

# **JOINT INITIATIVE ON PRIORITY SKILLS ACQUISITION (JIPSA)**

## **Towards a National Human Resources Database**

PROJECT : NATIONAL LEARNERS RECORDS DATABASE  
AUTHOR : SINGIZI CONSULTING  
DATE : OCTOBER 2008

### **TABLE OF CONTENTS**

1. Introduction.....	3
2. Rationale for a National Human Resources Database.....	6
3. A Synthesis of the Two Reports .....	8
3.1 Paper One .....	8
3.2 Paper Two .....	14
4. Concluding Recommendations .....	21
4.1 Considering Scenarios for the NLRD and National Human Resources Database .....	21
4.2 Specific Areas for Development .....	22
4.3 Possible Ways of Taking these Activities Forward .....	23
5 References .....	30

Appendix 1 - A review of the National Learner Record Database (NLRD) with regards to its potential role in supporting HRD planning

Appendix 2 - Towards a National Human Resources Database: the national learners' records database assessed

## **ABBREVIATIONS AND ACRONYMS**

CIDB	Construction Industry Development Board
DoE	Department of Education
EMIS	Education Management Information System
EPP	Employment Promotion Programme
ETQA	Education and Training Quality Assurance
FETMIS	Further Education and Training Management Information System
HEMIS	Higher Education Management Information System
HEQCIS	Higher Education Quality Committee Information System
LFS	Labour Force Survey
NEIMS	National Education Infrastructure Management System
NHRDS	National Human Resources Development Strategy
NLRD	National Learners Records Database
NQF	National Qualifications Framework
NSFAS	National Student Financial Aid Scheme
OFO	Organising Framework for Occupations
SACMEQ	Southern and Eastern African Consortium for Measuring Educational Quality
SAQA	South African Qualifications Authority
SETA	Sector Education and Training Authority

### 1. Introduction

This report provides a synthesis of two research papers<sup>12</sup> commissioned by the Employment Promotion Programme (EPP)<sup>3</sup> on the review and suitability of the National Learners Records Database (NLRD) to become the National Human Resources Database. The objectives of the studies determined by JIPSA together with South African Qualifications Authority (SAQA) were set out as follows:

1. To ascertain from the policy and decision makers the relevant information that they require
2. To evaluate the extent to which the NLRD already does collect or can be adapted to collect this data
3. On the basis of this analysis and a general evaluation of features such as flexibility, capacity and systems infrastructure, the overarching study is intended to ascertain which available database can be or has the potential to be the national education and skills database or whether such a national database is desired/required
4. Based on this assessment there would then be a need to determine suitable funding mechanisms

The two studies each had distinct areas of foci. The first study focused on the needs of policy and decision makers and the way in which institutions involved in providing data to the NLRD related to the NLRD<sup>4</sup>. The second report considered the NLRD itself in more

---

<sup>1</sup> Singizi Consulting (January 2008). A Review of the National Learner Record Database (NLRD) with regards to its potential role in supporting HRD planning: an analysis of the NLRD in terms of stakeholder perceptions and the status of education and training databases in South Africa.

<sup>2</sup> Development Policy Research Unit (June 2008). Towards a National Human Resources Database: the National Learners' Records Database assessed.

<sup>3</sup> The EPP is a tripartite body represented by Government (Presidency), organised labour and organised business. The EPP is funded through the Department of Foreign and International Development (DFID) of the United Kingdom and is managed and coordinated through the Development Policy Research Unit (DPRU).

<sup>4</sup> Singizi January 2008

depth and reviewed whether the NLRD meets a set of designated criteria: appropriate unit of analysis; comprehensiveness and currency of data<sup>5</sup>. This review took into account both the original objectives of the NLRD and possible new requirements that would be placed on the NLRD if it became the National Human Resources Database. Both reports then made recommendations in terms of the viability of the NLRD becoming the National Human Resources Database and outlined the different implications flowing from these recommendations.

This report begins with an explanation of the rationale for a National Human Resources Database and then provides a summary of the two reports and the key findings contained in these reports. It concludes with an overarching set of recommendations which suggest the way in which the NLRD could be strengthened and which indicate its possible relationship with a National Human Resources Database.

---

<sup>5</sup> Development Policy Research Unit June 2008

## 2. Rationale for a National Human Resources Database

South Africa is in the process of adopting the revised Human Resource Development Strategy for South Africa to cover the period from 2009 to 2014. The Strategy outlines the needs for certain coordinative mechanisms to drive the different strategic priorities. It also spells out the extent to which the success of the strategy is reliant on information. An excerpt from the strategy highlights the emphasis on the need for credible data.

*“Quarterly reports will be prepared by responsible institutions and agents for each of the Special Programmes included in the HRDS-SA. These reports will serve primarily as “early warning” signals and will guide implementation and remedial action where required. Targets will be monitored on an annual basis, and the results published in a consolidated Annual Report on the implementation of the HRDS-SA. This annual report will be complemented by the annual reports that will be prepared by each of the departments and agents responsible for the main sub-systems of the education and training system. These reports will serve primarily to review performance and revise targets, where necessary. They will also serve as a key accountability mechanism for the HRDS-SA. A major review, based on systematic evaluation studies and impact assessments, will be conducted every five years. These will include reviews by each of the main sub-systems within education and training, but will substantively rely on independent assessments, which include assessments using alternative data sources (such as those from STATS SA) to those used by line departments<sup>6</sup>.”*

This emphasis resonates with the findings of previous reports, for example the Human Resource Development Review 2008<sup>7</sup> which confirms that credible data is crucial for good evidence led research and policy formulation and suggests that there are currently a number of limitations relating to the data that is currently available.

They observe that in the present system the Department of Labour is,

---

<sup>6</sup> Department of Education (2008). Draft National Human Resource Development Strategy for South Africa: 2009 – 2014

<sup>7</sup> Kraak, A and K Press (2008). Human Resource Development Review: Education, Employment and Skills in South Africa, HSRC Press

*“Heavily reliant on the 23 SETAs for data collection on the NSDS, and has little power to ensure data integrity and quality in the production of such SETA data...”*

They add that as a further concern that,

*“...The task of researching off the HEMIS database is complicated further by the fact that it is retrospectively amended from time to time, as DOE officials audit the enrolment numbers provided by institutions...”*

These difficulties with information is captured most eloquently in one of the two papers<sup>8</sup> which states that,

*“The information base supporting skills provision and skill development for South Africa remains inadequate. Given the centrality of the skills challenge, noted in a multitude of forums, policy documents and public pronouncements, the fact that the country’s information base in this area is deficient is often overlooked. Indeed, often critical information is of low quality, incomplete and, in many instances, non-existent. As a result, many relatively simple yet vitally important data queries are difficult to obtain and estimates are often of dubious quality.”*

This section highlights the importance of a National Human Resources Database and suggests that the need for credible information will be increased going forward. It also points to concerns that have been identified previously about the extent to which it is possible to access comprehensive human resource data. Critically it highlights the importance of exploring whether the NLRD can play this role, the manner in which it relates to other databases in the system and whether there are areas that are not being effectively collected in any of these systems.

---

<sup>8</sup> Development Policy Research Unit. June 2008

### 3. A Synthesis of the Two Reports

As indicated, two papers were commissioned as part of this project. The first paper considered the perceptions of policy makers and role players pertaining to the types of data that they felt would be important for a National Human Resources Database<sup>9</sup>. The paper also engaged with those players that currently forward data to the NLRD. The second paper undertook a review of the NLRD itself and evaluated it both in terms of its original objectives and in relation to anticipated needs of a National Human Resources Database.<sup>10</sup>

#### 3.1 Paper One<sup>11</sup>

This paper focused on the needs that policy makers and other role players articulated for data to inform HRD Planning and also took into account the perspectives of those involved in collecting any component of this data and in particular players that submit data to the NLRD.

##### 3.1.1 *Summary of the Key Findings*

The paper outlines the views expressed by interviewees pertaining to usage and suggests that these can be clustered into three broad categories which are:

- Economic, developmental and social planning
- Understanding labour market trends
- Labour market supply

##### 3.1.2 *Existing Databases*

The report highlights the number of existing databases, a summary of which is provided below:

##### Labour Market Data

- StatsSA is responsible for a number of key surveys including with regards to employment and job creation (employment and earnings data which provides data on employment and earnings and the Labour Force Survey (LFS)) and

---

<sup>9</sup> Singizi Consulting. January 2008

<sup>10</sup> Development Policy Research Unit. June 2008

<sup>11</sup> Singizi Consulting January 2008



demographics (population census, mid year population estimates, health and vital statistics)

- In addition to StatsSA the Department of Labour and the SETAs have attempted to track and trace learners and the nature of the opportunities that they access
- The Department of Education is also putting in place a learner tracking system which allows the Department to understand what options learners pursue after completing their FET programme
- Government departments also collect some of this data pertaining to their sector. As examples, the Department of Public Works, together with the Construction Industry Development Board (CIDB) collects and analyses certain trends in the construction sector. The Department of Environment and Tourism has undertaken a number of studies on related issues
- Many of the industry associations, union research departments and bargaining councils also track trends regarding wages and salaries. Some also keep records about vacancies and employment trends
- An increasing number of professional bodies are commissioning skills audits and maintaining a database of members' skills and occupations, so that they can make informed inputs to sector industrial and skills planning processes

### Supply Data

#### *Schooling*

- Education Management Information System (EMIS) consists of a series of censuses, and deals largely with learner information (and other aspects of schools)
- National Education Infrastructure Management System (NEIMS) is a new system intended to provide a live database of physical infrastructure
- Persal, containing extensive information on the human resources within the public sector and including data pertaining to the post, a description of the necessary and desirable competencies for the job; desirable characteristics for employment and promotion within an occupational category

- Basic Accounting System (BAS) which is a database managed by Department of Treasury that aims to provide a single integrated financial management and accounting platform, to allow for comparability of financial information which provides considerable information about human resource development activities implemented by government at all three spheres
- Matriculation Board – has data on examinations drawn from the provinces and captured through SITA
- Systemic Evaluation which aims to provide a national framework for the evaluation of the education system at the key transition stages of Grade 3 for the Foundation Phase, Grade 6 for the Intermediate Phase and Grade 9 for the Senior Phase. There are also other similar assessments such as SACMEQ (the Southern and Eastern African Consortium for Measuring Educational Quality), TIMSS and PIRLS which are also international learner testing programmes which allow us to gauge the quality of schooling in South Africa against international benchmarks
- In addition, the Department is intending to introduce 'LURITS', the Learner Unit Record IT System. The programme will allow individual learners to be tracked as they move through the system which will allow for a more accurate picture of throughput within the system

### *FET Colleges*

- Further Education and Training Management Information System (FETMIS) - This collects student specialisation and educator data in a Department of Education (DoE)-determined format, and produce standardised reports

### *Higher Education*

- HEMIS - HEMIS is a DoE system which feeds off operational systems of the higher education institutions in the country. HEMIS data includes information on: student enrolments, course registration, student results, graduations, throughput rates, retentions, residences, academic staff members and employees, bursaries and loans from the National Student Financial Aid Scheme (NSFAS), available space, lecture rooms and venues, exam timetables, qualification profiles of staff, achievements and awards for researchers etc.

### *Skills development*

- Sector Education and Training Authority (SETA) information systems (Education and Training Quality Assurance (ETQA) databases) collect information on enrolments and achievements in National Qualifications Framework (NQF) qualifications and unit standards, legacy qualifications, learnerships, and apprenticeships
- The Sector Skills Plans developed by the SETAs are required to provide an assessment of existing provision covering initial training and continuing training systems, issues of qualification standards, training facilities and equipment, the quality of trainers, accreditation, funding, location, gaps in provision, and general deficiencies in the output of the education/training system
- The Department of Labour collects aggregated information on learnership enrolment and achievement, and on artisan enrolment and achievement (this includes data provided by INDLELA which is responsible for Trade Tests and in particular for Section 28 tests that may not go through the SETAs)

The report also indicates other related data sources such as Professional Councils, Independent Examinations Board (IEB), UNISA, UMALUSI and also specifically mentions the work that SAQA is doing to directly gather data from the private higher education institutions.

### *3.1.3 Locating the NLRD*

The report then locates the NLRD within this system and reiterates the Mission of the National Learner Records Database which is to:

*“Develop and maintain a Management Information System for the NQF that provides reliable, relevant, up-to-date, accurate, user-friendly, and accessible education and skills development data for the people of South Africa, and which contributes to national growth and development.”*

The paper notes that the NLRD already receives data from: UMALUSI, Department of Education (HEMIS), Private higher education institutions, SETA and professional ETQAs and the IEB.

The report suggests that in terms of data pertaining to the *economy and society* and the *labour market data* while critical for human resource planning and development is outside of the remit of the NLRD.

### 3.1.4 *Challenges with Regards to the System*

With regards to data pertaining to *Supply*, the report understood this to be an area that the NLRD already has a significant responsibility for. However, it highlights a number of challenges pertaining to these data sets, including:

- learner tracking which limits information about learner placements upon programme completion or the percentage of graduates that practice and secure professional registration and an absence of data relating to the pathways that individuals navigate upon completion of their studies.
- the absence of common and agreed upon definitions and methodologies for data even within a sector
- that much of the data is reliant on self-reporting (and in some cases concerns were raised about the possibility of providers inflating numbers was highlighted)
- issues related to the timing of data collection and submission, the number of transaction activities and the resultant time lags.

There were also challenges raised pertaining to the data within the NLRD such as gaps in areas where arrangements are not yet in place (such as with INDLELA for artisan data, FET College data and schooling data from the DoE), limited capacity within the institutions supplying data onto the NLRD and the issue of different definitions resulting in certain problems with the data, and confusion as to whether apprenticeship data should be submitted. While the possibility of analyzing throughput is there for some of the data this is not possible for other sectors. The report notes that certain challenges regarding the NLRD were, at the time of the research, being addressed by SAQA.

With regards to the question of whether the NLRD could become the “uber- database” a number of concerns were raised which centered on a reluctance on the part of sectors to lose control over their own data.

### 3.1.5 *Key Recommendations*

The report concludes by providing a set of key recommendations, which are captured below:

The concept of a Human Resource Development database is not a feasible option.

Data required for economic and social planning should not be included in the NLRD and should not be a focus of the NLRD. Instead StatsSA should create fields in its existing datasets that could provide data related for this kind of planning. This will require support to ensure both StatsSA and business are able to provide such data through the NHRD process

The Human Resource dimension of the labour market should remain outside the scope of the NLRD.

- The NLRD must consider the possibility of supporting departments that provide labour market trends and data;
- The report recommends a process to be facilitated with StatsSA to ensure that certain data that is currently not captured is provided for through the creation of new fields in certain of the survey instruments;
- The need to building on existing capacity with the NLRD which allows for the integration of data submitted by various professional councils and provides an analysis as to which graduates practice in their field and ultimately register with the relevant professional body.

The NLRD must become the central repository for data on labour supply.

There is a need for an agreement on the nature of data to be submitted to the NLRD, that includes fields, definitions and time frames for the collection, collation and submission of data. It is further recommended that SAQA enters into a Memorandum of Understanding with StatsSA that would allow for the promotion of a national quality of standards in collecting data.

SAQA should support the publication of an annual report that provides an analysis of:

- Throughput rates for different pathways and qualifications and
- The number of graduates leaving the system within a occupation

The NLRD should continue to allow accessing verified full records of individual learner's qualifications.

The NLRD should continue to provide information on qualifications with;

- A detailed listing of all qualifications. This will allow for an improved analysis of the learning pathway for the individual and relate to the Organising Framework for Occupations (OFO)
- A comprehensive list of providers providing these training programmes that could help planning
- Minimum entry requirements for a particular qualification offered by training service providers

Consideration should be given to sectors having to produce reports similar to those provided by NIEMS and HEMIS. This could be a special project supported by the NLRD.

### **3.2 Paper Two<sup>12</sup>**

This paper provided an evaluation of the NLRD against its original objectives and in terms of the envisaged requirements of a possible National Human Resources Database.

#### *3.2.1 Summary of the Key Findings*

This paper began with an explanation of what it understands to be the key criteria against which a National Human Resources Database should be evaluated. These are:

- Firstly, a National Human Resources Database should have the individual as its unit of analysis
- Secondly, a national human resources database should be comprehensive. In other words, the database should contain, at the very least, the qualifications of all South Africa's citizens. Furthermore, the database should be comprehensive in its coverage of all qualifications that individuals may have obtained during their lives

---

<sup>12</sup> Development Policy Research Unit. June 2008

- Thirdly, a national human resources database should be current. In other words, as far as possible, the information contained in such a database should be up to date, allowing for a reasonable lag
- Fourthly, in order to be useful, a national human resources database should represent the current stock of skills available to the labour market. Thus, while the database needs to be current in terms of the flow of new skills accrued, it also needs to be current in terms of individuals exiting the system

The paper also notes that while these are the main criteria against which the NLRD will be evaluated there are various other requirements for consideration. They include the need for a National Human Resources Database to include: some level of employment information; a minimum amount of geographic or locational data; anticipated changes in both demand and supply of the relevant skills. The paper acknowledges that there may be other requirements for a national human resources database but that it is felt that those listed are the most appropriate for the purposes of this report.

### 3.2.2 *Reviewing the capacity of the NLRD*

The paper then proceeds to evaluate the NLRD: it begins with a description of the NLRD and its processes and outlines the methodology employed to verify the capacity of the system. The research made a number of observations in this regard:

The paper specifically highlights the usage of the Edu.Dex tool which it notes is issued to all ETQAs. It comments that this ensures that before ETQAs are allowed to upload their data to SAQA, the data must be approved by Edu.Dex as complying with the set minimum standards. Once Edu.Dex passes the data as having fulfilled the requirements, ETQAs are then allowed to upload the data to SAQA. They observe that while Edu.Dex is sufficiently flexible in that it requires only that the proposed data submission be in a standard format, thus allowing ETQAs to utilise any software system they choose, it has zero tolerance on data quality and discriminates between fatal and non-fatal errors. Fatal errors concern non-compliance with load specifications, for example the file should be in fixed-width format. Another type of fatal error occurs where the data of birth provided does not correspond to the first six digits of the national ID number, or where gender takes any value other than male or female.<sup>13</sup> Once Edu.Dex has found fatal errors in an

---

<sup>13</sup> The NLRD load specifications (version 2.0) are detailed in a 41 page document (South African Qualifications Authority 2008b). The document also details best practice for validating and

ETQA's proposed submission, these errors would need to be rectified and the revised submission would require retesting and approval by Edu.Dex. Non-fatal errors tend to revolve around quality issues: the discovery of a non-fatal error does not prevent the data from being submitted to SAQA, but does provide ETQAs with feedback as to data quality issues, such as where data is entered as 'unknown' where it should have a proper value. The paper concludes that through the usage of the Edu.Dex tool SAQA appears to be making a significant and positive impact on the process of uploading data from data providers to the NLRD, thereby benefiting data quality and integrity within the NLRD. However, the paper notes two limitations with regards to the system: Firstly, Edu.Dex is limited in its ability to find duplicate individuals in the database (i.e. one individual is recorded more than once in the database's person table). A second limitation is the fact that, because Edu.Dex 'talks' to ETQAs, cross-ETQA training providers can be problematic and it may be difficult to identify duplicates within the database. As such, the NLRD team will always be required to check on duplicates, whether they are persons or providers.

The paper also emphasizes that in an effort to reduce the existence of missing or unspecified values in the data, the NLRD has developed a tool called "Get Best Data", which will attempt to ensure that the NLRD maintains a copy of the most recent and best data for a given individual. This is also considered a positive development.

### 3.2.3 *Reviewing the NLRD Against the Criteria Proposed for a NHRD Database*

The paper reviews the NLRD against the above-mentioned four criteria and the key findings are replicated below:

- The NLRD is, by its very nature, focused on the individual as the unit of analysis, at least as far as is relevant to our requirements. It is noted that the person table has one person per record and that the learner table has one record per qualification. However, the paper suggests that this is easily rearranged and that a national human resources database based on the NLRD would arguably best be structured along very similar lines to the NLRD, making analysis of qualifications or of individuals straightforward.

---

extracting data. The procedures are too detailed to describe here, but in terms of ID numbers, for example, see Appendix A.



- The paper argues that the criterion of comprehensiveness is not fulfilled by the NLRD and provides two reasons for this assertion. Firstly, it illustrates through an analysis of the data that the historical coverage of the NLRD is deficient. Secondly, the paper highlights that the current coverage of the NLRD is deficient in that there are still segments of the education and training system that are not submitting data. Specifically, major gaps are the lack of data regarding matric qualifications and the FET sector. The paper observes that SAQA is aware of these limitations and is working to address them. However, at the time of research they remain as gaps in the data sets.
- With regards to the currency criterion the paper acknowledges that it will only be possible to really answer this question with a more thorough review of the data uploads over time, so as to assess the rate at which data is being updated relative to the dates contained in the data submissions (e.g. dates of completion of qualifications). It notes that the NLRD receives updated information from the ETQAs at least twice a year which, it suggests, is sufficient for a National Human Resources Database. Nonetheless, the paper found that the data post-2004 is incomplete and suggests that this is an area of concern given that this is three years' worth of data (excluding 2008) that has not yet been uploaded into the NLRD. The paper comments that while HEMIS is identified as a major component of the missing data, it suggests that the public higher education system is unlikely to account for the full drop of over 470 000 qualifications between 2004 and 2005. The paper also notes that while data uploads may be performed frequently, it is the currency of the data from data providers that will determine the currency of the NLRD and, by extension, the national human resources database. The paper raises a concern as to how current the data being uploaded by data providers is, both at the level of upload to the NLRD as well as the level of upload to the ETQAs.
- With regards to the fourth criteria - the ability of the database to reflect the actual stock of skills available to South Africa - the paper notes that the other three criteria discussed thus far all have a bearing on whether or not the database can accurately reflect the current stock of skills. They suggest that limitations with regards to coverage and currency mean that the NLRD is not able to provide an accurate picture of the current stock of skills. More

important, the paper suggests that for a database to provide information on the available stock of skills is dependent on its ability to track: the ageing of individuals in the labour force, 'permanent' emigration of South Africans to other countries (or the return of foreign residents in South Africa to their home countries), and deaths. The paper observes that the NLRD has been set up with another set of objectives and currently does not account for any of the three factors. Thus, at present the NLRD does not meet this criterion. It notes though that the ageing of individuals in the labour force could be estimated from the NLRD through extracting only individuals within a specific age category from the NLRD, for example. With regards to emigration it notes that this type of data gap is easily addressed in terms of the structure of the database, but suggests that the practicalities of filling this data field are complex given the non-reporting of emigration, with the gap between official StatisticsSA data and receiving country data continuing, if not widening.<sup>14</sup> It notes that information on deaths is more straight-forward and indicates that data on both emigration and death is housed by the Department of Home Affairs that has a mandate to monitor this data.

With regards to the additional criteria mentioned the paper makes the following observations:

- With regards to employment information it questions what the source of such data would be and whether, given current systems, the employment data in a database of this type will be updated regularly or consistently enough to be considered current
- The paper raises similar concerns about locational information
- With regards to information on flows of individuals through the education and training system the paper argues that while this is not currently being undertaken by the NLRD it asserts that this should be possible given that the NLRD includes data on registration and completion of unit standards by learners

---

<sup>14</sup> According to Myburgh (2004: 124), "in 1997 SSA data only captured a third of those who emigrated, a tendency which seems to have increased over time: in 1987 SSA captured a more respectable 65 per cent".

### 3.2.4 *Key Recommendations*

The paper provides a number of key concluding comments which are summarized below:

The paper concludes by reiterating the point that the evaluation of the NLRD has been undertaken against criteria which are not consistent with the purpose for which the NLRD was established as the evaluation focused on the extent to which the NLRD in its current form was consistent with the requirements of a National Human Resources Database.

It emphasizes that the NLRD is systemically robust and that considerable care is taken to ensure that data submissions, received from a large number of institutions using a wide variety of computer systems, are of decent quality.

However, the paper notes that the NLRD does not meet all of the criteria associated with a National Human Resources Database. Nonetheless, the NLRD is fully compatible with a possible National Human Resources Database insofar as it uses the individual as its unit of analysis. It also suggests that in terms of biannual uploads from data providers currency of the data could be maintained but the report highlights practical issues pertaining to the application of this criterion. It also noted that, in terms of comprehensiveness, the NLRD has a number of data gaps and makes a specific recommendation about the need to support the NLRD to explore alternative sources of data from the public higher education system including the possibility of extending the Higher Education Quality Committee Information System (HEQCIS) to the public sector institutions. The paper also observed that the NLRD currently does not contain information about the stock of available skills though it highlights areas where this could be possible.

Taking all of this into account, the paper argues that though the establishment of a national human resources database would be substantially facilitated by using an existing database as a master list, it argues that the NLRD is currently not in a position to properly fulfill this role and suggests that this conclusion may remain valid for at least another decade.

The paper further argues that as the NLRD was not set up with all of the objectives relevant to national human resource planning that it would be preferable if the national human resources database is created as a separate institutional 'home'.

It suggests therefore that rather than becoming the National Human Resources Database the role of the NLRD could then be to serve as a master dataset – on the proviso that the concerns raised in this paper are addressed. Thus the NLRD would focus on those fields in which it is already collecting data, as well as certain additional fields where proper linking with identified official data sources is secured. The NLRD data could then feed into the National Human Resources Database. The paper notes that this assumes close cooperation between SAQA and the institution housing the National Human Resources Database.

However, the paper notes that for a National Human Resources Database to be established this will require that consensus be reached across government departments and state agencies about the desirability of such a database, which will need to be accompanied by a commitment to facilitate linkages between the databases under their supervision and the prospective database.

The paper also indicates that, in establishing a National Human Resources Database, it will be valuable to learn from the NLRD experience. It highlights that the NLRD illustrates the importance of the human element and specifically that the quality of the data is obviously enhanced when those individuals at the ETQAs and other data providers understand and appreciate the processes. It notes that SAQA and the NLRD have tried to facilitate data submissions and promote buy-in from those involved but that staff turnover at data-providing institutions impacts on continuity, particularly where institutional memory in terms of the requirements of the NLRD is not strong. It suggests that this should be an important lesson for the proposed human resources database.

Finally, the paper also observes that, given the analysis above, the establishment of a national human resources database is likely to be a longer-term undertaking. It proposes that, using the NLRD as the master list, the key is to identify the most useful databases to link to and ensure that those databases comply with minimum data quality standards and link easily to the NLRD. Just as important is ensuring that the institutional relationships are cultivated and strengthened in anticipation of the National Human Resources Database.

### **4. Concluding Recommendations**

This final section of the paper provides some overarching perspectives about a possible role for the NLRD and its relationship to a National Human Resources Database. This is based on internal deliberations within the JIPSA team and are provided here as initial thoughts to allow for fuller discussion and engagement.

The findings from these reports indicate the current mandate of the NLRD and suggest that the current system of the NLRD is robust and has credible mechanisms to ensure the integrity of the data within the system. Both the reports acknowledge that the NLRD was set up with a purpose quite distinct from that of a National Human Resources Database.

Taking this context into account, it is observed that the NLRD does not contain data relating to broader aspects of human resource planning such as employment data, locational data as well as data which allows for an analysis of the current stock of skills.

This report also notes that, since the beginning of this process, significant progress has been made with regards to the National Human Resource Development Strategy, and proposals are in place which articulate how these processes will be taken forward. Any discussion which considers the nature of the NLRD and the National Human Resources Database will need to be located within the context of these broader discussions.

#### **4.1 Considering Scenarios for the NLRD and National Human Resources Database**

It is noted that two scenarios are possible to address this.

In the first scenario, the NLRD is expanded to include these additional fields and to integrate the different databases from Department of Home Affairs and Stats SA. This assumes that the NLRD simultaneously addresses those areas where there are data gaps pertaining to supply and develops the capacity to track the flow of learners. The NLRD would also require considerable additional capacity in terms of ability to analyse different types of data and undertake planning related activities. These would substantively change the mandate of the system and would extend beyond the work that SAQA as an apex organization is involved with. In effect, in this scenario the NLRD would assume the functions of a National Human Resources Database.

In the second scenario, the NLRD will focus primarily on data relating to its original mandate. Thus the focus of the NLRD in the next period will be on consolidating existing functions and ensuring that it is able to meet the criteria outlined with regards to these data sets. This includes, for example, ensuring that there is a clear mechanism for schooling and further education and training (FET) data to be submitted to the NLRD and that a consistent mechanism for higher education data is applied (possibly through the extension of the HEQCIS). It would also ensure that sufficient capacity was built in the system to ensure that the data met the criteria of both comprehensiveness and currency and that it was able to track the flow of learners through the system. In this second scenario, a separate National Human Resources Database would be directly accountable to the National Human Resources Development Strategy (NHRDS) and the relevant structures that are established to drive this strategy across government. The nature of this database would require careful consideration, and it may in reality operate as a virtual database which relies entirely on data from the NLRD as a key database as well as other databases. This would be determined in terms of the indicators that are contained in the NHRDS and the function of the database would primarily be on the analysis of this data so that effective reporting against the indicators could take place. This would allow South Africa to generate a report which sets out the status of NHRD in the country on an annual basis. This report suggests that this second scenario is more likely to find acceptance by the different role players and that it will ensure that the NLRD remains focused so that it is able to address those areas that have been identified as requiring further strengthening. The work on the NLRD could then take place in parallel to the broader discussion on the nature of the NHRD database and would not need to wait for these decisions to be finalized.

### **4.2 Specific Areas for Development**

Within the context of the second scenario, it is anticipated that the following sets of activities will be important:

It is noted that the NLRD is very reliant on the capacity and willingness of those institutions that input data into the NLRD. This raises the concern that where this data is not provided as required then this impacts adversely on the ability of the NLRD to ensure that its data is comprehensive or current. The need to enhance the capacity of data providers and build strong relationships between these institutions and NLRD is

therefore identified as a key area that should be taken forward to address certain of the concerns highlighted by the reports.

The reports also suggest that there are areas such as flow of skills which are consistent with the mandate of the NLRD but which the NLRD does not yet monitor and report on. It is suggested that this could be an important area in which the NLRD is further strengthened.

These reports also emphasise the need for the NLRD to play a critical role in addressing certain system issues such as the need for consistent definitions to be applied in the sector. This includes the need for an agreement on the nature of data to be submitted to the NLRD and suggests the possibility of SAQA entering into a Memorandum of Understanding with StatsSA that would allow for the promotion of a national quality of standards pertaining to the collecting of data.

Implicit and explicit in these reports is the emphasis on the need to establish relationships between users of the data and the NLRD as well as ultimately between the NLRD and the National Human Resources Database. This suggests that discussions about the form that a National Human Resources Database will take should take into account the learning from the NLRD. These discussions should also consider the relationship that will need to be established with the NLRD as well as other databases.

### **4.3 Possible Ways of Taking these Activities Forward**

1. There will be a need to reach agreement about how the findings and recommendations in this report are processed.
2. Decisions about the NLRD and the National Human Resources Database will then need to determine the nature of support that need to be put in place.
3. There should possibly be a coordinative mechanism established which has is coordinated and includes the major stakeholders and role players in this process. This process should then reference developments taking place with regards to the NHRDS.

## **APPENDIX 1**

TITLE : A REVIEW OF THE NATIONAL LEARNER RECORD  
DATABASE (NLRD) WITH REGARDS TO ITS  
POTENTIAL ROLE IN SUPPORTING HRD  
PLANNING

AUTHOR : SINGIZI CONSULTING

DATE : JANUARY 2008



**JOINT INITIATIVE ON  
PRIORITY SKILLS ACQUISITION  
(JIPSA)**

**A review of the National Learner Record Database  
(NLRD) with regards to its potential role in supporting  
HRD planning:**

**An analysis of the NLRD in terms of stakeholder  
perceptions and the status of education and training  
databases in South Africa**

PROJECT : NATIONAL LEARNER RECORD DATABASE  
AUTHOR : SINGIZI CONSULTING  
DATE : JANUARY 2008

**TABLE OF CONTENTS**

1. Introduction and Background .....	26
2. Methodology.....	27
3. What kind of information is required for HRD Planning.....	28
4. Where is the Data Currently Being Held? .....	34
5. The current role of the NLRD within this data landscape: focusing on supply data	40
6. What are the challenges in relation to accessing information on skills and education?.....	45
6. Broad recommendations .....	52
7. In conclusion .....	59

---

## **1 Introduction and Background**

This report has been commissioned by the Employment Promotion Programme (EPP). The EPP is funded by DFID and managed by DPRU under the auspices of the Presidency, Labour and Business. The study was conceptualised within JIPSA to allow for an improved understanding of the data that is available to support Human Resource Development Planning and specifically to explore the possible viability and desirability of the National Learners' Records Database (NLRD) playing the role of a 'national education and skills database.

The objectives of the study were determined together with SAQA. It was agreed that this component of the project would focus on:

1. Understanding the nature of the existing database(s) including the sources of data, the manner in which data is collected, verification of data as well as how the data is submitted to the NLRD (or an alternate body such as the DoL)
2. To establish the successes and challenges pertaining to this process and whether there are any additional fields of information that would be valuable to collect (from the perspective of the data providers and stakeholders)

On the basis of this analysis and a general evaluation of features such as flexibility, capacity and systems infrastructure, the overarching study is intended to ascertain which available database can be or has the potential to be the national education and skills database or whether such a national database is desired/required. Based on this assessment there would then be a need to determine suitable funding mechanisms.

This report has the following components: An indication of the methodology that was followed within this study; An overview of the perspectives that the different stakeholders have about the role of HRD and the specific issues that they wish to track; The data requirements suggested by these diverse issues; The existing databases that collect this data; The manner in which the NLRD interacts with these systems; the challenges pertaining to accessing and using data from these databases, including the NLRD, that are reported by stakeholders; The report makes recommendations about the potential

---

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

role for the NLRD and indicates what would be required to enable the NLRD to fulfill this role.

### **2 Methodology**

The activities that were followed as part of this process included:

- Interviews with individuals responsible for a component of HRD Planning. This includes individuals from government departments as well as the social sector (business and labour). These interviews focused on ascertaining the information that these sectors require; the implications of this in terms of the data fields that are required, and determining the existing databases that these sectors use to collect this data.
- Interviews with individuals that collect and submit data to the NLRD. However it should be noted that while we originally intended to speak to 7 ETQAs we were only able to secure 4 interviews with ETQAs. The difficulty experienced in accessing these interviews is likely to relate to the time of year in which the research was being undertaken rather than a response to the nature of the study. Where we were not able to secure interviews we relied on data secured in previous research processes such as the SETA Review, a report on the NLRD commissioned by SAQA conducted by CEPD, as well as data provided by SAQA and related bodies such as HESA;
- Interviews with individuals that collect data that were seen to be relevant within this context – such as Stats SA; and,
- Review of documents which outline the current data collection mechanisms.

It is noted that at the same time as this process was taking place, there was research and consultation process taking place with regards to the review of the NHRD Strategy. Singizi, together with Bobby Soobrayan, was directly involved in this process and the discussions and engagements that formed part of that engagement have also fed into this report. This was considered opportune as it ensures that the analysis pertaining to the information needs of a national education and skills database is able to take into account the review of the NHRD Strategy and the way in which the NHRD strategy is being re-conceptualized.

---

### **3 What kind of information is required for HRD Planning**

This section provides a synopsis of the kinds of information that is required for HRD Planning: it begins by indicating who the potential users of this data are and outlining the kind of information that they may need, and then considers the data required to enable these users to make an analysis in a way that addresses their requirements.

#### **3.1 Who are the Users?**

The key users (current and potential) of HRD related databases, together with their rationale for requiring this data, can broadly be listed as follows (note that this is not a comprehensive list but highlights the key users):

Economic, development and community planners who rely on an education and skills base to support their broader development plans,

At a national level: Government, coordinated by the Presidency, focus on the way in which HRD can support the achievement of the country's economic and social growth and development goals as outlined in Government's Programme of Action and other strategic documents such as ASGISA, in a manner that is consistent with the global agenda of realizing the Millennium Development Goals (MDG). Thus for example in the POA – Government commits itself to ensuring that priority skills for the economy are developed and sets the following skills related objectives, *“Co-ordinate the acceleration of skills development for design, engineering, artisans, which are critical for advanced manufacturing, construction and cultural industries, Increase the number of available artisans by resolving the issue of training routes, Increase output of engineers by developing capacity of relevant faculties in universities, Identify and resolve system challenges which affect performance of initiatives, including SETAs, Review targets and strengthen integration and coordination of National Human Resource Development Strategy (including NSDS and JIPSA) to ensure long-term provision of core and critical skills especially in priority sectors.”*

Drilling down, the Department of Trade and Industry, as the coordinators of the economic, investment and employment cluster, has developed an Industrial Strategy which aims to ensure the development and implementation of this national strategy, as outlined in the POA, in a manner that is characterized by high levels of economic growth, generates employment and reduces levels of inequities. As part of this commitment the

---

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

DTI is responsible for the formulation of sector strategies. One component of these strategies is the imperative to address the skills requirements of the sector.

Within each sector, other line departments focus on economic and social growth within a sector, for example the Department of Transport as part of its commitment to transforming the taxi industry has the following objective, *“employees in the taxi industry including customer relations, driver training in taxi and buses - Development of skills strategy for the taxi industry and integrate bus sector initiatives, TETA and centres of development to do training courses, Commencement of training,”* or the Department of Environment and Tourism which states that one aspect of creating conditions for sustainable tourism growth and development for the benefit of all South Africans is to ensure that *“4707 new tourist guides are trained over a 2 year period and 420 new international training opportunities are secured.”*

Further, Provincial and local government need to plan and monitor HRD related activities in terms of the Provincial Growth and Development Strategies and Integrated Development Plans respectively. For example Gauteng’s vision, contained in its PGDS, is to *“To ultimately create a better life for all our citizens through: long term sustainable growth of our provincial economy, meeting the socio-economic development needs of our people, creating jobs, and addressing unemployment and poverty.”* It states that to achieve this vision there is a need to overcome multiple challenges, including the imperative to address, *“the quality of the pipeline for skills required in key economic and social sectors and to leverage human capital as a source of competitive advantage.”*

Business is both a partner in all of the above mentioned initiatives and focuses on the skills that they require to grow and develop to ensure that they are competitive both nationally, and where relevant internationally. Organised labour focuses on both the skills implications of these changes and the demand that this places for skills – and emphasise that they consider the manner in which ECD and schooling will provide foundational learning for their members.

Education and skills planning – Government as well as NGOs, private sector, and parastatals as well as bodies such as the SETAs.

Key players here include the Department of Labour, together with the SETAs, who need to extrapolate the critical and scarce skills that are likely to be required by the society and the economy, and determine whether the projected and actual supply will be able to

---

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

meet the demand. Related to this, the Department of Labour and the SETAs, establish the kinds of incentives that should be put in place to enable individuals to access relevant programmes that address the identified skills. A crucial aspect of this process then involves establishing whether these programmes enable learners to access economic opportunities and the kind of support that learners require to access these opportunities: as part of this process, there is a need to understand whether individuals that attain qualifications are using these qualifications in the labour market and the relationship between skills development and productivity.

The other key player is the Department of Education: the Department needs to understand long term economic and social priorities so as to inform its programme mix and to ensure that learners are able to successfully navigate these learning pathways. The Department also needs to develop plans pertaining to provision which include: a human resource plan for the system (what kinds of practitioners will be required), an analysis of the physical infrastructure required as well as other resource requirements, and the costing structure that can be put in place to enable learners to access these different learning programmes. This includes the need for an analysis on effectiveness and efficiency levels within the system.

Other players' needs resonate with these two key government departments; for example, business needs to determine what training they will support and the level of investment in training they will make and ensure that these feed into the Workplace Skills Plans and in turn their Sector Skills Plan. They also need to establish whether there will be a sufficient pool of skills to recruit from and whether this pool will be 'trainable'. Organised Labour needs to determine the skills needs of its members and ensure that these inform Workplace Skills Plans.

Researchers – educational institutions, research bodies, parastatals, NGOs, private sector

This group needs access to data that they can use to answer different questions that they may be seeking to understand. The nature of the data that this group requires will differ depending on the purpose of the research and the client, but what is paramount is that the data is valid and reliable and can be used to address HRD imperatives required for policy development and evaluation.

People who require individual learner records – employers, individuals and providers

---

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

Within this context the primary need is for information on learner achievement in relation to qualifications/unit standards; this relates to the need for a place that learners, employers and providers can access information on previous (i.e. completed) learning and current (i.e. registered) learning. Interviewees indicated a need to be able to access a consolidated official record of all learner achievement from GEC up and the need to validate and benchmark foreign qualifications.

### **3.2 What Information do they Require?**

The information that is reportedly required is clustered into three main areas within the context of this report. It should be noted that the terms we use have different meaning in different contexts and so for ease of analysis, the terms used are defined in relation to how they are used within this report.

- Economic, development and social - in this context is defined as the broader economic, developmental and social imperatives and the way in which these shape the demand for education and skills that will be required in the next 5, 10 to 20 years. That is, the strategic documents that frame governments overarching vision and objectives for South Africa, such as the government's Programme of Action, as well as macro economic strategies and industrial strategies speak to the broader growth and development priorities for South Africa and highlight the implications for the kinds of education and skills development that is required. Thus the implications of these trends for the nature of the education and skills demand needs to be clearly understood, and fed into HRD planning. This – at least in purpose – relates to the development of the Sector Skills Plans.
  - Labour market - in this context is defined as the way in which education and skills are recognized and used within the labour market and the indications of the patterns of demand. The data requirements within this context focus on a number of related areas: (1) What happens to learners when they graduate from programmes (including an analysis of the number of employed and unemployed graduates by field of study, level of education and institution)? What are the placement rates? What type of economic opportunities do they access? What changes does this have on the socio-economic status of learners? (2) What is the level of mobility (turnover) within a company but outside of the specific occupation - or to another company? Who is leaving the labour market and for what reasons – health, immigration? What
-

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

is the attrition rate? (3) What wage and salary rates do graduates at different levels receive? What are the trends in specific occupations and sectors? (4) Who is active in the labour market (including levels of literacy and educational and skills profile within the different sectors)? (5) Which vacancies are businesses struggling to fill?

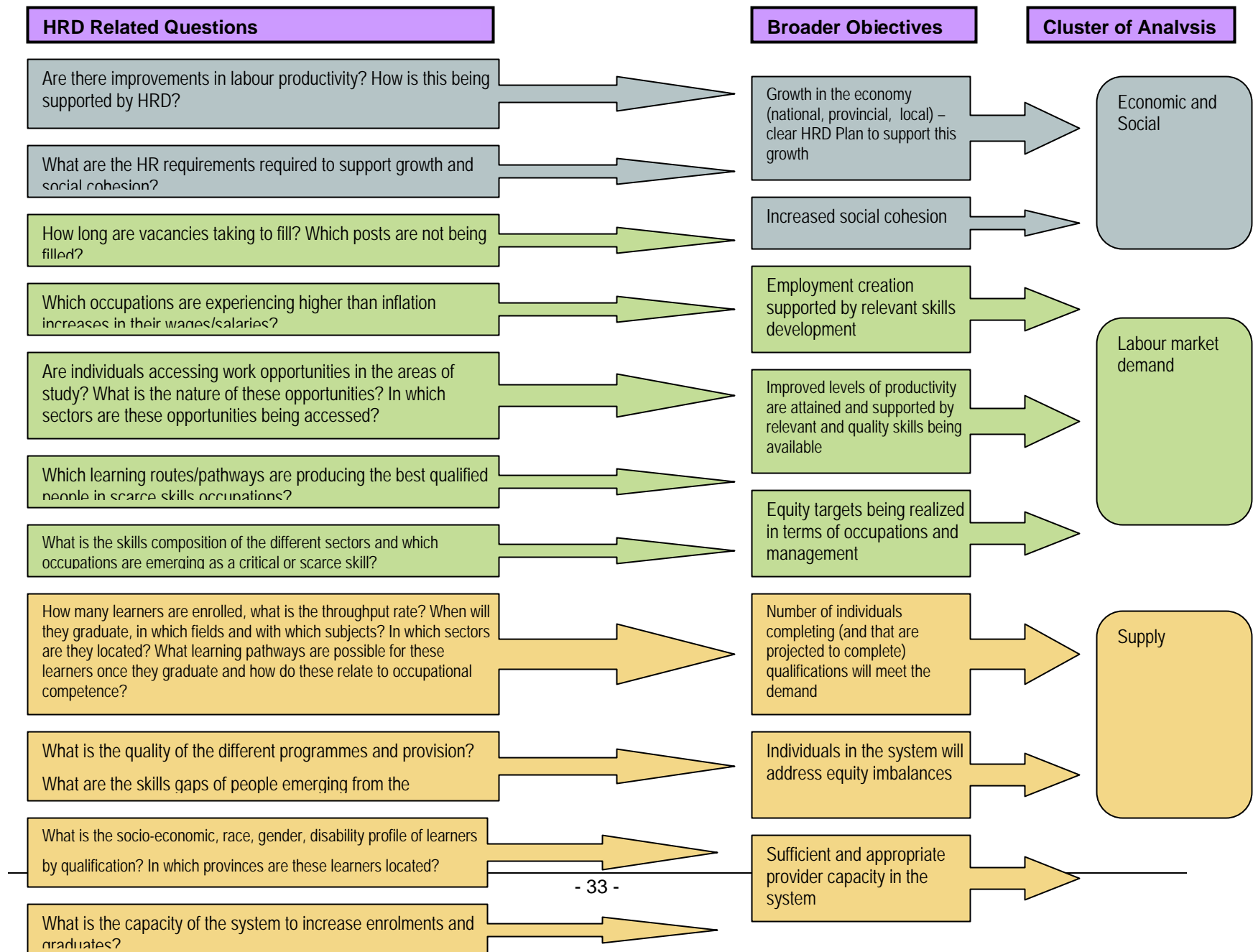
- Labour market supply – in this context focuses on the learners that are being produced within the education and training system. This includes: (1) The number of learners enrolled in different programmes, throughput rates, and the numbers that graduate. This information needs to be understood in terms of the field of learning, the standards or subjects and qualifications achieved as well as race, gender, disability, socio-economic status, geographic location (learner and provision) as well as the provider and the sector that is supporting the learner where relevant. Further, this information is required at both the level of the individual learner as well as at an aggregated level (2) The nature of provision (This includes both institutional learning, including public and private providers as well as workplace providers) – which providers offer what programmes? What capacity is in the system to escalate provision where required (human (staff) characteristics and responsibilities, infrastructure and financial resources available)? What is the scale and quality of this provision? Other quality indicators that may be relevant within a particular sub-sector of the education and training system – for example in higher education there is a need to understand the research capability of the institution (3) What are the funding arrangements for these programmes and for the learners participating in these programmes (private, government (DoE, NSF, NSFAS, etc, international)? (4) What qualifications are available in the system? What are the points of articulation between qualifications? How do these qualifications relate to occupations as defined by the Organising Framework of Occupations?

To support HRD planning there is a need to consider these three clusters of information in order to answer the key questions of the users of HRD information. The kind of data that is required within each cluster is illustrated on the following page through a set of indicative questions that the different constituencies are asking of the system (column one), the broad objective that is being sought is then provided in column 2, and the level at which these objectives are operating is provided in column 3.

---



## A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning



#### **4 Where is the Data Currently Being Held?**

With regards to the **economic and social data** this is primarily collected through processes that are in place in Stats SA as well as data collected by industrial associations. Specifically Stats SA indicates that it provides data in terms of economic growth (GDP, Industry and trade statistics, Financial statistics and National Accounts); Price Stability (CPI, Producer Price Index, Income and Expenditure Survey); Service Delivery (Community Survey, General Household Survey, Poverty Survey); It is anticipated that this data will continue to primarily fall within the realm of Stats SA and industry bodies as well as bodies such as the National Productivity Institute. Responsibility for this data is therefore not considered within the context of this report as it is considered unlikely that the NLRD will begin to collect data that relate to these data fields. Rather, the ways in which this data can more easily be analysed as part of Sector Skills Planning processes needs to be explored, and this is already part of a separate process within the Department of Labour.

**Labour market data** is less institutionalized and responsibility for this data is dispersed across a range of institutions with many data fields currently not being populated with data or links made between data fields in different databases .

- Stats SA is responsible for a number of key surveys including with regards to Employment and Job Creation (Employment and Earnings data which provides data on employment and earnings and the LFS) and Demographics (Population Census, Mid year population estimates, health and vital statistics). The General Household Survey collects data which can compare level of education as well as access to economic opportunities and wage and salaries that the individual is receiving. The Labour Force Survey probes this in more depth and establishes both the qualification and field of study and then explores whether the individual has access to economic opportunities, the nature of that employment and the wage/salary that the individual receives. These surveys all include data in terms of learner profile (race, gender, disability, socio-economic status, and province).
- In addition to Stats SA the Department of Labour and the SETAs have attempted to track and trace learners and the nature of the opportunities that they access. Tracing of learners post-qualification has proved to be extremely difficult, and there has been limited and

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

inconsistent implementation of this DoL requirement of SETAs. The Department of Labour is also working with the HSRC to monitor vacancies; and they indicate that they are currently establishing a database which can map the qualifications that learners attain against occupational categories specified within the Organizing Framework of Occupations (OFO). This is intended to also assist with tracking the extent that scarce skills are being addressed. Further, it is anticipated that certain occupations – such as artisans – will be required to register on the database and provide updated information about employment status and continuing learning and development. The Department of Labour also has data re wage rates but this data has not been kept in a manner that easily allows for the tracking of trends and this is an area that is also receiving attention within existing processes.

- The Department of Education is also putting in place a learner tracking system which allows the Department to understand what options learners pursue after completing their FET programme. An interviewee from the Department of Education reports that this is modelled on NSFAS which also tracks those learners that received loans when they graduate from their higher education programme. The possibilities for tracking learners after higher education programmes are also being reviewed. Databases of alumni often have details of the current employment of graduates and many faculties attempt to maintain contact with graduates, particularly at Masters and Doctorate level.
  - Government departments also collect some of this data pertaining to their sector. As examples, the Department of Public Works, together with the Construction Industry Development Board (CIDB) collects and analyses certain trends in the construction sector, the Department of Environment and Tourism has undertaken a number of studies on related issues.
  - Many of the industry associations, union research departments and bargaining councils also track trends regarding wages and salaries. Some also keep records about vacancies and employment trends.
  - An increasing number of professional bodies are commissioning skills audits and maintaining a database of members' skills and occupations, so that they can make informed inputs to sector industrial and skills planning processes.
-

## A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning

---

Currently the NLRD does not keep data that relates to the cluster of indicators relevant to the labour market, although SAQA has suggested that this could be a possible role for the NLRD and that certain of the professional councils have indicated a willingness to place their professional register data onto the NLRD and that they would have the capacity to capture this data across professions. This would allow for an analysis of the percentage of learners that graduate who then continue to practice and ultimately attain professional registration status. The viability of the NLRD playing a more extended role relating to labour market information is considered in the recommendations section of this report.

The research has highlighted that in relation to the supply data pertaining to education and skills, if there is one thing we are not short of, it is data collection points! There is no shortage of activity, but there is perhaps a shortage of data that can be used to meet the purposes of the users. This section provides an overview of the databases that are repositories for data pertaining to the supply of learners at different points in the pipeline. The following section then considers how these information sources interact with the NLRD.

The most comprehensive data collection is currently being done in the **Department of Education**, which uses the following data sources. These are organized in terms of the different sectors for ease of analysis<sup>15</sup>:

### 4.1 Schooling

- EMIS consists of a series of censuses, and deals largely with learner information (but also other aspects of schools). Data for EMIS is primarily collected through a Headcount Survey. The Headcount survey is a nationally standardised one-page survey that collects learner and educator numbers from all public and independent ordinary schools at the beginning of each school year. The Headcount has operated for many years, and currently constitutes the chief source of enrolment data for the schooling system. There is also the Annual Survey which collects data through a school booklet (currently about sixty pages long though the revised version will be implemented in 2008) and a one-page educator survey (each educator fills in one form). Provincial Education Departments (PEDs) run the survey in all public and independent ordinary schools, and capture the data, and the DoE consolidates the data into a national database.

---

<sup>15</sup> Drawn primarily from 'The overall monitoring map of the Department of Education', p. 3

---

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

- NEIMS is a new system intended to provide a live database of physical infrastructure – this provides a comprehensive picture of the conditions of facilities, utilization of sites, the availability of basic services on site, site security, site access, type of communication available, number of learners per classroom, numbers of learners per educator, availability of resources such as libraries, laboratories, computers, desks and chairs, and facilities for people with disabilities.
  - Persal (contains extensive information on the human resources within the public sector and includes data pertaining to the post (in terms of job title, salary and occupational classification) as well as a description of the necessary and desirable competencies for the job; indicators of competencies; desirable characteristics for employment and promotion within an occupational category and Basic Accounting System (BAS) which is a database managed by Department of Treasury that aims to provide a single integrated financial management and accounting platform, to allow for comparability of financial information. It therefore provides considerable information about human resource development activities implemented by government at all three spheres. .
  - Matriculation Board – has data on examinations drawn from the provinces and captured through SITA
  - Systemic Evaluation - This programme is a national programme and evaluates students' cognition and assesses progress towards the attainment of the goals of access, equity and quality. This is done through a comprehensive questionnaire distributed to students, teachers, parents and district officials. Through Systemic Evaluation, the Department of Education aims to provide a national framework for the evaluation of the education system at the key transition stages of Grade 3 for the Foundation Phase, Grade 6 for the Intermediate Phase and Grade 9 for the Senior Phase. There are also other similar assessments which are administered across countries, which include: SACMEQ (the Southern and Eastern African Consortium for Measuring Educational Quality), which is a programme geared towards measuring inputs and outputs in representative samples of schools in various countries (South Africa joined the programme in time for the 2000 data collection run) and TIMSS and PIRLS which are also international learner testing programmes which allow us to gauge the quality of schooling in South Africa against international benchmarks.
-

## A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning

---

- In addition, the Department is intending to introduce 'LURITS', the Learner Unit Record IT System (schools will upload information on individual learners into LURITS using SA-SAMS) in 2008. This system is already in place in the Western Cape and will be extended to other provinces. The programme will allow individual learners to be tracked as they move through the system which will allow for a more accurate picture of throughput within the system.

### 4.2 FET Colleges

- **FETMIS** - This is designed to feed off the operational systems of the FET colleges. FETMIS is intended to collect student specialisation and educator data in a DoE-determined format, and produce standardised reports. At present this system operates as follows: the Department of Education distributes a CD to each of the colleges. The colleges ensure that all the data fields are completed and then submit this data to the provinces. The provinces then provide this data to the national Department of Education which then retains a full record of this data. This system is in an embryonic stage but has been approved and data has begun to be collected through this mechanism.

### 4.3 Higher Education

- **HEMIS** - HEMIS is a DoE system which feeds off operational systems of the HE institutions in the country. The DoE provides SAQA with the HEMIS data which includes information on: student enrolments, course registration, student results, graduations, throughput rates, retentions, residences, academic staff members and employees, bursaries and loans from the NSFAS, available space, lecture rooms and venues, exam timetables, qualification profiles of staff, achievements and awards for researchers etc. Of this data, only the following are included in the transmissions to the NLRD: student enrolments, course registration, student results, and graduations. It is therefore possible to calculate throughput rates and retentions from these.

As indicated the **Department of Labour** collects labour market data and in addition it also carries the responsibility for supply related data. This is primarily gathered through the SETAs as well as in relation to monies allocated through the National Skills Fund.

---

#### **4.4 Skills development**

- **SETA information systems** (ETQA databases) collect information on enrolments and achievements in NQF qualifications and unit standards, legacy qualifications, learnerships, and apprenticeships. They rely on providers uploading the data at the commencement of programmes (enrolments) and on completion (assessments of learner achievements). It is noted that the nature of these databases may change with the formation of the QCTO as this will in the future be responsible for quality assurance of trades and occupations across all the sectors of the economy. This will need to be taken into account in the formulation of the role of the NLRD as it is envisaged that it will have implications on the flow of data.
- **The Sector Skills Plans** developed by the SETAs are required to provide an assessment of existing provision covering initial training and continuing training systems, issues of qualification standards, training facilities and equipment, the quality of trainers, accreditation, funding, location, gaps in provision, and general deficiencies in the output of the education/training system.
- The Department of Labour collects aggregated information on **learnership** enrolment and achievement, and on **artisan** enrolment and achievement (this includes data provided by INDLELA which is responsible for Trade Tests and in particular for Section 28 tests that may not go through the SETAs).

#### **4.5 Other Data Sources**

- The Professional Councils manage information relating to professional registrations and continuous professional development. In some cases these registers allow for an analysis as to whether individuals are working in their professions, however many of these databases cannot provide this information. The professional councils that are also ETQAs (there are 6 of these) are able to provide this information to the NLRD. It is reported that the HPCSA is the only one of these councils that is still struggling to meet the NLRD's requirements.
  - IEB – has data on adult examinations which it administers and submits results to the NLRD. UNISA also has this data although it is not yet made available to the broader system.
  - Private higher education institutions submit data directly to SAQA at the request of the HEQC
-

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

- UMALUSI: retains records of individual matric performance in terms of an aggregated mark as well as a record of each subject per individual as well as records of individuals performance for each programme that individuals complete national assessments for (this has been the N programmes and will in the future be the NCVs). They rely on SITA which houses all the data received by the Provinces.
- SETA and band ETQAs over and above their role in terms of learner enrolment and achievement (listed above) also retain data about the accreditation of providers and the monitoring process that they undertake to ensure the quality of this provision.
- StatsSA has surveys that contain information that provides a picture pertaining to the supply of skills: For example the General Household Survey includes questions pertaining to levels of literacy, highest level of education achieved, if not studying then reasons why individuals are not attending an educational institution, if studying which type of institution do you attend, whether it's a distance education institution or not, and if not how long the individual has to travel to access the education institution, amount of fees, whether the individual has a bursary, questions pertaining to problems experienced in the learning institution. The LFS asks about levels of literacy, highest level of education and also establishes in which field this learning has taken place using the 12 learning fields defined within the NQF, it also asks re skills training received and the length and field of these programmes (also in terms of the 12 learning fields), the survey also probes the nature of the institution and the modality of the programme.
- The HRD Data Warehouse, managed by the HSRC, is a repository of research outputs in which summary data generated by researchers is presented in tabular format.

### **5 The current role of the NLRD within this data landscape: focusing on supply data**

This section outlines the stated aims and role of the NLRD and indicates which supply data SAQA indicates they do not currently collect. It also indicates the flow of data into the NLRD, however the challenges pertaining to the flow of data and for ensuring that a complete set of data is submitted to the NLRD and can then be analysed is only captured in the following section.

The Mission of the National Learner Records Database is to:

---



## A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning

---

“Develop and maintain a Management Information System for the NQF that provides reliable, relevant, up-to-date, accurate, user-friendly, and accessible education and skills development data for the people of South Africa, and which contributes to national growth and development.”

Thus, as per the mission statement SAQA indicates that the NLRD is an integrated information system to facilitate the management of the NQF. Its two main functions are described as being: to assist policy makers with comprehensive information to enable informed decision making and to provide learners and employers with proof of qualifications obtained. The two “streams” of the NQF broadly relate to unit standards & qualifications and to learner achievement.

The first stream enables users to search the NQF and establish which qualifications or standards exist in a particular field or occupation. This allows planners to determine which qualifications exist in a particular learning field or occupation on the different levels of the NQF. It includes qualifications that have been generated through the SGB processes as well as provider qualifications. There are, however, certain trade qualifications that lie outside of the NQF – SAQA reports that this is currently being addressed and that these qualifications will all be registered on the NQF. This information set also includes some information on who provides these qualifications although this is not complete (as not all ETQAs have transmitted this information as yet) and many qualifications do not have this information listed, or do not have all providers listed against the qualification. The extent that users report that they find value in this ‘stream’ and can use it within their planning process is reflected upon in the next section within this report.

The learner achievement stream focuses on data pertaining to learner achievement, and where possible enrolment. This data relates to qualifications from level 1 on the NQF with the exception of schooling where the GEC is currently excluded from the NLRD. As part of this second stream (learner achievements), the **NLRD** receives data from many of the data sources outlined in the previous section – specifically it receives data from:

- UMALUSI provides achievement data pertaining to schooling at an FET level.
  - Department of Education provides SAQA with HEMIS data on an annual basis – this covers learner achievement for public higher education institutions.
-

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

- Private higher education institutions submit data directly to SAQA at the request of the HEQC (the data is captured on a system specifically designed together with the HEQC).
- SETA ETQAs and professional ETQAs also submit learner data to the NLRD.
- IEB provides the NLRD with data on adult learning

SAQA states that it is committed to make this data available; the NLRD provides learners and other subscribers with learner achievement records where required. There have also been two publications which provide an analysis of the higher education data (for the period 1992 to 2001 and for the period 1995 to 2004). The NLRD also provides this analysis as a customized service to certain clients.

However, the NLRD has not yet achieved a complete set of data from SETA ETQAs and so it has not been possible to generate a publication which addresses supply related issues pertaining to occupational qualifications. Further, while the NLRD has data pertaining to ABET from the IEB it does not have this data from the adult learning centres or UNISA so cannot provide a complete report pertaining to ABET. With regards to schooling, the NLRD receives matric results from Umalusi however it does not receive data pertaining to enrolment and SAQA indicates that the records are not in a format that allows them to undertake an analysis of these results. Currently, the Department of Education takes responsibility for the publication of this data. The NLRD also does not include data pertaining to the FET Colleges, and this data is partially made available in different formats at present.

It is also important to note that at this stage this stream focuses on learner achievement (in terms of throughput and completion) but does not include information about provider capacity. This information as it related to the Department of Education's mandate is retained, analysed and reported by the Department of Education; for example, in the National Assessment Report which provides this analysis pertaining to schooling.<sup>16</sup> With regards to the providers that offer programmes against occupational qualifications, it is noted that in the process of accrediting providers to offer these programmes, the SETA ETQAs collect data pertaining to the existing capacity of providers in terms of facilities, resources and practitioners, but this data is generally not analysed within the SETAs in a manner that allows for planning in terms of projected

---

<sup>16</sup> Public Ordinary Schools) Contract EDO 305, National Infrastructure Management System, September 2007.

---

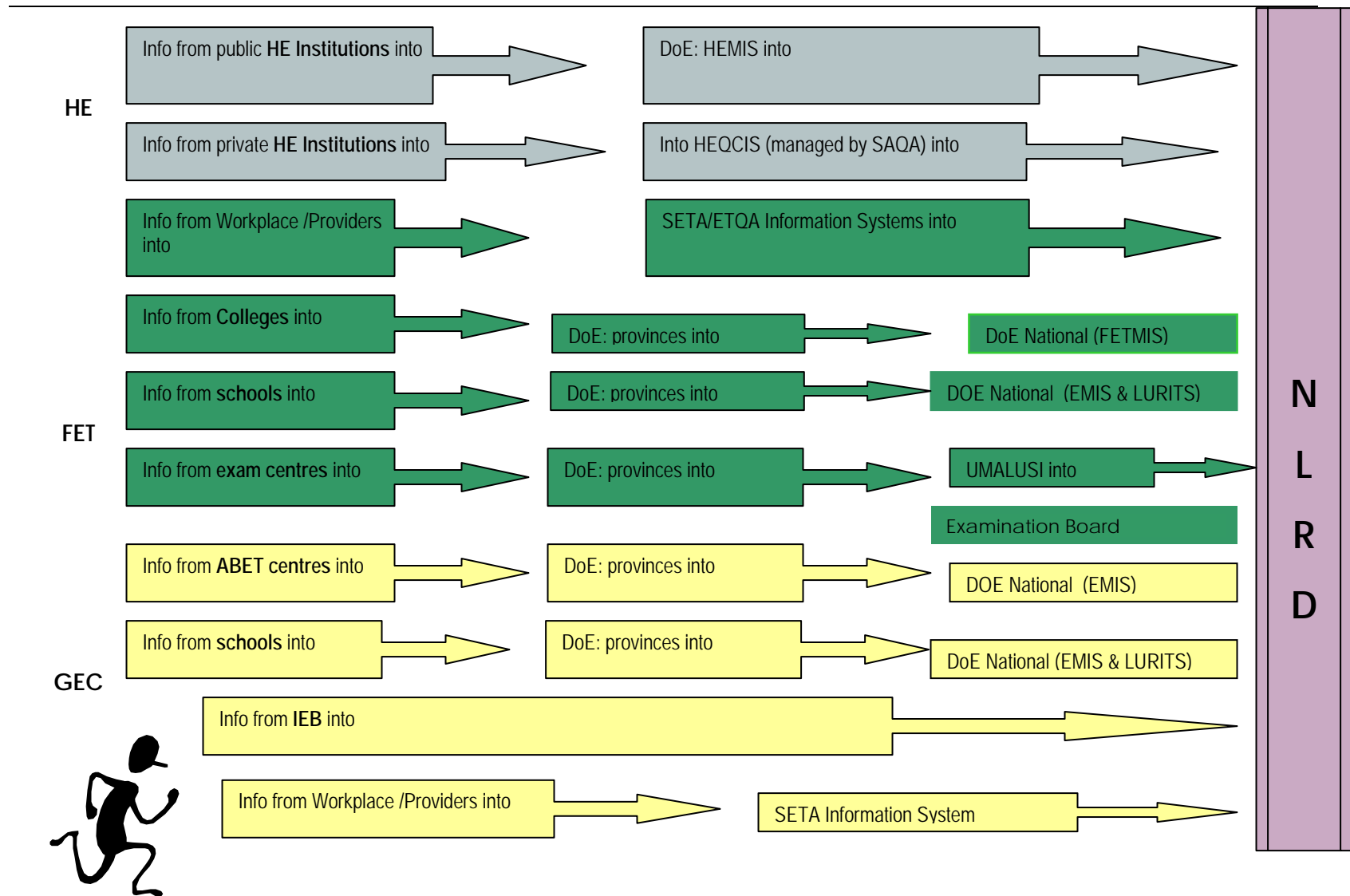
## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

capacity. This data is also not currently captured within the NLRD making analysis at this level impossible at present. Further, the capacity of workplaces to support the workplace component of these occupational programmes is also not captured within this context.

Some of the complexity of the flow of data (relating to supply) is illustrated by the following diagram:

## A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning



---

## **6 What are the challenges in relation to accessing information on skills and education?**

The framework that we are using in this report considers three areas in which data is required to understand HRD: the economic and social data, labour market demand data and data pertaining to supply.

As indicated previously data related to economy and society is crucial to enable planners to determine the scarce and critical skills that will be required to support growth and development. The report highlights that some of the data is already being used in the SSP processes although it is suggested that SETAs have not been able to access a consistent set of data that allows for an analysis of the skills required. Interviewees suggest that it would be valuable if additional datasets were available and that existing data be used more effectively. As stated previously, it is argued that this data falls outside of the remit of the NLRD and therefore specific challenges and recommendations relating to this data are not addressed in more detail in this report.

With regards to the data that in this report we have defined as labour market data the number of collection points for this data have been already highlighted in this report. However, despite this there continues to be real difficulties experienced by stakeholders in obtaining an accurate picture of labour market trends in the different sectors and the implications of these for human resource development planning. As with the economic and social data, while interviewees highlighted challenges pertaining to these datasets, many of these labour market issues are seen as outside of the scope of the NLRD - such as wage trends and vacancies. Therefore recommendations pertaining to the increased availability of these datasets are not listed within this report.

Supply is clearly the area in which the NLRD has thus far primarily focused, and a number of challenges pertaining to supply related data have been raised by interviewees. Specifically, an area in which challenges were raised relates to learner tracking: this includes an analysis of learner placements upon programme completion, and where relevant the percentage of graduates that practice and secure professional registration. Interviewees stated that the central challenge is the absence of data relating to the pathways that individuals navigate upon completion of their studies. The complexity and costs associated with collecting this data are widely acknowledged, however, this research process suggests that departments are in the process of developing these mechanisms internally and that Stats SA surveys, if slightly

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

adapted, also provide an opportunity to access this data in an aggregated form. It is therefore not recommended that the NLRD extend into this area and therefore the specific challenges pertaining to tracking of learners are not detailed in this report.

The one area which falls within the existing scope of the NLRD, and which the NLRD suggests they may wish to expand, is the registration of professionals. This role focuses only on those learners that complete programmes then practise and then seek registration. Currently this is limited to professions though as stated it may extend to artisans in the future. The professional councils already collect this data and will continue to do this as this is central to their mandate, however the challenge will be in encouraging these professional bodies to submit this data to SAQA, for while the professional councils that are ETQAs are likely to provide this data, and most already do, other professional bodies may be more reluctant to do so. It will also require of the NLRD that it can track the individual learners from graduation to registration and determine the percentage of learners that are seeking registration and how long this process is taking. This does not suggest significantly new activities for the NLRD but will require additional analytical capacity and as indicated – if this aspect is to be pursued there will need to be a process of convincing professional bodies that they should provide this data to SAQA.

With regards to supply data a number of critical issues were highlighted by stakeholders pertaining to this data – these, in the main, did not relate to significant gaps – but were much more related to the quality and comprehensiveness of the data. The following challenges were identified:

- It was indicated that there is an absence of common and agreed upon definitions and methodologies even within a sector, and that this makes the data received difficult to use and extends the time required to clean the datasets.
  - Interviewees stated that in many cases the data systems rely on self-reporting and it was indicated that this can delay the process, and requires an extensive period to verify the data, delaying the release of the data. It was indicated that it will be important to strengthen this process and to determine ways to improve the quality of the reporting and to build the verification steps into the process.
  - Linked to this, the timing of when data is collected and submitted and interviewees commented that this process is lengthened because of the number of transaction activities
-

## A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning

---

and the time lags at each point of the process. Respondents highlighted the difficulty of accessing data when developing plans and stated that there are often long delays in receiving the data and the datasets are even then often incomplete.

- Respondents also highlighted the difficulties caused by the inability of the various data-sources to 'speak' to each other, most respondents spoke about their own systems but had little understanding of other data systems and the manner in which they could relate to each other. This problem is largely related to the issue highlighted previously of the lack of a common 'language' in relation to skills supply issues, as well as the absence of protocols pertaining to the collection and housing of this data. It was indicated that this makes it difficult to compare data and to establish the causes where there is significant variance across data systems
- Also, as these data systems have different objectives and interviewees commented that this may lead to certain perversions, for example, it was suggested that where the data is used to determine funding allocations it may lead to the inflation of certain figures.

As an example of how these data challenges manifest in a sector, some issues highlighted in the IRIS project<sup>17</sup> are provided here. Some of the challenges raised in this report are listed here, as though they pertain to higher education data, the issues raised resonates with the challenges raised more broadly about supply data across the education and training system:

- An extensive problem raised by the majority of HEIs interviewed is the quality of the data that is captured by institutions. In many cases, data is captured incorrectly or omitted altogether, resulting in poor quality data with significant gaps, and hence inaccurate and misleading reports. Many of the interviewees used the expression "garbage in-garbage out" to describe this phenomenon. The main root of this problem is the use of untrained staff to

---

<sup>17</sup> Commissioned by the Southern African Research and Innovation Management Association (SARIMA), and funded by grants received from USAID, the Department of Science and Technology and the National Research Foundation, May 2005

---

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

input data. Examples of common errors identified include the incorrect assignment of roles by those entering data on supervisors and students, incorrect use of codes, and omission of key data such as the % contribution by the different authors in research papers. Another consequence of the inadequate training on the relevant information systems for data administrators is the lack of understanding of the workings and capabilities of the system as well as data entry requirements, i.e. what data needs to be entered and in what format. One of the interviewees pointed out that even when training has been provided to staff in their institution, the relevant persons either change roles or leave the institution and are replaced by staff members without relevant training.

- Allied to the above point is the complaint that the HEMIS data is not audited or validated often enough. The ITS has an in-built system for auditing and checking data but it does not appear to be well understood by the users and therefore is not extensively used.
  - Extraction and reporting of data has also been found to be fraught with challenges and inaccuracies. This is partly due to the poor quality of data in the systems but can also be attributed to ambiguities in the requests for data, which are open to misinterpretation by the data extractors. Different interpretations of questions can yield vastly different data sets. It has been pointed out that some institutions are not using the correct definitions when submitting data to the HEMIS and other stakeholders. The lack of integration of the systems at some institutions, even with the ITS, also makes it difficult to extract data for reports. Again this is due to a lack of understanding of the system by users. The lack of integration is especially true of research data. At present, many of the institutions are only able to produce research data in an Excel spreadsheet and are unable to link it to postgraduate student and staff information.
  - A few of the interviewees expressed difficulties in interpreting the definitions and extracting data for the HSRC's R&D survey, while the HSRC has found that data provided by different divisions of the institutions are often inconsistent. In many of the institutions, data for some sections of the survey are not readily available, even from faculty or departmental representatives.
  - Information on the total numbers and profiles of academic staff, for example for the R&D survey, can be a challenge to extract. This is due to the different methods and variables
-



## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

used for capturing staff data as well as the fact that the employment arrangements of the various institutions, including the type of appointments and the ranking systems, vary considerably. The growing trend in some institutions is to offer more contract and temporary appointments than permanent positions.

With regards to the NLRD a similar set of issues were raised, thus while interviewees highlighted the potential value of a system that could provide aggregated data in a manner that captures learners at different points within the pipeline they raised a number of issues as possible obstacles to the achievement of this objective – these can be divided into two main areas of concern: the first relates to the complexity of navigating through the multitude of systems, and the second – which is related – pertains to concerns about the large gaps of data on the NLRD and the implications that this has for the analysis required:

### **Navigating through the databases**

- Institutions are reluctant to ‘trust’ another institution with their data sets for the purposes of their own planning, and in this context while they recognized the potential role for an overarching data system they indicated that this would not be at the exclusion of their own data systems and their on-going role in analyzing their own data for the purposes of their own planning processes.
  - Departments stated that they were reluctant to provide data and then find that they had to wait for the analysis as they needed this to support their own planning processes.
  - Further, departments were concerned that they would then be required to pay for an analysis of their own data and in one case, there was also a concern that if the NLRD kept this data then this would mean that they would ultimately lose a revenue stream as they currently are able to charge for providing information about the data to employers that wish to verify results. SAQA reports that this is not the intention but the concern is highlighted as a number of departmental interviewees mentioned this as a concern.
  - It was noted that even the ETQAs which are required to provide the data to SAQA there are a range of different data systems in use, and ETQAs appear to have very different levels of capacity to capture and verify this data and to then submit this data to SAQA. Many of the
-

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

ETQAs outsource these functions and in some cases there appears to be very limited internal capacity to review the data and the outsourced agencies also appear to have varied understanding as to what is required and the definitions that should be applied. As noted previously, the advent of the QCTO is likely to have a significant impact on this landscape as it means there will be 3 main Quality Councils and therefore the possibility of greater coherence across the system.

### **Gaps in the data submitted to the NLRD**

- There are a number of areas in which arrangements are not yet in place for the data to be submitted to the NLRD. This includes data from the learners that complete the trade tests that are administered by INDLELA. This data is retained within Department of Labour and is not submitted to the NLRD. The NLRD also does not yet have all the data pertaining to FET Colleges and it is also reported that there continue to be some delays and inconsistencies pertaining to schooling data that SAQA receives; schooling data received also does not include data pertaining to enrolment and this data is kept within the Department of Education – rather than UMALUSI which provides the data to SAQA – and will increasingly form the focus of the work that the Department is doing in terms of LURITS.
  - The ETQAs are required to submit data, but as is already indicated capacity constraints and the varied definitions and functions within these data systems result in continued concerns being raised by interviewees about: incomplete records being submitted to the NLRD, delays in uploading data and a perceived lack of recognition of ‘non-NQF’ qualifications which ETQAs suggest explains some of the data that they do not submit – as they indicate that they do not provide data against qualifications that they are not accredited to quality assure. Interestingly this is not considered wholly accurate by SAQA suggesting a continued misperception as to what should be reported to SAQA.
  - In particular artisan qualifications and an uncertainty as to whether this data should be submitted to the NLRD or not. Interviewees suggest that as a result of this confusion some SETA ETQAs do not submit their data pertaining to apprenticeships thereby widening the gap in the data that is available on the NLRD pertaining to trade qualifications.
  - There was also a concern that while the NLRD is designed to capture the enrolment data from SETA ETQAs to allow for an analysis of throughput rates in terms of the different
-

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

programme types, this data is not consistently provided and only some of the SETA ETQAs provide this data. Further, interviewees indicated that not all enrolment data resides with SAQA making trends regarding throughput in some sectors impossible to determine, and it was noted that in some sectors the capacity to capture individual learner records is only now being developed which means that it is not yet possible to undertake this analysis except at an aggregated level. These challenges impact on the ability of policy analysts to model projected supply and to understand where the bottlenecks in the system are.

- With regards to the NLRD stream which focuses on maintaining the content of qualifications and standards on the NQF, stakeholders indicated that while it is easy to navigate the searchable database (on the SAQA website) and establish which qualifications are registered, the articulation relationship between these qualifications remains difficult to determine – for example, if a user wishes to understand what qualification a learner must undertake to become an electrician it is not possible to search for an electrician as the trades themselves are not yet linked to the NQF, instead the user can search the electrical qualifications that have been generated by the electrical engineering and construction SGB – this yields a result of 12 qualifications. It is difficult for the user to understand which of these will lead to the trade. Further, as these qualifications are all on levels 1- 4 on the NQF it is difficult for the user determine the progression arrangements that will enable the learner to then access a technician programme at level 5 on the NQF. Interviewees suggest that these challenges pose difficulties for learners but also make modeling more complex for planners.
- Further, as raised previously, the NLRD does not currently provide data that supports an analysis of the capacity of providers in the system, and while this data is available for schooling, FET Colleges and higher education it is difficult to access this data in a consolidated form for occupational qualifications. This information is even more disparate in terms of the workplace provision where SETAs have different systems and keep different types of information about the workplaces.

SAQA reports that they have put in place a number of measures to support the improvement of these datasets and that these have yielded real improvements in the system. Some of the initiatives that have been undertaken by SAQA to strengthen this capacity include: the implementation of an Information Administrator Learnership at ETQAs, the appointment of a

---

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

Data/Learnership Coordinator at SAQA who both supports the learners within the learnership and assist the ETQAs with data submission, and the development of a league table for ETQAs. In addition to this, to ensure the validity of the data SAQA has also introduced a 'minimum standard for data transmission' and the 'Edu Dex data testing solution'. SAQA indicates that where the learners have continued with the learnership, the data from the ETQAs in which these learners are based, has improved. SAQA also reports that the Data/Learnership Coordinator has been able to assist SETA ETQAs.

However, SAQA and the ETQA interviewees highlight continued difficulties with this process and it is acknowledged that one of the problems that SAQA faces is that although data submission is a requirement for ETQA accreditation, SAQA relies very much on promotion, workshops and persuasion. Some ETQAs have consistently not supplied usable data and they have not been sanctioned in a manner that ensures that this data is provided. This issue is exacerbated where the datasets lie outside of the ETQAs and SAQA must rely entirely on a shared vision of the NLRD to access the data. Unless these issues are addressed then the added value of the NLRD to national human resource planning that is envisaged in this report will not be achievable.

### **7 Broad Recommendations**

The NLRD is presently a repository of and data source for information on skills supply with a very limited role being played in Labour Market data insofar as it relates to the registration of professionals. This report has highlighted the myriad of sources of this data and some of the complexities that are encountered in ensuring that this data is submitted to SAQA in a manner that ensures both that the data is timeous (available within a year), complete and that the data is verified. Further, while by its nature, an 'NLRD' must hold information in terms of individuals – it is a national database of learner records - to be useful for research processes, the data should ideally be available at the level of individual records, and at a level which can be aggregated, in order to show broad trends. This report has highlighted some of the difficulties that have been experienced in ensuring that complete data is submitted into the NLRD within the agreed upon time frames and this final section of the report highlights the recommendations made by the team regarding the extent that *“the available database can be or has the potential to be the national education and skills database.”*

---

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

### **7.1 Recommendation One**

With regards to the data required to determine the demand flowing from economic and social growth it is recommended that this not become a focus for the NLRD. Instead, it is suggested that this data could possibly be collected by adding certain fields within existing processes being facilitated by Stats SA and by ensuring that the data gathering mechanisms within the industry associations are developed and used consistently. Both Stats SA and business have indicated a willingness to consider how to ensure that this data is consistently collected and made available as part of the NHRD processes being conceptualized within the Department of Education under the auspices of the NHRD Task Team. This is being undertaken in a manner that takes into account the work that the Department of Labour is doing to ensure that there is alignment between SSP and Industrial Strategy processes.

### **7.2 Recommendation Two**

With regards to the HR dimension of labour market issues - which is critical to providing a picture of the competence and qualifications of individuals working in the labour market and the manner in which the labour market is absorbing these individuals (wages, salaries and vacancies) - there are areas in which data is not yet being collected, and other areas in which data exists but it is not yet available in a consolidated manner. At the point of this research, stakeholders generally seem committed to supporting these data collection activities directly and both the Departments of Education and Labour are putting in place individual learner tracer systems. The Department of Labour and Stats SA are both actively collecting other data pertaining to human resources in the labour market and while this data is not all captured and analysed for these purposes there is openness to addressing these imperatives.

It is recommended that in the light of these developments, this data should not become an area of focus for the NLRD as this would involve competing with other systems, and would require a scale of activity that is entirely different from the existing capacity within the NLRD.

It is further recommended that instead of undertaking these activities, it will be preferable for the NLRD to consider whether it is possible for them to provide support to the departments that are implementing these monitoring activities - and the capacity implications of this will be considered - and to work with Stats SA to ensure that the relevant fields are included in their existing surveys so as to ensure that more comprehensive data becomes available. This would ensure that data becomes available in a way that supports an analysis of existing skills and

---

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

qualifications within the labour market, as well as other labour market issues and trends which point to the extent that equilibrium between supply and demand is being achieved.

Further, it is recommended that in the light of the indication that certain professional councils are already submitting their registration data to the NLRD, and others have indicated a willingness to submit their registration data onto the NLRD, it will be useful for the NLRD to integrate these datasets and to explore ways in which it can support an analysis as to the number of graduates that are practising and going on to attain professional registration. It may also be possible that a similar arrangement is made for the Department of Labour's artisan register. However, it is not recommended that this be prioritized in terms of resource allocations but that this builds on the existing capacity within the NLRD. Thus these activities could continue where there is willingness to provide the information but should not become mandatory for councils at this stage.

### **7.3 Recommendation Three**

It is recommended that the strategic focus for the NLRD should be to ensure that it can play a key role as a central repository for data on supply. This should allow for an analysis which tracks individual learners, in terms of race, gender, disability, locale and the sector that is supporting the learning where relevant, in terms of:

- Individual records of learners that complete their FET in schools with an analysis per subject area;
  - Individual records of learners that enrol and complete the NCV - this should provide this information in terms of qualifications and subject areas and duration of time that it has take to complete the qualification
  - Individual records of learners that enrol and complete occupational learning programme - this should provide this information in terms of qualifications and subject areas and duration of time that it has take to complete the qualification
  - Individual records of learners that complete programmes in the higher education band - this should provide this information in terms of qualifications and subject areas and duration of time that it has taken to complete the qualification
-

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

In addition, it is recommended that SAQA should support the publication of an annual report which provides an analysis of this data in a manner that supports an analysis of: (i) the throughput rates for different learning pathways and qualifications, and (ii) the numbers of graduates exiting the system within a particular occupation where relevant. This should provide a detailed picture of the learning routes that learners traverse and should allow planners to project potential supply of qualified individuals into the labour market. This could then be mapped against career pathways outlined within the Organising Framework of Occupations (OFO).

### **7.4 Recommendation Four**

Implicit in the previous recommendations is the understanding that there are a plethora of databases which collect and analyse this information and that each agency that collects data has specific and unique needs in relation to their data. Essentially, there are as many data fields and data sources as there are areas of research and areas in which planning is required – for example, someone doing planning on the positioning of new ABET sites is going to need very specific data in relation to socio-economic factors, physical infrastructure, access to facilitators, ability to generate sponsorship, the need for any specific skills in particular geographic areas etc. As it is virtually impossible to predict the nature of the research studies in relation to education and skills supply, it is virtually impossible to ‘design’ a database that can cover all of these eventualities. Further, the data sources have evolved over time and it is doubtful whether there would be a true cost-benefit in trying to bring them together. Finally, as indicated the departments responsible are likely to resist such a change and this will also likely adversely impact on the ability of a central database to effectively play this role.

Thus as per the previous recommendation that the NLRD serve as a central repository for data it is further recommended that there is a need for agreement to be reached about the nature of the data that is submitted onto the NLRD – including fields, definitions, and time frame for the collection, coordination and submission of data. To support the implementation of this recommendation it is suggested that SAQA, as the apex organization, should explore the possibility of forming a relationship – possibly a Memorandum of Understanding - with Stats SA. This could indicate the role that SAQA could play in the promotion of the national quality standard outlined by Stats SA and of ensuring that the data that is loaded onto the NLRD is consistent with this standard. This process should be supported by the fact that the key

---

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

departments responsible for providing this data have already been involved in discussions about the standard, and the Department of Education already has an MoU with Stats SA and is implementing certain of the steps contained in the standard. The emergence of the QCTO could also be seen as an opportunity within this context as it may be possible to ensure that this system is set up in a manner that takes these principles and criteria into account. However, even with this, it should be recognized that this role will take considerable persuasive abilities within SAQA to ensure that the agreed upon datasets are submitted in a manner that can be loaded onto the NLRD.

### **7.5 Recommendation Five**

In addition, it is recommended that the NLRD continue to enable individuals, employers and providers to access verified full records of learning of individual learners. This would allow learners to provide prospective providers or employers with a full record of learning rather than having to return to each institution of learning and would be both more cost effective and ensure the integrity of data so that providers and employers can trust the records.

In practical terms, the recommendation is that a 'learner' is recognised as such on the NLRD when they have completed a qualification at NQF Level 1 (Grade 9 at school or ABET level 1). As the learner progresses on their learning pathway, information on enrolment and achievement is then captured from the various data sources into their learning record.

### **7.6 Recommendation Six**

A further recommendation is that the NLRD should also continue to provide information relating to qualifications and this information should include:

- A detailed listing of all the qualifications that are available and the major fields of study linked to each qualification. This should allow for an analysis of the learning pathway and where relevant this should relate to the OFO. Therefore this recommendation would need to take into account the work that has been completed within the Department of Labour pertaining to the OFO and the mapping of occupations in terms of qualifications and associated curriculum.
-



## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

- A comprehensive list of the providers that offer these programmes (it is noted that the NLRD has some of this information but there is not yet sufficient information in this area to support planning)
- The minimum entrance requirements for qualifications and for programmes offered by particular providers (where there are additional requirements).

### **7.7 Recommendation Seven**

Finally, with regards to data pertaining to the capacity of providers, it is recommended that the feasibility of sectors being supported to produce reports of a similar depth and quality as the NIEMS report be explored. It is noted that HEMIS already allows for such an analysis and that this serves as an input into the higher education planning process. However, while the SETA ETQAs should be collecting this data pertaining to their providers there is very limited analysis of this nature available in terms of providers supporting occupational learning – the NLRD could consider a special project which supports them to undertake such an analysis which could improve planning in the future.

### **7.8 So what is Needed?**

This section builds on the recommendation that SAQA establish an arrangement with Stats SA and considers what will be required to ensure compliance with the Stats SA Draft Certification Framework (key aspects of the standard are attached as an annexure to this report). On the assumption that the NLRD wishes the statistics that it releases to be recognized as Scale Five statistics - SAQA needs to ensure that the data complies with the following quality criteria, as articulated in Stats SA's Draft Certification Framework:

- The agency or organ of state wishing to have their statistics certified must be a member of the National Statistics System.
  - Statistics to be certified must satisfy the stipulations of Section 3 of the Act regarding the purpose and principles of national statistics, namely:
  - To produce statistics that will inform planning, decision-making, and monitoring or assessment of policies, decision-making or other actions of organs of state, businesses, other organizations or the public;
-

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

- To protect the confidentiality of the identity of the respondents as well as the information provided in the process of collecting national statistics;
- To meet all the quality dimensions of national statistics to the satisfaction of the Statistician-General.

The standard specifies the concomitant capacity that will then be required of a body that is responsible for statistics that are recognized on scale 5: it must demonstrate the capacity, in terms of human resources, infrastructure and equipment, to sustainably produce statistics over a period of at least five years (Draft Certification Framework, Stats SA).

It is noted that the NLRD operates with a very limited number of staff, and that there is not the internal capacity to support this level of statistical analysis or to provide the level of support that may be requested by agencies to enable them to improve the quality of the data. SAQA will therefore need to be able to make a case for additional resources if it is to be able to play this role, this will include both a budget to enable SAQA to ensure that there are complete sets of data collected, that it can be analyzed and made available, and that SAQA can deploy the relevant level of statisticians and data experts to support the agencies that provide the data so that the data can be provided in accordance with the standard. This does not suggest that SAQA will employ these individuals but it should have a pool of people that have the relevant expertise and the resources to be able to deploy them.

### **7.9 Red Flags**

In implementing these recommendations, there are a number of areas that will need to be monitored:

- The issue of who is accountable for what in relation to the provision of data for the NLRD needs to be very carefully outlined, and agreements managed.
  - The manner in which SAQA will support departments and other agencies with their data collection processes so that they adhere to the Stats SA standards will need to be carefully managed – this includes the need to ensure that a balance is struck between being supportive on the one hand, and ensuring that bodies such as ETQAs which are accredited by SAQA are required to comply with requirements as a condition of accreditation, for example that they are ‘forced’ to meet data submission deadlines.
-

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

- Currently some agencies charge for the use of information from their databases (eg. Umalusi). It has been indicated that in this case (between SAQA and Umalusi) the funding protocols have been worked out in a manner that is satisfactory – this issue needs to be monitored to ensure that it does not negatively impact on relations between the NLRD and the data providers.

### **8 In Conclusion**

It is noted that this report suggests that the concept of a national HRD database is not feasible, and recognizes that different aspects of HRD data will need to be the responsibility of a number of institutions. It suggests that while there are certain roles that the NLRD could play in labour market data – such as the analysis of registration data this should not be the focus of the NLRD. Rather, it suggests that the NLRD should play a central role in supply related data, and in summary, on the understanding that the challenges identified in this report are addressed, it is recommended that SAQA, and specifically the NLRD, could:

- Serve as the central repository for enrolment and completion data:
- Support the publication of an annual report which provides an analysis of the supply pipeline.
- Work with Stats SA to consider what is required to ensure compliance with the SASQAF standard, and to then provide further momentum to the process of complying with the standard, and support increased compliance across all the bodies that will be supplying them with data.
- Provide a verification service for individuals, companies and providers on individual qualifications, through holding complete and up to date learner records.
- Enable learners, providers and planners to determine possible learning pathways against the OFO, and to determine where (at which providers) they can access these programmes.
- Finally SAQA could, through its accreditation function, support ETQAs to analyse data pertaining to providers so as to be able to produce reports which provide an indication of provider capacity within particular sectors.

This suggests that SAQA needs to develop a plan to ensure that it has the requisite resources to: (i) work with the suppliers of the data to ensure that there is compliance with the SASQAF

---

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

standard, (ii) support high levels of data analysis, (iii) and develop annual publications which make this information available. It is suggested that SAQA could explore whether an arrangement could be made – possibly within the context of the NHRD Strategy - to explore ways in which this activity can be supported at this level rather than on a cost recovery basis. This would involve a discussion as to how this supply data could support the monitoring of the NHRD in terms of the agreed upon supply indicators.

---



---

**Annexure 2**

**Fields to be captured within the NLRD**

In order to play these roles the NLRD needs to take into account a range of variables. The bulk of these are already included in the NLRD fields, although not all are completed by the relevant providers resulting in incomplete data being submitted to the NLRD. Further, the imperative to consider the sector that is supporting the learner is a new dimension that is proposed.

1. Learner Identification (this can be taken directly from the current NLRD fields for learner ID – see Annexure 1)
    - 1.1 ID number
    - 1.2 Full name
    - 1.3 Date of Birth
    - 1.4 Race
    - 1.5 Gender
    - 1.6 Disability status
    - 1.7 Socio-economic status (including employment status)
    - 1.8 Address and contact details (all available)
    - 1.9 Nationality
  2. Learner Achievement – GEC
    - 2.1 Certificate (ID, Provider, Assessor, Enrolment date, Achievement date)
    - 2.2 Provider (schooling or adult – if so programme and provider)
  3. Learner Achievement – FET
    - 3.1 Qualification (ID, Provider, Assessor, Enrolment date, Achievement date)
-

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

- 
- 3.2 Unit Standard (ID, Provider, Assessor, Enrolment date, Achievement date)
  - 3.3 Other qualification (name, provider, contact details of provider)
  - 3.4 Provider and sector (where relevant)

### **4. Learner Achievement – HE**

- 4.1 Qualification (ID, Provider, Assessor, Enrolment date, Achievement date)
- 4.2 Certificate/module or part completed qualification (ID, Provider, Assessor, Enrollment date, Achievement date)
- 4.3 Other qualification (name, provider, contact details of provider)
- 4.4 Provider and sector (where relevant)

### **5. Current learning underway**

- 5.1 Qualification (Band, ID, Provider, Assessor, Enrollment date, Anticipated completion date)
  - 5.2 Unit Standard (ID, Provider, Assessor, Enrollment date, Anticipated completion date)
  - 5.3 Other qualification (name, provider, contact details of provider)
  - 5.4 Provider and sector (where relevant)
-

**Annexure 3**

**Stats SA Draft Certification Framework**

Stats SA highlights that national statistics are certifiable on a five-point scale as follows:

- **Scale Five - Official Statistics:** Statistics are designated as official statistics, if they are sustainable, meet the quality requirements as stipulated by Stats SA, and are relevant beyond the organ or agency that collected them.
- **Scale Four - Significant Statistics:** Once-off statistics, which satisfy the quality requirements and are relevant beyond the organ or agency that collected them, are designated as significant statistics.
- **Scale Three – Departmental Statistics:** Statistics which satisfy all the quality criteria, but which are relevant only to the department collecting them, may be classified as departmental statistics. Departmental statistics may be once off or sustainable.
- **Scale Two – Acceptable Statistics:** These are statistics which meet most, but not all, the quality standards. They are designated as acceptable to the extent that, despite their limitations, some deductions can be made, and they can therefore be utilized.
- **Scale One – Questionable Statistics:** Statistics are labeled as questionable if the methods used to collect and compile them are suspect, or if conclusions that could be drawn from the sample are inappropriate. This would include uncontrolled measurement error, reporting standards not defined, problems of overgeneralization or systematic error.

The quality dimensions of national statistics to the satisfaction of the Statistician-General are indicated below:

*Relevance:*

The statistics to be certified must meet the real needs of clients whose activities significantly influence the welfare of the public; and this has to be demonstrated by specifically identifying the users or clients. In addition, the following must be in place:

- A system for monitoring clients' needs;
  - A system for programme or activity review;
-



## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

- A system for determining priorities;
- A system for monitoring performance, especially evidence that processes are in place and that measures of usage such as client satisfaction have been established.

### *Methodological soundness*

The production process of the statistics to be certified must demonstrate good practice in relation to the following:

- Concepts, definitions and classifications;
- Methods of data collection, which must be aligned with international recommendations or agreed methodology by Peers, aligned with international or national classifications of variables, and aligned with international or national frameworks for analysis (e.g. SNA93);
  - Summarization (analysis) of data;
- Reporting of results;
  - Consistency between different data processes amenable for time and spatial series analysis;
- Consistency between different domains of study at the same point in time amenable for cross-sectional analysis.

### *Accuracy*

The statistics to be certified must be demonstrably accurate in the sense that, within acceptable levels of sampling error, the statistics correctly describe phenomena they are designed to measure. The relevant agency or organ of state must demonstrate best practice in survey design, survey implementation and assessment of accuracy.

- The survey methodology must be fully and comprehensively documented, indicating specified targets or assumptions for key estimates.
  - A system for monitoring and correction of errors during implementation must be demonstrated.
-

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

- After the survey there must be a demonstrated assessment whether the survey design was carried out properly, including staff debriefings.
- Levels of coverage of the target population and sub-populations must be demonstrated, preferably with a listing frame.
- Sampling error must be assessed, including standard errors or methods for deriving them.
- Non-response rates or the proportion of estimates imputed must be reported.
- Other accuracy or consistency problems (such as a widely misunderstood question) must be reported.

### *Timeliness*

The statistics to be certified must be timely in the sense that they are available before or when they are needed.

- Release dates for major releases must be announced in advance.

### *Accessibility*

There must be a system in place to inform clients when statistical information is available, where they can locate it and how they can import it to their own working context.

- Salient aspects of every official statistics must be made available in the public domain.
  - There must be a clear policy governing dissemination of statistical information, with particular reference both to pricing policy and to meeting the public good information needs. Where information is intended for sale, there must be published a pricing policy as well as a price schedule.
  - There must be a corporate catalogue system to identify information available to users and to assist them to locate it. The catalogue must also cover information from individual programmes.
-

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

- There must also be in place a corporate delivery system to provide access to information through specified distribution channels and formats.
- There must be put in place usage and user satisfaction measures for the catalogue and delivery system.

### *Interpretability*

The agency whose statistical information is to be certified must demonstrably facilitate user understanding through provision of metadata for the information produced.

- The agency must have a policy in place to inform users on what basic information they need in order to interpret data. Such basic information must include:
- Concepts, definitions and classifications underlying the data;
- Methodology used to collect and compile data;
- Measures of accuracy of the data;
- Metadata needed to describe information required to describe each data holding;
- Direct data interpretation commenting on each release on primary messages in the information.
- The agency must put in place a process to seek user feedback on usefulness and adequacy of metadata.

### *Coherence*

Clear grounds must be spelt out for data comparisons resulting in data reconciliation or explanation of differences. Such grounds must include demonstrated use of:

- Standard frameworks, concepts, variables and classifications for all subject matter topics the agency measures so that measurement is consistent across all programmes and quantities being estimated bear known relationships to each other;
  - Common frames, methodologies and systems for data collection and processing in order to eliminate inconsistencies between data sources (for example, common
-

## **A review of the National Learner Record Database (NLRD) with Regards to its Potential Role in Supporting HRD Planning**

---

formulations of questions when data on the same variable are collected in different surveys).

- The agency must demonstrate the actual practice both of comparing data sets and of integrating data from different sources.

### *Integrity*

There must be in place a system that facilitates for integrity of the statistics to be certified; in particular, the system should provide for the statistics not to be influenced by any individual or organization.

- The agency should publish terms and conditions, including confidentiality, under which dissemination of statistics is undertaken.
  - Conditions for internal government access to data before release must be identified.
  - Ministerial commentary when data are released must be identified.
  - Information must be provided on revisions, and advance notice of major changes in methodology must be given.
-

## **APPENDIX 2**

TITLE : TOWARDS A NATIONAL HUMAN RESOURCES  
DATABASE: THE NATIONAL LEARNERS'  
RECORDS DATABASE ASSESSED

AUTHOR : MORNÉ OOSTHUIZEN  
UNIVERSITY OF CAPE TOWN. DEVELOPMENT  
POLICY RESEARCH UNIT

DATE : JUNE 2008

## **Joint Initiative on Priority Skills Acquisition (JIPSA)**

### **Towards a National Human Resources Database: The National Learners' Records Database Assessed**

PROJECT : NATIONAL LEARNERS RECORDS DATABASE

AUTHOR : MORNÉ OOSTHUIZEN  
UNIVERSITY OF CAPE TOWN. DEVELOPMENT  
POLICY RESEARCH UNIT

DATE : JUNE 2008

## **TABLE OF CONTENTS**

List of Acronyms

Acknowledgements

1. Introduction .....	1
2. A national Human Resources Database .....	2
3. The Potential Role of the NLRD .....	4
3.1 The National Learners' Records Database	4
3.2 The Individual as the Unit of Analysis	11
3.3 Comprehensiveness	11
3.4 Currency of the Data	21
3.5 Ability to Represent the Stock of Skills	22
3.6 Compliance with Other Criteria	24
4. Conclusion .....	26
5. References .....	30

## **LIST OF FIGURES**

Figure 1: Number of Qualifications over Time, by Race – 1900-2008.....	12
Figure 2: Current Age Distribution of Pre-2005 Qualification Achievers, by Race and Gender .....	14
Figure 3: Distribution of Individuals by Number of Qualifications in the NLRD .....	15
Figure 4: Distribution of Qualifications Across Age at Completion, by Race.....	17

---

**LIST OF ACRONYMS**

ABET	Adult Basic Education and Training
DoE	Department of Education
DoHA	Department of Home Affairs
DoL	Department of Labour
ETQA	Education and Training Quality Assurer
FETMIS	Further Education and Training Management Information System
HEMIS	Higher Education Management Information System
HEQCIS	Higher Education Quality Committee Information System
HSRC	Human Sciences Research Council
NLRD	National Learners' Records Database
NQF	National Qualifications Framework
SAQA	South African Qualifications Authority
SARS	South African Revenue Service
SASSA	South African Social Security Agency
SETA	Sector Education and Training Authority
SGB	Standards Generating Body
SMME	Small, Medium and Micro Enterprise
UIF	Unemployment Insurance Fund



---

## **ACKNOWLEDGEMENTS**

The author would like to thank Samuel Isaacs, Yvonne Shapiro, the staff of the Directorate: National Learners' Records Database, and Kirstin Barth of Praxis Computing, for their help and advice during the research and writing of this report. Also, the comments of Haroon Bhorat and assistance of Sumayya Goga are much appreciated.

## **1 Introduction**

The information base supporting skills provision and skill development for South Africa remains inadequate. Given the centrality of the skills challenge, noted in a multitude of forums, policy documents and public pronouncements, the fact that the country's information base in this area is deficient is often overlooked. Indeed, often critical information is of low quality, incomplete and, in many instances, non-existent. As a result, many relatively simple yet vitally important data queries are difficult to obtain and estimates are often of dubious quality.

It is, therefore, critical to any appropriate human resource strategy for South Africa that a high quality, integrated human resource database is constructed. There appears to be a general acceptance that the creation of an effective, high-quality information base on education and training issues in South Africa is both critical and highly desirable. Given the various databases that exist in this arena and their respective strengths and weaknesses, it is suggested that the National Learners' Records Database (NLRD), housed within the South Africa Qualifications Authority (SAQA), should serve as the master dataset on skills, into which other databases and datasets may be incorporated. Specifically, data held by the Departments of Education and Labour, assuming it is in a compatible format, will be merged into the NLRD, as well as any other relevant data held by other institutions.

The purpose of such a national human resources database would be to inform planning surrounding skills and human resources generally, to identify scarce skills and potential problems relating to the skills pipeline, and to facilitate matching of skills supply with skills demand. No doubt, these are ambitious objectives and it is unlikely that any system set up would be able to fulfil them immediately.

This report has been commissioned as a review of the National Learners' Records Database (NLRD) with a view to its potential role as a master dataset on skills in South Africa. It must be noted that the NLRD was originally set up with very different objectives in mind and, irrespective of its performance relative to its own objectives, it is likely that the employment of this database as a master dataset on skills would require some changes or additions, particularly in terms of data collected. Thus, the analysis below takes as its point of departure this potential role that the NLRD may be required to fulfil.

The report is structured as follows. Section 0 looks at what would be required from a national human resources database in order for it to function effectively. These requirements are then held as criteria against which the NLRD will be evaluated. This

evaluation is presented in section 0, after a brief overview of the NLRD. Section 0 then concludes, with recommendations as to a way forward.

## **2. A National Human Resources Database**

Before the question of the suitability of the National Learners' Records Database for use as a national human resources database can be answered, it is critical to define exactly what is required from such a database. It is these abilities and capacities that are required of the database that will determine the extent to which the database can be used and the nature and scope of any changes or additions required. The characteristics required of the national human resources database have, however, not been detailed upfront and, consequently, there is little official guidance as to how the NLRD is to be assessed. In the absence of official criteria, a set of suggested criteria for a usable and useful national human resources database is presented below, against which the NLRD will be assessed in this paper.

Before the criteria are detailed, however, it is important to note that, throughout this paper, the term 'skills' is used in the sense of qualifications, rather than in an occupational sense. This is an important distinction, since qualifications do not map one-to-one against occupations or vice versa, while the link between occupational categories and the education system is indirect. In other words, a certain qualification is but one route into a given occupation, while a given qualification can often entitle an individual to work in any number of occupations. It is asserted here that basing a national human resources database within a framework of qualifications, rather than occupations, is a more sensible choice, aligning it directly with data coming out of the education system.

Firstly, a national human resources database should have, as its unit of analysis, the individual. Without individual-level data, the ability of the national human resources database to provide specific, clear and targeted answers to questions posed in the course of policymaking is considerably weakened.

Secondly, a national human resources database should be comprehensive. In other words, the database should contain, at the very least, the qualifications of all South Africa's citizens. A strong case can be made for the database to include the qualifications of all residents of South Africa, i.e. including non-citizens, but for our

## **Towards a National Human Resources Database: The National Learners' Records Database Assessed**

---

purposes we will opt for the weaker requirement of all citizens.<sup>18</sup> Furthermore, the database should be comprehensive in its coverage of all qualifications that individuals may have obtained during their lives.

Thirdly, a national human resources database should be current. In other words, as far as possible, the information contained in such a database should be up to date, allowing for a reasonable lag. Importantly, the more information there is in the system regarding flows of learners through the educational and training systems, the longer the 'reasonable' lag can be. Irrespective of the amount of information on flows, however, delays in the updating of the database should be minimised.

Fourthly, in order to be useful, a national human resources database should represent the current stock of skills available to the labour market. Thus, while the database needs to be current in terms of the flow of new skills accrued, it also needs to be current in terms of individuals exiting the system. This is important in terms of emigration, although not necessarily critically important. Specifically, though, it is vitally important that a national human resources database is updated to account for deaths, as these constitute permanent reductions of the available skills.

These are the main criteria against which the NLRD will be evaluated. However, there are various other requirements which, although less critical to the functioning of a national human resources database, are nevertheless desirable for consideration.

First, a national human resources database, for many applications, needs some level of employment information. Specifically, the availability of skills and skills shortages cannot be accurately estimated without taking into account current employment levels. To put this in context, the country may have a large number of individuals with a certain qualification, but without information on how many of these are currently employed, it is not possible to determine whether or not there is a surplus or a deficit for that given qualification.

For any policymaking at a sub-national level, such a database would require a minimum amount of geographic or locational data. Without locational information, it is impossible to say anything with a reasonable degree of confidence on regional skills

---

<sup>18</sup> It is important to note, however, that at least for foreign citizens *legally* residing in South Africa, the fact that employers generally seek to have independent local verification of foreign qualifications means that such information is available at least for those seeking or holding formal sector jobs.

---

supplies or shortages. The data requirement here is not overly ambitious or burdensome: location in terms of district council would be excellent, although provincial level information is also not entirely inadequate.

Finally, any accurate picture of current skills shortages must be informed by anticipated changes in both demand and supply of the relevant skills. In other words, the policy options that are appropriate in the face of temporary skills shortages, where anticipated changes in supply and demand of the specific skills will work to eliminate the shortage relatively quickly, are different from those that are appropriate where such changes are not expected. From the perspective of the national human resources database, therefore, anticipated flows of skills should arguably be accounted for. This would entail the inclusion of individuals currently within the education and training system in such a database, providing necessary information and promoting more efficient and effective policymaking.

Naturally, there may be other requirements for a national human resources database and other criteria against which candidate master databases could be evaluated. However, the criteria listed above have been chosen as being most appropriate for the purposes of this report.

### **3. The Potential Role of the NLRD**

#### **3.1 The National Learners' Records Database**

##### *3.1.1 The Database*

The National Learners' Records Database is the "electronic management information system that facilitates the management of the National Qualifications Framework (NQF) and enables the South African Qualifications Authority to report accurately on most aspects of the education and training system of South Africa" (South African Qualifications Authority 2008a). The NLRD therefore contains a vast amount of information, namely:

- SAQA and its sub-structures, including Standards Generating Bodies (SGBs) and their areas of competence and member details;
- Unit standards and qualifications registered on the NQF including exit level outcomes and assessment criteria;

## **Towards a National Human Resources Database: The National Learners' Records Database Assessed**

- 
- Accredited Education and Training Quality Assurance bodies (ETQAs) and the unit standards and qualifications for which accreditation has been granted;
  - Accredited providers and registered assessors;
  - SAQA-appointed Moderating Bodies (to be implemented as and when required);
  - The records of learners who achieve the outcomes of unit standards or qualifications registered on the NQF and their achievements (South African Qualifications Authority 2008a).

Thus, the data that would be required by the national human resources database is a subset of the data contained in the NLRD. The database is constructed as a relational database, basically consisting of a variety of tables that are related to one or more other tables in some defined way. Thus, for example, the learner achievement table is related to a qualification table (through a qualification ID) as well as to a person table (through a person and learner ID). Structured in this way, the NLRD is therefore very simply adjusted and adapted as changing requirements and realities may dictate.

Apart from the attainment of qualifications by individuals, the NLRD details accreditation and registration. SAQA accredits ETQAs, who are then able to accredit training providers. ETQAs, therefore, exhibit a level of 'ownership' of training providers: training providers may often be accredited by a given ETQA and only provide training within that ETQA's field of responsibility and it is then that ETQA's responsibility to obtain data from that provider and submit it to SAQA. In many other instances, however, a training provider may be accredited by more than one ETQA since it is active across various ETQAs' fields, with the training providers then having a 'home' ETQA depending on the range, field and scope of their training programmes. Irrespective of the number of ETQAs that accredit a training provider, the providers report to the individual ETQAs that 'own' the qualifications for which the provider is providing training. In other words, in terms of data and reporting, the links between training providers and the ETQAs are driven by qualifications, with each qualification being the responsibility of only one ETQA. The NLRD currently contains the details of approximately 30 000 providers and 2.5 million learners. According to the NLRD's current database administrator, Kirstin Barth of Praxis Computing, the

---

NLRD is “a very smart” database, and there has been international interest in emulating the system for their local purposes.

Two main sources of information have been utilised in this research: discussions with Yvonne Shapiro of the NLRD and Kirstin Barth, the database administrator from Praxis Computing; and investigations of the live database. The discussions have been extremely useful in framing our conceptualisation of the NLRD, its functioning and the interactions between it and the databases of the various ETQAs that submit (or do not submit) data to the NLRD. The investigations of the database, on the other hand, have, apart from anything else, helped to confirm the implications of the various issues raised in the discussions on the actual data.

Our primary method of interaction with the database has occurred via Microsoft Access: linking tables from the NLRD into Access, running various queries to obtain relevant data and then exporting the data generated by these queries into STATA format for further analysis. The data used in this report was generated at the NLRD in Pretoria on 26 May 2008. Given the dynamic nature of the database, this means that subsequent queries of the database may yield differing statistics to those presented below.

Our methodology in extracting the data was as follows. From the learner achievement table, we extracted the fields containing the learner identification number and the date of completion, excluding those records where the unit standard code was not blank and those records where the completion code was not equal to two. Non-blank unit standard codes indicate that the record refers to a unit standard rather than an actual qualification. Since we are interested only in qualifications, we therefore exclude records with non-blank unit standard codes. A completion code of two denotes that the qualification has been completed by the learner concerned. We did not extract any further details of the qualifications themselves as analysis of these does not really fall within the scope of the report. From the person table, we extracted the fields containing the person identification number, the learner identification number (merged in from the person designation table), gender, race, and date of birth. Data from the learner achievement table and the person table were then merged to create a dataset. Since not all persons in the person table are learners, the STATA merge variable yielded only ones and threes, indicating that all learners in the learner file were matched to persons in the person file, while there were some persons in the person file that were not matched to learners in the learner file – an expected result.

---

### *3.1.2 The Data Updating Process and the Edu.Dex Tool*

ETQAs are required by law to submit the required data on training and education within their respective fields of responsibility. The minimum requirement for uploads is twice a year – once in July/August and once in January/February – although ETQAs are free to upload at any time more often. The submission of data to the NLRD is the means by which the NLRD maintains its currency, but while it is clearly an essential process, it does pose extremely serious risks to the integrity of the NLRD if the process is not managed properly. The data submission from ETQAs contains information on providers, learners, qualifications and unit standards

Broadly, the process by which data is uploaded from ETQAs to the NLRD and then added to the existing database is as follows. ETQAs collect and store data obtained from the training providers for all the qualifications for which the relevant ETQA is responsible. The data is subjected to a set of checks before it is uploaded to SAQA, who then run further checks. Once SAQA is sure that the uploaded data is of sufficient quality, it is uploaded into the production environment.<sup>19</sup> What is therefore clear is that the quality of the data in the NLRD is only as good as the quality of the data it receives from the ETQAs and the thoroughness of the checks performed on the data prior to the incorporation of the data into the production version of the database.

All uploads are date-stamped and before it is uploaded into the NLRD, checks are run on three levels, namely file content, business rules and layout.

Developed by Praxis Computing ([www.praxis.co.za](http://www.praxis.co.za)), Edu.Dex is the tool introduced by SAQA and utilised by the various ETQAs to ensure that the data submitted to SAQA complies with the NLRD's requirements. Edu.Dex is a "custom designed application that ensures that data transmitted within the Education and Training sector of South

---

<sup>19</sup> It is perhaps useful at this point to distinguish between the *production* environment and the *development* environment. The production environment is the live database, the current best version of the database from which any outside queries are extracted or analyses performed. However, in many instances it is useful to have a test version of the database into which data can be uploaded or in which technical tests can be run without fear of compromising the integrity of the live database. This test database is referred to as the development environment. Generally, the difference between the production and development environments will be that the latter is merely a clone of the former at some point in the past.



---

Africa conforms to the education and training exchange (Edu.Dex standard) standard. The primary objective for the development of the Edu.Dex standard is to ensure that any learner registration related data that is transmitted from one information system to another system in batch format complies to the ... NLRD ... Load specifications" (Praxis Computing 2008). As of April 2008, Praxis Computing are involved with HEMIS and FETMIS and have six SETA clients.

The Edu.Dex tool is issued to all ETQAs and, before they are allowed to upload their data to SAQA, the data must be approved by Edu.Dex as complying with the set minimum standards. Once Edu.Dex passes the data as having fulfilled the requirements, ETQAs are then allowed to upload the data to SAQA. Edu.Dex requires only that the proposed data submission be in a standard format, thus allowing ETQAs to utilise any software system they choose. In this characteristic, it is similar to HEMIS. However, Edu.Dex differs from HEMIS in that HEMIS allows dirty data into the system where Edu.Dex would not. Edu.Dex has zero tolerance on data quality and discriminates between fatal and non-fatal errors. Fatal errors concern non-compliance with load specifications, for example the file should be in fixed-width format. Another type of fatal error occurs where the data of birth provided does not correspond to the first six digits of the national ID number, or where gender takes any value other than male or female.<sup>20</sup> Once Edu.Dex has found fatal errors in an ETQA's proposed submission, these errors would need to be rectified and the revised submission would require retesting and approval by Edu.Dex. Non-fatal errors tend to revolve around quality issues: the discovery of a non-fatal error does not prevent the data from being submitted to SAQA, but does provide ETQAs with feedback as to data quality issues, such as where data is entered as 'unknown' where it should have a proper value. Simply put, non-fatal errors relate to the 'nice-to-haves' and Edu.Dex will output reports on all these errors, providing ETQAs with feedback on where their data could be improved beyond the minimum standards. Data submissions are also tested to ensure that each learner has an achievement and that each achievement has a learner.

---

<sup>20</sup> The NLRD load specifications (version 2.0) are detailed in a 41 page document (South African Qualifications Authority 2008b). The document also details best practice for validating and extracting data. The procedures are too detailed to describe here, but in terms of ID numbers, for example, see Appendix A.

## **Towards a National Human Resources Database: The National Learners' Records Database Assessed**

---

Once a proposed data submission is passed by Edu.Dex without any fatal errors, ETQAs are free to upload the data to SAQA. The proposed submission contains all records in its system in eight files, namely *Providers*, *Provider Accreditations*, *Person Information*, *Person Designation*, *NQF Designation Registration*, *Learner Achievements and Enrolments – Learnerships*, *Learner Achievements and Enrolments – Qualifications*, *Learner Achievements and Enrolments – Unit Standards*. SAQA requires all records to be included in the data submissions as this “is the most robust way to operate, as it ensures that all updates are transmitted to the NLRD” (South African Qualifications Authority 2008b: 40). This does not mean that the NLRD is populated from scratch with each upload of data, but instead the database’s load procedures will check whether or not a record has changed since the previous submission, with only new or changed records being incorporated into the NLRD. The inclusion of all records held by the ETQAs means that each submission represents a complete snapshot of their respective databases at a different point in time. Encrypted data submissions are possible via two methods. Firstly, an external staging area exists where each data supplier has a login and password and data is transmitted to SAQA via a secure FTP-like service. Alternatively, data suppliers may physically send CDs or USB drives containing the data submissions to SAQA. SAQA’s first choice, though, is the use of the external staging area.

Version files of data submissions are kept with a record of which version of Edu.Dex ran on that particular submission. Submissions are rejected if the version of Edu.Dex is not sufficiently current, i.e. the version of Edu.Dex used to test the submission may not be more than two days old.

The Edu.Dex tool appears to be making a significant and positive impact on the process of uploading data from data providers to the NLRD, thereby benefiting data quality and integrity within the NLRD. There are, though, two limitations that need to be kept in mind. Firstly, Edu.Dex is limited in its ability to find duplicate individuals in the database (i.e. one individual is recorded more than once in the database’s person table). Individuals are identified as such according to numerous identifying numbers stored in the NLRD (such as South African ID numbers, student numbers, and HSRC database identifiers). Not all individuals will have information in all of these fields, but generally they will have national ID numbers, making linking data for that individual straightforward. However, it is possible that an individual may in one record have only a national ID number as a unique identifier, while in another record, that same individual has only, say, a student number. Even though this is the same individual,

## **Towards a National Human Resources Database: The National Learners' Records Database Assessed**

---

Edu.Dex is unable to make the link between the two records because there is insufficient data to enable an automated linking process. This is complicated, for example, by the legacy of HEMIS data that does not include names. A second limitation is the fact that, because Edu.Dex 'talks' to ETQAs, cross-ETQA training providers can be problematic and it may be difficult to identify duplicates within the database. As such, the NLRD team will always be required to check on duplicates, whether they are persons or providers.

In an effort to reduce the existence of missing or unspecified values in the data, the NLRD has developed a tool called "Get Best Data", which will attempt to ensure that the NLRD maintains a copy of the most recent and best data for a given individual. Essentially, the tool checks the data in an incoming record field by field. If it finds a missing value in the incoming field, it checks the existing information in the relevant field in the NLRD and will not allow an overwrite of the existing information if a value exists. For example, if in a new upload an individual's gender is listed as unknown, this tool will check whether the NLRD contains a value for gender for the individual and, if it does, it will maintain that gender value and not overwrite it with the missing value from the new record. Further, if the tool finds conflicting information in a given field in the existing and incoming records, it checks the date stamp on each record and, where the incoming data has a more recent date stamp, it will overwrite the existing information with the incoming information. For example, where an incoming record has a different address for an individual from the address currently stored in the NLRD, the tool will check which is the most recent data based on the date stamp and, where the incoming record is more recent than the existing data, it will overwrite the existing address information with the incoming address information. However, where the date stamp of the existing data is more recent than that of the incoming record, the existing data will not be overwritten.

Once the data has been passed through Edu.Dex and uploaded to SAQA, and SAQA is happy with the data, it is first uploaded into the development environment and then later into the production environment. Even though ETQAs upload all their data to SAQA, only records that have changed will be overwritten with the new data and new records will be added to the dataset. Where an existing record is identical in both the NLRD and the data submission, no action is taken. Since learners are the service focus of SAQA, only serious problems with a data submission would prevent it being uploaded into the production environment. Normally, if less serious problems exist, the data will still be uploaded into the production environment, where the problems

---

will be resolved. This is done to ensure that learners are not turned away from the NLRD empty-handed because of a relatively minor data problem in the batch of records containing their qualification, when the verification of their qualification may determine whether or not they are offered a job.

### **3.2 The Individual as the Unit of Analysis**

The NLRD is, by its very nature, focused on the individual as the unit of analysis, at least as far as is relevant to our requirements. The person table has one person per record. The learner table has one record per qualification, but this is easily rearranged. A national human resources database based on the NLRD would arguably best be structured along very similar lines to the NLRD, making analysis of qualifications or of individuals straightforward. A relational database is also very flexible making the incorporation of additional data from outside sources relatively simple. Thus, the current focus of the NLRD as far as learners are concerned and the required focus of a national human resources database on the individual as the unit of analysis are perfectly compatible and the NLRD fulfils this criterion.

### **3.3 Comprehensiveness**

#### *3.3.1 Evidence from the NLRD*

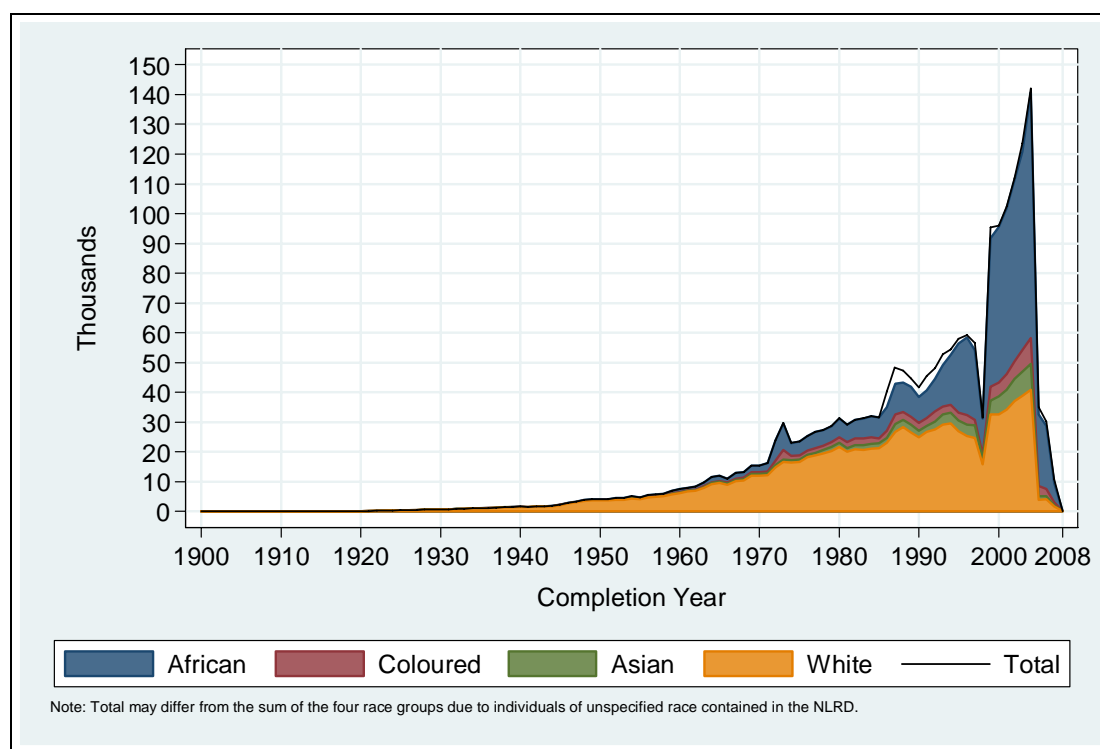
By comprehensiveness, we mean that the NLRD should contain the qualifications of all South Africa's citizens and/or residents in order for the national human resources database to fulfil its objectives. Our research leads us to believe that the NLRD is, for various reasons, currently not comprehensive enough in order to satisfy this requirement.

The data in a database that contains the qualifications of all South Africans will have a very specific profile. Firstly, in terms of qualifications attained over time, *ceteris paribus* and specifically with no change in the population size or structure or in educational and training behaviour, that a graph of the number of qualifications over time should be a horizontal line, where time is on the horizontal axis. Assuming changes in terms of population size and structure as has been experienced in South Africa, one would expect an upward-sloping line, while allowing for changes in educational access would probably result in a more steeply upward-sloping line.

However, the pattern of qualifications obtained over time as derived from the NLRD does not comply with the expected pattern (see Figure 1). Instead, the overall pattern is one of virtually no change from 1900 to the mid-1940s, followed by a relatively

## Towards a National Human Resources Database: The National Learners' Records Database Assessed

rapid rate of growth (albeit from a very small base of 2 368 qualifications in 1945) of approximately 6.5 percent per annum until 1996. By 1998, the number of qualifications obtained in that year fell by almost one-half to 31 515, only to leap to around 95 000 in 1999, 112 000 in 2002 and 142 000 in 2004. The implication is that, considering only the pre-2005 data, more than one in three qualifications contained within the NLRD have completion dates in the six years from 1999 onwards, signalling an important shortcoming in terms of temporal comprehensiveness from the perspective of the national human resources database.



Source: Own calculations, NLRD (2008).

**Figure 1: Number of Qualifications over Time, by Race – 1900-2008**

The number of qualifications obtained falls dramatically between 2004 and 2005, a drop of 107 000 or 75 percent of the 2004 total. Part of the reason for this relates to delays in receiving the information on public higher education from the CHE. In fact, communication with Yvonne Shapiro on 30 June 2008, confirmed that the NLRD had only just received the 2005-2006 HEMIS data. The HEMIS data is discussed in section 0 below, but the issue is flagged here. This highlights the fact that the figure is really based on two sets of data, namely the HEMIS data and the data from the ETQAs. While the drop related to the non-incorporation of post-2004 HEMIS data is clear, what is not clear is the underlying trend for the data being submitted by the

---

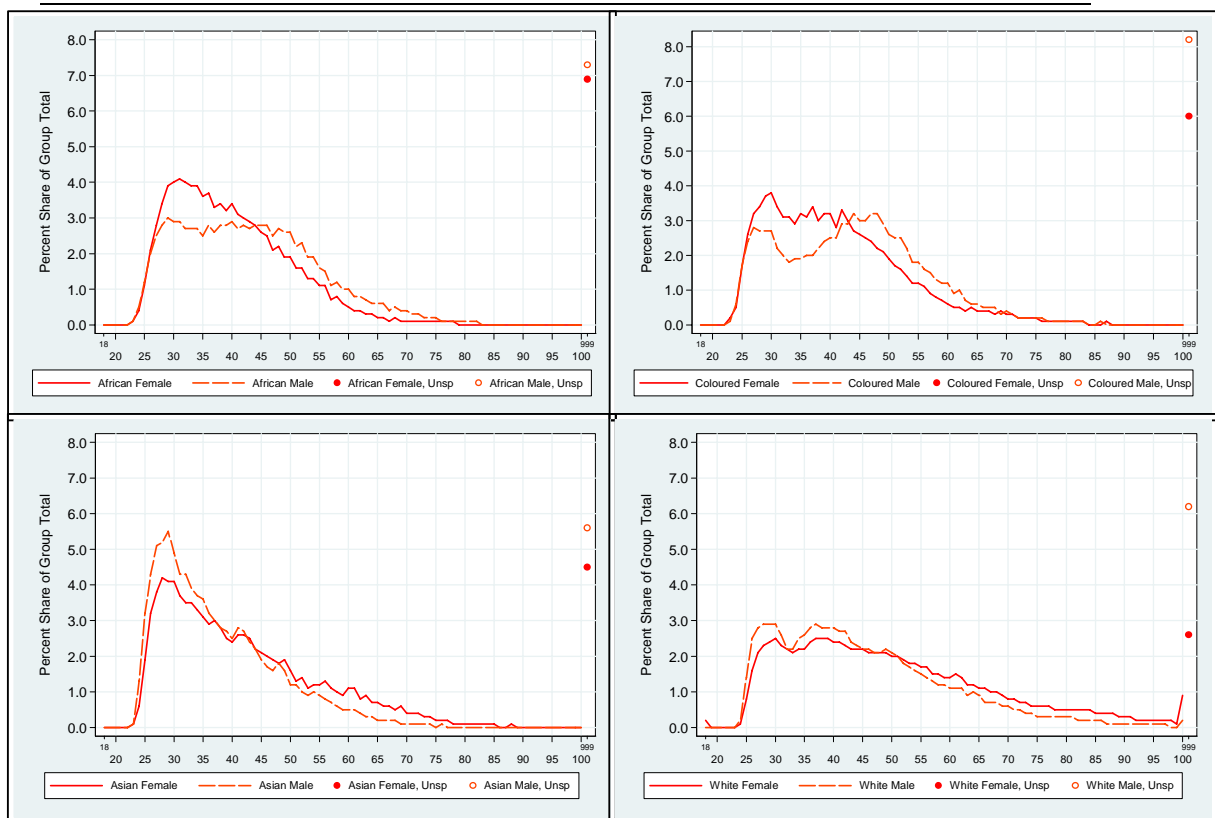
ETQAs, with most ETQAs performing well in terms of complying with the minimum data requirements.

Information on coverage gaps over time can also be inferred from the age distributions of qualification achievers contained in the NLRD.<sup>21</sup> Figure 2 presents the age distribution of qualification achievers by race and gender, with age on the horizontal axis and the percent share on the vertical axis. The surge in qualifications recorded in the NLRD from 1999 onwards is clearly visible in the graphs for all race-gender groups, apart from African males. There are several key points that emerge from these figures. Firstly, the heights of the graphs for Africans, Coloureds and Asians are consistently higher than those of Whites for younger ages, and consistently lower than those of Whites for very high ages. As a result, the slope of the graphs for White males and females is considerably flatter than for the other race groups. This pattern is consistent with, firstly, the more rapid expansion of participation by Africans, Coloureds and Asians in education and training at institutions submitting data to the NLRD, and secondly, more substantial improvements in the capturing of qualifications from institutions with relatively high proportions of African, Coloured and Asian qualification achievers. Whites dominate the NLRD in terms of qualifications, particularly prior to the 1990s, accounting for more than 50 percent of qualifications recorded in the NLRD in each year until 1994, and falling to around one-third in 1998. This is largely due to the bias towards Whites of the historical data contained in the NLRD and to higher participation rates for Whites in education.

---

<sup>21</sup> This report will refer to individuals who have obtained a qualification as 'qualification achievers'. This is opposed to the term 'graduates', which traditionally has a much narrower interpretation, being individuals who have obtained a qualification from a tertiary education institution.

**Towards a National Human Resources Database: The National Learners' Records  
Database Assessed**



Source: **Own calculations, NLRD (2008).**

Notes: **Age 18 includes all individuals aged younger than 18 years. ge 100 refers to all individuals aged 100 years or more, while 999 refers to individuals of unspecified age.**

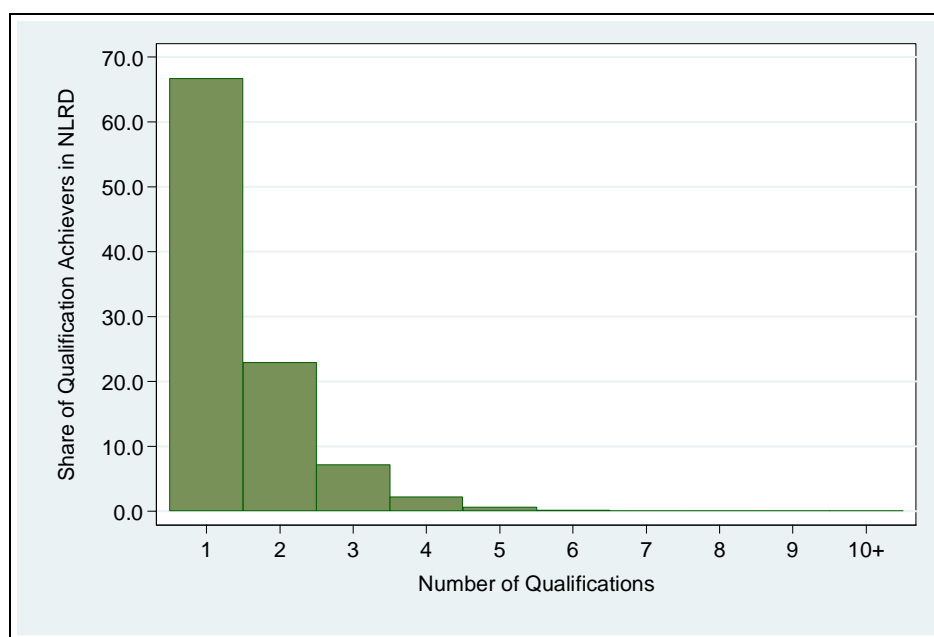
**Figure 2: Current Age Distribution of Pre-2005 Qualification Achievers by Race & Gender**

Secondly, the rise in the education and training participation rates for females is evident for Africans and Coloureds, where the peaks of the female graphs are significantly higher than for their male counterparts. Interestingly, this is not evident for Asians or Whites. Questions regarding complete coverage emerge from the graph for Coloured males, where there is a significant dip in the proportion of individuals aged around 32 to 37 years. This pattern is echoed in the graphs for White males and females. Apart from these types of anomalies, however, without historical information on participation rates in education across races, it is not possible to disentangle the effects of improved capturing of qualification data, increased participation rates and changes in the distribution of qualification achievers across institutions whose data is or is not contained in the NLRD.

The implication of the patterns in the figures above are that, amongst males, African, Coloured and Asian qualification achievers contained in the NLRD are younger

relative to their White counterparts. Thus, while 39 percent of Asian males, 37 percent of African males and 35 percent of Coloured males are no older than 35 years of age, this is true of only 23 percent of White males. A similar pattern is evident using 45 years as the age cut-off. This is, however, not true for females. The only exception is Asian females, of whom 49 percent are no older than 35 years of age, compared to between 25 and 29 percent for the other races.

In terms of internal consistency, there are various possible checks that can be performed. Two of these checks are presented below. The first check investigates how many qualifications in the achievement table in the NLRD attach to specific individuals in the learner table. Figure 3 presents the distribution of individuals in the NLRD by the number of qualifications attained (and recorded in the NLRD). The figure confirms an expected pattern, with the majority of learners (67 percent) having attained one qualification.



*Source: Own calculations, NLRD (2008).*

*Notes: This figure refers only to qualifications obtained prior to 2005.*

**Figure 3: Distribution of Individuals by Number of Qualifications in the NLRD**

Between one-fifth and one-quarter (23 percent) of individuals in the NLRD have two recorded qualifications, while a further seven percent have three qualifications. This downward trend continues such that a mere 0.2 percent have six qualifications. Overall, there are on average 1.48 qualifications per learner stored in the NLRD.



---

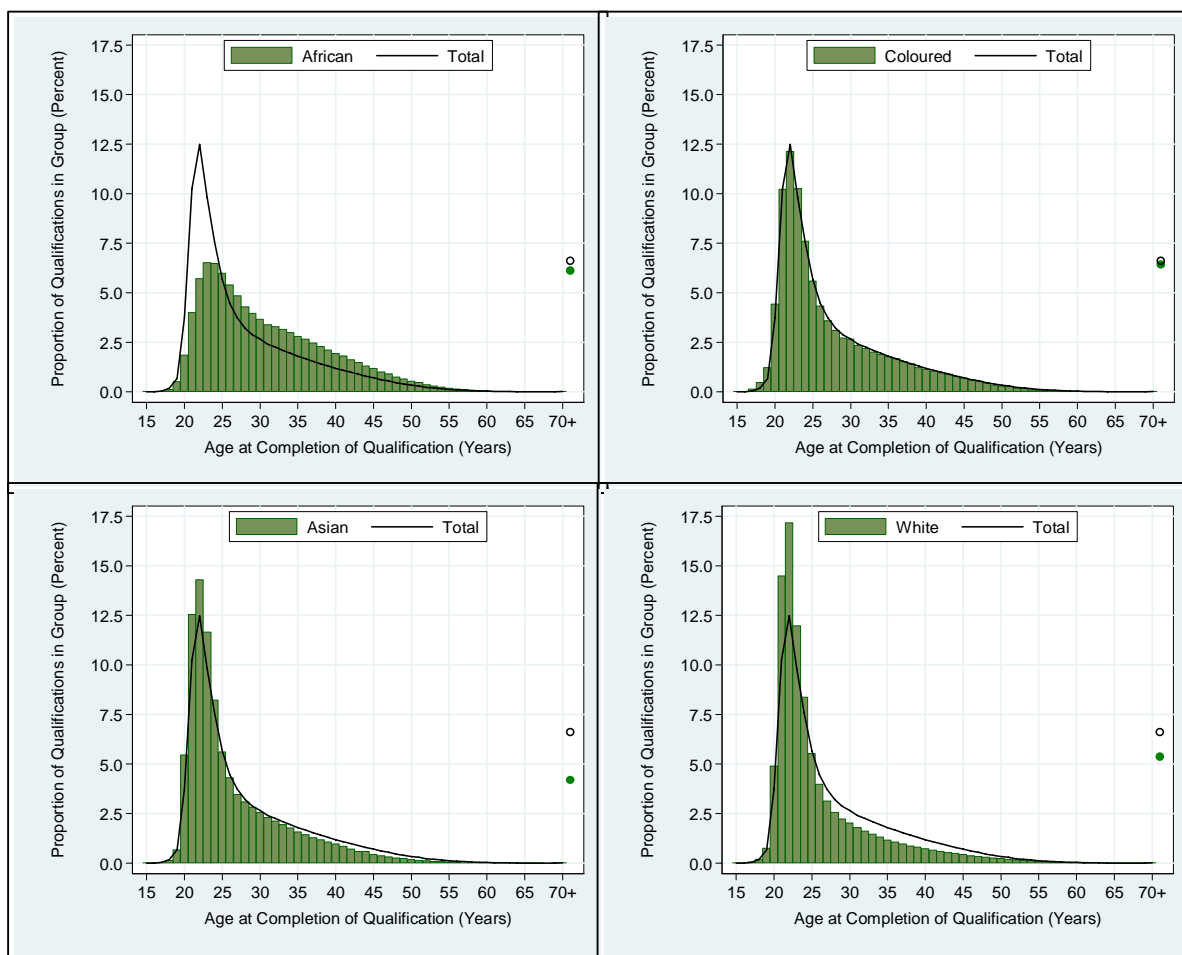
A second check is that of the age of individuals at completion of their various qualifications. Individuals' ages at the time of the completion of the qualification are calculated as the difference between the completion date in the learner achievement table and the date of birth in the person table. In , the distribution of qualifications attained prior to 2005 is presented across individuals' age at completion. Each race group is presented individually with the overall distribution (including all race groups and those with unspecified race) overlaid. The overall pattern shows a rapid rise until the age of 22 years, when 12.5 percent of qualifications were achieved. The proportion of qualifications falls again quite rapidly until the late twenties, whereafter it declines more gradually. In all, one-half (50.4 percent) of all qualifications were obtained by individuals under the age of 26 years, with three-quarters (76 percent) obtained by age 34.

The patterns by race are broadly similar, although there are important differences. Asians and Whites are similar in that a greater proportion of qualifications are obtained at ages 19 through 22 than the overall pattern, while lower proportions are observed from 25 years upwards. For Asians the distribution peaks at 14 percent, while for Whites the peak is at 17 percent, both at age 22. Thus, for Asians and Whites, two-thirds of qualifications are obtained by ages 27 and 26 respectively, substantially more than is the case overall. The pattern for Coloureds is very similar to the overall pattern, with the proportions of qualifications attained at each age never differing by more than 0.7 percentage points from the overall proportions. For Africans, the pattern is different. The distribution peaks at a higher age than the overall distribution, with 6.5 percent of qualifications achieved by Africans being achieved at the age of 23 years, and a further 6.5 percent being achieved at the age of 24 years. Thus, the distribution peaks at an older age, while at a lower proportion than any other race group. Consequently, the proportion of qualifications obtained by Africans above the age of 30 years, at 40.5 percent, is substantially higher than the 26 percent overall. The effect of these differences is that, for Africans, less than one-third (31 percent) of qualifications are obtained by age 25, and two-thirds (66 percent) by age 34. Although Africans and Coloureds in the NLRD are more likely than other groups to have unspecified ages at completion, the difference is relatively small and insufficient to significantly affect the distribution.

While these patterns are quite different, a number of reasons can be suggested. Firstly, given their historically lower rates of progression through the schooling system, Africans are likely to be older than individuals of other race groups when

## Towards a National Human Resources Database: The National Learners' Records Database Assessed

enrolling for, and consequently when completing, a given qualification. This may be compounded by higher rates of repetition in the post-school education system. Secondly, a large number of learnerships, for example, are targeted at the mid-



lower end of the skills spectrum where, due to the legacy of inferior education provision and opportunities for Africans, a large proportion of Africans are located. This, combined with very high rates of youth unemployment amongst Africans may serve to raise the proportion of qualifications obtained aged in the thirties and forties.

**Source:** Own calculations, NLRD (2008).

**Notes:** This figure refers only to qualifications obtained prior to 2005.

**Figure 4: Distribution of Qualifications Across Age at Completion, by Race**

In essence, the NLRD is constrained in its ability to collect historical data and it is only really from 2000 onwards that, it appears, a reasonably comprehensive picture of qualifications achieved is captured in the database. The constraints that the NLRD faces have little, if anything, to do with the NLRD's capacity. Instead, the constraints

---

appear to have more to do with the nature of the historical data. It must be noted that in probably most instances historical data is not available in electronic format, and where historical data is available electronically, it is often in an archaic format, somewhat complicating its incorporation into the NLRD.

### *3.3.2 Data Sources affecting Comprehensiveness*

#### *The HSRC Register of Graduates and HEMIS*

An important historical data source that has been incorporated into the NLRD is the Human Sciences Research Council (HSRC) data on university graduates, the so-called Register of Graduates. This data covers the period from 1914 to 1998 and is the NLRD's only source of higher education data prior to 1998. However, this data covers only *university* graduates. Initially, the HSRC collected this data by voluntary mail return: graduates were contacted by post and requested to update their information and return the completed form to the HSRC. Interestingly, the HSRC had a relatively high response rate despite collecting data in this way. From 1986 onwards, the method of data collection was changed, with the HSRC receiving electronic downloads from the universities. In this process, the HSRC was able to draw on its extensive IT resources and align the widely varying data submissions to the HSRC's required format. The questionnaires were still administered by the HSRC and individuals that actually responded to the questionnaire were labelled 'active participants'. In effect therefore, the HSRC data incorporated within the NLRD is patchy until 1985, with much better coverage from 1986 onwards.

In 1999, higher education data collection moved to HEMIS (Higher Education Management Information System). HEMIS data, though, informs the funding processes around public higher education. Specifically, data submitted by universities to HEMIS is used to allocate per student subsidies from the higher education budget. As a result, the data submissions are vulnerable to the incentives that the financial motive may introduce into the process. For example, there are financial incentives for public higher education institutions to inflate their student figures. Another issue is that, where the race of a student is unknown, some universities may be resetting students with race 'unknown' to 'African' in their HEMIS submissions because of the resulting financial benefit of higher African enrolments. This creates problems in the NLRD as individuals are allocated conflicting racial designations.

## **Towards a National Human Resources Database: The National Learners' Records Database Assessed**

---

At the heart of it though, because HEMIS is a funding database, it is the Department of Education's responsibility and duty to ensure that the data is correct. The extensive auditing of HEMIS data submissions is time-consuming and often results in institutions having to resubmit data. It is only once they are happy with the data that HEMIS is willing to submit it to the NLRD. Apart from the time delays associated with this process, it is arguably advisable that the NLRD should not be drawing data from sources where the incentives are not aligned to the provision of accurate data.

SAQA has recently been commissioned by the CHE to establish HEQCIS, the Higher Education Quality Committee Information System, to collect information on private higher education to complement their coverage of the public higher education system. This database is expected to be expanded to collect an increasing amount of data over time. What this means is that private higher education data is, as yet, not included within the NLRD. Should this process prove successful, and sufficient buy-in is evoked from stakeholders, HEQCIS could easily be used to collect information on the public higher education sector, which is a considerably smaller sector in terms of number of institutions (there are about 100 private higher education providers, compared to 23 universities), thereby benefiting data quality.

There do appear to be some gaps in the coverage of university graduates from public universities at the time of the changeover between the HSRC register and the HEMIS data, although this appears to be located within specific universities. Some random checking found, for example, a gap in coverage of graduates from the University of Stellenbosch completing their qualifications between 1997 and 2000. This type of gap, though, should be relatively straightforward to correct, at least to the extent that institutions maintain their own historical data electronically.

At the level of the individual, this latter problem certainly requires going back to institutions to get them to resubmit historical data. However, at a more aggregated level, it is possible to get a decent picture of the flows of graduates. This has been done, for example, in the publication entitled "Trends in Public Higher Education in South Africa 1995 to 2004" (South African Qualifications Authority 2007), which used various techniques and interpolation to fill in data gaps at a more aggregated level.

### *Umalusi*

Umalusi is an ETQA so it should, theoretically, be uploading to the NLRD. Umalusi is the quality assurer for the general and further education and training system. It holds information on learners' performances in the matric exams and issues matriculation

## **Towards a National Human Resources Database: The National Learners' Records Database Assessed**

---

certificates. However, data from Umalusi is not uploaded into the NLRD. According to Yvonne Shapiro, two unsuccessful attempts have been made by Umalusi to upload the data. Thus, although it is noted by some that the NLRD includes around 5.5 million matric records, these records are held separately from the NLRD.

One complicating factor with Umalusi data is that it does not have one record per learner, but rather one record per certificate. Therefore, if you have matriculated and a certificate is issued to you, you have one record in the dataset. If you request the certificate be reissued, a new record is created for you and there will therefore be two records for you in the dataset. SAQA has had some engagement with the Department of Education (DoE) to request that it be allowed to use the DoE's exams section as the source of matric qualifications data. From the NLRD's perspective, the DoE would be a better source of this data, particularly given that Umalusi receives its data from the DoE and the Umalusi data is therefore a subset of the DoE data. However, according to Yvonne Shapiro, the State Information Technology Agency (SITA) previously quoted R100 000 to provide about 15 years' of data, a download that apparently has already been done for the DoE, possibly as a standard backup, and would still require some work from the NLRD's side in terms of formatting for the NLRD. The NLRD has subsequently, after much deliberation, decided to proceed with their request, but with certain amendments to facilitate uploading into the NLRD. A letter was apparently sent to SITA in May 2008 by Joe Samuels, the Deputy Executive Officer of SAQA, and SAQA awaits their reply.

### *Other Data*

Two other data sources namely FETMIS, covering the FET sector, and a UNISA-held database on adult basic education and training (ABET) have also not been uploaded into the NLRD. SAQA is of the opinion that the data in these databases is currently not in a suitable condition for incorporation into the NLRD. In terms of FETMIS, there appear to be issues with the data at the individual-level, although the problems are not as pronounced at an aggregate level.

Another important historical data source is the legacy artisan data held by the Department of Labour (DoL), a dataset which the NLRD is eager to incorporate. This data is stored in a very outdated format and would require some work, probably a day or two, to convert into a format compatible with the NLRD. SAQA has offered to send someone in to the DoL to convert the data, but they have not yet been granted access to do so.

---

### **3.3.3 Conclusion on the Level of Comprehensiveness**

From the perspective of the national human resources database, the NLRD, in our view, does not fulfil the comprehensiveness criterion for two reasons. Firstly, the historical coverage of the NLRD is deficient. In other words, coverage prior to the late 1990s is far from complete. This is immediately evident from Figure 1, which shows a massive surge in the number of completed qualifications around the turn of the century. This is partly the result of the NLRD's inability, through no apparent fault of their own, to access certain legacy data, such as the artisan data and data on matric results. It must be noted that there is only so much that can be done to correct this gap. It is probable that large sections of historical data may not be retrievable without significant investments of time and money, if at all, given that many institutions would not have this data available electronically.

Secondly, the current coverage of the NLRD is deficient in that there are still segments of the education and training system that are not submitting data. Specifically, major gaps are the lack of data regarding matric qualifications and the FET sector. As noted, SAQA does recognise these gaps and is, over time, working to close them.

Given that the objective of the national human resource database would be to provide information on the stock of skills available to the South African economy, it is evident that the gaps in the coverage of the NLRD, both over time and across institutions, are a serious limitation.

### **3.4 Currency of the Data**

The need for current, up-to-date information for effective policymaking cannot be disputed. The NLRD, as noted, receives updated information from the ETQAs at least twice a year, although ETQAs are free to make more frequent submissions. Biannual updating of the database is arguably sufficient for a national human resources database, although should more frequent updates be deemed necessary the implications for the NLRD must be noted.

As noted, the data post-2004 is incomplete, as confirmed by the steep drop-off in the number of qualifications obtained per year (see Figure 1 above). This is an area of concern, given that this is three years' worth of data (excluding 2008) that has not yet been uploaded into the NLRD. While HEMIS is identified as a major component of the missing data, the public higher education system is unlikely to account for the full drop of over 470 000 qualifications between 2004 and 2005.

---

A final issue in terms of the currency of the data is that, while data uploads may be performed frequently, it is the currency of the data providers' data that will determine the currency of the NLRD and, by extension, the national human resources database. The question is, therefore, how current is the data being uploaded by data providers, both at the level of upload to the NLRD as well as the level of upload to the ETQAs. This question, though, can only really be answered by a more thorough review of the data uploads over time, in order to assess the rate at which data is being updated relative to the dates contained in the data submissions (e.g. dates of completion of qualifications).

### **3.5 Ability to Represent the Stock of Skills**

Given the intended uses of the national human resources database, the ability of the database to reflect the actual stock of skills available to South Africa at any given time is fundamental. The various criteria discussed thus far have an important role in determining whether or not the database can accurately reflect the current stock of skills. However, key to this ability is whether or not the database makes provision for reductions in the available stock of skills arising from three main sources, namely ageing of individuals in the labour force or working age population, 'permanent' emigration of South Africans to other countries (or the return of foreign residents in South Africa to their home countries), and deaths.

It is important to recognise the divergence, here, between the objectives of the NLRD and those of a national human resources database. The NLRD does, and should, maintain a full historical record of the data on learning and qualifications. In other words, as far as the data allows, the NLRD stretches back into time. It acts as a repository of data, irrespective of whether or not a person is part of the labour force, or whether or not they are still alive. In contrast, a national human resources database should not contain data of individuals who cannot be considered part of the available stock of skills. Simply put, the NLRD would theoretically constitute the data universe, while the national human resources database would constitute a subset of that data.

Given that the NLRD has been set up with another set of objectives, it currently does not account for any of these three factors that would have a diminishing impact on the overall stock of skills. The first source, namely the ageing of individuals in the labour force or working age population, is simple enough to implement. Clearly, at some point an individual is too old to realistically be considered part of the labour

---

force. Official definitions define individuals as out of the labour force at the age of 64 or 65 years, although this is arguably too low an age from the perspective of a human resources database in a skills scarce country. Determining the upper bound age for inclusion in the human resources database is more an administrative question than one which can be answered through analysis. Perhaps 70 years of age is a reasonable upper bound, but in some occupations this may be too low given that individuals may often still be active in their occupations at that age, while one must realistically consider the likelihood of attracting individuals aged 70 years or older back into the workforce, even if only in some kind of consultative or skills transfer-type capacity. Irrespective, once a decision has been reached, the national human resources database can be easily set up to extract only individuals within a specific age category from the NLRD, for example.

Emigration is a notoriously difficult phenomenon to measure in South Africa as in many other countries. This is largely the result of emigrating individuals not notifying authorities when they leave the country that they actually are emigrating. This has led many researchers and analysts to rely on data from receiving countries rather than any official South African statistics (see, for example, Kaplan *et al.* 1999; Myburgh 2004). Emigration serves to reduce the supply of skills in South Africa. Despite the various programmes that aim to attract South Africans living abroad back to the country, some of which have seen some success, it would be unwise to include the skills of permanent emigrants as part of the stock of skills available to the South African economy.<sup>22</sup> Currently, the NLRD does not include data on whether an individual has emigrated or not. As noted, this type of data gap is easily addressed in terms of the structure of the database, but the practicalities of filling this data field are complex. Theoretically, filling the field should be straightforward: information on emigration is held by the Department of Home Affairs, and the national ID number would constitute the link between the DoHA data and the NLRD. However, as noted, there are problems in terms of non-reporting of emigration, with the gap between official Statistics South Africa data and receiving country data continuing, if not

---

<sup>22</sup> Once again, it would be relatively straightforward to incorporate in a database a distinction between skills held by South Africans living in South Africa, and skills held by South Africans living abroad. In fact, this may be a desirable feature to include. However, this distinction and its usefulness would be highly dependent on the accuracy of the underlying data regarding temporary and permanent emigration.



---

widening.<sup>23</sup> Setting up an accurate database of the stock of skills available to South Africa is, at least partly, dependent on getting a decent handle on, at least, permanent emigration.

While accounting for both the ageing and emigration of individuals in the database is important, possibly even more important is accounting for their deaths. In terms of the NLRD's suitability as a master dataset for the national human resources database, its current lack of information on deaths is a critical flaw, but one which can be rectified with a link to data held by the Department of Home Affairs. The NLRD currently does have a field for date of death, but this is not populated. To the extent that registration of deaths is (near) universal and the data is well-maintained by the DoHA, a link via the national ID number would be essential for the database to fulfil its objectives.

### **3.6 Compliance with Other Criteria**

Sections 0 through 0 have detailed the main criteria that would arguably be essential to ensuring that the proposed national human resources database is useful and effective in achieving its objectives, specifically in its ability to represent the current stock of skills available to South Africa at a given point in time. This section presents two further data requirements that, although not absolutely essential, would significantly enhance a national human resources database.

#### **3.6.1 Employment Information**

Knowing how many individuals in the country possess a particular skill is only part of the supply-side answer to the greater question around skills and skills shortages. The other part of the puzzle is information on employment status. When considering the national human resources database, it is likely that one of its main uses will be to match individuals with specific skills to existing vacancies. This type of process of matching individuals to vacancies would be vastly facilitated if the database held information on the employment status of individuals.

Two main issues arise in terms of incorporating employment information into a national human resources database. Firstly, what would be the source of such data? The second issue is the fact that, given current systems and a reasonable

---

<sup>23</sup> According to Myburgh (2004: 124), "in 1997 SSA data only captured a third of those who emigrated, a tendency which seems to have increased over time: in 1987 SSA captured a more respectable 65 per cent".

---

assessment of possibilities in the near future, it is highly unlikely that employment data in a database of this type will be updated regularly or consistently enough to be considered current. Since there is no national employment register, no universal mechanism by which employers (including SMMEs and the self-employed) are required to report the details of every employee, there are probably two potential data sources that would start to reveal this type of information.

One option is the Unemployment Insurance Fund (UIF). Given that employers are required to contribute to the UIF for employees, one possible way for a national human resources database to determine individuals' employment status is by connecting to the UIF, using national ID numbers as the link. As far as it is comprehensive and exhaustive, UIF data could shed light on whether an individual is employed, or (recently) unemployed. Technically, this data field could be updated several times per year, depending on requirements and whatever practical issues may arise, making it potentially reasonably current. A second option is income tax data from the South African Revenue Service (SARS), although access to this data is likely not to be forthcoming. With SARS's record of individuals' incomes, it would be possible to identify whether an individual was employed during a given tax year. This data, though, would probably only be updated on an annual basis and would not indicate current employment status. Considering each of these data sources and their potential, UIF data is likely to be the best suited for this purpose. It must be noted, however, that insofar as these databases only capture formal sector activities, the database will be mute on the issue of informal sector employment.

In terms of whether an individual is currently retired, it may also be possible to cross-reference data on pensions, for example, from the Department of Social Development and/or the South African Social Security Agency (SASSA). In fact, in time, when the compulsory national retirement scheme is implemented, that database may be an excellent source of data in terms of employment status, but also for other information, such as location.

### *3.6.2 Locational Information*

Sub-national locational information, while clearly not an essential part of a database of this type, can contribute to a more refined, nuanced understanding of skills shortages. There is no perfect mechanism by which working-age individuals with a particular type of skills will automatically be distributed identically to the distribution of job opportunities for that type of skills and, consequently, regional skills shortages may exist where nationally no such shortage exists.

---

As in the case of employment information, it is not automatically clear from where locational information of this nature should be sourced. It is most likely that address information for individuals would be best suited for use in the derivation of locational information, probably at a provincial or district council level. Although address information is held by the Department of Home Affairs, it is unlikely to be updated regularly enough due, at least partially, to a reliance on self-reporting of address changes by South Africans without any clear legal obligation to do so or punishment where individuals have failed to comply. An alternate source is, again, the South African Revenue Service, who have relatively up-to-date address information for all taxpayers. As noted above, however, this data source may be very difficult to access, and its coverage is not complete. Whatever the data source, individuals' full address details would be unnecessary: the national human resources database requiring, at its most disaggregated level, locational information according to district council.

### *3.6.3 Information on the Flow of Skills into the Labour Force*

As noted, the inclusion of information on the anticipated flows of skills into the labour force is an important requirement for efficient and effective policymaking, particularly as far as it allows for clearer differentiation between temporary (or short-term) skills shortages and more structural, longer-term skills shortages. Given the NLRD's mandate, it is uniquely positioned to collect information on flows of individuals through the education and training system. This is particularly relevant given that the NLRD now also includes data on registration and completion of unit standards by learners.

It should be a relatively straightforward exercise to use data on attainment of unit standards to predict the inflow of individuals qualified in specific fields into the labour market. This data would also be extremely useful as a type of 'early warning system' in critical fields, allowing policymakers and other users of the database to monitor qualification flows and identify potential problems. This would facilitate more proactive policymaking regarding encouraging enrolments in these fields, as well as identifying bottlenecks or systemic problems.

## **4 Conclusion**

Addressing the issues surrounding skills shortages and skills development remains key to the acceleration of sustainable and equitable growth in South Africa. As in many other instances where the market is unable to ensure a matching of supply and demand, one of the problems in the labour market relates to informational

## **Towards a National Human Resources Database: The National Learners' Records Database Assessed**

---

asymmetries and a general lack of information. It is with the view towards a better understanding of skills issues and the facilitation of effective and efficient policymaking that the establishment of a national human resources database has been suggested.

Central to a database of any kind is information. Without reliable, comprehensive, accurate and timely information, a national human resource database will be unable to benefit policymakers and other users of the database. Noting the various data sources that exist in South Africa and the competing demands for resources, the establishment of a national human resources database would be substantially facilitated by using an existing database as a master list, providing a backbone into which other data sources may be linked and merged. Given its central role in storing data on qualifications, therefore, the National Learners' Records Database is a natural first choice to fulfil this master list function. It has been the objective of this report to evaluate the current suitability of the NLRD for this role and identify any potential issues, gaps or additional data required from which the national human resources database can be created.

Our conclusion on the suitability of the NLRD to act as a master list for the proposed national human resources database is that, all things considered, the NLRD is currently not in a position to properly fulfil this role. Further, current indications are that this conclusion may remain valid for at least another decade. This timeframe depends specifically on the extent of the problems surrounding the incorporation of full historical data. Even as the NLRD attempts to integrate more and more of the historical data, it will reach a point where the obstacles to integrating this data (such as where the data is in hardcopy only, where it is held by an unknown number of institutions, or where the data may have been misplaced or irretrievably lost) may render the inclusion of these data sources in the NLRD unviable. It must be remembered that the creation of a national human resources database in 2008, for example, would require full qualifications data going back to 1963 at least, further if it were to include data on individuals aged over 65 years in 2008. The only remedy to this problem, from the national human resources database perspective, is the aging of the missing data out of the database while maximising the stream of current data into the database.

Given that the NLRD has been set up for a very specific purpose and the fact that national human resource planning is not part of their mandate, the national human resources database should not be created by augmenting the NLRD and subverting

## **Towards a National Human Resources Database: The National Learners' Records Database Assessed**

---

its prime objectives. Instead, the NLRD data should feed into the national human resources database, and the latter should have a separate institutional 'home'. Close cooperation between SAQA and the institution housing the human resources database would be essential.

If government is at all serious about wanting to have a national human resources database in the future, the NLRD is arguably best placed to serve as a master dataset. Having addressed the data coverage issues, it would be in the position to be able to satisfy most of the criteria for use as a master dataset. In some instances, with proper linking to official data sources, other criteria can be met. Accordingly, immediate actions are required to support the NLRD and its data incorporation processes, thereby facilitating the establishment of a national human resources database at some point in the future. Consensus must be reached across government departments and state agencies on the desirability of a national human resources database, along with a commitment to facilitate linkages between the databases under their supervision and the prospective database.

As already noted, the objectives of the NLRD are not necessarily aligned to those of a potential national human resources database and, as such, the criteria against which the NLRD has been evaluated in this report are not meant to inform a specific opinion on the state of the NLRD itself in isolation. Given our analysis of the data and discussions with Yvonne Shapiro, we believe that the NLRD is, systemically, robust. Considerable care has been taken to ensure that data submissions, received from a large number of institutions using a wide variety of computer systems, are of decent quality. SAQA is fully aware of their responsibilities in terms of data quality and integrity and have therefore ensured that precautions are taken to ensure that substandard data is not accidentally uploaded into the production version of the NLRD. With the majority of ETQAs submitting data that meets the NLRD's minimum requirements, it has been possible for SAQA to further refine their requirements and improve overall data quality.

The NLRD was evaluated against four main criteria, namely unit of analysis, comprehensiveness, currency and ability to represent the stock of skills. With the individuals as its unit of analysis, the NLRD is fully compatible with a possible national human resources database, while the minimum requirement of biannual uploads from data providers means that the currency of the data is maintained. However, in terms of comprehensiveness, the NLRD has a number of data gaps that, if it were used as the master list, would seriously compromise the integrity of a

## **Towards a National Human Resources Database: The National Learners' Records Database Assessed**

---

national human resources database. Not only are these gaps present in the historical data, but also in current data. Chief amongst these is the lack of matric data in the NLRD, although there are other important data sources such as the legacy artisan data. The situation surrounding the data on matric qualifications should be resolved and SAQA should be supported in securing access to learner-level data from the Department of Education's Examinations Section for incorporation into the NLRD.

An important issue that affects currency of the database is the NLRD's current reliance on HEMIS as a source of public higher education information. Because of the extensive auditing of the HEMIS data by the Department of Education, in line with its responsibilities as a government department spending public funds, HEMIS submissions can take substantial amounts of time. Consequently, by mid-2008, HEMIS data from 2005 had yet to be uploaded to the NLRD's production environment. It is recommended that the NLRD be supported in its exploration of alternative sources of data from the public higher education system. Chief amongst these are the extension of HEQCIS to the public sector institutions, a process that would separate submissions to the NLRD from submissions related to funding, or directly accessing information from the institutions themselves. The former is arguably a preferred route, given the resource demands on the NLRD and SAQA that a direct interface with public higher education institutions implies. The issue of currency is an important one and, should the national human resources database be established, it would certainly result in more pressure on data providers and systems to ensure timely and accurate data submissions.

One of the key functions of the national human resources database is to provide an indication of the stock of available skills. While the NLRD does contain age data for the vast majority of learners stored in the database, it would still only be able to provide an approximation of the stock, even discounting any data gaps there may be. It is in the area of mortality and migration that specific links to official data are essential. A link into the databases of the Department of Home Affairs should be facilitated in order to populate fields detailing emigration, immigration and deaths.

One of the challenges facing the NLRD currently, and something that would arguably be experienced in the setting up and management of a national human resources database, is the human element. Data quality is enhanced when those individuals at the ETQAs and other data providers understand and appreciate the processes and, accordingly, SAQA and the NLRD have tried to facilitate data submissions and promote buy-in from those involved. However, staff turnover at data-providing

---

institutions impacts on continuity, particularly where institutional memory in terms of the requirements of the NLRD is not strong. This should be an important lesson for the proposed human resources database.

Given the analysis above, the establishment of a national human resources database is likely to be a longer-term undertaking. Using the NLRD as the master list, the key is to identify the most useful databases to link to and ensure that those databases comply with minimum data quality standards and link easily to the NLRD. Just as important is ensuring that the institutional relationships are cultivated and strengthened in anticipation of the database.

## **5 References**

- Kaplan, D., Meyer, J.-B. and Brown, M. (1999), "Brain Drain: New Data, New Options", *Trade and Industry Monitor* 10: 10-13. Available (online): [http://www.commerce.uct.ac.za/Research\\_Units/DPRU/TradeMonitorPDFs/TIMVOL11.PDF](http://www.commerce.uct.ac.za/Research_Units/DPRU/TradeMonitorPDFs/TIMVOL11.PDF). Accessed: 12 June 2008.
- Myburgh, A. (2004), "Explaining Emigration from South Africa", *South African Journal of Economics* 72(1): 122-148.
- Praxis Computing (2008), *Praxis Website*. Available (online): <http://www.praxis.co.za>. Accessed: 9 June 2008.
- South African Qualifications Authority (2007), *Trends in Public Higher Education in South Africa 1995 to 2004*. Available (online): <http://www.saga.org.za/show.asp?include=docs/reports/hetrenz/2005/main.html>. Accessed: 10 June 2008.
- South African Qualifications Authority (2008a), *South African Qualifications Authority Website*. Available (online): <http://www.saga.org.za>. Accessed: 30 May 2008.
- South African Qualifications Authority (2008b), *Specifications for Load Files for the National Learners' Records Database: Version 2.0*.

## **APPENDIX A**

Example of data validation processes undertaken by Edu.Dex (extract from "Best Practice for Validating and Extracting Data", South African Qualifications Authority 2008b: 36-37):

"5. Ensuring the integrity of National ID and Birth Date (and Gender):

5.1 Any value indicated as being a South African National ID should be validated by the information system to ensure all of the following:

- 5.1.1 The string contains exactly 13 characters.
- 5.1.2 The allowed characters are [the numbers 1234567890].
- 5.1.3 The string does not contain four zeros from position 1 to 4.
- 5.1.4 The string does not contain four zeros from position 7 to 10.
- 5.1.5 The string does not contain ten of the same numbers (e.g.

1111111111).

5.2 The information system should automatically populate the **date of birth** field by deriving it from the National ID. If the user opts to change the data of birth the system should generate an error. In this case, the user must undertake one of the following actions:

- Indicate that the value input into the National ID field is in fact not a National ID, or
- Change the National ID, or
- Leave the data of birth as is.

The information system should also check the data of birth field to ensure that it is feasible. At present, ETQA systems usually only relate to people of age at least 15, so the information system must not allow a date of birth generating a person age of less than 15.

5.3 The information system should automatically populate the **gender** by deriving it from the National ID. If the user opts to change the gender the system should generate an error. In this case, the user must undertake one of the following actions:

- Indicate that the value input into the National ID field is in fact not a National ID, or
- Change the National ID, or
- Leave the gender as is.

If these requirements cannot be achieved for data capturing, then the program that extracts the data for the NLRD must ensure that the extracted data fields conform to these standards."

---