

2020

The Workplace Skills Plan and Annual Training Report: An Initial Review of the Data to Determine a Way Forward

Labour Market Intelligence
research programme



higher education
& training
Department:
Higher Education and Training
REPUBLIC OF SOUTH AFRICA



2020

The Workplace Skills Plan and Annual Training Report: An Initial Review of the Data to Determine a Way Forward

Labour Market Intelligence
research programme

Authors:

Morné Oosthuizen and Tim Köhler

Development Policy Research Unit (DPRU), School of Economics, University of Cape Town

Corresponding author:

Morné Oosthuizen (morne.oosthuizen@uct.ac.za)



© Published in 2020 by:

Department of Higher Education and Training

Private Bag X174

Pretoria 0001

www.dhet.gov.za

Contents

PART 1

INTRODUCTION AND BACKGROUND	1
------------------------------------	----------

PART 2

THE DATA	2
-----------------	----------

2.1 Potential analysis using WSP/ATR data	3
2.2 State of the current data	6
2.3 Potential analysis using existing WSP/ATR data	9

PART 3

CONCLUSION AND PROPOSED WAY FORWARD	15
--	-----------

List of tables

Table 2.1: WSP/ATR submissions over time, as received from the DHET, by SETA	8
Table 2.2: Review of WSP/ATR submissions	10

PART 1

Introduction and background



The Workplace Skills Plan (WSP) and Annual Training Report (ATR) are important tools for collecting information pertinent to skills planning in South Africa. However, to date, their potential has arguably not been realised, and there exists no national-level analysis of this data.

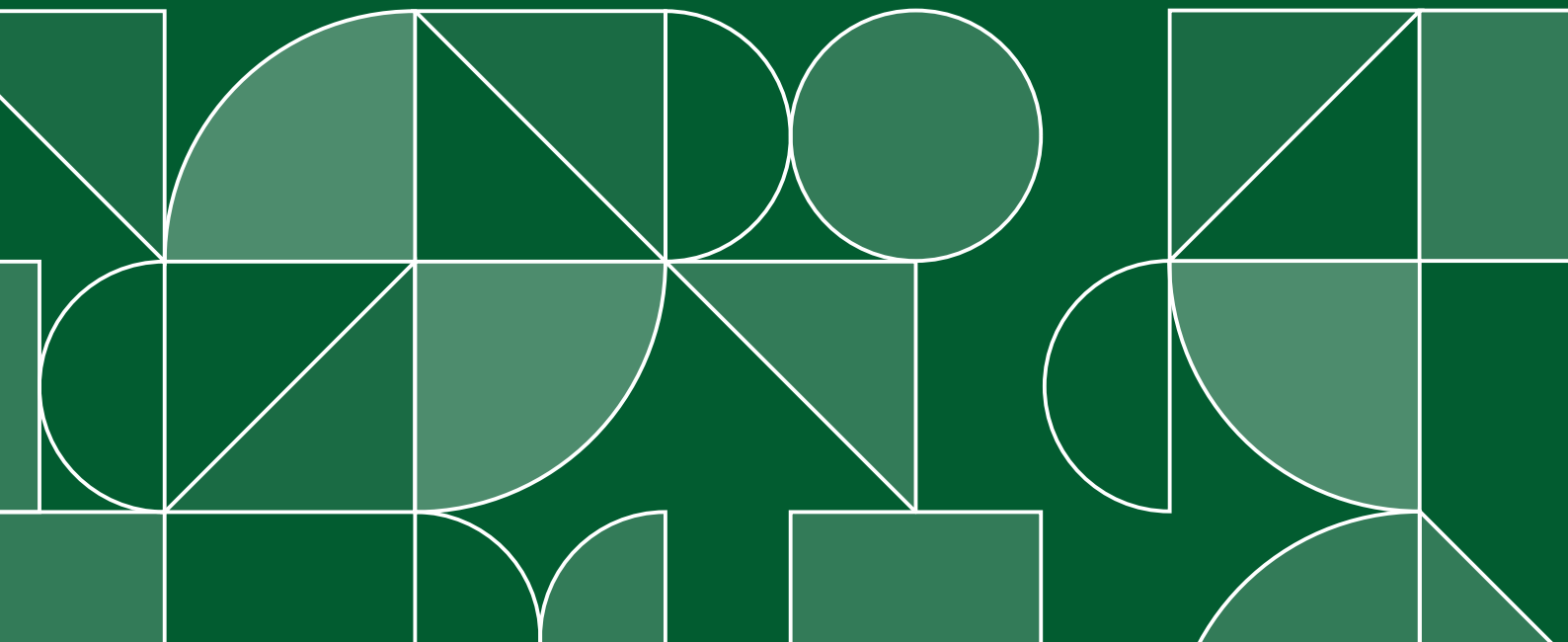
The minimum requirements for the WSP/ATR data submissions are detailed in Annexure 2 of Government Gazette No. 35940, published on 3 December 2012. These requirements are technically *minimum* data requirements and differ slightly for private and public sector employers. Data are submitted by skills development levy-paying employers to their respective Sector Education and Training Authority (SETA), which collates the data and submits them to the Department of Higher Education and Training (DHET). This process occurs on an annual basis. However, beyond Annexure 2, there seems to be little by way of guidelines or requirements pertaining to the process of data submission. Thus, for whatever reason, the quality of this data is viewed with some scepticism.

Given that considerable resources are invested by SETAs and the DHET in the process of collecting this data, and that employers are incentivised by a 20 percent rebate on the Skills Development Levy for submitting the data, it stands to reason that the submissions must be valuable to SETAs and the DHET. The concern is that this value is not properly reflected in the use of the data.

The aim of this brief report is twofold. First, it provides a review of the WSP/ATR data submissions between 2016 and 2018, highlighting the key issues with the data as received. Second, it recommends a way forward based on the findings of the review.

PART 2

The data



2.1 Potential analysis using WSP/ATR data

A major design constraint of the Annexure 2 data collection requirements is that data must be collected in tabular form. There are various arguments for and against the collection of data in this format, but the decision to collect summarised data necessarily limits the types of analysis that are possible. This decision has been debated in various fora and through various processes, and we do not address it further here.

While the data collected is at the level of employers, it is difficult to see any useful employer-level analysis that can be performed using it – that is, unless one is simply interested in tracking the performance of a particular employer.¹ The key reason for this is that the data are lacking in critical details pertaining to employer characteristics – such as financial performance, capital-intensity, the extent to which the employer may straddle industries, and whether the employer engages in international trade – any of which would allow an analysis of skills development in relationship to these characteristics. Naturally, if such information might be readily merged into this dataset – from the South African Revenue Service or other administrative databases – there would be potential for additional analysis.

As things stand, the data collected through the WSP and ATR can be used in various ways. These include the estimation of:

- Employment by occupation, demographic characteristics, and location;
- The volume of planned training of the employed and unemployed by occupation and level;
- The number of employees actually trained by occupation, demographic characteristic, and location;
- The number of people who completed training by occupation, educational institution, and level;
- The percentage of payroll spent on training across employers;
- The volume of planned PIVOTAL (professional, vocational, technical, and academic learning) training of the employed and unemployed by occupation, PIVOTAL programme, National Qualifications Framework (NQF) level, demographic characteristic, and location;
- The number of employees who actually received PIVOTAL training by occupation, PIVOTAL programme, NQF level, demographic characteristic, and location;
- The number of verified beneficiaries of PIVOTAL training by occupation, PIVOTAL programme, NQF level, demographic characteristic, and location; and
- The number of employed and unemployed individuals who received PIVOTAL training by occupation, PIVOTAL programme, education institution, and level.

¹ Tracking an individual employer over time may be useful in at least three ways. First, by doing so it may be possible to detect data problems or the need for additional assistance from the SETA if there are significant variations in particular variables over time. Second, it would allow the DHET or the SETAs to assess the extent to which planned training materialises and begin to identify serial underperformance, which may reflect any number of problems relating to the employer's business environment, their ability to supply the required data, or bottlenecks in their training system. Third, if a set of useful indicators could be identified, it would be possible to set up a system whereby employers received a report on their performance based on these indicators. If properly designed, this may help build support from employers for better-quality data submissions.

In terms of assessing trends in employment and skills development, perhaps the most useful level of aggregation possible within this data relates to geography (based on the local municipality fields in Sections B through F of the template), industry (based on the Standard Industrial Classification [SIC] code requested in Section A), and SETA. It is also possible to analyse particular aspects of the data according to occupation, based on occupational codes. However, in order to do this kind of analysis the data must be of good quality.

Data quality can be defined in different ways and from subjective and objective perspectives. Further, good-quality data can be argued to have a number of specific characteristics, including but not limited to accuracy, completeness, credibility, usefulness, timeliness, and accessibility, among others. We focus here on three criteria that bear most relevance to the question of what type of analysis can be conducted using the data already collected. The data criteria selected – completeness, accuracy, and consistency – are each discussed in further detail below.

CRITERION I: THE DATA MUST BE COMPLETE

In order for data users to make optimal use the WSP/ATR data, it stands to reason that the dataset must be complete, since any gaps in the dataset will diminish the ability of an analyst to see the “full picture”. Ideally, the dataset should cover the universe of relevant training in the South African economy, with the obvious caveat that this only includes employers legally required to submit WSP/ATR data.

There are various levels to the notion of completeness. At the most fundamental level, all employers legally required to submit data to the WSP/ATR should be providing the information to their SETAs. Without this basic function in place, completeness is impossible to achieve. However, it is clear that currently not all employers are submitting the required data and, though this is a common problem in data collection efforts of this nature, SETAs do not consistently reflect the extent of this failing in their submissions.

A further aspect of completeness relates to the extent to which employers provide complete information. Obvious as this point may be, employers should be filling in *all* of the fields listed in the template, and all this information should be passed on by the SETAs to the DHET.

Indeed, the transmission of data from SETAs to the DHET brings us to our final point about completeness. It is imperative that all SETAs submit their collated data to the DHET. Without this, it is impossible to derive a national-level picture of the training conducted among levy-paying employers in South Africa. Addressing this problem requires that the DHET actively interrogate the data submissions from SETAs in a timely manner, and engage with SETAs when issues with the data are identified.

CRITERION II: THE DATA MUST BE ACCURATE

Clearly, it is important that collected data be accurate. Without accurate data, compelling analysis is not possible and the data are rendered useless from the perspective of research that aims to inform policy.

In order for the entire dataset to be accurate, each employer should be submitting accurate information, as requested by the template. An important determinant of accuracy is the manner in which employers provide data and this, in turn, is impacted by the extent to which they feel that the data collection has a worthwhile purpose. A further factor affecting accuracy is the extent to which questionnaires are phrased in unambiguous terms that are clearly understood by respondents.

Focusing on accuracy also highlights the central role that SETAs occupy in the data collection process. SETAs bear responsibility in this regard, in terms of motivating employers to make their submissions; providing clarifications to employers who are uncertain about how to fill in the template; and performing a data-checking or verification function. SETAs are best placed to fulfil these functions given their role as intermediaries between employers, on the one hand, and the DHET and the broader skills development system, on the other.

Linked to this data-checking function, SETAs also coordinate the collation of data submissions for their respective sectors. Any errors that might occur during this process will compromise the accuracy and completeness of the data that the DHET receives, and this again highlights the importance of the DHET's role in overseeing the entire process and evaluating the data that is submitted. Therefore, data accuracy requires close co-operation between the DHET and the SETAs, and between the SETAs and individual employers, ensuring that information flows properly between employers (as data providers) and the DHET (as data requester).

CRITERION III: THE DATA MUST BE CONSISTENT

The idea of data consistency broadly has two components: consistency across actors, and consistency across time. A problem with either component has the potential to significantly compromise the quality of the collected data as a whole.

Consistency across actors speaks to the need for a common understanding and a common approach to WSP/ATR data collection from employers, SETAs, and the DHET. Without this type of consistency, different employers or different SETAs may have different interpretations of key terms, of aspects of the data collection instrument, or even of the relative importance of individual questions or sections of the template.

To improve consistency across actors (and therefore within the dataset), the need for clarity is again paramount. This can be promoted by including, for example, detailed and clear instructions and delineations of reference periods, definitions of key terms, and even pre-completed example rows in the tables. Further, it should be clear to employers that where a response is zero (for example, if an employer has no disabled Asian employees), they should fill in zero and not leave it blank.

Consistency across time requires ensuring that any changes to the data collection instrument do not create problems for analysts wanting to compare data between two or more time periods and is also critical to the function of analysing trends and making accurate comparisons over time. On this point, a key advantage of the WSP/ATR is that the instrument has not been subject to significant change over time, either through additions or subtractions of individual items or through changing definitions. Changing definitions are particularly challenging to deal with and typically introduce breaks in the time series of data that make accurate time comparisons difficult, if not impossible.

Consistency also refers to various aspects of process, including the type of support offered by SETAs to employers in making their submissions. As part of the annual data collection process, submissions made by SETAs to the DHET should be consistent with each other (and with the data collection instrument) and should facilitate the creation of a national dataset. This requires that a clear template be provided to SETAs for the purpose of data submissions to the DHET so that, once the data has been checked by the DHET, it is a straightforward process to create the final national dataset.

2.2 State of the current data

In order to assess a potential national-level analysis of workplace skills training in South Africa using the WSP/ATR data that has already been collected, a review of the individual submissions was conducted. The initial plan had been to use the individual submissions to create Stata datasets to evaluate, but it soon became clear that this would not be possible for the vast majority of the datasets. Instead, we have reviewed the datasets in the format in which they were provided to us, and have identified the key issues across them. These issues are outlined below. It is important to note, however, that individual datasets do not necessarily suffer from all of these issues simultaneously.

Variable identification problems. Data submitted by the SETAs to the DHET are collected on the basis of a data collection instrument (or template). This means that the fields in the submission should clearly correspond to the fields in the instrument. However, this is not the case in a number of submissions, making it impossible to identify which question particular parts of the submission are responding to. This is particularly problematic for those sections of the instrument that are identically formatted (for example, Sections B1, C1, and D1 are identically structured with identical fields, as are Sections E1, E2, F1, and F2) and where SETAs have not properly identified which section a particular sheet corresponds to.

Submission organisation problems. Almost every submission to the DHET has been made in the form of an Excel spreadsheet. These spreadsheets do not have a prescribed format and, as a result, heavy reliance is placed on SETAs' clearly identifying the links between the data collection instrument and the contents of the spreadsheet. A number of submissions do not clearly identify which section of the data collection instrument is linked to each sheet within the submission.

Missing or additional variables. Apart from issues related to actually identifying variables or sections, submissions do not always match the data collection instrument. Specifically, submissions either include only a subset of the required fields, or include additional fields that do not form part of the instrument at all. Missing fields compromise the completeness of the final dataset, while the inclusion of additional fields suggests that submissions may not be receiving sufficiently close attention.

Private–public differentiation. Submissions do not always contain the necessary information – a dummy variable – to distinguish between private and public employers. Because public and private employers use slightly different, but largely overlapping, questionnaires, such a dummy variable is important in discerning whether information is missing because the employer omitted to provide it or because the question is not part of the relevant questionnaire.

Inconsistent units of analysis. The data collection instrument has several units of analysis, with the most basic being the employer. Some data (for example, the employment summary) are collected by Organised Framework for Occupations (OFO) code within a given employer, meaning it should be possible to use the OFO code as a unit of analysis by aggregating across employers. Currently, while most submissions are provided at the employer-level, some are provided as employee-level data. While employee-level data may be more useful in a number of respects, it is clear that collecting all WSP/ATR data at the employee-level is simply not feasible. Importantly, this type of inconsistency introduces various challenges while offering little in the way of benefits for the DHET. Employee-level data, for example, breaks the structural link between the data collection instrument and the data submission and requires that the DHET consolidate the employee-level data into the standard employer- and OFO-code-level formats. It further places strong reliance on the SETA to clearly identify relevant fields in their submission. Since the DHET has access to employee-level data for only a subset of SETAs, this type of data is not particularly useful for the DHET's purposes in terms of the WSP and ATR. The choice here for the DHET is whether to stipulate a format for both types of submissions (employer- and employee-level data) so that the two are easy to consolidate, or whether to stipulate a single format for data submissions at the level of employers and require SETAs to consolidate their employee-level data before submission.

Incomplete submissions from employers. Employers do not seem to be completing all sections of the data collection instrument, which compromises the completeness of the final dataset.

What are the implications, then, of these and other problems with the existing WSP/ATR submissions in terms of the three criteria highlighted in Section 2.1?

CRITERION I: THE DATA MUST BE COMPLETE

Working on the assumption that the data provided to us represents all the submissions to the DHET over the 2016–2018 period, it is clear that the current set of submissions covers only a subset of the data universe in any given year (Table 2.1). The table below shows that, for 18 out of the 21 SETAs, there was at least one data submission provided over the three-year period, while just 10 SETAs provided submissions in each of the three years. For the INSETA, MQA, or Services SETA, no data at all were provided by the DHET. Five SETAs have submitted two years' worth of data, while three have only submitted data for a single year (2016). This varying coverage on its own makes sensible national-level analysis impossible and renders comparisons over time based on the full annual datasets meaningless. This alone is a critical flaw.

TABLE 2.1: WSP/ATR submissions over time, as received from the DHET, by SETA

	2016	2017	2018	TOTAL
AgriSETA	X	X	X	3
BANKSETA	X			1
CATHSSETA	X	X	X	3
CETA	X	X		2
CHIETA	X	X	X	3
ETDP SETA		X	X	2
EW SETA	X			1
Fasset	X	X	X	3
FoodBev SETA	X	X	X	3
FP&M SETA	X	X	X	3
HWSETA	X	X		2
INSETA				0
LGSETA	X		X	2
merSETA	X			1
MICT SETA	X	X		2
MQA				0
PSETA	X	X	X	3
SASSETA	X	X	X	3
Services SETA				0
TETA	X	X	X	3
W&R SETA	X	X	X	3
Total	17	14	12	43

CRITERION II: THE DATA MUST BE ACCURATE

It is difficult to determine the accuracy of the submitted data at this point, given that the data are between two and four years old already. Indeed, assessment of data accuracy is beyond the scope of this review. However, given that oversight and rigorous assessment of the final submissions do not seem to have been prioritised, there is clearly potential for significant problems with data accuracy.

CRITERION III: THE DATA MUST BE CONSISTENT

There are also significant problems in terms of the consistency of the data. While this review of the data does not consider the extent to which there is, or is not, a common understanding and approach across employers, SETAs, and the DHET, several inconsistencies have been identified.

Comparing almost any two submissions, it is clear that there are inconsistencies with the way in which SETAs submit their data. These inconsistencies range from the most basic level, such as the structure of the Excel spreadsheet, to more complex levels, such as whether the information refers to employer- or employee-level data. Further complicating the compilation of a national dataset from the submissions is the fact that individual SETA submissions are inconsistent over time (i.e., the format of the submission changes for a given SETA between years). It should not be the DHET's responsibility to wrangle the data submissions into a suitable format for integration into a national-level dataset. Instead, through clear upload specifications and appropriate engagements with the SETAs, data should arrive at the DHET in a format that makes analysis simple and straightforward.

The problem of missing or additional variables within SETA data submissions and incomplete submissions from employers further compromises consistency.

Table 2.2 presents an overview of the problems found in the various submissions as received from the DHET. The first four problems – variable identification problems, submission organisation problems, missing or additional variables, and private–public differentiation – are critical problems from the perspective of using the data for analytical purposes. The latter three are issues relating to integrating SETA submissions into a national dataset (inconsistent format and separate datasets) and the conciseness of the datasets (repeated information).

2.3 Potential analysis using existing WSP/ATR data

Given the problems of falling submission rates over the three years, comparisons using a combined national-level dataset are not possible. Further, given the problems evident in most of the submissions with respect to identifying which variable in a given dataset corresponds to which question in the data collection instrument, it will be difficult to generate a robust cross-sectional analysis even within a single year, let alone for all three years.

While it may be possible to use the existing data to construct a very limited overview of aggregates for 2016, we would be missing information for one-quarter of the SETAs. Moreover, even though only 4 out of 21 SETAs did not provide information in 2016 (nearly one-fifth of the total), these SETAs cover relatively large and important sectors (education and training, insurance, mining, and services). Furthermore, it must be noted that of the 17 SETAs that submitted data for 2016, 12 have variable identification problems and 9 have submission organisation problems. These problems will be difficult to resolve given that the data are now four years old and the SETAs may no longer be able to assist in addressing them.

Simply put, the investment of time and resources into constructing a dataset for 2016 is unlikely to yield any tangible benefit for the DHET, given the age of the data and the fundamental problems in the submissions.

The only other option would be to exclusively select SETAs that do have comprehensive information for a given year and combine their data. However, this would not allow for any meaningful trend analysis and, frankly, would be of very limited use.



TABLE 2.2: Review of WSP/ATR submissions

SETA	YEAR	VARIABLE IDENTIFICATION PROBLEMS	SUBMISSION ORGANISATION PROBLEMS	MISSING OR ADDITIONAL VARIABLES	PRIVATE-PUBLIC DIFFERENTIATION	INCONSISTENT FORMAT	SEPARATE DATASETS	REPEATED INFORMATION	COMMENT
AgriSETA	2016	X	X		X		X	X	Datasets provided in folders by firm size, each folder containing multiple Excel files corresponding to specific sections of the WSP/ATR. Even if clearly structured, it requires considerable effort to create a single dataset.
	2017	X	X		X			X	
	2018	X	X		X			X	
BANKSETA	2016			X					Dataset includes only employment summary.
	2017	-	-	-	-	-	-	-	Unit record submitted to the DHET but not shared with DPRU due to challenges with creating unique identifiers.
	2018	-	-	-	-	-	-	-	Unit record submitted to the DHET but not shared with DPRU due to challenges with creating unique identifiers.
CATHSSETA	2016	X	X	X					
	2017	X	X					X	Structure of this dataset and variable names different compared to 2018 and 2016 submissions.
	2018	X	X	X				X	
CETA	2016								Good organisation of data.
	2017	-	-	-	-	-	-	-	No data in folder: only PIVOTAL list available, but no completed WSP/ATR data.
	2018	-	-	-	-	-	-	-	No submission.

SETA	YEAR	VARIABLE IDENTIFICATION PROBLEMS	SUBMISSION ORGANISATION PROBLEMS	MISSING OR ADDITIONAL VARIABLES	PRIVATE-PUBLIC DIFFERENTIATION	INCONSISTENT FORMAT	SEPARATE DATASETS	REPEATED INFORMATION	COMMENT
CHIETA	2016	-	-	-	-	-	-	-	Data not in Excel format.
	2017	-	-	-	-	-	-	-	Data in folder is for 2018, and only includes minimal information on employer profile, employee profile, and hard-to-fill vacancies.
	2018	X	X	X			X		Datasets provided in folders by firm size, each folder containing multiple Excel files corresponding to specific sections of the WSP/ATR. Even if clearly structured, it requires considerable effort to create a single dataset. Submission includes data for more than one year.
ETDP SETA	2016	-	-	-	-	-	-	-	Data submitted but not consolidated and therefore not used for this research.
	2017	X					X	X	Datasets provided in folders by firm type (private, public, small, and PBE). Even if clearly structured, it requires considerable effort to create a single dataset. Good organisation of data with respect to the structure of the WSP/ATR.
	2018	X					X		Datasets provided in folders by firm type (private, public, small, and PBE). Even if clearly structured, it requires considerable effort to create a single dataset. Good organisation of data with respect to the structure of the WSP/ATR.
EW SETA	2016	X	X				X		
	2017	-	-	-	-	-	-	-	Unit record submitted to the DHET but not shared with DPRU due to challenges with creating unique identifiers.
	2018	-	-	-	-	-	-	-	Unit record submitted to the DHET but not shared with DPRU due to challenges with creating unique identifiers.



SETA	YEAR	VARIABLE IDENTIFICATION PROBLEMS	SUBMISSION ORGANISATION PROBLEMS	MISSING OR ADDITIONAL VARIABLES	PRIVATE-PUBLIC DIFFERENTIATION	INCONSISTENT FORMAT	SEPARATE DATASETS	REPEATED INFORMATION	COMMENT
Fasset	2016	X					X		
	2017	X	X	X					
	2018	X		X					
Foodbev SETA	2016	X		X			X		
	2017	X							Data in 2017 folder appears to be for 2018.
	2018	X							Data in 2018 folder appears to be for 2019.
FP&M SETA	2016	-	-	-	-	-	-	-	File cannot be opened.
	2017	X	X						Data in 2017 folder appears to be for 2018.
	2018	X	X						Data in 2018 folder appears to be for 2019.
HWSSETA	2016	X	X	X			X		
	2017	X	X	X					
	2018	-	-	-	-	-	-	-	Unit record submitted to the DHET but not shared with DPRU due to challenges with creating unique identifiers.
INSETA	2016	-	-	-	-	-	-	-	Unit record submitted to the DHET but not shared with DPRU due to challenges with creating unique identifiers.
	2017	-	-	-	-	-	-	-	Unit record submitted to the DHET but not shared with DPRU due to challenges with creating unique identifiers.
	2018	-	-	-	-	-	-	-	Unit record submitted to the DHET but not shared with DPRU due to challenges with creating unique identifiers.

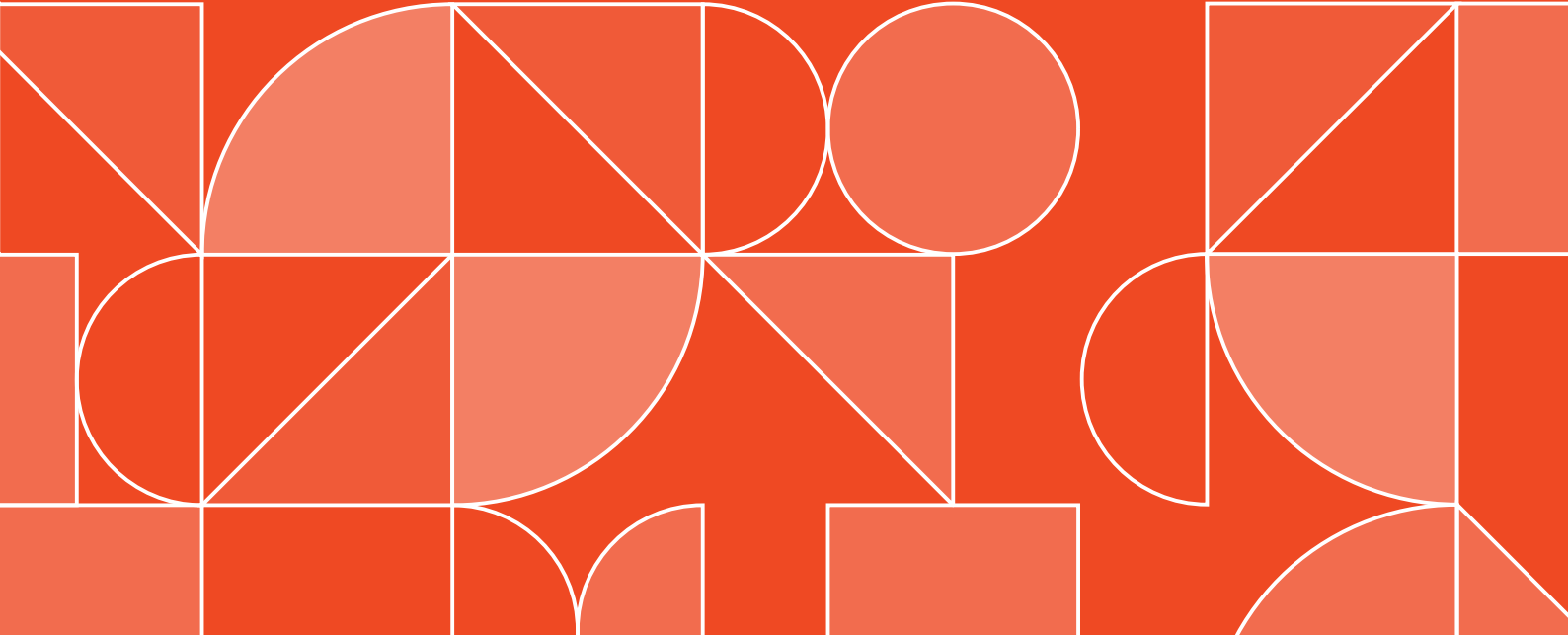
SETA	YEAR	VARIABLE IDENTIFICATION PROBLEMS	SUBMISSION ORGANISATION PROBLEMS	MISSING OR ADDITIONAL VARIABLES	PRIVATE-PUBLIC DIFFERENTIATION	INCONSISTENT FORMAT	SEPARATE DATASETS	REPEATED INFORMATION	COMMENT
LGSETA	2016		X	X					
	2017	-	-	-	-	-	-	-	Unit record submitted to the DHET but not shared with DPRU due to challenges with creating unique identifiers.
	2018	X							
meSETA	2016	X	X						
	2017	-	-	-	-	-	-	-	Unit record submitted to the DHET but not shared with DPRU due to challenges with creating unique identifiers.
	2018	-	-	-	-	-	-	-	Folder empty.
MICT SETA	2016	X	X						
	2017	X	X				X		
	2018	-	-	-	-	-	-	-	Unit record submitted to the DHET but not shared with DPRU due to challenges with creating unique identifiers.
MQA	2016	-	-	-	-	-	-	-	Unit record submitted to the DHET but not shared with DPRU due to challenges with creating unique identifiers.
	2017	-	-	-	-	-	-	-	Unit record submitted to the DHET but not shared with DPRU due to challenges with creating unique identifiers.
	2018	-	-	-	-	-	-	-	Unit record submitted to the DHET but not shared with DPRU due to challenges with creating unique identifiers.
PSETA	2016	X		X			X		Good organisation of data with respect to the structure of the WSP/ATR.
	2017	X		X					Good organisation of data with respect to the structure of the WSP/ATR.
	2018	X					X		Good organisation of data with respect to the structure of the WSP/ATR.



SETA	YEAR	VARIABLE IDENTIFICATION PROBLEMS	SUBMISSION ORGANISATION PROBLEMS	MISSING OR ADDITIONAL VARIABLES	PRIVATE-PUBLIC DIFFERENTIATION	INCONSISTENT FORMAT	SEPARATE DATASETS	REPEATED INFORMATION	COMMENT
SASSETA	2016	X	X						
	2017	X	X	X					
	2018	X	X						
Services SETA	2016	-	-	-	-	-	-	-	Unit record submitted to the DHET but not shared with DPRU due to challenges with creating unique identifiers.
	2017	-	-	-	-	-	-	-	Unit record submitted to the DHET but not shared with DPRU due to challenges with creating unique identifiers.
	2018	-	-	-	-	-	-	-	Unit record submitted to the DHET but not shared with DPRU due to challenges with creating unique identifiers.
TETA	2016	X	X						
	2017	X	X	X	X				
	2018	X	X		X				
W&R SETA	2016	X			X				Good organisation of data with respect to the structure of the WSP/ATR.
	2017	X		X					Good organisation of data with respect to the structure of the WSP/ATR.
	2018	X		X	X				

PART 3

**Conclusion
and proposed
way forward**



The intention of this brief report has been to summarise our findings in terms of the state of the WSP/ATR submissions received by the DHET from the SETAs and to ascertain what useful research may be conducted using this data.

The review identified completeness, accuracy, and consistency as three key criteria for good quality data and found that the current set of WSP/ATR submissions has serious problems with respect to completeness and consistency. At the same time, there is reason to suspect that there may also be issues with respect to accuracy, given that oversight by the DHET over the submissions has been weak.

We acknowledge the DHET's concerns that our findings imply that the resources dedicated to collecting this data thus far have yielded little. However, at this point, remedying the situation would require a considerable additional investment with an uncertain return given the age of the data. At the very least, it seems that any analysis of this data would require substantial engagement with SETAs to remedy the data problems.

Our view, however, is that the WSP and ATR can still be used to provide useful information to the DHET. To this end, we propose that the DHET work to ensure that good-quality data can be harvested from the 2020 submissions, and we are happy to fit into this process wherever it would be useful. This would require four key actions on the part of the DHET:

1. The DHET should actively engage with SETAs regarding the importance of the 2020 submission and provide them with a proper Excel template that the SETAs will be required to use;
2. The DHET should ensure that the submitted data are properly checked and, where required, should return problematic submissions to the SETAs and engage with them to remedy the problems;
3. The analysis of the 2020 data should present a baseline for future analysis and, importantly, should aim to present the data in ways that are useful to the SETAs (over and above what is useful from the DHET's perspective);
4. The DHET should follow through with this process into the future so that it becomes institutionalised within the DHET and the SETAs.

For their part, the SETAs must strengthen their efforts to collect high-quality data from employers. This is critically important as, once the data reaches the DHET, it is too late to rectify problems that occurred during data collection. Good-quality data requires strong partnerships between the DHET and the SETAs, and between the SETAs and employers. The role of SETAs within these partnerships is absolutely crucial for the collection of data that is regular and consistent. There is also ample scope for SETAs to innovate with methods that simplify and streamline data collection, which could then be shared among stakeholders to ensure that the entire system may benefit.

Unfortunately, given the focus of this research, it is not possible to draw strong conclusions as to the problems experienced by SETAs in the process of collecting data from employers. It is clear, however, that SETAs operate within widely differing contexts, particularly with respect to the number of employers from which they are required to collect data and those employers' capacity – both in terms of human resources and technical know-how – to provide the data required. This is an important issue that deserves further attention as part of the process of strengthening data collection processes.

As noted previously, the responsibility for wrangling 21 SETAs' inconsistently formatted data submissions should not be the DHET's. Instead, SETAs should be submitting their data according to a set of clear criteria and/or templates. This is the type of process that the South African Qualifications Authority has followed in terms of the data submitted to the National Learners' Records Database and, while there were early teething problems, the process has largely been institutionalised within SETAs. However, this type of process does require a concerted effort and an uncompromising focus on data quality.



DPRU CONTACTS

Programme Leader: Professor Haroon Borat – haroon.bhorat@uct.ac.za

Programme Manager: Ms Kezia Lilenstein – kezia.lilenstein@uct.ac.za

DHET CONTACTS

Programme Leader: Ms M. Khuluvhe – Khuluvhe.M@dhet.gov.za

Project Secretariat: Ms M. Ramasodi – Ramasodi.M@dhet.gov.za