Insights From a Long-term Research Collaboration

Léo Czajka, Florence Kondylis, Bassirou Sarr and Mattea Stein

Summary

As they increasingly adopt digital infrastructure, public administrations worldwide are increasingly collecting, generating and managing data. Empirical researchers are, at the same time, collaborating more and more with administrations, accessing vast amounts of data, and setting new research agendas. These collaborations have taken place in low-income countries in particular, where administrative data can be a valuable substitute for scarce survey data.

However, the transition to a full-fledged digital administration can be a long and difficult process, sharply contrasting the common leap-frog narrative. Based on observations made during a five-year research collaboration with the Senegalese tax administration, this qualitative case study discusses the main data management challenges the tax administration faces. Much progress has recently been made with the modernisation of the administration's digital capacity, and adoption of e-filing and e-payment systems. However, there remains substantial scope for the administration to enhance data management and improve its efficiency in performing basic tasks, such as the identification of active taxpayers or the detection of various forms of non-compliance. In particular, there needs to be sustained investment in human resources specifically trained in data analysis. Recently progress has been made through creating – in collaboration with the researchers – a 'datalab' that now works to improve processes to collect, clean, merge and use data to improve revenue mobilisation.

Keywords: administration; data management.

Léo Czajka, Louvain Institute of Data Analysis and Modelling in economics and statistics (LIDAM).

Florence Kondylis, Development Impact Evaluation, World Bank.

Bassirou Sarr, Ministère des Finances et du Budget, République du Sénégal.

Mattea Stein, University of Naples Federico II and Centre for Studies in Economics and Finance (CSEF).

Contents

	Summary	3
	Acknowledgements	5
	Acronyms	5
1	Introduction	6
2	Context and organisation	9
	2.1 Institution	9
	2.2 Information system at the DGID	10
3	Type 1 information: who are the taxpayers?	12
	3.1 Unique taxpayer identifier	12
	3.2 Taxpayer registry	14
	3.3 Contact information	16
4	Type 2 information: how much do taxpayers earn or own?	17
5	Type 3 information: what complementary information is available on the taxpayer?	19
6	Type 4 information: how can we use the taxpayer’s declaration history?	20
7	Type 5 information: what does the law say?	21
8	Creation of a datalab	21
9	Conclusion	24
	References	25
Tables		
Table 1	GovTech maturity indexes and selected indicators for Senegal and comparable countries	10

Acknowledgements

We thank all members of the Direction Générale des Impôts et Domaines (DGID) who we have been working with. In particular we thank the Director, Mr Bassirou Samba Niasse, Deputy Director General Abdou Aziz Gueye, former general director, Mr Cheikh Ahmed Tidiane Bâ, Mr Mor Fall, Mr Serigne Mabeye Fall, Mr Amadou, Mr Abdoulaye Badiane, Mr Oumar Diop Diagne, the Director of the Intelligence department, Mr Cheikh Mouhamed Hady Dieye, the director of the IT department, Mrs Yama Kouyate Diaby, IT experts, Mr Oumar Ngom and Mr Mamadou Diangar, the director of the fiscal services directorate, Mr Lamine Sy, the director of the medium size taxpayer directorate, Mrs Ndeye Aïssatou Ndao, and the director of the large taxpayer directorate, Mr Abdoulaye Diagne. Mr Aliou Sow and Mr Mamadou Guisse also provided helpful administrative support over the years. All errors are our own.

Acronyms

ANSD	Agence Nationale de la Statistique et de la Démographie
CEI	Citizen engagement index
CIT	Corporate income tax
CTAD	Cellule d'Analyse et de Traitement des Données
DGID	Direction Générale des Impôts et Domaines (Senegalese Tax Administration)
GTEI	GovTech enablers index
LTC	Large taxpayer centre
MTC	Medium taxpayer centre
NINEA	National Identification Number of Enterprises and Associations
PIT	Personal income tax
PSDI	Public service delivery index
RPC	Regulated professions centre
SIGTAS	Standard Integrated Government for Administration System
VAT	Value added tax

1 Introduction

Information is central to tax collection. Tax administrations need it to identify taxpayers, assess their taxable base, compute their tax liability, detect non-compliant declarations, and enforce effective payment. With the rise of digital technology, the cost of collecting, storing, sharing and exploiting information has dramatically decreased. Accordingly, tax administrations from low-income countries have increasingly modernised their digital infrastructure, and accelerated the collection, production and use of data. Yet, the transition towards a high-performing, fully digitised system is a complex process – and probably more so in contexts with high prevalence for the informal sector in the economy, irregular access to electricity and the internet, and taxpayers having little familiarity with digital systems.

Over the last five years the authors have worked in collaboration with the Direction Générale des Impôts et Domaines (DGID, Senegalese tax administration) on a set of research projects. For these projects we have interviewed and worked with many members of the administration,¹ digitised vast amounts of data, helped design web applications to file some declarations online, created a uniform template for registry management, audited the quality of various data sources produced by the administration, and, finally, assisted the administration in the creation of a dedicated datalab within its intelligence department.

The duration and depth of this collaboration has been essential to our understanding the inner workings of the tax administration – on all issues relating to data creation, management and use – which inform our collaborative research agenda. Without continuous interaction with practitioners actually working within the institution, we believe our knowledge would have remained very superficial, if not completely missing key problems.

Contributing to a growing literature on the use of digital technology in tax administration,² we provide an in-depth look at the way the Senegalese tax administration collects, generates, stores and exploits information. In doing so we highlight the numerous data management challenges the administration faces, which prevent it from efficiently performing basic tasks – such as identifying active taxpayers, or detecting simple non-compliant behaviour.³ We identify recent improvements, propose further specific solutions, and describe the ongoing collaboration at the datalab to improve data management and exploitation. The objective of this case study is to gather and present the most relevant insights that we have learned during this five-year exercise, with a view to benefiting future similar collaboration.

Recent major improvements in data generation and management practices at the Senegalese tax administration include rolling out e-filing and e-payment systems across its large and medium-sized taxpayer centres since 2019, increasingly replacing – for two of these centres – a *modus operandi* that relied on the digitisation of paper-based declarations and in-person payments by taxpayers (Section 2).

¹ The DGID has been extremely collaborative throughout the years. We have had an opportunity to work with and interview several technical advisors of the General Director, as well as members of 11 tax centres (Large Tax Unit, Middle Tax Unit 1, Middle tax Unit 2, Regulated Sectors centre, Dakar Plateau, Parcelles Assainies, Ngor-Almadies, Grand-Dakar, Dakar Liberté, Pikine-Guédiawaye, and Rufisque), and five departments - the Directorate of Control and Tax Intelligence, the Information System Directorate, the Directorate for Tax recovery, the Communication Department and the Modernisation Department.

² See Okunogbe and Santoro (2022) for a recent review of this literature.

³ Related analysis was conducted by the Uganda Revenue Authority on the specific issues of the taxpayer registry quality (Mayega et al. 2019), and overall quality of information in taxpayer declarations (Mayega et al. 2021).

Yet, despite these improvements, we identify various areas in which current data generation and management remains sub-optimal, and prevent the administration from performing basic tasks efficiently (Sections 3-7):

1. The systems to create and attribute unique taxpayer identifiers are robust, but there are still issues with their use in practice – in particular, they are not used consistently in reports submitted by third parties, the system only covers firms (including self-employed) and associations, and no tax identifiers exist for employees and individuals earning other forms of income.
2. Various issues are detected with respect to the taxpayer registry, including a central registry that is out of date (with 40 per cent of taxpayers listed never having submitted a tax declaration, and an additional 20 per cent not having declared in the last four years), and the use of several parallel non-synchronised registries (plagued by issues such as duplicate entries) for day-to-day operations at the tax centres.
3. The administration experiences difficulty in physically locating taxpayers, as available contact information is incomplete – an issue compounded by the fact that only a few areas in Senegal's main cities have well identified addresses.
4. The large-scale exploitation of third-party information to help the assessment of a taxpayer's taxable base has only recently started as part of the research collaboration, and has not yet been fully mainstreamed to the systematic detection of various forms of non-compliance, and opportunities to systematically use other available information are currently not exploited.
5. Complementary information on taxpayers, such as sector of business and tax regimes is often inaccurate or incomplete, and information on taxpayers' declaration, compliance and audit history is not used to flag inactivity and non-compliance systematically.

Current advances and further scope for improvement exist on several dimensions, which we discuss in sections 3-8:

1. Existing data creation and management tools can be improved, such as harmonising and cleaning the taxpayer registry while adding a better data storage design (e.g. standardised address entry), and eliminating errors in online forms (e.g. for value added tax (VAT) filing). Work is ongoing, but there are some limits to improvements due to the structure of the existing tax administration software.
2. Improved protocols could be put in place to integrate data quality management in operational procedures, such as for systematic address and phone number verification, and cross-checks to identify information for taxpayers registering for the new online interfaces.
3. By creating e-filing options for forms that are currently still paper-based, the digitisation of forms is accelerating, but there is more scope to digitise procedures (especially with respect to the auditing process). However, the administration will not be able to fully switch to e-filing in the foreseeable future, due to insufficient IT access among a large segment of (potential) taxpayers.
4. Work is ongoing to test the use of new information sources – this includes in particular the systematic exploitation of third-party information, with similar approaches possible for tax history information. Another promising advance is increasing communication with taxpayers via email and phone – here the administration could take advantage of the substantial consolidation of individuals' phone numbers in the last couple of years due a large expansion in mobile payment in the country (propelled by the COVID crisis and lowered costs). Based on successful data exchange partnerships with the Customs Authority and Public Procurement Agency, the tax administration has become very proactive in building new relationships with other institutions, which will provide valuable new information in future.

5. To enable progress on the aforementioned issues and exploit the vast amount of data generated and collected by the administration, it is essential to invest over a long period in human resources specifically trained in data analysis. An important step was taken with the recent creation of a datalab at the DGID, together with the researchers (Section 8). This works on several of the issues and improvements discussed here, through research-oriented as well as operational projects. This has recently been named the Cellule d'Analyse et de Traitement des Données (CTAD). However, it still works on temporary and project-specific budgets. A budget allowing permanent recruitment of the needed human resources would need the Ministry of the Economy and Finance to officially recognise the data unit as a core element of DGID's structure.
6. Further progress requires broader reforms, some beyond the sphere of the tax administration. These include the creation of a system of tax identifiers linked to individuals, and the creation of a general cadaster that harmonises addresses (the latter is currently ongoing for parts of the capital city Dakar).

Finally, we also should stress that, even though some of the proposed solutions may seem relatively simple, our experience suggests that making substantive improvements within the administration is a long and very complex task that requires changing longstanding practices and sustained coordination with legal experts, tax inspectors, computer scientists, centre managers, and various front office staff.

Methodology

This case study synthesises numerous working documents we compiled while working on implementation of our projects during our long collaboration with the DGID. We only discuss here issues that we have been directly confronted with. This helps to guarantee the accuracy of what we describe, and should reinforce the relevance of our recommendations. But it also reduces our scope – the areas covered are inevitably shaped by the specific needs of our research agendas. In particular, we do not elaborate much on other crucial dimensions, such as human resource management (recruitment, training and incentives), the DGID's collaboration with other tax collecting institutions (Treasury and Customs), or the processes by which the administration chooses and develop new technological instruments.

The following discussion is structured according to the different types of information the tax administration needs in order to collect tax revenue from taxpayers. This allows us to cover all types of information we have been collecting for our projects, and closely reflects the processes by which we identified the different issues exposed here. We classify information into five groups:

Type 1 information is everything that strictly relates to the identity of the taxpayer: fiscal ID, name, postal and email address, telephone number, bank account information, as well as all other IDs produced by other institutions.

Type 2 information includes any detail on the amount and nature of the taxable base (type of capital, type of income, tax credit, etc.).

Type 3 information is all complementary information about the type of taxpayer, including legal form and economic sector for firms, age and marital status for physical persons. These elements can be useful to the DGID for tax monitoring, but are not always strictly necessary.

Type 4 information comprises all meta-information on the tax history of the taxpayer related to non-compliance, including changes in the aforementioned information and interaction with the administration. This information is termed as meta, because a pre-requisite is having information of the previous types.

Type 5 information is about the tax law itself. This is necessary to determine a taxpayer's tax liability based on the other available information.

In section 2 we present the institutional context in Senegal, and the structure of the information system at the DGID. In sections 3 to 5 we discuss how Type 1 to 3 information is collected and maintained by the DGID, and analyse the challenges the administration faces with respect to effective data management. Sections 6 and 7 elaborate on the remaining types of information. Section 8 describes how a datalab was set up at the DGID to work towards improving data management, and section 9 concludes.

2 Context and organisation

2.1 Institution

Senegal is a low-income country in West Africa, with a population of 16.5 million and stable democratic institutions. Most economic activity takes place in the agglomeration of Dakar, the capital and an important harbour in the region. In 2019, its tax-to-GDP ratio was 16.6 per cent – more than 3 percentage points below the long-term regional target of 20 per cent. Tax collection is split between three institutions – the Treasury (local taxes and taxes on civil servants), the Customs Authority, and the DGID. Our collaboration has been exclusively with the DGID. It derives most of its tax revenue from VAT, personal income tax (PIT) and corporate income tax (CIT). It directly manages around 30,000 taxpayers (mostly firms or self-employed), and we estimate it indirectly collects taxes from an additional 350,000 individuals through withholding PIT collected by firms. Thus the overall tax base covers less than 5 per cent of the adult population. Administratively, the DGID is split into 20 tax centres: one for large taxpayers (LTC), two for medium-sized taxpayers (MTC1 and MTC2), one for regulated professions, such as doctors, lawyers and architects (RPC), and finally 16 geographically-based centres. However, tax collection is heavily concentrated – with approximately 900 taxpayers, the LTC alone accounts for 80 per cent of all tax revenue collected by the DGID; the smallest 16 centres account for less than 10 per cent.

Table 1 gives a few indicators on technological adoption, government capacity and revenue for Senegal and 17 other comparable sub-Saharan countries, according to the GovTech Maturity Index (GTMI). Figures in columns 1-5 give each countries' indexes, based on Dener et al. (2021). In terms of GTMI, Senegal ranks poorly compared to the rest of the world (122nd out of 198 countries), but relatively well compared to the chosen sample of comparable countries (7th out of 18). Splitting GTMI into its four sub-components, we observe that, compared to the 17 similar sub-Saharan African countries, Senegal's Core Government Systems index (CGSI – 17 indicators measuring government-wide digital infrastructure) is mediocre. Its GovTech enablers index (GTEI, 16 indicators characterising technology-friendly regulatory and policy environments), its Citizen Engagement Index (CEI, 6 indicators capturing digital public participation), and its public service delivery index (PSDI, 9 indicators measuring digital services including e-filing and e-payment), are above the sample's mean. Ghana, Uganda and Rwanda significantly outperform Senegal in three out of four dimensions, and are probably what Senegal should be aspiring to in terms of digital transformation.

Figures in columns 6-11 are averages over 2017-2019, the most recent period for which complete information is available. Senegal outperforms most of the comparable countries in all dimensions here. Its GDP per capita is 13 per cent higher than the sample mean. It has collected most tax revenue on average over the period (16.1 per cent), closely followed by

Rwanda (15.9 per cent). Senegal's total revenue (adding government's property income, social security and grants) over GDP is greater than the sample's average. It also has the highest statistical capacity, the second highest share of population with access to the internet (39.1 per cent), and ranks fifth in terms of mobile connectivity.

Table 1 GovTech maturity indexes and selected indicators for Senegal and comparable countries

	GovTech Maturity Indexes (world rankings)					GDP per cap.	Total revenue (% GDP)	Tax revenue (% GDP)	Statistical capacity (0 - 100)	Mobile subs. (per 1,000) internet	% of pop with
	GTMI	CGSI	PSDI	CEI	GTEI						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Ghana	68	43	86	24	119	5435	13.9	12.1	75.5	134.0	51.3
Uganda	76	85	112	54	70	2161	13.1	11.7	70.0	58.4	15.2
Rwanda	99	80	87	129	85	2138	23.2	15.9	68.1	79.1	25.8
Côte d'Ivoire	104	137	132	68	112	5142	15.0	12.2	69.2	144.0	36.9
Nigeria	106	88	141	116	93	5069	7.6	4.6	57.7	93.1	33.7
Burkina Faso	108	104	150	101	88	2150	19.7	15.2	66.3	101.3	19.1
Senegal	122	138	130	119	103	3333	19.6	16.1	75.9	109.4	39.1
Benin	132	120	128	164	131	3257	13.7	10.2	64.8	87.3	22.4
Cameroon	140	144	143	148	122	3707	15.6	12.5	53.3	80.0	33.7
Togo	159	181	149	143	154	2097	17.2	12.5	71.8	77.9	19.6
Mali	163	126	162	154	178	2277	19	13.9	61.8	118.9	24.3
Madagascar	164	151	158	156	172	1558	13.3	10.3	52.5	48.7	15.0
Liberia	170	146	177	138	185	1469	27.8	11.7	56.6	35.9	22.1
Guinea	173	175	168	155	167	2559	14.9	12.9	55.1	100.9	23.6
Niger	174	156	184	131	182	1215	17.8	10.3	68.1	55.4	
Guinea-Bissau	177	177	179	178	150	1896	15.8	9.3	50.3	86.3	19.3
Gambia	179	165	187	181	144	2181	18.5	10.4	70.0	120.2	34.5
Mauritania	180	128	186	177	188	5216	21.3	14.0	58.8	104.7	37.6
Sample av.	139	130	148	130	136	2937	17.0	12.0	63.7	90.9	27.8

Countries are ordered by GTMI. Columns 1-5: ranks are computed by the authors based on indexes produced by Dener et al. (2021), out of a total of 198 countries. The GovTech Maturity Index (GTMI) is the average of 4 indexes. The core government systems index (CGSI - 17 indicators) captures the key aspects of a whole-of-government approach, including government cloud, enterprise architecture and other platforms. The public service delivery index (PSDI - 9 indicators) uses existing data from the UN Online Service Index as well as new indicators on online portals, e-filing services, e-payment capabilities, and more. The citizen engagement index (CEI - 6 indicators) measures aspects of public participation platforms, citizen feedback mechanisms, open data and open government portals. The GovTech enablers index (GTEI - 16 indicators) captures strategy, institutions, laws and regulations, as well as digital skills and innovation policies and programmes to foster GovTech.

Columns 7-8: Authors calculation from the International Monetary Fund's World Revenue Longitudinal Data. All figures are averages over the period 2017-2019. Columns 6 and 9-11: authors' calculation from World Development Indicators. All figures are averages over the period 2017-2019.

GDP per capita is in 2017 \$ PPP.

2.2 Information system at the DGID

Since 2006 the DGID has adopted and increasingly deployed an integrated tax administration system called Standard Integrated Government for Administration System

(SIGTAS).⁴ This multi-module software package allows storing the content of tax declarations, computing tax liability, processing payments and producing receipts. It is the backbone of the DGID's digital system. A user interface allows administrators to examine the situation of taxpayers case-by-case, or based on some pre-programmed criteria (e.g. a list of taxpayers that are corporations). As of early 2022, SIGTAS was being used in all tax centres apart from one (the centre in the region of Thiès). This major transformation has taken place over years, in order to not disrupt longstanding practices too abruptly, and to let members of the administration learn and adapt. It has led, among other things, to a proper data revolution, making information available at an unprecedented scale at DGID.

Now its digital infrastructure is more mature and resilient, the administration is accelerating its digital transition, adding more modules to the core system and developing fully-fledged web applications:

- **ETAX:** until 2018, all tax declarations were submitted on paper and had to be manually put into SIGTAS; two centres (the LTC and MTC1) then adopted e-filing and e-payment via a new SIGTAS module, ETAX.
- **Mon Espace Perso (MEP)** was developed by a Senegalese company and is similar to ETAX, in that it also allows taxpayers to submit their basic tax statements online. As the use of ETAX is restricted to large and medium taxpayers, MEP was proposed for all other taxpayers during 2020 through a broad communication campaign. However, MEP does not yet have an online payment solution – taxpayers registered on this platform make an online declaration, but have to go to their tax centre to pay in person. The government plans to develop a general payment solution for all taxpayers (large and small), and all tax collecting institutions (Treasury, Customs and DGID).
- **sen-etafi** is a platform where, since 2020, all large and medium firms declare their annual financial accounts. It provides valuable information – not only to assess the CIT, but also to study the economic and financial situation of Senegal's largest firms. The resulting database is accessible to the DGID and other institutions, such as the National Statistical Agency, the Ministry of the Economy and Finance, and financial regulators.
- **E-services** allow taxpayers to declare online some complementary statements for which no module has been developed yet in ETAX or MEP. These include appendix documents that provide detailed information on payments to employees and service suppliers. E-services was recently augmented with an encoding interface that will be used by the administration to digitise a wealth of information still declared on paper, such as public contracts and property transfers.
- **property census application:** in collaboration with another team of researchers, DGID has developed a digital tool to locate and value properties, and identify property owners liable to property tax (assessed by DGID, collected by the Treasury). This application is now operational and has already been used to identify thousands of properties in the region of Dakar. For more information about this project, see Knebelmann et al. (2021).

All the data generated by and stored in these systems (with the exception of sen-etafi and the property census app) can be explored by SQL requests generating tables. These tables are themselves stored in a central data portal, where data shared with DGID by the Customs

⁴ This system was developed by the Canadian company SOGEMA, which also operates in 8 other African countries - Nigeria, Ethiopia, Rwanda, Bénin, Mali, Burundi, Chad and Liberia.

Authority, the Public Procurement Agency and the National Statistical Agency is also available.

Despite these recent technological developments, DGID still needs to process countless hard-copy documents for two main reasons. First, online declaration is only legally mandatory for the LTC. Efforts to strongly encourage online declaration for MTC1 have been successful, but adoption remains low in the other 18 centres. Second, not all documents have a corresponding online form. While the second obstacle is likely to be overcome soon, the first is not. A broad switch to e-filing in small taxpayer centres is not imminent, and cannot be made mandatory in the near future. The experience of tax practitioners we interviewed suggests that most taxpayers do not have a smartphone, let alone a personal computer, and that the internet is not yet reliable in all regions (an issue for both taxpayers and the administration). These constraints impose limits on the potential cost-saving benefits from e-filing for both taxpayers and the administration.

In parallel to these centralised systems, fiscal centres resort to many ad hoc systems (on paper or in spreadsheets) to store and access a wide range of information. These parallel systems typically emerge, or survive, when the available solution (mainly SIGTAS) is incomplete, inefficient or mis-used. We discuss extensively below why separate tax registries in the form of locally maintained spreadsheets emerged, and continue to be used by tax centres to keep track of taxpayers' status (active or inactive) and contact information. There are two other important areas where these parallel arrangements are used. First, DGID's version of SIGTAS offers relatively poor functionality to monitor audit procedures. Information related to audits is thus processed outside the above-mentioned systems, mainly on paper. Second, some payments – a very small proportion by our estimation – are still only processed on paper/spreadsheets, without passing through SIGTAS.

3 Type 1 information: who are the taxpayers?

3.1 Unique taxpayer identifier

The key element in a functional taxpayer identification system is a unique fiscal identifier number, created for each taxpayer using robust production and attribution protocols. It will allow to uniquely define physical and legal entities, but only as long as it is always used in this way. This means that throughout the tax system, the same identifier cannot be used to refer to different taxpayers, and different identifiers cannot be used to refer to the same taxpayer, and, notably, that all reporting by and on specific taxpayers should routinely require reporting the corresponding unique fiscal identifier. In other words, the system will work well if all actors use it well, all the time.

The main unique identifier numbers found in Senegalese tax data are: the National Identification Number of Enterprises and Associations (NINEA), National Identity Card Number, and Register of Commerce and Personal Property Credit Number. These systems – which do not cover exactly the same group of economic actors – were designed by different institutions to uniquely identify physical and legal entities. The tax administration uses the NINEA as its main fiscal identifier, as its directly managed taxpayer population consists mainly of firms and self-employed individuals. This practice is entrenched into fiscal law. The NINEA is generated by the National Statistical Agency, and was initially created to identify enterprises and associations, independently of whether they were registered for tax or not. This system is robust, the creation and attribution of NINEAs by the National Statistical Agency is consistent and unique, and newly-enrolled taxpayers are given a NINEA before

any tax declaration or payment – at least in the vast majority of cases. This is a good example of how a well-functioning digital solution can positively impact data quality.

However, while the creation and attribution of unique tax identifiers is robust, issues can occur in their use. This weakness of data management practices is not specific to Senegal, and has been documented by Ligomeka (2019) in the case of Malawi. First, to maintain the uniqueness of tax identifiers, throughout the tax system the same identifier must be used to refer to one taxpayer only, and a taxpayer should never be referred to by different NINEAs. In practice, we found the following issues can occur:

- after a business merger both original NINEAs tend to remain in use for some time, as not all users are immediately aware of changes;
- after a change in legal form some taxpayers ask for a new NINEA, even though the business continues to exist;
- individuals obtain a NINEA registering a firm under their personal name, while later – or for certain interactions – using a business name, which can also change over time; or
- typos in the NINEA (made at a tax centre at some point) can become entrenched, and become a firm's de facto NINEA.

To our knowledge these issues affect a minority of cases, but illustrate how a well-functioning digital tool (the production of NINEA by the Agence Nationale de la Statistique et de la Démographie - ANSD) needs to be accompanied by clear and duly enforced procedures, guaranteeing consistent use across all actors (taxpayers, tax officials and other institutions).

In this respect, the move to online declaration has dramatically improved data quality overall, by allowing tax officials to force taxpayers to use correct identifiers. However, imperfect design of the application may have created new challenges. Taxpayers were, for instance, able to sign up for the E-service by only using their NINEA – in some cases they introduced typos in their NINEA, or added zeros at the end rather than the beginning. For future declarations these taxpayers then had to use this incorrect NINEA to log on to E-service, and the administration cannot link the information to other sources (losing some of the advantages of online declaration). This issue could have been avoided by improving the protocols for registering for E-service, and requiring a cross-check using additional identifying information available in the administration's taxpayer registry (see Section 3.2).

Second, so that the administration can use the information, reporting by and on specific taxpayers should routinely require reporting the corresponding unique tax identifier. Information reported by taxpayers is routinely stored with their unique tax identifier, whether it reaches the administration through online or paper-based declarations. However, the same is not true for information on taxpayers. An example encountered in our research project work relates to third-party payment declarations that, according to the tax law, should have been reported with payment recipients' unique fiscal ID. In practice this was not enforced, and a laborious and less accurate matching process using other identifying information from the registry was needed. It is impractical for the administration to do this in its operational routine. Opportunities to fully exploit the information available to the administration are then missed. One element of our research project consists of cross-checking and increased enforcement of payment-recipient declaration of the fiscal ID. In the longer term, improved design of the web application – such as additional automatic checks and warnings to taxpayers – should contribute to improving data quality.

Finally, while the current system to create and attribute unique tax identifiers is robust, it borrows from an identification system that was originally designed to identify enterprises and associations – legal entities, rather than individuals per se. This reflects the fact that the

administration is organised to only collect taxes via firms, either directly or through withholding taxes (see Section 2.1). Individuals who want to submit a PIT statement are required to register as some form of firm. Indeed, all individuals declaring PIT run a business or rent properties. Neither employees, nor individuals earning other types of capital income (interest, dividends or capital gains), have tax identifiers. In the long run, to enrol physical persons on a large scale and enforce the progressivity of PIT, the DGID will need to either use the national ID card, or design a new identification system specifically for individuals.

3.2 Taxpayer registry

To be able to interact with taxpayers, the administration needs a taxpayer registry that links the abstract fiscal identifier with other information that may or may not also uniquely identify physical and legal entities: name, phone number, email address, postal address and bank account details. Obviously, this information needs to be kept up-to-date, which relates to a key but often implicit feature of a registry – the physical and legal entities listed must be real (active and alive).

The administration is currently not able to produce an up-to-date, complete and duplicate-free list of all active taxpayers. In fact, with the exception of the centres for large and medium firms – which have both more resources and fewer taxpayers to manage – such a table can rarely be produced even at the centre-level. This situation is due, on the one hand, to low quality of the registry data stored in SIGTAS, which results from a few poorly designed features, resource and structural limitations, and inadequate use of the application at different levels. On the other hand, it is due to ad hoc work-around solutions adopted by the centres in response. We discuss four issues in more detail.

First, SIGTAS has no module to supervise the management of paper declarations, and no fields to record additional identifiers such as paper folder IDs. However, centres have to receive, digitise, store, and be able to find these paper documents later for verification or audits. In the absence of a suitable SIGTAS module, centres process this information on separate spreadsheets. Rather than using taxpayers' NINEAs to identify physical folders (one folder per taxpayer in principle), centres have adopted different folder numbering systems for practical reasons – to facilitate folder storage and access, they have assigned increasing numbers to physical folders, augmented with a few letters that encode relevant geographical or sectoral distinctions.⁵ However, these folder IDs cannot be recorded in the central SIGTAS registry.⁶ This makes it difficult to link SIGTAS information back to the corresponding paper declarations, and reduces the practical usefulness of the SIGTAS registry for the centres.

Second, as a work-around to be able to link information in SIGTAS to the corresponding paper folder, the centres started to diverge from the centralised SIGTAS registry by building locally managed spreadsheets with taxpayer names, NINEA and folder ID. These spreadsheets were increasingly used to store and update taxpayer contact information – thus turning into full registries that do not interact with SIGTAS, but have become more and more central in the day-to-day work of the centres.

This work-around has major drawbacks. Most importantly, a centre's spreadsheets are not shared via a server, but locally stored as different non-synchronised versions on, typically, several computers within the same centre. In most cases, tax centres split their registry into a

⁵ NINEAs are of limited use for these practical purposes as they are partly random and 'meaningless'.

⁶ It is not clear if this is the case because adding this field was not possible, or costly, or not deemed necessary. In the past, most centres added this information in the field intended for the taxpayer's name, but this solution caused other issues and was increasingly abandoned.

few management units (a maximum of six), each of which cover taxpayers that share a specific sector or location. This approach implies a further decentralisation of the information. When we started working with the centres to help them improve their registry management, we discovered cases of taxpayers having been assigned to more than one unit within the same centre. These duplicates lead to different physical folders with different folder IDs assigned to the same NINEAs, which inevitably leads to inconsistent appraisal of the taxpayer's situation across management units. Spreadsheet registries also suffered from inadequate design and low data quality. NINEAs were not stored systematically, allowing the existence of tens, sometimes hundreds, of undetected duplicates within the same spreadsheet (and thus assigned different folder IDs). Differences in information across management units made it difficult to aggregate it at the centre level, and, even more significantly, differences across centres prevented the administration from being able to obtain the total number and list of active taxpayers in the country by aggregating across centres.

Third, the use of separate spreadsheet registries by the centres has been further reinforced by growing inaccuracies in the SIGTAS registry. An inspection of the historical record of SIGTAS reveals that about 40 per cent of the 100,000 taxpayers it contains have never submitted a tax declaration, and an additional 20 per cent have not declared in the last four years. By exploiting information about past declarations, the administration could help tax centres distinguish those recently active among the group of pseudo-active SIGTAS taxpayers, but this has not yet been implemented. Importantly, SIGTAS itself would have probably been better designed if it had incorporated from the start a module allowing access to information on the last signs of activity more easily. Unfortunately now, from the point of view of the centres, the SIGTAS registry does not accurately represent the list of taxpayers they actually manage, which further incentivises them to rely on their own spreadsheet registries, entrenching the discrepancies between the two sources, and sometimes even exacerbating a general sentiment of mistrust towards part of SIGTAS information.

Fourth, from the point of view of the administration as a whole, keeping the SIGTAS taxpayer registry up-to-date is no simple task – difficulties come from structural and resource constraints. Taxpayers are enrolled into the SIGTAS registry based on voluntary request or as a result of on-site audits.⁷ Following registration, however, taxpayers may simply not submit any actual declarations – as is the case for 40 per cent of SIGTAS registry taxpayers, either to avoid paying taxes or because they have no revenue to declare.⁸ The burden then is on the administration, which has to find and reach out to non-declaring taxpayers – this is no easy task, given resource constraints and the quality of the available information.⁹ As a result of weak enforcement, some taxpayers can go several years without declaring, and yet be considered active by the system. A subset of firms concerned may have become truly inactive (permanently or not), but they very rarely report this spontaneously to the administration, which therefore cannot easily distinguish them from active but non-compliant taxpayers. A final difficulty seems to come from the fact that, even where some evidence of

⁷ To our knowledge it is not possible that taxpayers could have been added without any interaction with the administration, based on third-party sources only.

⁸ Mascagni et al. (2022) investigate reasons behind a different but related issue - in the case of Rwanda half of CIT declarations declare zero revenue, expenses and tax. The authors highlight as part of the explanation intensive new-taxpayer registration efforts by the authorities, combined with taxpayers being confused about the rules.

⁹ Non-compliance at this stage is sufficiently high to make it challenging for the administration to contact each of the taxpayers concerned by phone. If they refuse to declare, do not respond, or the phone number in the registries is out of date, there are only two remaining solutions. If the registries have an up-to-date bank account number, the administration can send a notification to the bank and force it to transfer money from the account, based on an estimate of the amount that would have been due if the taxpayer had declared. Yet, bank account numbers are rarely available. Then, the only remaining solution is to visit the taxpayer. At this stage, however, addresses are too often imprecise or out of date, and the administration's own courier service is not able to make this solution viable for all non-declaring taxpayers (see Section 3.3).

permanent inactivity has been collected, the taxpayers' account cannot be closed in SIGTAS if they still owe money to the DGID.

Overall this illustrates how seemingly small failures in the IT infrastructure can have extremely far-reaching consequences in a context where information is scarce and enforcement costly. Work towards overcoming some of the taxpayer registry issues described here is ongoing in the research collaboration – this is discussed in more detail in Section 8.

3.3 Contact information

By law, communication with taxpayers is only legally binding if it is done by post. Consistent addresses for all taxpayers are thus central to the DGID's tax enforcement capabilities. However, even though major improvements have been made recently, there is no systematic, complete and consistent land and housing registry in Senegal. Only a few areas of the major cities have well-identified addresses. This dramatically increases the cost of locating taxpayers. The administration does not rely on the national postal service, but has its own courier service, with two or three couriers per centre.

Taxpayer address information at the DGID can be found in four places: SIGTAS, the centre-level spreadsheet registries, the national registry of NINEAs (kept by the National Statistical Agency), and sometimes the courier's own notebooks or phones. Obviously, these sources are not necessarily consistent with one another. Due to poor data storage design (e.g. in the spreadsheet registries, all address elements are typically stored in one column), even when the information is the same, these sources are very rarely identical. This issue is even more relevant for addresses, the majority of which cannot be made precise anyway because of Senegal's incomplete addressing system. Our experience suggests that, furthermore, the address information itself frequently differs significantly between the sources, probably due to lack of systematic updating.

In the long run, ongoing projects to produce more systematic addressing information in Senegal will contribute to overcoming the constraints in locating taxpayers to deliver legally binding communications, especially in the region of the capital city Dakar. In the meantime, we believe that a short-term solution to improve the address information available to the DGID is to put in place a protocol that requires contact information to be verified each time an agent of the administration is in contact with a taxpayer, and each time a new declaration (which contains address information) is filed online or digitised.

Communications via email and phone cannot be used for legally binding procedures, yet these elements of contact information can be very useful for delivering a wide range of messages (including sending reminders of tax deadlines, obtaining information on the taxpayer, and suggesting adoption of e-filing and e-payment).¹⁰ There are usually only email addresses in registries for large and medium firms, but phone numbers are more available. Large telecommunication campaigns are becoming a more commonly used tool at DGID. Furthermore, the pilot of part of our project suggests that having a taxpayer's phone number(s) increases the chance of locating them if there is ambiguous address information. Finally, their format (much more systematic than addresses) and quasi-uniqueness make phone numbers an excellent substitute identifier when attempting to merge data sources that lack unique tax identifiers.

¹⁰ Mascagni and Nell (2022) show that, in the case of the Rwandan tax administration, email and SMS are highly effective message delivery methods.

Similarly to addresses, the quality of phone numbers available can be improved by putting in place a protocol that requires contact information to be verified at each contact between taxpayer and administration. This is a good time to put effort into updating phone numbers, as over the past couple of years mobile payment has taken off in Senegal, seemingly due to the combined effect of substantially slashed transaction costs, due to a new entrant in the market, and the authorities encouraging contactless payment methods due to the COVID health emergency. This spreading of mobile payment has led to a consolidation of individuals' phone numbers (previously it was common to change numbers and have several phone numbers with different providers), and to a clearer definition of 'the business phone number' (as opposed to the private numbers of the owner or of employees in charge of input purchases, sales, accounting/taxes, which may differ depending on the purpose of the interaction).

Overall, and as discussed for the tax identifiers, inadequate digital tools alone will not guarantee the quality of the data they collect. Well thought-out procedures must be put in place to verify and better guarantee the accuracy, quality and consistency of the contact information received.

4 Type 2 information: how much do taxpayers earn or own?

For the purposes of assessing a taxpayer's taxable base and computing the tax liability, the administration needs information on the amount and nature of both income earned and wealth owned. The administration has relatively few options for collecting this information.

The first option is to trust the taxpayer's declaration about their level of income and assets. There is plenty of opportunity to evade here, and this can only be made less likely by regular (threat of) audits – or due to general tax morale. The second option is to use direct observation of what the taxpayer owns and spends. Property tax, for instance, is convenient in information-poor environments, precisely because land and housing are easily observable (leaving aside issues of attribution and valuation). The third option is to collect information from third parties (employers, employees, suppliers, clients, banks and other institutions). Research has shown that third-party reporting is crucial for accessing reliable information on taxable bases (Pomeranz 2015; Kleven et al. 2011), can be useful for tax enforcement (Brockmeyer et al. 2019), and might have been pivotal for increasing tax-to-GDP ratio in the long run (Jensen 2022; Kleven et al. 2016).

The type of information source(s) DGID uses to assess tax bases depends both on the nature of the tax and type of taxpayer. For CIT or similar tax instruments,¹¹ the administration essentially applies two methods. For very small non-registered firms (small shops visible from the streets), it carries out an annual shop-to-shop collection exercise, which assesses total liability based on the observable characteristics of each establishment. At the other end, for duly tax-registered incorporated firms and self-employed individuals, DGID mainly derives its tax liability estimates from self-declared statements. VAT is only collected from (and remitted by) tax-registered firms, again based on self-declaration. Tax-registered employers withhold and declare payroll tax on employees, who almost never submit their own declaration. The same is true for dividends or interest. To date, tax-registered incorporated

¹¹ Mainly turnover tax for small firms, and PIT for the self-employed.

firms are the only source of information about labour income and investment income. Finally, tax-registered firms also withhold some tax from informal or foreign service suppliers.

In addition to self-declared information, the administration collects a wide range of third-party information declared by other taxpayers, or obtained from other institutions. Despite recent progress, these complementary sources are still mainly used on a case-by-case basis during document audits. Their systematic exploitation is one of the most promising avenues to strengthen tax enforcement and increase tax compliance – this is the central focus of our research collaboration. The relevant sources of third-party information are the following:

- **payments to service suppliers and rent payments:** all firms must withhold a 5 per cent tax on every transaction paid to unregistered domestic service suppliers. This tax instrument was adopted as an attempt to indirectly tax the informal sector. A similar 5 per cent withholding tax exists for rent payments to unregistered landlords of properties. Withholding firms must declare aggregate amounts every month, but on a quarterly and yearly basis they also have to list all of their landlords and service suppliers (formal and informal) with identification information, amounts paid, and, for the latter, the type of services paid for. By law, landlord and supplier identification information should include: NINEA (if it exists, personal ID otherwise), address and name; for rent payments, the rented property's address should also be provided. Until 2020 this disaggregated data was always declared on paper, thus preventing any automated use of the information. Tax inspectors did use it to try to identify non-compliant service suppliers (either large informal or under-reporting tax registered ones), but always on a case-by-case basis.

Our first project with the administration consisted of digitising the quarterly and yearly data on payments to suppliers and landlords submitted by large and medium firms since 2012. Once we started to exploit this data, it became evident that information on identification was often partial (in particular missing the NINEA), or of poor quality (in particular with respect to the address information, but also incorrect NINEAs). These issues – specifically missing or incorrect NINEAs – significantly complicate the large-scale exploitation of this data to identify non-compliant behaviour, as that requires merging it to the suppliers' self-declared revenue. We then assisted the administration in the design of an application allowing firms to submit these declarations online. This application has been deployed on E-services for large and medium firms in 2020 and 2021 respectively. This solution saves on the cost of digitisation and storage of paper declarations, and has contributed to considerably improving the quality of data by imposing rigorous formats and field validation rules. This has made it significantly easier to cross-check and exploit the data for tax enforcement purposes.

- **VAT transactions:** every month VAT-registered firms are required to submit a list of all purchases made from other VAT-registered firms, all purchases made from foreign firms, all VAT-exempt export sales, and all sales made to VAT-exempt firms. VAT-exempt firms must also submit a list of payments made to VAT-registered firms that have charged them VAT to be able to claim it back (either as a tax credit or reimbursement). Until the adoption of ETAX, there was no official form for declaring this information. An online declaration module for large and medium firms was introduced with ETAX. However, the quality of the data generated by this module is poor – some fields that would help in the systematic exploitation of the data are missing, filing instructions are not sufficiently clear and strict, and very few automatic data checks have been programmed. This was exacerbated by IT issues that occurred during the roll-out phase – firms that were unable to submit online emailed their forms directly to their tax manager. Given the wealth of information this data source contains, and following the success of the E-service module

to declare payments to service suppliers (see above), we are now helping the administration improve the collection of VAT transaction data.

- **other third-party sources:** over recent years, the DGID has set up ambitious partnerships with the Customs Authority and the Public Procurement Agency. The DGID now has access to regularly updated data at transaction level on exports, imports and public procurement. Preliminary examination of this data suggests it is of good quality, and should be extremely useful to help detect under-reporting taxpayers. Systematic exploitation of this data has just started, and some questions remain on how to interpret its content. For example, in public procurement data it is not clear whether the *date* field refers to the date of the contract signature, the bill, the effective transaction or the end of the contract. To make progress on this type of issue, a more direct line of communication needs to be established between the individuals exploiting the data in the tax administration and those producing it in partner institutions. Based on these successful partnerships, the tax administration has become very proactive in building new relationships with other institutions.

5 Type 3 information: what complementary information is available on the taxpayer?

By Type 3 information, we mean all complementary information about the taxpayer type that can be useful for tax monitoring, but which is not always strictly necessary. The most useful Type 3 information are the legal form, the tax regimes they are subject to, and the economic sector.

First, information on the legal form is hardly ever missing, and is very consistently stored and used. We believe this is because you have to know the taxpayer's legal form at every stage of the tax collection process. The nomenclature is strict, well-known by tax practitioners, and with a limited number of possible values. There is only one value per taxpayer, and it rarely changes over their lifetime. When the legal form does change, however, some taxpayers seem to request a new NINEA, which can complicate data merge across time.

Second, based on characteristics, as well as choices, taxpayers get registered into different tax regimes using SIGTAS. Some legal entities have to register for all main types of taxes (CIT, VAT, and withholding of PIT for employees and informal/foreign service suppliers). Others may be exempt from VAT and/or CIT, either permanently or during an investment phase. There are others who are exempt from all these taxes, as they are subject to a simplified flat-rate turnover tax. The general picture becomes even more complicated once we take into account the many sector- and transaction-specific taxes. Knowing all the taxes a taxpayer is or could be subject to is very complex, as the convoluted fiscal law makes it difficult to map all possibilities.

Until recently, the IT department did not extract the list of tax regimes by taxpayer from the general database – programming the SQL request is complex, and it may not have been considered worth the effort. Nevertheless, once an extraction was made, it revealed that the underlying programming of SIGTAS does not reflect legal inter-dependencies across tax regimes, and allows registration into legally inconsistent tax regimes. This will need a careful inspection, and possibly revision of the list of tax regimes that taxpayers are currently under.

Finally, information on the taxpayer's sector of business activity is also important to the administration. Not only are some taxes and fiscal policies sector-specific, but knowing the type of economic activity of a taxpayer allows the inference of crucial information, especially when combined with other details (location, duration of business activity, legal form etc.). It can be used to produce, among others, estimates of turnover, number of employees, type of client (businesses or individuals), type of capital, and vulnerability to certain shocks.

Information on the sector in SIGTAS derives from what is encoded when the taxpayer first registers for a NINEA. The nomenclature employed is based on a previous NAEMA classification,¹² with a few ad hoc add-ons. However, this information is missing for about 30 per cent of taxpayers. In parallel, some sectoral information is also stored in the centres' spreadsheet registries. This information does not follow an official nomenclature, but instead tax practitioners use the usual description (such as owner, pharmacy or restaurant) to reflect the information useful to them on a daily basis. This is often mixed with what would correspond to one or several levels of the NAEMA classification (class, group, sub-group etc.). One of the research projects is currently working on harmonising the business sector information and filling in gaps.

6 Type 4 information: how can we use the taxpayer's declaration history?

By Type 4 information we mean all meta-information derived from the tax history of the taxpayer relating to non-compliance, changes in information Types 1-3, interaction with the administration, and so on. This is the least exploited source of knowledge at the DGID for two reasons. First, by definition, it needs enough information on the other types, starting with consistent use of unique tax identifiers. Second, it is more complex to produce and manipulate.

The DGID's centralised database records all registrations, declarations and payment-related operations processed by SIGTAS, ETAX and Mon Espace Perso. Based on the data stored there, the DGID could produce various indicators that would help improve overall data quality and serve as new tools to systematically detect non-compliance. First, the ability to compare current declared amounts to past ones would help detect implausible values. Second, providing the number of missed deadlines, missing declarations or missing payments since registration would not only help to identify inactive taxpayers, but could also flag active taxpayers with a strong record of non-compliance. Third, counting the number of consecutive and total years with a null or even negative tax base (net VAT, profits, turnover, wage bill, etc.) could be useful for similar purposes. Here we believe much could be done in the short term to pick these low-hanging fruits.

Finally, a taxpayer's audit history would be a key source of information for future compliance monitoring. However, audit-related information is not part of the DGID's centralised database, and is instead mostly still kept on paper, in a physical folder for the taxpayer held at their tax centre. An application is being developed to start digitising the information generated in the auditing process. Until this can be used, it will be difficult to determine whether and when a taxpayer has been audited.

¹² NAEMA is the nomenclature of economic activities adopted by the member states of Afristat, an international organisation aiming to contribute to the development of economic, social and environment statistics and statistical capacity. Senegal is among the founding members.

7 Type 5 information: what does the law say?

A clear and complete understanding of fiscal law is required to adequately identify informal economic actors that should be paying taxes, register new taxpayers and make sure that they are enrolled for each type of tax that applies to their situation, design tax declaration statements, compute tax liability, enforce actual tax payment, and conduct audits. The source for this information is the tax code, which represents the main documentation about fiscal rules available within the administration. Heads of department and tax inspectors know the law very well – before entering the tax administration they are trained specifically on this. But a significant number of the administration’s practitioners did not receive this training. Given the complexity and scope of the tax code they have limited direct access to the information, and have to rely on explanations from those who have been trained.

Not all members of the administration need to understand the entire tax code in all its detail. However, it would be beneficial to many of them to have ready-made summary notes on key elements of the tax code. To our knowledge, there is currently only one such document that is uniformly available within the administration – a regularly updated summary of all tax statement deadlines. Producing, in addition, notes on each type of tax and the main procedures in the tax code could help in various ways. First, administrators without a background in law would not need to ask colleagues for legal help as frequently. Second, and maybe even more importantly, the production of these notes would bring to light, and hopefully help reconcile, differences in interpretation and practice across the administration.

Throughout our work with the administration, we have experienced these differences in interpretation and practice in many regards. For instance, interpretation differs regarding sufficient conditions for a physical person to have to declare PIT – some experts claim that any physical person earning a monetary revenue of any kind is supposed to submit a PIT statement, others propose that this should only be the case when the person earns some income that has not been already subject to a withholding tax, and others claim that the declaration must be submitted as soon as the person has more than one income source. Further, an example of the different practices that we encountered is that, within the boundaries of the law, tax practitioners can have different opinions about the appropriate medium to communicate with taxpayers (letter, phone or email), the content of the message, and the best timing for follow-up procedures.

8 Creation of a datalab

The culmination of the research collaboration – to date – has been the creation in 2021 of a datalab within the DGID’s intelligence department. The datalab hosts the research projects’ data analysts, anchoring the projects more firmly within the DGID. It also implements operational projects that are specifically designed for the DGID. Several of the latter are directly based on the lessons learned from the research projects, and aim to address some of data management issues highlighted in this paper. In this section we discuss the process of creating the datalab, and describe its current activities.

Since the beginning of the collaboration, the research project has relied on the expertise of local consultants, recruited with research funds but based at the tax administration. The first to be hired was an IT and database expert, who supervised the data entry operation; they were subsequently taken on by the administration in a permanent capacity. The second hired was an economist-statistician – the modernisation unit of the DGID started co-funding their

salary after it began to see the value of their data analysis output. An additional economist-statistician followed on another research project.

As the collaboration advanced, it became clear that the DGID needed these types of expertise for its own purposes. Despite collecting and generating vast amounts of data, the DGID lacked the human resources to exploit it to its potential. The data analysis produced by the research analysts for the purposes of our research projects is likely to have contributed to convincing the administration of the timely value of these specific skills. Thus, in September 2020, the DGID hired two data analysts with very complementary profiles (one specialised in economics and statistics, the other trained in programming and database architecture), under the supervision of the head of the intelligence department.

New budgets for three research projects in collaboration with the DGID allowed the recruitment of three additional data analysts in May 2021. Increasingly, this team has defined processes to store and access data, write scripts, produce reports, communicate results, and collaborate with researchers, other departments and fiscal centres. In November 2021, this emerging datalab was internally acknowledged by the DGID as the Cellule d'Analyse et de Traitement des Données (CTAD), and has since continued to grow. As of April 2022, it consists of ten individuals – eight data scientists (including trained economist-statisticians) and two data managers. However, it is still at a tenuous stage. Three of the staff members are paid via research project budgets, and specifically work on these projects. Six have been recruited thanks to a technical assistance budget earmarked for this type of work. But only one is a civil servant with a permanent contract. The datalab has in the past months suffered from the departure of two of its staff members, possibly due to the insecure status. The datalab is still a proof of concept, but only official recognition of the data unit as a core element of DGID's structure by the Ministry of the Economy and Finance would pave the way for a specific budget endowment and permanent recruitment.

Currently, the datalab works on two types of projects – research-oriented projects, financed by research funds and led by an external research team, and operational projects that are specifically designed for the DGID, sometimes based on lessons learned from the research projects. The centralisation of all research projects at the datalab guarantees that lessons learned and output produced by the projects are immediately shared with all datalab members, thus allowing considerable economies of scale and capacity-building.

There are five research projects ongoing at the moment. Our own project exploits data on payment to service suppliers, described above, to estimate the impact of direct and indirect interventions on under-reporting firms linked to large and medium firms. Another project exploits all available sources of information to identify partially-compliant rich individuals, induce them to comply with the law, and, ultimately, better enforce the progressivity of PIT. A third project studies the impact of housing property registration and taxation in the region of Dakar, using a new data collection and management application designed for the project, and now adopted by the DGID. A fourth project, now almost complete, tested algorithmic audit selection of firms on audit results. The fifth and final project studies market concentration and its impact on wages in the formal sector – while, unlike the others, this project does not directly address issues of interest to the DGID, it directly benefits from the considerable data work required.

The programme of the datalab for the coming years is very ambitious, and there are several operational projects. We describe here the activities that relate to the observations made in this paper and to which we have contributed (by highlighting the need for intervention and helping in implementation, including through research project data analyst time):

- **systematic reporting of anomalies:** as data is being analysed for the first time on a large scale, the data analysts often encounter data issues. The datalab now records every data anomaly found in a centralised database shared with the IT department. Users can specify the source and type of the problem (implausible, inconsistent, or missing value, wrong format), indicate its prevalence, provide examples and suggest ways of improvement.
- **registry harmonisation:** a reliable registry is crucial for our research project. Therefore, we had been working on registry harmonisation, in collaboration with the director overseeing the 16 small-firm centres, well before the creation of the datalab. This is an activity that has naturally transitioned to the datalab's operational projects, due to its clear operational value. We have worked to design and deploy a unified registry template to increasingly replace the idiosyncratic formats of the centres' spreadsheet registries, a short-term solution to some of the issues discussed in Section 3.2. The unified template contributes to more efficient management of the registry (removing duplicates, increasing accuracy and completeness of information, better identifying inactive taxpayers, etc.). The format has been revised many times to find the best balance between the need for harmonisation at the DGID-level, and centre-specific demands. It has now been adopted in many, but not all, of the centres. The adoption process is slow because it involves replacing longstanding practices that had been followed by a large group of tax administrators based in different centres.
- **improvements of use-case reports:** building from the data on declarations and payments collected daily, use-case reports available on DGID's data portal allow tax administrators to verify in a few clicks the situation of a group of taxpayers based on specified criteria. In collaboration with the IT department, the datalab is working to make new types of reports available that will allow a more elaborate and relevant use of the raw data. We have been working, for instance, on a new type of report to monitor non-declaration, suggesting adding information on the date of the last declaration. This would allow the centres to target their actions on different types of failure to declare (lateness of the current declaration while the previous one is there; a declaration was made but some time ago; the taxpayer has never declared). We believe several reports can be improved in this vein, making the most of still under-exploited Type 4 information.
- **sectoral monographies:** some elements of the fiscal system are specific to certain sectors of the economy, and some sectors are economically strategic. Thus there are regular demands from the administration to specifically investigate the tax performance of these sectors. Members of the datalab contribute to these studies by preparing the relevant data and producing descriptive statistics that may later inform decision-making at a high level.
- **report on tax collection:** each year, the directorate of tax collection within DGID produces a very detailed report on the amount of tax collected by type of tax, centre and month. We have merged these reports into one central database that the datalab now regularly updates and explores to study the evolution of the performance of centres and tax instruments. This work allowed, for instance, precise measurement of how concentrated tax revenue is among the largest firms, to identify what types of tax bring the least revenue, and to assess how performance varied across regions in response to the COVID shock.
- **other projects:** at the suggestion of an enthusiastic top managing team, several other operational projects have been initiated by the datalab over the last year that we are not directly involved in. These include the extraction of information about unknown taxpayers

using web-scraping techniques, design of an application to digitise a selection of strategic documents not taken into account by existing applications, monitoring the design and deployment of new data related applications by the IT department, participation in the household property census, and consolidation of the data for the land and housing registry.

9 Conclusion

Based on observations made during a five-year research collaboration with the tax administration in Senegal, this paper provides detailed insights on the challenges the administration faces to manage and exploit the wealth of information that it increasingly generates and collects as digitisation intensifies. We explain why, despite recent major improvements, the tax administration experiences difficulty in identifying taxpayers, contacting them, accurately measuring their taxable base, and detecting non-compliant declaration behaviour. We discuss some potential solutions, and describe the recent creation of a datalab within the administration as a first step towards creating an effective data management and analysis infrastructure.

Obviously, creating a datalab is not in itself enough to solve all of the data management issues highlighted in this paper. However simple some of the proposed solutions may look, we must stress that making substantive improvements within the administration is a long and complex task that requires coordination with legal experts, tax inspectors, computer scientists, the centres' managers, and various front office staff – and changing longstanding practice.

References

- Brockmeyer, A., Smith, S., Hernandez, M. and Kettle, S. (2019) 'Casting a wider tax net: Experimental evidence from Costa Rica', *American Economic Journal: Economic Policy* 11(3): 55-87
- Dener, C., Nii-Aponsah, H., Ghunney, L. and Johns, K. (2021) *GovTech Maturity Index: The state of public sector digital transformation*, World Bank Publications
- Jensen, A. (2022) 'Employment structure and the rise of the modern tax system', *American Economic Review* 112(1): 213-34
- Kleven, H., Knudsen, M., Kreiner, C., Pedersen, S. and Saez, E. (2011) 'Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark', *Econometrica* 79(3): 651-692
- Kleven, H., Kreiner, C. and Saez, E. (2016) 'Why can modern governments tax so much? An agency model of firms as fiscal intermediaries', *Economica* 83(330): 219-246
- Knebelmann, J., Pouliquen, V. and Sarr, B. (2021) 'Developing and adopting new digital tools for property tax reform in Dakar, Senegal', International Centre of Taxation and Development, blog
- Ligomeka, W. (2019) *Assessing the Performance of African Tax Administrations: A Malawian Puzzle*, ICTD Working Paper 14, Brighton: Institute of Development Studies
- Mascagni, G. and Nell, C. (2022) 'Tax Compliance in Rwanda: Evidence from a Message Field Experiment', *Economic Development and Cultural Change* 70(2): 587-623
- Mascagni, G., Santoro, F., Mukama, D., Karangwa, J. and Hakizimana, N. (2022) 'Active Ghosts: Nil-filing in Rwanda', *World Development* 152: 105806
- Mayega, J., Ssuuna, R., Mubajje, M., Nalukwago, M. and Muwonge, L. (2019) *How Clean Is Our Taxpayer Register? Data Management in the Uganda Revenue Authority*, ICTD African Tax Administration Paper 12, Brighton: Institute of Development Studies
- Mayega, J., Waiswa, R., Nabuyondo, J. and Nalukwago, M. (2021) *How Clean Are Our Taxpayer Returns? Data Management in Uganda Revenue Authority*, ICTD African Tax Administration Paper 24, Brighton: Institute of Development Studies
- Okunogbe, O. and Santoro, F. (2022) *The Promise and Limitations of Information Technology for Tax Mobilisation*, ICTD Working Paper 135, Brighton: Institute of Development Studies
- Pomeranz, D. (2015) 'No taxation without information: Deterrence and self-enforcement in the value added tax', *American Economic Review* 105(8): 2539-69