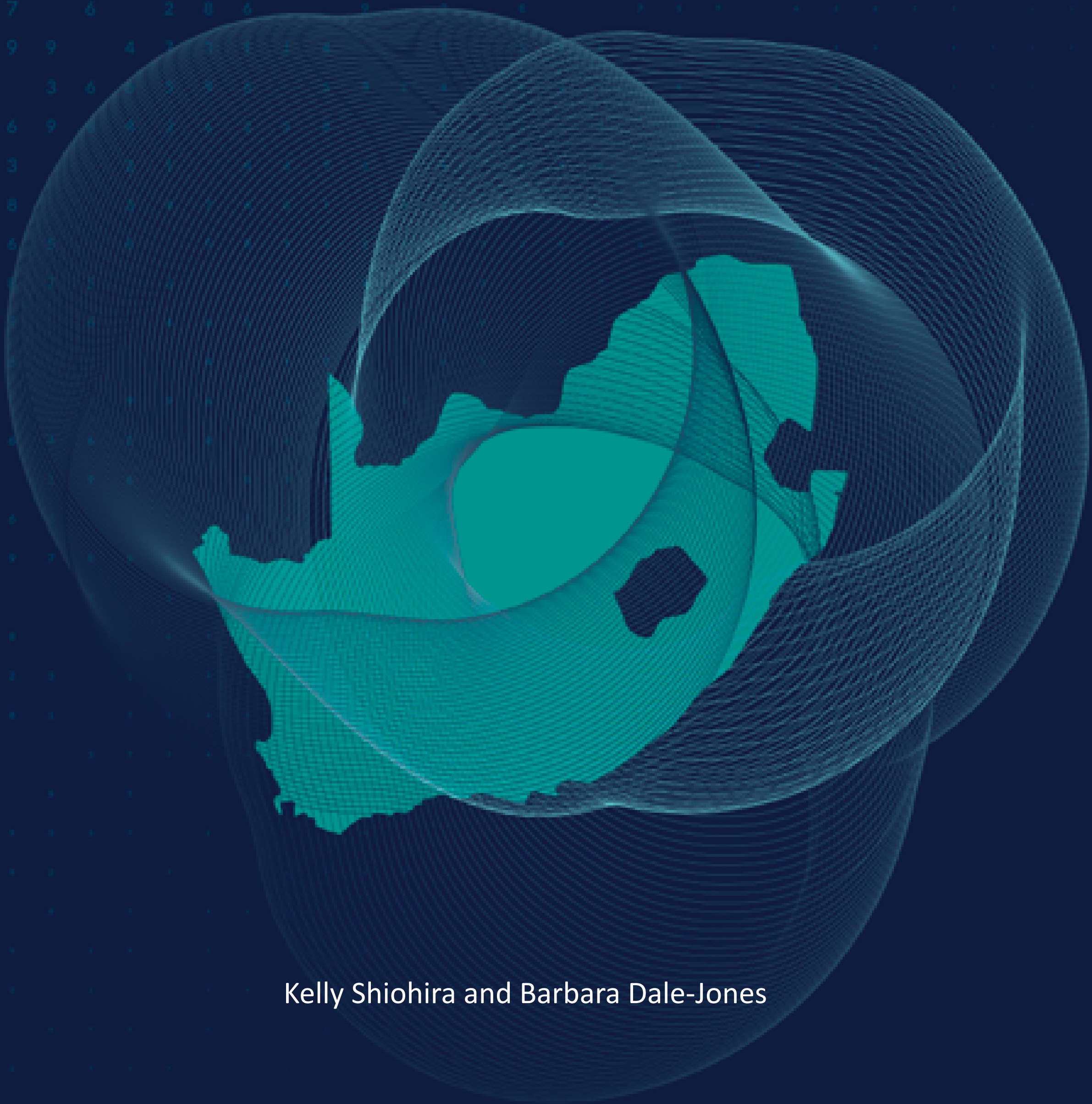
INTEROPERABLE DATA ECOSYSTEMS

An international review to inform a South African innovation



Kelly Shiohira and Barbara Dale-Jones

First published (2019)

JET Education Services
Education Hub
6 Blackwood Avenue
Parktown
Johannesburg, 2193
www.jet.org.za

merSETA
merSETA House
95 7th Avenue, Cnr Rustenburg Road
Melville
Johannesburg, 2109
www.merseta.org.za

How to cite this publication:

Shiohira, K. & Dale-Jones, B. (2019). *Interoperable data ecosystems: An international review to inform a South African innovation*. Johannesburg: JET Education Services and merSETA.

INTEROPERABLE DATA ECOSYSTEMS

An international review to inform a South African innovation

Kelly Shiohira and Barbara Dale-Jones

Contents

List of Figures

Preamble

Interoperability, data and ecosystems are three concepts that, to date, have not been used together enough internationally, and even less so in South Africa.

We now stand at the cusp of a new era in which we have an over-supply of data, while our ability to harvest the data simply does not keep up with the new approaches that artificial intelligence in a second machine age places at our disposal. Many countries, including South Africa, have a range of datasets that cover many critical aspects of their education and training systems – some of these may be well-established but perhaps archaic in design and based on outdated software, while others may be well-designed, containing pristine data that is not available anywhere else in the national system. These datasets are mostly government-owned and controlled, very often in a siloed manner, and mostly with access only given to a select few. The problem with this situation in South Africa, and the country is not unique in this regard, is that this lack of interoperability leads to a weak national data ecosystem, made up of only a few willing partners cooperating across datasets, often only because they are legally compelled to do so. As a result, the ability of the national system to link a latent workforce (the supply side) with existing and new opportunities (the demand side) is severely constricted, if functional at all. All of this is while the country struggles with high unemployment, and even higher youth unemployment.

On one level, a more interoperable data ecosystem may well contribute to linking supply and demand, but there is also much more to consider. New analytical techniques applied to big data can be used in ways not possible before: What if a career advisory service could guide school-leavers towards jobs in demand at the time they will graduate? A young unemployed graduate could see where there are employment vacancies in their immediate areas in real-time? An employer could plan the opening of new vacancies to coincide with the graduation of strong candidates with the relevant qualifications? A government department could design accurate scenarios for workforce development? There is a long list.

In all of this, and more so in developing countries like South Africa, government is the slow and steady gatekeeper, but the keys have become rusty and the lock and chain much too flimsy. Innovative people and companies will not wait for government to catch up, as we can see in the increase of the many platforms that link supply and demand across the world (think Uber for Work, Airbnb, and many others). While public-private partnerships that harness these new technologies have huge potential to support emerging economies, if left uncoordinated, they often result in increased inequality – something that South Africa can ill afford.

This international review attempts to explore some of these debates as part of a larger South African initiative that aims to ultimately give South African citizens the ability to make informed labour market decisions that lead to employment. International readers will also find value in the analysis as the move towards interoperable data ecosystems is certainly a global narrative that can only be ignored at your own peril.

Acknowledgements

The authors would like to thank the Manufacturing, Engineering and Related Services Sector Education and Training Authority (merSETA) for the opportunity to engage in the research and development of this paper. In particular, Dr More Manda and Sebolelo Nomvete made significant contributions to the conceptualisation and support of this project.

Bangani Ngeleza contributed his considerable facilitation skills, particularly in developing the PSET digital ecosystem theory of change.

Dr Lori Foster contributed immensely through her constructive engagements and review of the paper.

Appreciation is also especially given to the individuals who participated in interviews and therefore contributed their time, advice and considerable knowledge to this paper. It was a genuine pleasure learning about the wide range of significant efforts in this field.

Finally, this paper would not have been possible without the support of the determined and committed team at JET Education Services. Special thanks to our collaborators, James Keevy, Rooksana Rajab and Tadiwanashe Murahwi, as well as our editor extraordinaire, Maureen Mosselson.

About merSETA

The merSETA is one of 21 Sector Education and Training Authorities (SETAs) established to facilitate skills development in terms of the Skills Development Act of 1998 (as amended). The 21 SETAs broadly reflect different sectors of the South African economy. The merSETA encompasses Manufacturing, Engineering and Related Services.

About JET Education Services

JET Education Services is an independent non-governmental organisation in South Africa which works with government, the private sector, international development agencies and education institutions to improve the quality of education and the relationship between education, skills development and the world of work.

About the authors

Kelly Shiohira holds a Master of Science Degree in International Educational Development from the University of Pennsylvania and a Master of Arts degree in Applied Linguistics from Rhodes University. Kelly has worked internationally in the fields of language and literacy acquisition, teacher development and technology for educational improvement, and has applied these areas of expertise to curriculum and policy design, research design and implementation, strategic planning and monitoring and evaluation in the education sector. Kelly is a published author in both English and Japanese and has presented at conferences internationally on topics including curriculum design, integrating ICT into the classroom, AI and skills development, differential education in second-language learning, cultural adaptation, materials development in indigenous languages and evaluating the impact of ICT in the classroom.

Barbara Dale-Jones holds a Master of Arts degree in English Literature. She is one of the founders and directors of both The Thunderbay Collective and The Field Institute. She is passionate about transforming systems and organisations to address the challenges of the emerging future. Barbara is extensively experienced in the field of education. She was previously the CEO of BRIDGE, a key NPO in education in South Africa. As well as having expertise in running organisations, Barbara is a digital transformation specialist (she worked in digital transformation consulting at Dimension Data's Global Digital Practice) as well as a knowledge management practitioner (she has consulted widely in the field of knowledge management and learning organisation design). Barbara ran an e-learning company from 2005–2009, and has also worked in publishing and lectured in English at Rhodes University.

Acronyms and abbreviations

4IR	Fourth Industrial Revolution
ABS	Australian Bureau of Statistics
AI	artificial intelligence
API	Application Programming Interface
AS	autonomous systems
CHESSIC	Chinese Credentials Verification Authority
DBE	Department of Basic Education
DDD	Data Driven Districts
DHET	Department of Higher Education and Training
DPME	Department of Planning Monitoring and Evaluation
DSP	Demand-Side Platforms
DTPS	Department of Telecommunications and Postal Services
EAC	East African Community
EdTech	educational technology
EPA	The United States Environmental Protection Agency
EU	European Union
FICCI	Federation of Indian Chambers of Commerce & Industry
GSBPM	General Statistical Business Process Model
HEIs	higher education institutions
HiMSS	Healthcare Information and Management Systems Society
ICTs	information and communication technologies
IEEE	Institute of Electrical and Electronics Engineers
IFR	International Federation of Robotics
ISO	International Organization for Standardization
JET	JET Education Services

LANs	local area networks
LMS	learner management systems
merSETA	Manufacturing, Engineering and Related Services Sector Education and Training Authority
MOOCs	massive open online courses
MSDF	Michael and Susan Dell Foundation
MVP	minimum viable product
NGOs	non-governmental organisations
NSTF	National Science and Technology Forum
ODINE	Open Data Incubator Europe
OECD	Organisation for Economic Co-operation and Development
OGC	Open Geospatial Consortium
PSET	post-school education and training
SABS	South African Bureau of Standards
SA-SAMS	South African School Administration Management System
SDGs	Sustainable Development Goals
SDMX	Statistical Data and Metadata eXchange
SIS-CC	Statistical Information System Collaboration Community
SSP	Supply-Side Platforms
STEM	science, technology, education and maths
UIS	UNESCO Institute of Statistics
UNECE	United Nations Economic Commission for Europe
UNESCO	United Nations Educational, Scientific and Cultural Organization
WANs	wide area networks
WEF	World Economic Forum

SETTING THE SCENE

INTRODUCTION

Over the last fifty years, governance has been repeatedly challenged to respond to continuous technological innovations of the public, private and academic sectors. With these innovations come new opportunities, and the successful early engagement of these prospects can provide governments and populations with new, important resources and insights to navigate the shifting landscapes of global and local economies.

The pace of technological advancement has only increased, with new innovations in hardware (such as tablets and smartphones) enabling software which has disrupted entire industries. In August 2011, Marc Andreessen published an article in the *Wall Street Journal* titled, “Why Software is Eating the World”. Using examples like Amazon’s domination over brick-and-mortar bookstores, the establishment of Netflix, and the rise of Google as a marketing platform, Facebook, Skype, Twitter, Flickr, Netflix, iTunes, LinkedIn, Zynga, PayPal and others, Andreessen uses the metaphor of software “eating” industries one after the other, from photography to direct marketing to telecommunications to finance, and argued that survival in the new age requires a company to transform from service-based to software-based. He notes that even primarily physical industries, like the automotive industry and retail, are increasingly reliant on software. As to why, Andreessen explains that “six decades into the computer revolution, four decades since the invention of the microprocessor, and two decades into the rise of the modern internet, all of the technology required to transform industries through software finally works and can be widely delivered at global scale”. A large part of this is related to cost: Andreessen notes that cloud computing was pioneered as early as 2000 – available at 100 times the 2011 price.

Although the conversation now centres on “disruption” rather than on ingestion, it is difficult to refute Andreessen’s conclusion. Since the publication of his article, transportation and hospitality have been added to the list of disrupted industries, with software platforms such as Uber, Lyft and Taxify revolutionising transportation and platform-based Airbnb, barely a whisper in 2011, in 2018 booking out 2 million rooms a night. In *The Platform Revolution: How Networked Markets are Transforming the Economy*, authors Parker, Alstyne and Choudary note with particular interest the astonishing and pervasive trend illustrated by these companies – they don’t own the physical assets, they supply. What disruption essentially reflects is an underlying transformation in the nature of an industry’s value chain through the introduction of platforms. These platforms provide value-add through the application of a filter which connects consumers to relevant products in a type of marketplace in which the ability to draw both ample supply and demand, and the effectiveness of the filter which connects demands to relevant supply, anticipate a company’s success.

The ability of platform companies to draw revenue is indisputable, but even more profound, from a policy perspective, are the potential effects of disruption on economies and workforces. The ripple effects of these new types of public-sector companies include, in some cases, new models of employment and revenue generation, at times displacing or eliminating existing industries. Taking Amazon as an example, from the consumer perspective, the elimination of the additional costs associated with brick-and-mortar stores – managers, cashiers, cleaners, stockers, shopfitters, supplies such as display cases and so on – and the convenience of home delivery are positive aspects.

However, from the perspective of employees at brick and mortar book stores, the situation looks quite different. At the same time, the continuing decline of print industries has enabled new points of entry, from authors who self-publish on e-platforms to a widening pool of contributing journalists unbound by traditional affiliations or locations.

The swiftness with which industries and even entire sectors can be disrupted requires agility on the part of companies, individuals and policy-makers. Employees are tasked with ongoing upskilling and reskilling to maintain their productivity and employment, while companies are tasked with the continuous investigation and adoption of new practices, which range from the adoption of new software to integrating new methods of consumer engagements like targeted advertising. For policy-makers, the scope, scale and uses to which data and new technologies are put to use require swift responses to not only protect citizens but also ensure appropriate education for a changing labour market. Unfortunately, in most countries, including South Africa, policy changes are unable to keep up with global technological trends.

TOWARDS A SOUTH AFRICAN INNOVATION

In South Africa, the Manufacturing, Engineering and Related Services Sector Education and Training Authority (merSETA), in collaboration with government and a well-established non-profit research organisation, JET Education Services (JET), is in the process of developing and establishing an integrated and interoperable digital ecosystem for the post-school education and training (PSET) sector. The main objectives of this innovation are to strengthen integration across current and new systems/platforms; strengthen collaboration to improve efficiencies, governance and management; harvest rich knowledge and intelligence; enhance strategic planning and decision-making; be responsive to the needs of the labour market and national priorities; and strengthen, integrate, coordinate and improve efficiencies in the governance and management of PSET. Overall, the initiative intends to ensure that data sets in the South African PSET sector will be interoperable, well-synchronised and used effectively as a source of information for planning and improving efficiency in the PSET system. This international benchmarking review of best practices is a first step towards the creation of the digital ecosystem for the PSET sector.

The stated purposes of the international benchmarking review are to:

1. Determine the enablers and inhibitors of interoperable data management and usage systems in PSET internationally and within contexts similar to South Africa;
2. Provide evidence of best practice and learnings in terms of approaches in designing, developing and maintaining complex data systems;
3. Explore opportunities for partnership with or usage of existing systems or platforms.

In the course of providing information on comparable or related international projects, this paper is also meant as a foundation for policy-makers and other stakeholders who may engage the PSET digital ecosystem. As such, it explores the importance of the "Fourth Industrial Revolution" (also known as 4IR) in order to contextualise the vision of the project in relation to current social, political and economic changes. The paper then discusses three primary opportunities to leverage these changes in the service of education: platforms; the creation of interoperable data systems; and artificial intelligence. The discussion of each of these opportunities includes relevant frameworks, enablers and inhibitors and examples of good practice. The paper concludes with recommendations for building a system such as the conceived PSET digital ecosystem.

METHODOLOGY

In order to develop this review, desktop research was conducted on topics which are central to the formation of a digital ecosystem in the modern era:

- The context of the Fourth Industrial Revolution and its impact on both broad economic productivity and individual sectors and organisations;
- The broad structures and use cases[1] of artificial intelligence;
- Interoperability and interoperability frameworks;
- Data standards and their role in interoperability as well as examples of data standards;
- Platforms and the uses of platforms in education; and
- Open data and its value.

Types of sources consulted included peer-reviewed academic publications, published books, policy documents, print media such as newspapers, conference proceedings, websites of national and international statistical organisations and their associated resources, websites promoting standards and platforms, webinars and similar publications by related organisations, and so-called "grey literature" on the topics published by think tanks, non-governmental organisations (NGOs), academics, graduate students and/or other individuals.

In addition to the literature, interviews were conducted with representatives of international organisations and practitioners in order to better understand current international best practices in the systemic flow and use of data. Participants were identified based on knowledge of policies, platforms, systems, operations and/or other aspects relevant to the research. In some cases, interviewees were prior contacts of the researchers, but in many, the request for an interview was the first contact. Interview questions covered the following topics: views on interoperability; general information on platforms and systems in use; system beneficiaries; data procedures and protocols; human resources and skills; and best practices and lessons learnt.

In total, 19 organisations and individuals were contacted for interviews. Of these, one individual and four organisations did not respond (the Open Skills Platform at the University of Chicago, the IMS Global Learning Consortium, the FutureSkills Platform of India, and CB Insights). Two organisations, Statistics New Zealand and the Alan Turing Foundation in the United Kingdom, responded but were unable to set up interview dates. Between the months of January and April 2019, 13 interviews were completed. The results of these interviews informed further development of the literature review.

Although this report is intended to have a primarily external focus, it is written with the context of the PSET digital ecosystem in mind (see Figure 1). In particular, the intersection between the international benchmarking and the preceding activity of identifying existing data systems as well as subsequent activities, such as developing the architecture, integration specifications and minimum viable product (MVP) descriptions, have been considered in the framing of this report. For this reason, some South African context is included in the first section, and the conclusions and recommendations are focused towards the South African context.

Although this report is intended to have a primarily external focus, it is written with the context of the PSET digital ecosystem in mind

1 A use case is a set of actions or events that outlines the software requirements for how a user will interact with the system to achieve a given goal. The establishment of a use case thus requires the definition of a goal or set of goals users may achieve.

DIGITAL ECOSYSTEM THEORY OF CHANGE

The establishment of a robust digital ecosystem hinges on the ability to solve a number of challenges, including technology and capacity gaps, access to data and data interoperability. However, if managed, the PSET Digital Ecosystem will provide South African citizens and government agencies access to a wider range of data which can be utilised in more meaningful ways to make decisions about education, skills and labour.

Early project activities include stakeholder analyses, consultations and national and international research which seeks to determine the feasibility, advantages and critical success factors for developing the relevant networks and ancillaries for such a system.

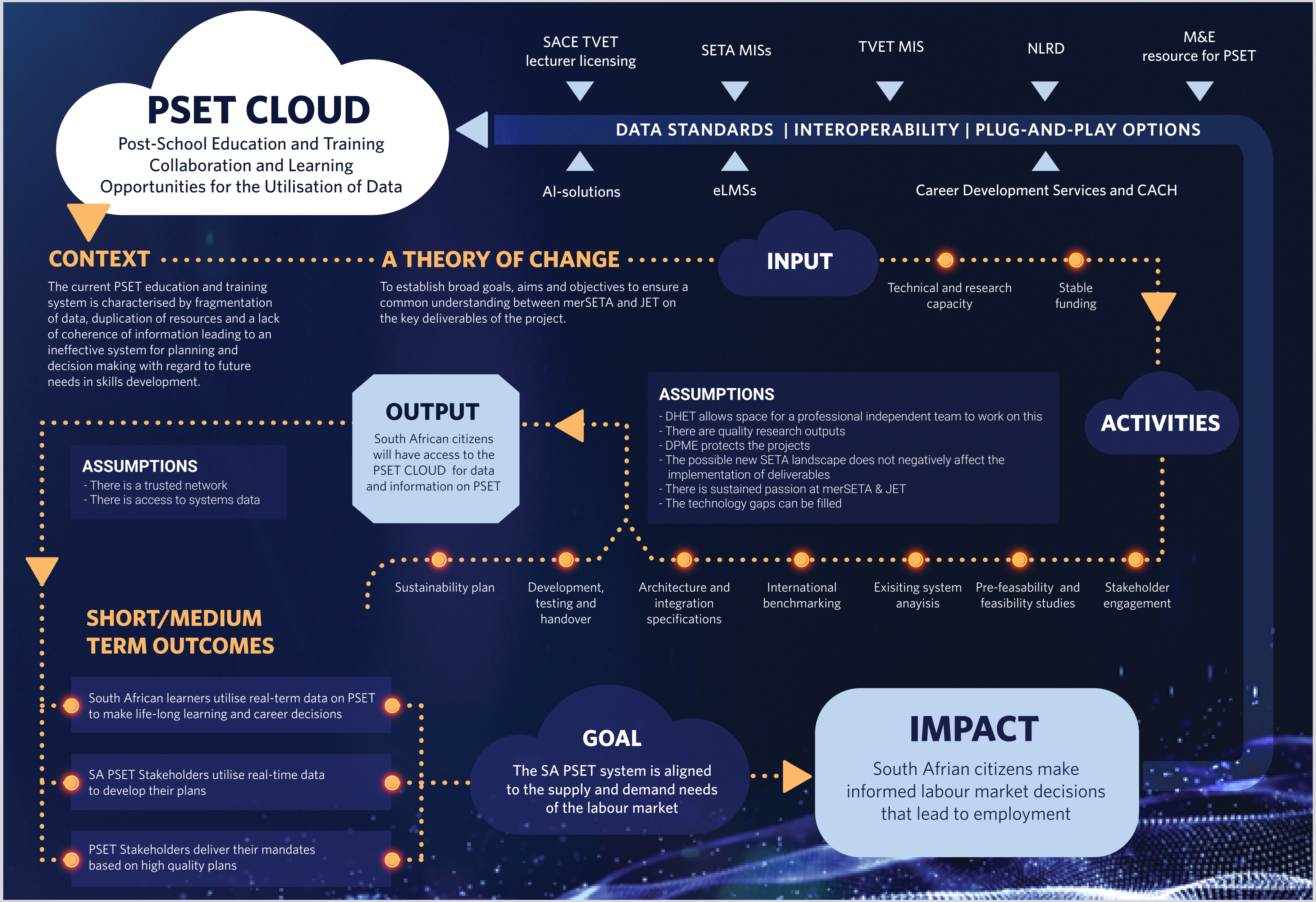


Figure 1: Digital ecosystem theory of change (JET and merSETA, 2019)

ADAPTING TO A TECHNOLOGY-DRIVEN ERA

This paper is premised on the understanding of the PSET sector as a complex set of organisations which deliver, manage, oversee, quality assure and provide credentials for higher education, vocational education, adult education, further education and skills development. These organisations include public and private education providers, qualifications authorities and credentialing agencies, workplace-based learning or work experience programmes, and government departments engaged in oversight of these functions. The acronym "PSET" is used in this paper as this is the term used to describe this sector in South Africa. One of the primary purposes of the PSET sector is to provide education and training which enables participants to engage in society economically, socially and, in a democratic system, politically. As such, it is important that PSET systems in all nations ensure that they remain both effective in their delivery and relevant in their offerings.

Currently, a number of transformations are occurring in both national and international labour markets which especially affect the PSET system as one of the direct points of contact between education and labour. The rise of platform technologies and their ability to disrupt entire industries, the ongoing development of the "gig economy", increasing advancements in robotics and artificial intelligence, and the development of linked innovations such as "wearable" health monitors in the form of accessories, smartphones and their slew of applications, the Internet of Things, blockchain and so forth are set to have profound impacts on both the labour market to which education responds and the delivery of education itself.

The following section explores the Fourth Industrial Revolution, a term used in this paper as a blanket label to cover these and other technological innovations, their introduction into various sectors and the potential impacts on labour and education. The discussion on the Fourth Industrial Revolution provides a broad context which outlines the social, political and economic changes which have resulted in a perceived need for and advantages of having a PSET digital ecosystem in South Africa. The New Zealand framework for education practice developed in response to 4IR challenges and opportunities is provided as an example.

THE FOURTH INDUSTRIAL REVOLUTION

Over the years, there have been shifts in the processes of production and the way factories operate caused by changes in technology and the pursuit of efficiency, known as industrial revolutions. Lee et al. (2018) postulate that over the years the development that has been fostered by each industrial revolution has been: mechanical technology in the first revolution; electrical technology in the second revolution; and information technology in the third revolution.

The Fourth Industrial Revolution, as the technological revolution enabled by advancements in connectivity, robotics, artificial intelligence (AI) and cloud computing has come to be known, remains a controversial concept. While some go as far as questioning whether there is a revolution worth being called the Fourth Industrial Revolution in the first place (Lee et al., 2018), in 2017, the World Economic Forum (WEF) declared that the Fourth Industrial Revolution has been progressing since the start of the 21st century, using recent diverse technologies unique to the 21st century as evidence. These technologies are generally characterised by a fusion of the physical, digital and biological spheres (Schwab, 2017). The Fourth Industrial Revolution is further set apart from its predecessors due to its speed (transformation will take years rather than decades) and scope (the impending changes will affect everyone) and, more importantly, the need to take a systems approach (the changes will be far more extensive and deeper) (WEF, 2017).

The Fourth Industrial Revolution can be described as a revolutionary change characterised by "a much more ubiquitous and mobile internet, by smaller and more powerful sensors that have become cheaper, and by artificial intelligence and machine learning" (Schwab, 2017, 7). During this latest industrial revolution, innovations such as advanced robotics, AI, cloud computing and digital fabrication (3D printing) are combining to change the way factories and businesses operate. Computing devices and sensors are connected to the internet, allowing devices to exchange and act on data, a concept known as the "Internet of Things" (Wortmann & Flüchter, 2015). Cyber-physical systems technically assist humans in complex tasks and have the autonomy to make simple decisions without human input (Marr, 2016). For instance, big data analysis may be used to forecast the outbreak of deadly diseases and track epidemics; 3D printers can be used in refugee camps to produce cheap, custom-built prosthetic limbs; and AI can read digital scans more accurately than doctors (Kituyi, 2018). In short, the Fourth Industrial Revolution is all about advanced computer technologies working with and for humans.

Recent advancements linked to the Third and especially Fourth Industrial Revolutions are largely rooted in *digitisation*. The integration of advanced digital-based and intelligence-based machines and platforms (Lee et al., 2018) requires digitisation: a conversion of data so that it can be processed by and stored in computer systems (Bloomberg, 2018). Digital technologies are rapidly spreading, and the digital transformation of societies is emerging consequently, including in Africa (Ndemo & Weiss, 2017). Ndemo and Weiss (2017) categorise the deployment of digital technologies into two substantive categories, namely: for the optimisation of organisational processes; and for the optimisation of market transactions.

The optimisation of organisational processes can be achieved through the digitisation of the internal processes in an organisation (for instance, enterprise resource planning software, accounting software, inventory management, cloud computing, etc.). Such technology requires an initially labour-intensive conversion of physical activities and records into mapped-out process flows and digital archives (Ndemo & Weiss, 2017).

Optimising market transactions is about innovations that assist an institution in its specific market functions. For instance, AI in the form of machine learning algorithms is being used to analyse large data repositories and provide advice in fields such as finance, health, recruiting and logistics in order to improve efficiencies (Ndemo & Weiss, 2017). Another example would be predictive data analytics in agriculture that can provide location-specific forecasts on crop productivity (Ekekwe (2017) & Ndemo (2017), cited in Ndemo & Weiss, 2017).

Not to be confused with digitisation, *digitalisation* encompasses operations, redesign of products and services and closer interaction with customers. The redesigning of products and services allows greater responsiveness and interaction and also allows tracking of activity and results (Bloomberg, 2018). When analysed, data received from such products can then show how well the products are functioning. Efficiency will thus be affected positively by digitisation, to the point of physical products being displaced by digital products. The digitisation of operations will increasingly transform how organisations and institutions do business and operate their productions and even affect society (Lee et al., 2018).

As in any other industry, education can be disrupted to increase efficiencies and deliver improved services and supports. In the context of education, there are efforts towards digitisation in the form of learner management systems (LMS), and the digital revolution opens new possibilities for forecasting, assessment, planning, curriculum delivery, monitoring and tracking (Wakefield, 2017). However, while investment in educational technology is robust, the sector lags behind marketing, finance and even health (CB Insights, 2018a), likely due to a combination of factors including policy and, most importantly, the lack of direct financial returns on investments. As a result of its unique factors, in many contexts, education systems are reliant on direct government intervention to change.

In addition to sector divides, a developed-developing gap is also emerging. Generally, the global North has made enormous strides in technological growth, especially in computing power (Maharajh, 2018). While the United States still dominates in terms of the number of AI startups and total equity deals, that share is decreasing (CB Insights, 2019), and scholars like Schäfer (2018) argue that the European Union (EU) could lead the Fourth Industrial Revolution. There have been comprehensive reforms to address the Fourth Industrial Revolution in European countries such as Germany, France and the United Kingdom (Liao et al., 2018), focused on investment in digital infrastructure and strengthening the market by reducing barriers to the trade in goods, information, services and applications (Schäfer, 2018). In 2018, the European Commission undertook a number of related initiatives, including a declaration of cooperation on AI, a policy on AI, an AI roadmap and plans for a coordinated action plan (European Commission, 2018).

However, Asian countries in particular, as well as some developing countries, are also investing heavily in the Fourth Industrial Revolution. Singapore and Taiwan, as well as China, have invested significantly in the development and use of AI (Stratus, 2017). In fact, the International Federation of Robotics (IFR) predicted that in 2019 China would be responsible for 40% of worldwide robotic sales, a notable increase of 27% over the country's 2015 contribution (Gonzalez, 2018; IFR, 2017); and as of 2018, China is the biggest shareholder of the robotic global market at a net worth of USD 30 billion (Wang, 2018). Across Asia, organisations are leveraging industrial Internet of Things solutions to provide greater flexibility and improved efficiency of services as well as lower production costs. They are developing new customer-relevant applications to bring to market and investing in long-term automation capabilities (Stratus, 2017). However, other countries in the region, such as the Philippines, and continents such as Africa and South America show varying degrees of diffusion in the adoption of the various 4IR technologies (Dadios et al., 2018). This is strongly affected by an uneven distribution of the "scientific and technological infrastructure necessary" for such adoption; the global South and particularly Africa have largely been unable to reap the benefits of the series of technological advancements enabled by modern electricity (Maharajh, 2018).

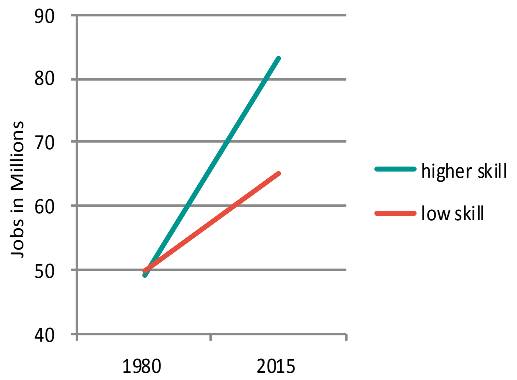
Regardless of geographical location, digital disruption is transforming the way career paths are structured, largely driven by two factors: the desire of workers for job security (Rakyan, 2017); and the shortage of skills in the science, technology, education and mathematics (STEM) fields (Donovan & Benko, 2016). In 2016, 54% of American workers recognised further upskilling as necessary for their careers (Pew Research Center, 2016), and according to a Deloitte survey, 39% of large-company executives say they were "barely able" or "unable" to find the skills their companies needed (Donovan & Benko, 2016). Much of the shortage is in the STEM fields (Rakyan, 2017), and increasingly, economic success is reflective of an ability to successfully engage in a trifecta of skills development and training, technology, and the world of work. This results in what Benko and Anderson (2010) described in *The Corporate Lattice: Achieving High Performance in the Changing World of Work*. The "corporate lattice" represents career paths that diverge from the traditional fixed-step "corporate ladder" imagery, creating a more flexible career development pathway which requires both lateral and vertical movement. As further evidence of the ubiquitous interest in continuing education, in 2017, the growing investment in educational technology (EdTech) was heavily skewed towards consumer and corporate-focused learning environments, with only 13% of AI for education investment going to primary and secondary education and a further 8% towards higher education (Puskar, 2018).

> Digital disruption is transforming the way career paths are structured

The progress of digitisation and the pace at which technology is created, replaced and retired has made traditional models of training and development difficult to maintain, and innovative solutions to skills development have to be found. This illustrates the need for ongoing changes in education, not only within the traditional education system but also throughout the new career path. Increasing levels of education will be necessary to keep up with digital disruption, and there is no reason to expect a dramatic departure from the historic trendlines showing a shift from low-skill to high-skill labour. According to research by the Pew Research Center (2016, cited in 1-Page Ltd., 2016), "as of 2015, 83 million Americans worked in jobs requiring an average or above-average level of preparation (including education, experience and job training)", a 68% increase over the number of jobs requiring average or above-average preparation in 1980. While the number of jobs needing below-average level preparation still rose from 50 million to 65 million, the rate of growth was less than half that of higher skill professions. And while these effects are expected to be less dramatic in emerging economies (Purdy & Dougherty, 2017), a study of future jobs in India by EY, the Federation of Indian Chambers of Commerce & Industry (FICCI) and the National Association of Software and Services Companies (NASSCOM) in 2017 determined that by 2022, around 46% of the workforce would be engaged in jobs that did not exist at the time of the study or would be deployed in similar jobs with radically changed skill sets (EY, FICCI & NASSCOM, 2017).

The dynamics of the potential of new technology, combined with the changing skills needs of new economies, create a particular challenge and opportunity in the education sector.

Figure 2: Growth in higher and low skill jobs, 1980–2015 (adapted from Pew Research Center, 2016)



The Fourth Industrial Revolution's contribution to the SDGs

The Sustainable Development Goals (SDGs) outlined in the 2030 Agenda for Sustainable Development adopted by all United Nations (UN) Member States in 2015 is an urgent call for action by all countries to work together in a global partnership to realise peace and prosperity for people and the planet, now and into the future. The 2030 Agenda is driven by 17 SDGs that address the global challenges and include those related to poverty, inequality, climate, environmental degradation, education, prosperity, and peace and justice (UN, 2015). The Fourth Industrial Revolution has the potential to transform and realign economies and societies (Herweijer et al., 2017). As described by the World Economic Forum, "New technologies are enabling societal shifts by having an effect on economics, values, identities and possibilities for future generations" (WEF, 2017, 3). With the combination of big data, advanced robotics, AI advancements, interoperability and the Internet of Things, the sustainable development goals are more within reach than they ever have been: innovations such as massive open online courses (MOOCs) and personalised learning AI could provide remote learning opportunities to improve education access and outcomes; prescriptive AI linked to soil sensors could greatly improve crop yields and decrease food poverty; and advancements in interoperability in healthcare could dramatically improve patient care and allow for new types of healthcare delivery to reduce maternal and infant mortality.

Seeing such potential, the UN believes that technology must be harnessed to deliver on the ambitious SDGs (see the examples cited above). However, the United Nations Educational, Scientific and Cultural Organization (UNESCO) proposes that the starting point for this change is quality education (SDG 4) and calls for a special focus on narrowing the gender gap, especially in the growth sector constituted by professions requiring a STEM background. Essentially, UNESCO's view is that the Fourth Industrial Revolution must be a development revolution (UNESCO, 2016).

As part of the initiative to leverage the Fourth Industrial Revolution for sustainable development, the first United National World Data Forum was held in Cape Town, South Africa in 2017. The work of this forum is being taken forward by the Collaborative on SDG Data Interoperability, which, among other initiatives, supported the development of a guide, *Data Interoperability: A Practitioner's Guide to Linking up Data in the Development Sector* (Morales & Orrell, 2018). The guide identifies five areas which it considers crucial to the development of interoperable data systems at scale:

- Interoperability, data management and governance;
- Canonical data[2] and metadata models;

2 A canonical data model can be described as a ''common data model'' that ''presents data entities and relationships in the simplest possible form". (https://www.techopedia.com).

- Classifications and vocabularies;
- Standardised interfaces;
- Linked data.

However, there are many intervening factors for developing countries with regard to using technology. Attempts to bring the benefits of technology to developing countries require a certain threshold of capacity to use and understand technology and its implications and to recognise the daily opportunities to make technology work for people. Most developing countries face the constraint of infrastructure. Technology products designed in developing nations are often not designed to meet the needs of the poor or those in remote areas. Affordability is also an issue: with roughly half the world living on less than four dollars a day, many potential users are too poor to afford any form of access to technology (Miah & Omar, 2012).

The Fourth Industrial Revolution and South Africa

In terms of technological advancement, South Africa is in a better position than a number of countries on the African continent, although a lot is to be done in areas of access, inter alia (Brown, 2017). South Africa has committed to an early step in the direction of an action plan to deal with the Fourth Industrial Revolution, and plans to establish a Presidential Commission on the Fourth Industrial Revolution are underway (Department of Telecommunications and Postal Services (DTPS), 2018).

However, the National Science and Technology Forum (NSTF) (2018) opines that the country might be underprepared and is not taking sufficient advantage of 4IR opportunities for establishing businesses and industries, despite its strong scientific base and innovative people. For example, a study by Deloitte found that in the manufacturing sector, "the adoption of Cloud solutions is currently more driven by consumers than businesses, with cyber-crime fears and privacy issues cited as main concerns by the latter" (Deloitte, 2016). The study did note, however, that manufacturers recognise that they could be making more use of 4IR technologies for processes such as "monitoring, controlling, tracking etc." (Deloitte, 2016).

Brown (2017) states that the inequality in South Africa translates to a digital divide in terms of access to technology, a challenge recognised by the *National Integrated ICT Policy White Paper* (DTPS, 2016). Brown (2017) purports that the advancement in information and communication technologies (ICTs) can help address the challenge by broadening access as visualised in the nation's *National Development Plan 2030*. However, to capitalise on these, "a massive skills reorientation will be needed" given the rapid shifts in the skill sets required by the economy (Brown, 2017, 3). New models in the South African education system are needed to cope with the new skills demand. Although the innovation that comes with the Fourth Industrial Revolution may be a boost to the education system, utilising these innovations is a challenge on its own. The challenges facing higher education institutions (HEIs) in South Africa include the contextual problems inherited from past educational practices (Bozalek & Boughey (2012), Leibowitz (2012) & Soudien (2012), cited in Ng'ambi et al., 2016) and preparing future generations of students for emerging technologies (Broekman, Enslin & Pendlebury (2002) & Veletsianos (2010), cited in Ng'ambi et al, 2016).

However, according to Ng'ambi et al. (2016), there has been movement in South Africa's higher education sector towards cloud-based ICT infrastructure and its associated benefits, including improved efficiencies, greater flexibility for academics and students and larger geographical reach. Investors are capitalising on the potential, with a marked increase in investment in EdTech, particularly in South Africa, Kenya and Nigeria – locations where a combination of ICT infrastructure and the lack of affordable quality public education may drive early adoption of learning technology (Puskar, 2018).

Responses to the Fourth Industrial Revolution in education: Disruption, education and skills

In a presentation at the Tertiary ICT Conference in 2017, Stuart Wakefield presented a vision for the New Zealand education system based on uses and responses to the Fourth Industrial Revolution. The system is designed to be data-driven and to put learner information at the centre of a network of supports such as career planning and parent outreach, areas which could also draw and feed information back into the system. An overview is presented in Figure 3.

Four facets of education are affected by digital disruption: administration; assessment; teaching; and learning. Wakefield outlined the potential of disruption to greatly reduce administrative costs in education. Some sample applications of AI technology at the level of individual institutions would be student application and selection procedures and processes; HR and talent management for lecturers and other staff; building maintenance; grading; and back-office records maintenance. If applied, these technologies could greatly decrease the costs of administration and, applied across institutions and larger geographical areas, offer a solution to some of the more prevalent issues such as transference of credentials, an issue which severely limits workforce mobility (Purdy & Dougherty, 2018). Wakefield's description of New Zealand's model, with students at the centre of a data-driven, interoperable system which tracks and monitors their progress through and across institutions within the system, is depicted in Figure 4.

In terms of assessment, Wakefield drew on the work of Puentedura (2010), who provided a differential model for uses of technology which ranged from transformative technology, which would enable new tasks or a redesign of tasks, to enhancement, or the use of technology to augment the features or capabilities of traditional classroom practice. Translating this to the practice of assessment, Wakefield suggested new methods of assessment practice which could both improve the quality and relevance of assessments as well as the capture and use of data generated by assessments.

In the realm of teaching and learning, New Zealand added two new technology-based subjects to its secondary school curriculum, one example of the many policies emerging to manage the risks of scarce skills in STEM fields.

The frameworks proposed by Wakefield (2017) outline the depth of consideration given to educational transformation in the era of the Fourth Industrial Revolution, encompassing the use of technology to improve efficiency and the availability of data across the education system as a whole as well as the inclusion of 4IR-related technologies and related skills within the curriculum. The importance of the student as the central point in a web of affected stakeholders is highlighted, with 4IR technologies forming core methods of engagement at multiple points across the system.

Figure 3: New Zealand model of educational supports based on digital processes (adapted from Wakefield, 2017)

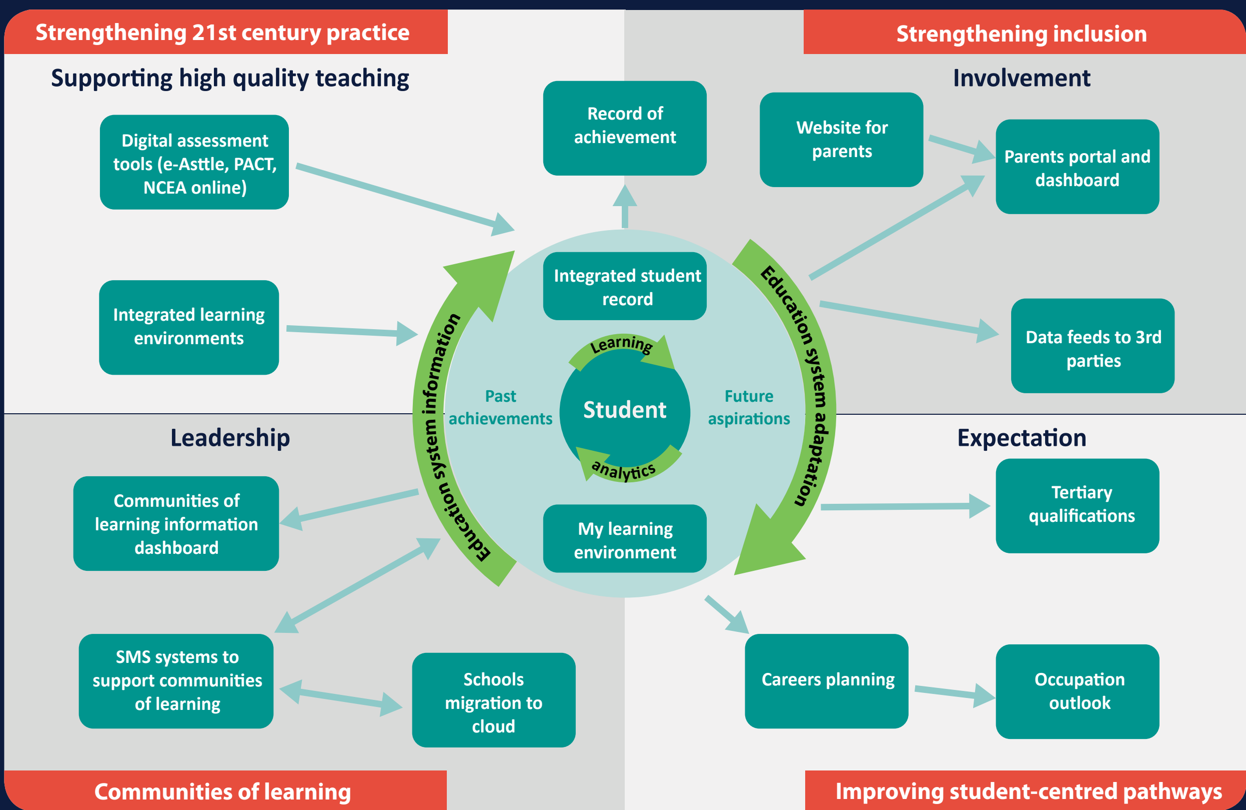


Figure 4: New Zealand student-centred management system (adapted from Wakefield, 2017)

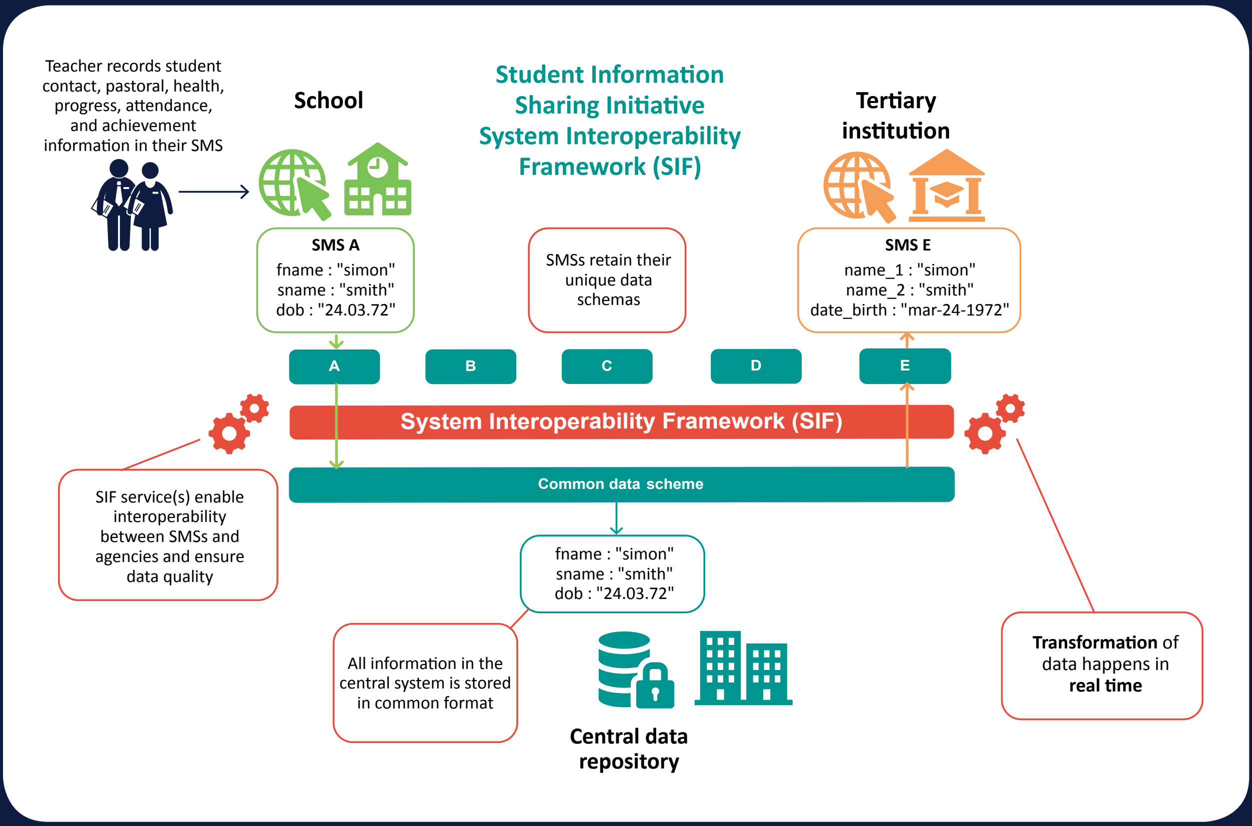
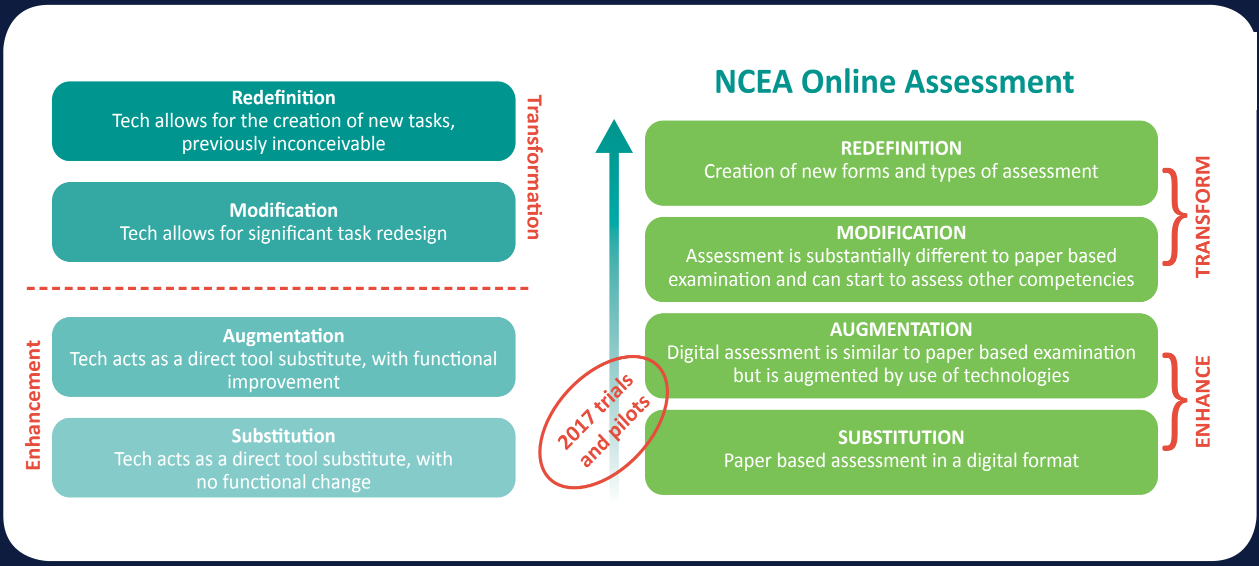


Figure 5: Puentedura's model of technology use in education (left) and Wakefield's interpretation for the field of assessment (right) (adapted from Puentedura, 2010 and Wakefield, 2017)



OPPORTUNITIES: LEVERAGING A TECHNOLOGY-DRIVEN ERA

As seen in the New Zealand framework, educational institutions and systems alike have begun to understand the potential of the technologies associated with the Fourth Industrial Revolution to transform the way education is delivered as well as the education which is delivered. On the side of the education curriculum, countries are increasingly investing in STEM and seeking interventions to reskill populations with the new skills which will be integral to their professions – or future professions – going forward. ICT competency at some level is now a nearly ubiquitous requirement for employment, and in recognition of this, the curriculum is adapting to include ICT competency components at both basic and higher education levels.

The purpose of this paper is related to the second aspect: the use of 4IR and related technologies to enhance the delivery of education. Technological advancements, from predictive maintenance to profiling to adaptive learning programmes and platforms, hold huge potential to increase both the efficiency and relevance of education. This section will discuss three key aspects which can and are being leveraged to improve education delivery: platforms; interoperable data systems; and artificial intelligence. Case studies are presented in conjunction with theoretical and technical descriptions related to these concepts.

PLATFORMS

The platform economy has risen sharply in recent years. As with other opportunities currently arising based on technological innovation, the innovation itself is only part of the equation – what has enabled rapid industry change is, at least in part, the ubiquitous internet, cloud computing, blockchain and, most importantly, widespread and constant internet access enabled by devices such as smartphones and tablets.

Platforms, in many instances, form the core interactions between humans and technology and provide new methods of connecting consumers and providers. The rising influence of platforms is evident in sectors from hospitality to finance to transport. In education, platforms often form the basis for engagement in massive open online courses (MOOCs) and other forms of distance education. In relation to the complex undertaking of building a PSET digital ecosystem, platforms are one prominent means through which various stakeholders in the sector can be connected, for example, connecting matriculants to educational opportunities or tertiary-level students to labour market opportunities. In recognition of their potential as an engagement component in the digital ecosystem, this section of the paper explores platforms as a strategy and provides examples of educational platforms.

Platforms as a strategy

According to McAfee and Brynjofsson (2017, 137), platforms work on the principals of "free, perfect, and instant", an assertion which encapsulates the properties of digital transactions – whatever is consumed for a one-time purchase price is available instantly, anywhere in the world, and identical copies of digitised information can be freely made. While this accurately describes the advantages of platform-based inventory against brick-and-mortar stores, this understanding is dependent on the definition of a platform as "a digital environment characterized by near-zero marginal cost of access, reproduction, and distribution" (McAfee & Brynjofsson, 2017, 137). However, the primary focus of this definition is on platform services which create a digital replica of a traditional store to sell their own products, an exchange of digitised goods and services through digital interactions. While this is certainly one use of platforms, many of today's platforms could be described as another type of digital marketplace, one with the added caveat that the marketplace is created by an individual or organisation which does not own the products.

For other disruptive platforms such as Uber (and its many competitors) and Airbnb, "instant" remains a factor, while "perfect" and "free" as defined by McAfee and Brynjofsson lose meaning, due in part to the exchange of services rather than digital products. In such settings, the principals have more to do with price, convenience (of which "instant" might rather be viewed as one component) and value-add. The advantages Uber has over traditional taxis are the ability to call a driver from any location without pre-arrangement and no need to wait or walk to a common pick-up point – and in most places the cost is highly competitive. The advantages Uber holds over its similar competitors have to do mostly with value-add, whether perceived or actual: Uber provides thorough vetting of its drivers, handles disputes well, provides training for drivers, and so on. The implication for education is that any platforms developed must add value over defined or traditional approaches for all stakeholders, which may include students, parents, policy-makers, education providers, the public sector, etc.

Platforms work by essentially connecting products and services to customers. The crux of this connection lies in attracting a balanced supply and demand (suppliers and customers) and an effective filter – a method through which only relevant products are shown to a potential customer. This could be based on simple proximity, customer input or a variety of factors including both input and circumstantial data (Parker et al., 2016). In the PSET sector, relevant questions for creating effective filters might include: What are the relevant factors which should be taken into consideration in connecting people to and supporting them through educational opportunities? What are the relevant factors for access to labour market opportunities?

Effective platforms have revolutionised a number of industries, including retail and services, transportation, education, hospitality, and arguably, the nature of work itself. Examples include:

- Craigslist, a platform on which goods and services can be exchanged (retail items and services such as home repairs, etc.);
- Uber/Lyft/Taxify and others, which connect drivers to ride-seekers across the world;
- Airbnb, which allows individuals to rent out rooms or homes;
- Coursera and its competitors, which connect educational institutions providing short courses or modules to students; and
- UpWork, a site which connects individuals (usually professionals such as graphic designers, editors, writers, etc.) to those in need of their services.

These are all examples which in some way intersect with the "gig economy", where, from the perspective of a worker, prospects are temporary, skills-based and on-demand. This state of affairs is enabled partially by digital technologies which "deconstruct jobs" into component parts (Zarkadakis, 2018) and partially by platforms – in theory (and quite possibly in practice), an entire company could generate products without hiring more than a handful of permanent employees, relying on task- or job-based labour to construct or develop the necessary deliverables. In all platform-based labour, some portion of the profits go back to the owners of the platform, and the traditional obligations of protections associated with employment, such as decent working conditions, health and retirement benefits and so forth, fall away (Zarkadakis, 2018). These concerns are examples of places in which current government policy may need adjustments or additional consideration, as policy was not created with these types of contingencies in mind.

How platforms work

A simple example of one use of a platform is in advertising. While many will have realised the increasing frequency of "targeted advertising" or advertising based on an individual's prior activities and/or demographic factors, most don't realise that this is dependent on platform interactions which are invisible to the average consumer.

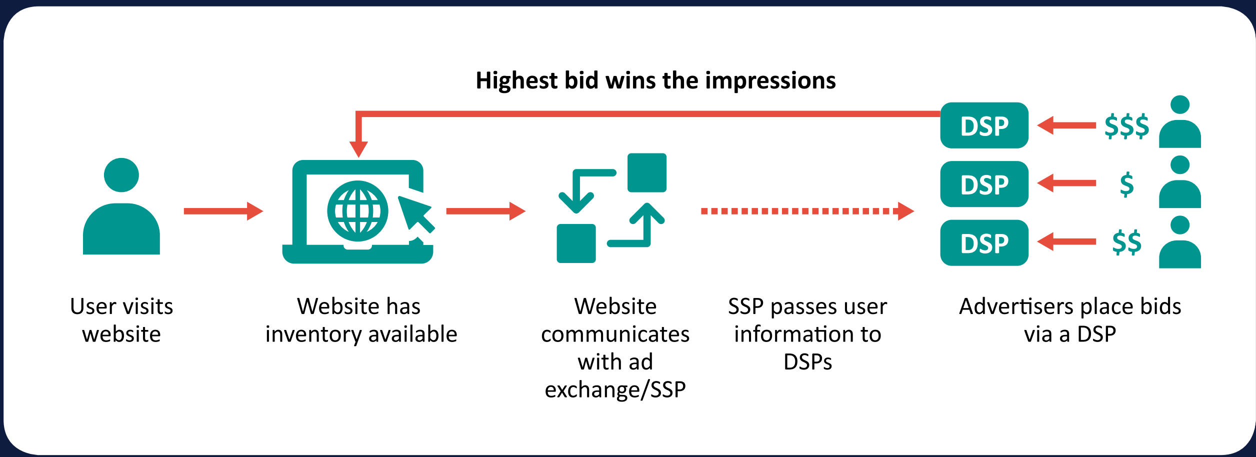
Demand-Side Platforms (DSP) and Supply-Side Platforms (SSP) are used to connect advertisers to a marketplace in which publishers list advertising inventory using a typical platform connection between buyers and sellers.

The key features of a platform industry are present in this example: Advertisers (consumers) and ad space providers (suppliers) are connected through independent third-party software, and the third-party software owns no products, only manages the exchange.

Initiatives of interest

In education, in addition to MOOC platforms such as Coursera and its competitors, governments are creating a different kind of platform to connect and guide education-seekers. This platform is not necessarily profit-driven but rather based on goals of improving economic participation and the relevance of workers' skills. These platforms can be developed by governments for use by the general population and/or leveraged by for-profit corporations.

Figure 6: Key features of a platform industry (adapted from Vrountas, 2018)



The AT&T talent overhaul

A good example of a platform leveraged by a for-profit corporation to improve workforce skills is the "talent overhaul" undertaken by the American company AT&T. The company is currently attempting a massive upskilling of their existing staff in conjunction with academic partners and largely through the use of ICT (Donovan & Benko, 2016).

The purpose:

The AT&T initiative was undertaken in response to changing labour needs: faced with the decision to either replace or retrain half of their staff and cognisant of the continually changing skills requirements for the field of telecommunications, the company chose to invest in current employees.

How it works:

Together with a restructure which broadened job specifications, increased performance expectations and reduced the number of job titles, AT&T invested USD 340 million over three years in employee upskilling, launching an online self-service platform for workforce development which includes virtual workshops and career planning tools; developing joint programmes with universities to fill reskilling needs; investing in employee tuition in formal and online courses; and enabling digital credentialing of workers (Donovan & Benko, 2016).

Lessons and relevance:

This initiative highlights a key feature of the new labour force – the ability and desire to learn is as important as a robust initial skill set, if not more so.

The talent overhaul initiative also highlights the importance of tighter collaboration between academia and industry, as it seeks to close a gap between education and training and work skills requirements. AT&T has commissioned universities to design programmes in response to the company's direct training needs and instituted a series of badges – essentially micro-credentials internally recognised by the company – in addition to degree programmes which may have wider audiences.

While the results of this workforce skilling experiment have yet to be publicised, the attempt demonstrates both the need for something like the PSET digital ecosystem initiative to support continuing development among workforce participants, a prominent feature of work in the era of the Fourth Industrial Revolution. Further, it provides insights into one method which can be explored to create tighter relationships between labour and education. Finally, this initiative speaks to the importance of platforms as a component of a more holistic vision and strategy.

AT&T is currently attempting a massive upskilling of their existing staff in conjunction with academic partners and largely through the use of ICT

The FutureSkills platform

The FutureSkills platform developed by NASSCOM launched in India in April 2019. The platform is a government initiative which connects individuals, with a focus on youth, to the new labour market through the identification of career pathways and, to some extent, the provision of relevant skills and credentials.

The purpose:

The FutureSkills platform is a direct response to the government's ambitions to increase its competitiveness in the international market in the realm of new technologies such as AI and robotics, cloud computing and related industries.

How it works:

The platform currently covers nine areas: Virtual Reality, 3D Printing, Artificial Intelligence, Cyber Security, Internet of Things, Robotic Process Animation, Big Data Analytics, Social and Mobile Platforms; and Cloud Computing. The Future Skills platform, therefore, has a direct focus on both career guidance and upskilling of the current younger workforce, up to age 35, although it makes no specific exclusions based on age.

The platform engages users in a "skilling journey", essentially a cycle of identifying career paths and the skills required for those paths, identifying current skill levels, learning from curated content compiled from a number of online platform providers (such as Coursera, YouTube, etc.), assessing proficiency through standardised testing and finally, certification and credentialing.

Figure 7: The FutureSkills platform Skilling Journey[3]

Lessons and relevance:

As a state-owned and financed initiative, the FutureSkills platform demonstrates one current international best practice in using 4IR technology to enhance the educational journey of citizens. The pathway, mapped from policy directives through learning to credentialing on the platform, provides insight into a key process flow which may be considered for the PSET digital ecosystem.

3 Adapted from http://futureskills.nasscom.in/

Data Driven Districts

"A standard makes complex systems run smoothly, coherently, and efficiently for all parties" (Janson, date unknown).

In 2012, the Michael and Susan Dell Foundation (MSDF) partnered with the South African Department of Basic Education (DBE) to launch the Data Driven Districts (DDD) initiative. The DDD initiative developed an easy-to-use, intuitive dashboard that allows education officials at a variety of levels to visualize South African School Administration Management System (SA-SAMS)[4] data immediately and graphically. DDD brings data collected from schools across the country to life visually, helping education officials to make informed and data-driven decisions.

The purpose:

The intention was to improve the way education data is collected in the basic education system and to ensure the translation of that data into clear, actionable insights for use by education officials.

How it works:

The majority of learner-level data in South African schools is housed in the SA-SAMS school information management system, which is owned by the DBE and provided to schools at no cost. However, extracting and consolidating this data at the provincial, district, and circuit levels has historically proven to be very challenging. Similarly, knowing what to do with the data has proved a challenge.

MSDF also introduced a validation toolset to measure and improve data submission quality. The combination of high-quality submission controls, immediate feedback, and measured and reported data scores has given administrators renewed ownership of their education data. The MSDF website currently says "so far, the DDD project has collected data from over 22,000 schools and more than 11 million learners each school term, with over 7,000 of these schools submitting updated data on a weekly basis. This is a strong start to reaching all of the country's 12 million learners" (MSDF, date unknown). It has been an impressive achievement in a short time. The MSDF measures not only adoption and reach, however, but also how timely and useful the data is as well as how clean it is, and also tracks changes in behaviour.

Lessons and relevance:

According to interview participants, some key lessons have been learned, including:

Getting started:

- Key objective/s must be clearly defined in advance, as must the problem to be solved. You must start at the strategic level.
- Time and energy must be invested to understand user needs.
- In a context where there are many users, it is challenging to identify what will add the most value.
- You need champions and advocates (although sometimes your advocates leave – and you need to continue to manage that).
- On-going stakeholder engagement is important, but especially so in the early stages.
- Super users, users who are really engaged, can help to define what good looks like and refine user requirements.

4 SA-SAMS, the South African School Administration Management System, is a government-provided school administration and management system

Challenges:

- The operationalising of a platform of this nature is long and slow.
- The biggest challenge is creating the capability for effective data management. In this regard, data standards for South Africa would be helpful.
- There is a need to continuously innovate and improve.
- The user community is critical to success – and innovation comes from users.

Government's role:

- Good leadership is required from government.
- Government's ability to operate quickly and take risks is constrained, while external funders can take risks and thus make ideal innovation partners.
- It is difficult to get government to take over complete ownership.
- National government relationships have proved more difficult to achieve than provincial ones.

MSDF warns that there are a lot of vested interests in both the private and government sector, and consequently, it is hard to build collaboration.

These platforms can be developed by governments for use by the general population and/or leveraged by for-profit corporations

INTEROPERABLE DATA SYSTEMS

Broadly speaking, *interoperability* is the ability of discrete computer systems or software to exchange and make meaningful use of shared data or other resources. In practice, this means that data can be used and processed in different applications by different users. Interoperability can also be defined as an approach which seeks to leverage data and networks to improve outcomes (Morales & Orrell, 2018), which can range from increased service delivery to social development, economic gains, etc.

Interoperability at the basic level of exchanging information has always existed in some form (Heubusch, 2006). The act of writing this report is an act of interoperability – data is being captured by a "system", in this case, a project team; the data can then be accessed and utilised by anyone who reads the report, given that the reader is able to utilise the same "standard", in this case, the English language. However, this is a highly constrained form of interoperability in that it requires human input for transmission and manual recording for further use. The interoperability of systems referred to in our current environment encompasses aspects of speed, consistency, access and use which go far beyond the constraints of interpersonal exchanges of information by leveraging technology and networks of organisations, individuals, government departments and other stakeholders with shared or overlapping objectives or needs. For example, interoperability of data exchange between soil sensors and cloud-based artificial intelligence software can yield predictive outputs which direct farmers to improve crop yields (Walsh, 2019). It is the achievement of this level of interoperability with which this paper concerns itself.

Today's interoperability is like electrical wiring – visible only by the outcomes attained. When a light switch is pressed, the light turns on, and we know the electrical wiring is functioning. When an individual opens an email app on their mobile phone, the emails they sent or received on their computer are visible – the two systems (the mobile phone and the computer) are interoperable and cooperating as desired, although the mechanism is, like the electrical wiring, invisible.

One of the advantages of interoperability is that it improves efficiencies and reduces error in data capture, processing and analysis of data by removing at least some of the manual contributions which would otherwise be necessary. The *.Stat and SIS-CC Strategy 2014 – 2019* developed by the Organisation for Economic Co-operation and Development (OECD) states that streamlined processes "reduce costs, shorten production cycles, develop single-source-publishing models or web-driven models, increase the capacity to innovate in response to users' demands, minimise manual operations and facilitate the creation of value-added products" (OECD, 2014, 6). A further advantage is that interoperability can reveal trends and patterns which would otherwise be invisible by increasing the number of fields and the number of data points available to a system.

For example, interoperability in health care could enable caregivers at a rehabilitation site to access relevant medical records from a primary caregiver at a hospital, reducing reliance on outdated technology such as fax, which requires manual re-entry of data received into a usable format for analysis and/or copious paper file records. Through tracking patient care records over the years from multiple practitioners, hospitals and/or hospital networks, trends in long-term symptoms from a specific medication may be identified, and the accuracy of frequency of symptom estimates may improve, outcomes which would not be possible with multiple, discrete, smaller datasets without significant effort. The aggregation of large data sets also enables new training and application opportunities for AI.

> Interoperability is the ability of discrete computer systems or software to exchange and make meaningful use of shared data or other resources

According to the World Bank Singapore Hub/Global Infrastructure Connectivity Alliance (2018), "all network centric concepts share the simple idea that information sharing is a source of potential value" which can be measured in terms of functionality, reliability, convenience and cost. Given these considerations, interoperability in its current manifestation is best described as both a technical and a business concept. The business aspects remain largely unchanged from historic patterns of information flow involving defined information, uses of the information and a human-centred purpose to the exchange. The technical aspects rest on the construction of the necessary infrastructure in the form of hardware, software and networks and the utilisation of data standards which define the form and structure of the information to be exchanged. Given its importance and potential in bridging data gaps in a complex environment like PSET, interoperability is viewed as a key enabler and one of the core principals which must be understood and pursued for the success of the digital ecosystem.

The following section discusses the types and levels of interoperability which must be considered, best-practice approaches for engaging in the development of an interoperable system, and challenges which must be overcome to achieve interoperability. In addition to frameworks, data standards are discussed as a core component of interoperability, and examples of organisations and collectives which have developed and use data standards are provided, together with learnings from these initiatives.

Frameworks for interoperability

There are multiple sub-descriptors of interoperability. Techopedia, an IT education website[5], describes two main types of interoperability: *Syntactic interoperability* is the ability of multiple systems to communicate and exchange data, regardless of whether or not they have shared programming languages[6] or user interfaces[7]; while *semantic interoperability* refers to the ability of discrete systems to understand and make meaningful use of shared resources by using common interpretations of data and services and common identifiers for individuals as well as for institutions. These common elements ensure the standardisation of data sets, based on the desired results defined by the users of the systems involved in the exchange (Shukla, Harris & Davies, 2010).

The Healthcare Information and Management Systems Society (HiMSS, date unknown) proposed four types of interoperability, essentially dividing syntactic interoperability and semantic interoperability each into two components. Syntactic interoperability is divided into *foundational* interoperability, which establishes the inter-connectivity requirements of the discrete systems, and *structural* interoperability, which defines the structure of data or standards used. *Semantic* interoperability is divided into a component of the same name, which speaks to the technical considerations necessary for the use of shared data, such as shared vocabulary, and *organisational* interoperability, which includes the non-technical considerations, such as the timeliness, accessibility and use of data by stakeholders.

The inclusion of organisational interoperability, in particular, is of paramount importance, as without a meaningful user interface and predetermined outcomes shared by all users, interoperability is meaningless. There is no point to electrical wiring if there are no lights in the house, or the lights are all inside one particular cabinet. In addition, use cases will define the auxiliary and supporting work such as presentations, test data and documentation necessary for data standards to be put into use (Open Data Incubator for Europe (ODINE), 2016).

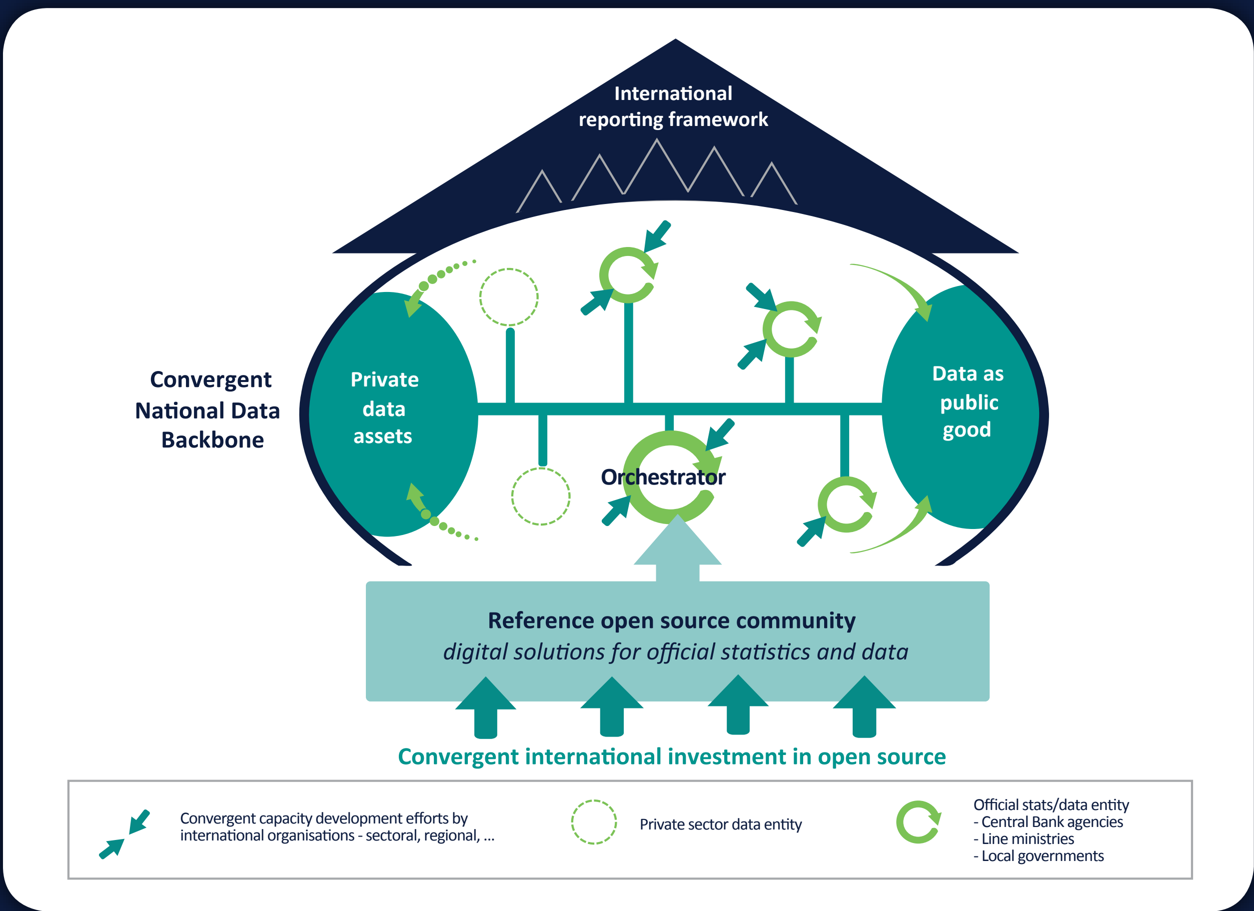
Goldstein, Gasser and Budish (2018) of the Berkman Klein Center for Internet and Society at Harvard University presented a similar framework, *The Data Commons Framework*. This framework makes explicit the layered aspects of data interoperability, beginning with technical infrastructure, the data itself and formats and labels associated with data – referred to as the narrow data commons. Built on these narrow data commons are broad data commons, that is, the organisational practices which encourage or mandate the required collaboration and stakeholder participation; the institutions, law and policy which manage the risks, benefits and availability of data; and the humans who feed into the development and preservation of other layers as well as make use of them (Goldstein et al, 2018). While the framework is not explicit in its division of types of technical interoperability, it does capture both the technical and the societal elements of an interoperable system, as well as the relationships between them, and serves as an important tool for understanding the hierarchies implicit in the development (infrastructure, standards and semantic architecture) and use (interactions between organisations, policy, broader society and individual users) of an interoperable system. The *Data Commons Framework* is displayed in Figure 8.

Another interoperability framework which presents these levels as they apply to a practical organisation is the *Statistical Information System Collaboration*

5 https://www.techopedia.com/
6 Programming language refers to a set of instructions which can be used to produce various outputs, used in computer programming.
7 A user interface refers to the intermediary (i.e., a keyboard, touch screen, controller, ATM, etc.) through which a human user interacts with computer software.

Figure 8: The Data Commons Framework (adapted from Goldstein et al., 2018)

Figure 9: National Data Backbone conceptual model (adapted from SIS-CC, 2018, 6)



Community (SIS-CC) model of the National Data Backbone, shown in Figure 9 (SIS-CC, 2018). This conceptual model was developed to address issues of operational silos and fragmented data ecosystems and to enable easy interactions with international reporting frameworks and requirements; therefore, the emphasis of this model is on the "broad" components, particularly organisations and institutions. The model calls for a central orchestrator, usually the National Statistics Institute, to coordinate data flows and access using the assets of a reference open-source community which adheres to a technical standard. The National Data Backbone conceptual model could be adapted to any complex system which operates with multiple organisations in a common ecosystem. Of particular interest is the movement and management of private data, an increasingly relevant aspect of data management and dissemination.

The SIS-CC (2018) notes further considerations in developing such a framework to be the cost of development and ongoing maintenance and support of user communities. Skills available are a further consideration. Component architecture enables countries, organisations and other stakeholders to use open-source assets while also tailoring and even creating their own resources, provided the necessary expertise is available within the various organisations to do so. Governance structures, as well as the roles and responsibilities of the national orchestrator, must also be considered and ideally mandated by policy, in the view of the SIS-CC.

Linking interoperability to the data value chain

As the *Data Commons Framework* implies, interoperability is closely linked with data processing and methods of data processing. Open Data Watch (2018) presents a framework for moving data from collection to impact in the form of a Data Value Chain, shown in Figure 10.

The Data Value Chain is useful in that it elucidates:

- The component processes involved in data collection from identification to processing;
- Publication processes such as analysis, release and dissemination of data, which can be comprised of various formats, from reports to policy briefs to charts and tables;
- Uptake of data by end-users such as policy-makers, universities and the general public;
- Impact of data, which speaks to how the uptake of data results in real change or advantage; and finally
- The importance of feedback mechanisms along the entire data value chain, which can take the place of formal monitoring and evaluation reports and statistics, user experience surveys, and so on, as appropriate for each link.

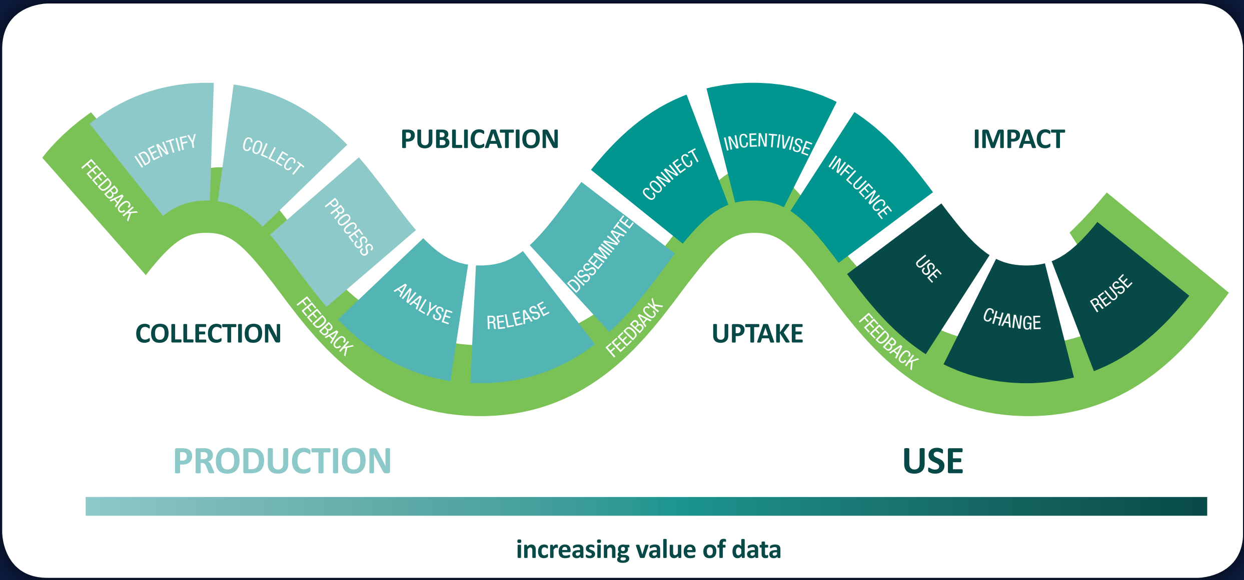
Interoperability from a technical systems perspective is primarily concerned with the production, the collection and publication of data, processes which are aligned to the "narrow" functionalities outlined in Goldstein et al.'s *Data Commons Framework* (2018). The "broad" functionalities which have to do with organisations, institutions and humans speak to the uptake and use aspects of the Data Value Chain.

> Interoperability from a technical systems perspective is primarily concerned with the production, collection and publication of data

Through interoperability, aspects of operations along the Data Value Chain can be improved. When interlinked organisations participate in discrete production methods:

- Data points may be collected multiple times by multiple institutions, leading to system strain.
- Inter-reliance on data requires data re-capture and results in inefficiencies.
- There is an increased chance of error and inconsistencies.
- Timelines are not always aligned, causing delays in dissemination.
- Uptake of produced and disseminated data can be ad-hoc across the system.
- Inconsistent user interfaces can limit beneficiary access.
- Macro-trends may be under- or over-stated.
- Uses of data points are obscure, raising privacy challenges.

Figure 10: The Data Value Chain (adapted from Open Data Watch, 2018)



Achieving even technical interoperability streamlines this process significantly, reducing the potential risks and challenges noted. While data collection, cleaning and quality assurance processes still must be undertaken, either at individual organisations or through a central agency responsible for this aspect, redundancies can be eliminated through planning, and system checks can be put in place on field values to assist in identifying non-conforming data. The data in aggregate can then be accessed through Application Programming Interfaces (APIs) to provide individual organisations or other beneficiaries access to the data they need from the central database or data lake. This has the potential to eliminate or at least reduce redundancies in collection and errors in data as well as provide timeous access to current data for multiple organisations with shared objectives or use cases.

Ensuring effective operational interoperability has to do with how the data produced is used. The possibilities for changes which can affect broad interoperability range from shifts in dissemination processes and access of beneficiaries to the creation of new departments specifically responsible for managing aspects of interoperability. One of the enabling aspects of both technical and operational interoperability is concerned with ensuring sound semantic interoperability and a robust semantic architecture.

Creating an interoperable system: The General Statistical Business Process Model

When speaking of interoperability, it is important not to divest the process of creating interoperable data and processes from the broader concerns of creating products from statistical data. The *General Statistical Business Process Model* (GSBPM) (United Nations Economic Commission for Europe (UNECE), 2019) is one framework for considering the steps inherent in the creation of statistics-based products, including a complex interoperable data system.

The GSBPM framework consists of eight primary phases: specifying needs; design; build; collect; process; analyse; disseminate; and evaluate, all of which contain sub-processes to be considered. Although the phases and

Figure 11: Interlinked organisations collect, process and disseminate data discretely

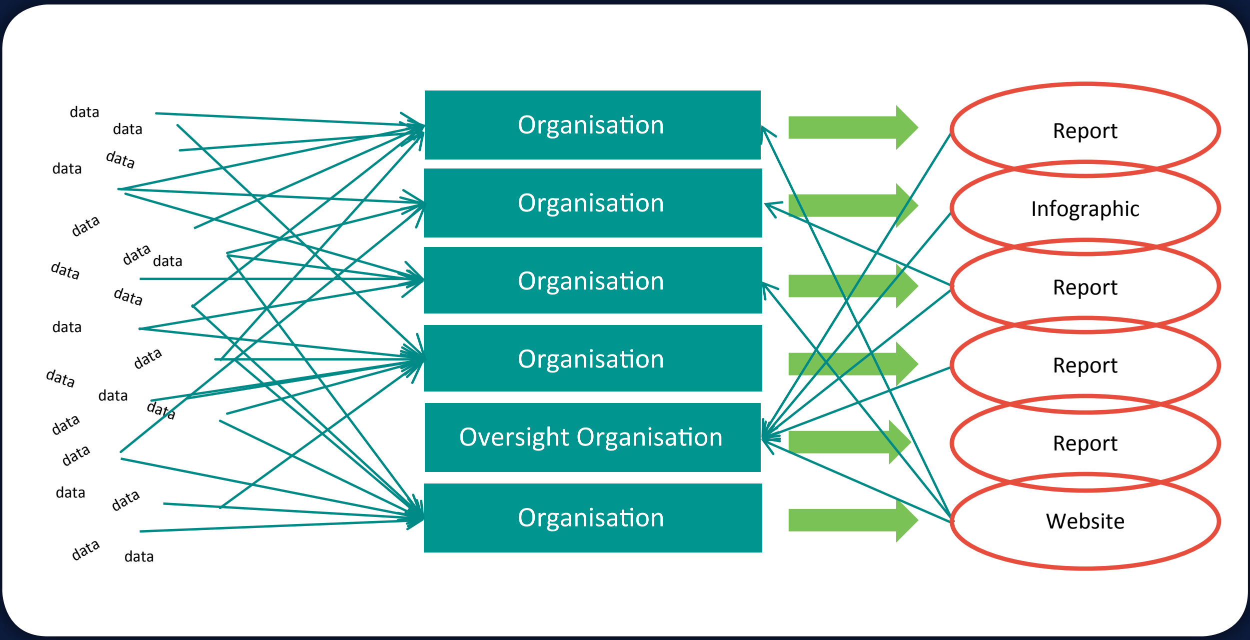


Figure 12: Data flow in an interoperable system of linked organisations

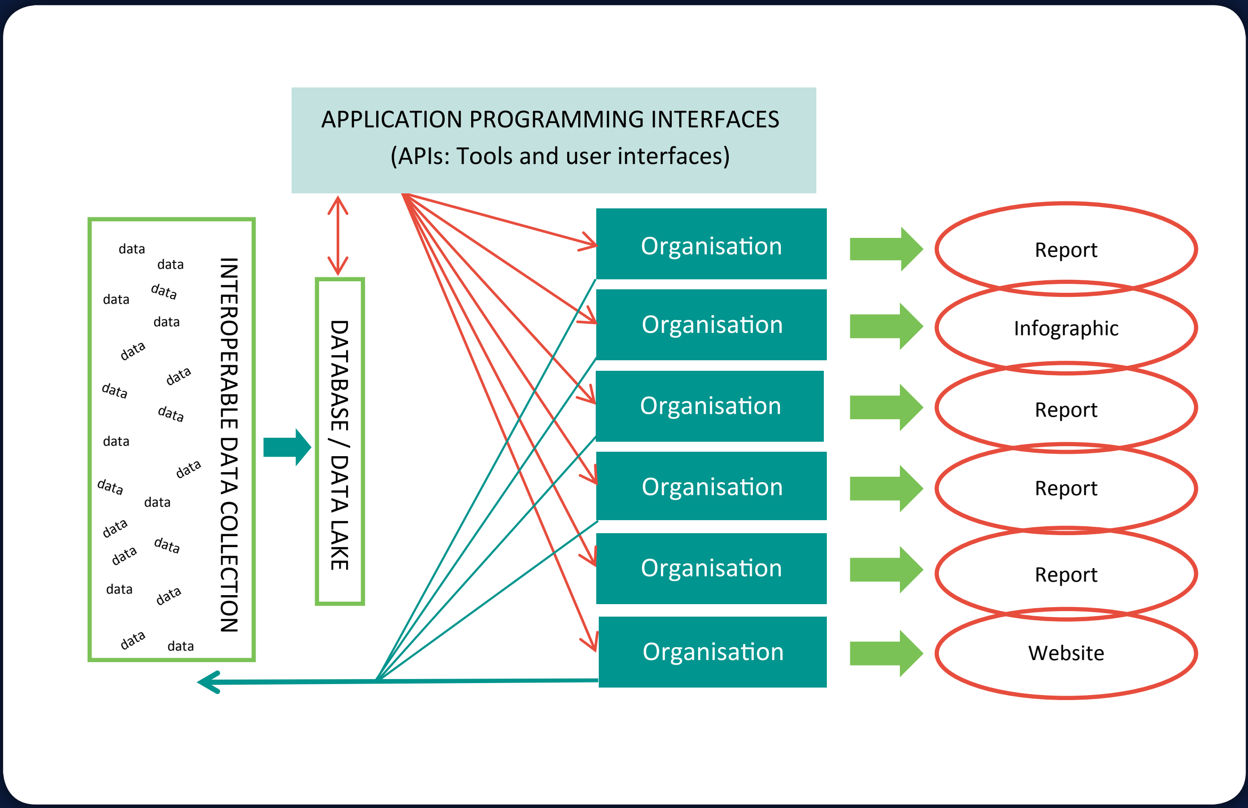


Figure 13: The General Statistical Business Process Model (adapted from UNECE, 2019, 10)

OVERARCHING PROCESSES							
Specify needs	**Design**	**Build**	**Collect**	**Process**	**Analyse**	**Disseminate**	**Evaluate**
1.1: Identify needs	2.1: Design outputs	3.1: Reuse or build collection instruments	4.1: Create frame & select sample	5.1: Integrate data	6.1: Prepare draft outputs	7.1:Update output systems	8.1: Gather evaluation inputs
1.2: Consult & confirm needs	2.2: Design variable descriptions	3.2: Reuse or build processing & analysis components	4.2: Set up collection	5.2: Classify & code	6.2: Validate outputs	7.2: Produce dissemination products	8.2: Conduct evaluation
1.3: Establish output objectives	2.3: Design collection	3.3: Reuse or build dissemination components	4.3: Run collection	5.3: Review & validate	6.3: Interpret & explain outputs	7.3: Manage release of dissemination products	8.3: Agree an action plan
1.4: Identify concepts	2.4: Design frame & sample	3.4: Configure workflows	4.4: Finalise collection	5.4: Edit & impute	6.4: Apply disclosure control	7.4: Promote dissemination products	
1.5: Check data availability	2.5: Design processing & analysis	3.5: Test production systems		5.5: Derive new variables & units	6.5: Finalise outputs	7.5: Manage user support	
1.6: Prepare & submit business case	2.6: Design production systems & workflow	3.6: Test statistical business process		5.6: Calculate weights			
		3.7: Finalise production systems		5.7: Calculate aggregates			
				5.8: Finalise data files			

their sub-processes contain interlocking dependencies and recursive loops, for ease of access, one common presentation is as a matrix, shown in Figure 13.

While the GSBPM does not provide answers to how interoperability can be achieved, it provides a series of phases in which interoperability should reflect. For example, data interoperability is only useful if shared needs of organisations can be identified: if there is no overlap in objectives, use cases or needs, an interoperable system may provide too little benefit as compared to costs. If interoperability is likely to provide shared benefits, then aspects of interoperability should be considered and included in the design and capacitation efforts related to phases, from design to dissemination.

If there is no overlap in objectives, use cases or needs, an interoperable system may provide too little benefit as compared to costs

Data standards

Data standards are the concrete outputs utilised to fulfil the requirements of achieving the structural interoperability or data interoperability foundational to the development of a complex interoperable data system, and they are utilised to ensure that data elements and their metadata[8] are reusable across the system. Data standards are, therefore, an important enabler of a system such as the planned PSET digital ecosystem. This section will discuss data standards as a part of interoperability and provide examples used across complex systems and in education.

The United States Environmental Protection Agency (EPA) defines data standards as "documented agreements on representation, format, definition, structuring, tagging, transmission, manipulation, use and management of data" (EPA, date unknown), and lists efficiency, access, transparency, comparability and consistency as benefits of employing data standards.

Three aspects of strong data standards are that 1) they are free and easily accessible to their communities; 2) data is not structured to suit one specific vendor, user or business model; and 3) they provide clarity on a system's objectives and primary use cases. No data standard can be all things to all people, so it is necessary to determine and prioritise the ways in which the standard is to be used and then construct the standard to meet those objectives (ODINE, 2016). Data standards must also be developed which can ensure shared representations of data and vocabularies across multiple contexts (Shukla et al., 2010).

Standards can be *crafted* or *adopted*, and one of the decisions which must be made is which avenue to pursue. Crafted standards, or "standards by nature" have the benefit of organised consultation led by technical experts and can be developed within a set time-frame. Adopted standards evolve naturally over a more extended period of time and may start with a few early adopters and then gradually gain wider use. The advantages to an adoption approach include that adoption requires less initial buy-in and is less likely to meet immediate opposition; on the other hand, putting off a formal process can lead to a lack of clarity or focus in the standard (ODINE, 2016). No matter which approach is used, the standard must be flexible enough that development can happen in iterative cycles and respond to increasing usage, new sources of data, policy and other changes in the environment or standard community *[Interview respondent, 2019]*.

The following are examples of organisations involved with standards:

- The **World Wide Web Consortium (W3C)**[9] is an international standards body engaged in the creation of the "semantic web", a massive undertaking with the goal of assisting machines to understand human language through the addition of metadata and using Natural Language Processing (NLP).
- The **International Organization for Standardization (ISO)**[10] is a global network of national standards bodies such as the South African Bureau of Standards (SABS). ISO creates standards for industries or other stakeholders using technical and area experts and consensus processes. ISO has a number of publicly available standards and is working towards international standards in a number of areas. To date, ISO has finalised 22,521 International Standards.
- The **South African Bureau of Standards (SABS)**[11] was created by the Standards Act of 1945, which was updated in the Standards Act of 2008. SABS oversees the development and sale of over 6,500 national standards in 36 areas. SABS was a founding member of ISO and is involved in the Southern African Development Community Cooperation in Standardisation (SADCSTAN). SABS does not currently engage in the education sector.
- The **Open Geospatial Consortium (OGC)**[12] is a non-profit dedicated to the development of open standards, or standards which are freely available for use. OGC currently has 13 national and regional forums which contribute from Asia, Oceania, Europe, North America and Northern Africa.

8 Metadata refers to the set of "auxiliary" data which describes a data point. For example, a photograph might have metadata which describes the day, time and location of the photograph.
9 https://www.w3.org/
10 https://www.iso.org/standards.html
11 https://www.sabs.co.za/
12 http://www.opengeospatial.org/

Standards and software of interest

Two examples of standards and one example of middleware[13] which support interoperability were included in the case studies as examples of best practice and approaches to ascertaining interoperability with regard to data structures. The initiatives outlined demonstrate different methods of achieving data interoperability, while organisational cases provide concrete examples of standard adoption and utilisation.

SDMX and .Stat[14]

The Statistical Data and Metadata eXchange (SDMX) standard is an open global standard used by UN agencies, the World Bank and many donors. SDMX-enabled tools can be reconfigured so that donors and stakeholders can access data automatically and independently, reducing the workload on staff.

.Stat is a suite of tools which use data recorded in the SDMX standard. .Stat is used globally by statistics organisations including the OECD, the UNESCO Institute of Statistics (UIS) and countries in Africa, Europe, Oceania and South America. .Stat provides options for internal data sharing within member organisations, machine to machine data exchange, access to data through a data browser and streamlining of data dissemination by allowing a range of publication options to operate from a single data source.

.Stat was made available to members outside the OECD in 2007 through bilateral agreements and is currently working off of a five-year strategy which concludes at the end of 2019.

The problem:

The mission of .Stat is to provide a set of tools which supports the data value chain through collection, processing and dissemination. The five-year strategy gives some insight into the objectives of .Stat:

1. Continually improve the ease and speed at which data and metadata can be accessed by users;
2. Open Data Dissemination and the development of an SDMX and open data strategy;
3. Create innovative methods for data dissemination which provide “richer and easier data experiences for users”;
4. More industrialised data dissemination and continuous streamlining of data dissemination processes, including improving metadata to enable more robust linkages between data and analytical content;
5. Implementation of and support for the implementation of international standards, specifically SDMX.

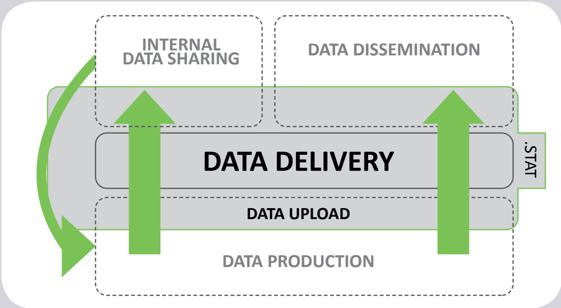
How it works:

Data is produced by individual members who therefore retain responsibility for and control of the quality and quantity of inputs. Data uploaded into .Stat can then be accessed by different levels of beneficiaries using a variety of processing and dissemination tools for internal use or broader stakeholders. .Stat beneficiaries are grouped into data producers and editors and data consumers. Producers and editors are usually internal stakeholders, while consumers may be internal or external.

13 Middleware is software which enables meaningful connection between discrete existing programmes (or between existing programmes and back-end data resources). It is often described as “software glue”.

14 https://sdmx.org/ Information on the education initiative can be found at https://sdmx.org/?page_id=8919

Figure 14: .Stat contribution to data processes (adapted from OECD, 2014, 12)



Each participating member of .Stat has its own brand and proprietary content, including ownership of their own data. For example, the OECD is the sole owner of the singular product, OECD .Stat.

.Stat can be located within the GSBPM under the areas of Build and Disseminate. There are three basic components to .Stat: an upload engine; delivery engine; and browser. The upload engine links organisational data production tools (such as Excel, Stata and custom-built tools) to .Stat. The data browser comprises the core comprehensive data exploration and download tool. The data delivery engine provides a range of data extraction services in standard formats and through APIs. As part of a future direction, the delivery engine will explore linked data services through the use of semantic web protocols to enable links between data and analysis. The core functionalities and contextual links are displayed in Figure 15.

The standard and platform is a community of practice (COP) based initiative. The COP in question, the Statistical Information System Collaboration Community (SIS-CC) was established by the OECD in 2010. The SIS-CC "was set up so that participating members could benefit from a broad collaboration, sharing experiences, knowledge and best practices, and to enable cost-effective innovation in a minimal time" (OECD, 2014, 4–5). The community plans a rate of growth of two to three members per year, which enables integration into governance and decision-making processes as well as sufficient mentorship and support for newly engaging institutions. For example, when the Tunisia National Statistics Office joined the .Stat community, one stakeholder noted that existing members were involved in efforts supporting robust data collection and modelling in the country.

In addition to such collaborations, community members contribute financial[15] and/or in-kind contributions towards .Stat, and tools which are developed by members are made available to all within the community. Currently, plans are underway to create more open access, but support for non-contributing members will be limited, particularly as opening access to non-paying members may result in faster rates of growth.

A workshop is held annually for community members and those investigating joining the community. The theme of 2018 was "Building National Data Backbones – Empowering Countries through Capacity Development and Technology". One of the core topics of discussion was a redesign of the frameworks involved in data creation. Previous conceptualisations relied on discrete departments undertaking individual solutions and capacity building, leading to fragmented production cycles and silos in operations which did not allow for true interoperability and its benefits. Additionally, the silo approach increased the level of complexity created at the top administrative levels in aggregating or synthesising information and resulted in a confusing user experience with the data, with even the potential for conflicting statistics on the same information. In addition, the emphasis on data production cycles in some cases led to decreased emphasis on the creation of data products, which may have resulted in sub-optimal or flawed data being disseminated. These factors, in addition to dynamics resulting from multiple stakeholders and capacity challenges, gave rise to a fragmentation of the national data landscape, as shown in Figure 15 (SIS-CC, 2018).

15 Financial contributions enable the development of coordination and support activities undertaken by the OECD, which also undertakes promotion and product development using internal funding (OECD, 2014:8).

Figure 15: .Stat interactions with SDMX and other data management processes (OECD, 2014, 13)

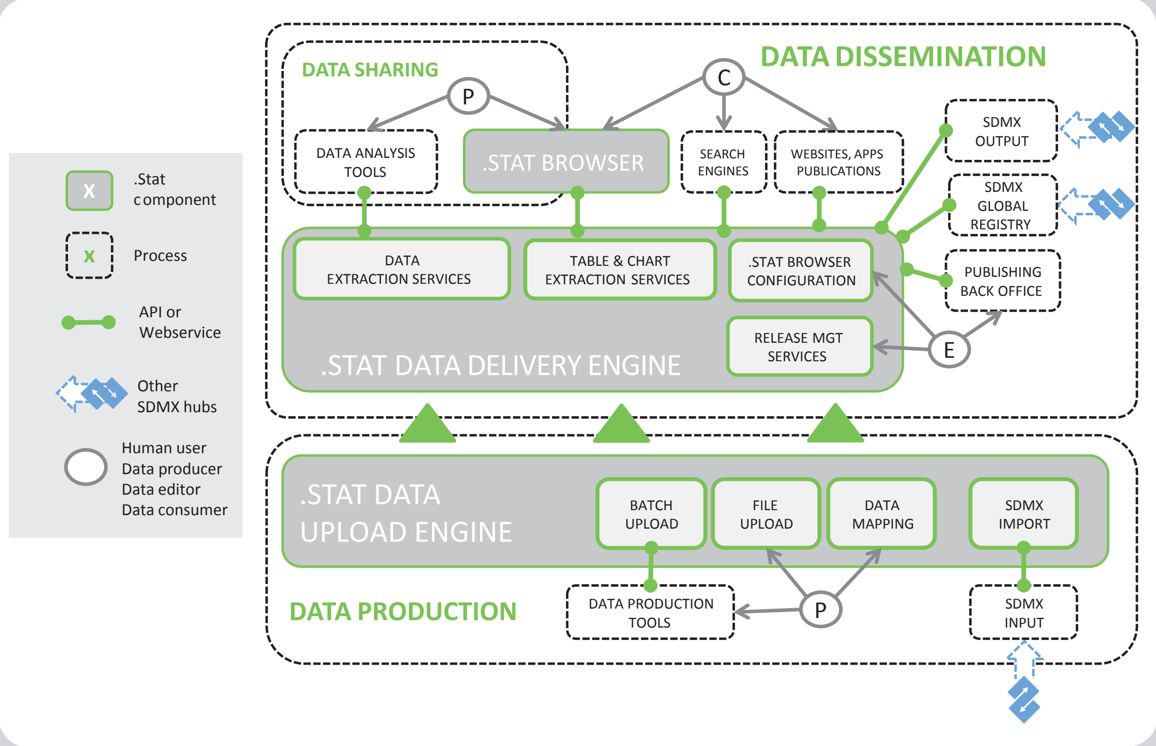


Figure 16: Fragmented data processes across an ecosystem (adapted from SIS-CC, 2018, 6)

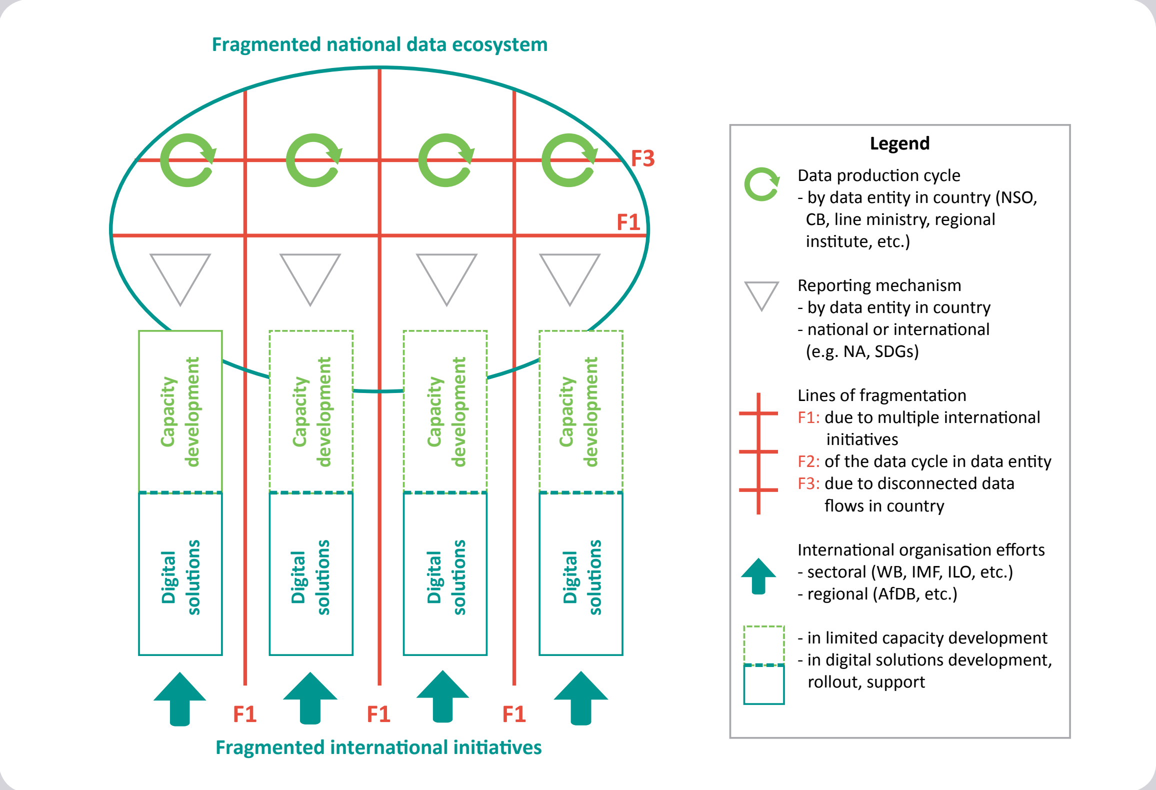
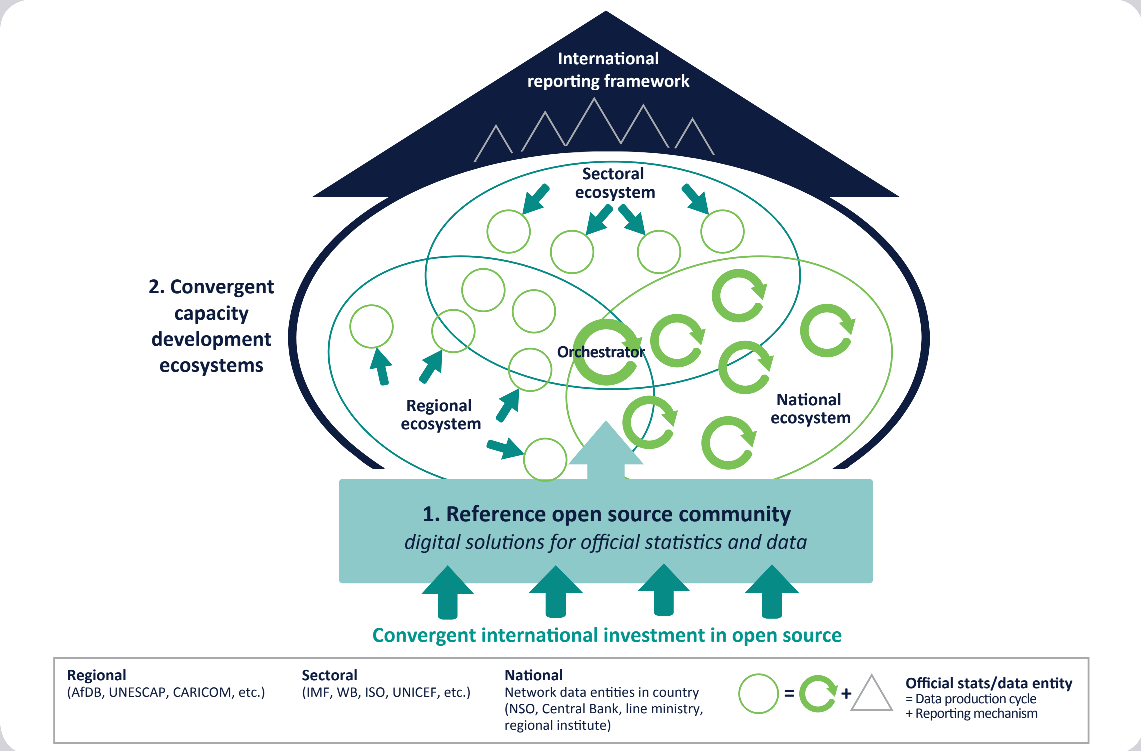


Figure 17: Leveraging a Reference Open Source Community for integrated data ecosystems (adapted from SIS-CC, 2018, 7)



In order to address this, the SIS-CC community engaged in the creation of a Reference Open Source Community to focus on pooling of resources, component-oriented architecture and coverage of the data cycle from creation of data products to dissemination. Component-oriented architecture allows for individual organisations to compile and leverage their own specific data platform suited to their needs, as well as to leverage international standards. The new model is presented in Figure 17.

Critical to the achievement of an interoperable system is an orchestrator, who plays the role of managing the information and its dissemination in each ecosystem and overall. As a stakeholder from the OECD indicated, the solution on the table is to have a number of producers of data, a smaller number of aggregating institutions and a central hub; the central hub both contributes to and draws from the Reference Open Source Community and is the core contributor to the international reporting framework.

SDMX and .Stat Use Case: UNESCO Institute for Statistics (UIS)

For UNESCO, the orchestrator is the UIS, which was described by an interview participant as having a “semi-autonomous relationship” with UNESCO. The semi-autonomous nature of UIS is important to maintain the reliability of data as well as the Institute's reputation, demonstrating the importance of the orchestrator operating independently from political agendas while still operating within policy frameworks and mandates. Interoperability is critical to the UIS, which exists to globally compare statistics and therefore needs the ability to merge sources of data without losing value.

A UIS stakeholder noted that in order to achieve interoperability, there is a need for a set of standards which deals with dimensionalised data or data points which have different associated dimensions (e.g. literacy rates may be tagged with the country and year). In the case of UIS, this standard is SDMX. A second and equally important requirement is the relationships necessary to access and receive the data as well as manage concerns about data use or the data itself, such as having lower literacy rates than a neighbour, and so on. These relationships are largely managed outside of the UIS in order to maintain its reputation and autonomy in its role as a statistics institute.

UIS data is sourced from different programmes within UNESCO, for example, education, science, culture and community units, and one of the achievements of the UIS was noted as the development of an automated capture tool which has improved data collection significantly. Each unit is responsible for its own data collection and processing to the point of upload to the system; to this end, each unit must have its own "orchestrator" who manages the collection and compatibility of data. On upload, tools look for abnormalities in the data and can flag some errors for re-entry; staff at the UIS also clean and dimensionalise the data at this point. Once the processing is complete and checked, the data is transferred to the dissemination database. The UIS manages two new releases and two refreshes of the database per year, which are scheduled as opposed to ad-hoc. UIS also responds to specialised stakeholders such as donors, the OECD and the World Bank through provisioning agreements which are built into their annual work plan and enable the UIS to provide specialised datasets or other outputs at a cost. This assists with funding the organisation and is one method of controlling data flows for multiple levels of access.

General releases and refreshes are made generally available to the public through .Stat and APIs which enable data visualisations and downloads and were developed to accept new data rather than being built on a static dataset – an important component contributing to later efficiencies. The cost of obtaining data through the UIS as well as the time necessary to change the data or dataset encourages users to rather make use of the API.

The dissemination team necessary to manage the system is small, with only six people: five with technical expertise and three managers (two have both skillsets). When necessary, external teams are pulled in for development or special projects through a partially automated human resources process. This team, which is considerably smaller than in previous years, is able to manage the same deliverables, in part due to the achievements in interoperability.

A UIS stakeholder noted the flexibility and shared tool aspects of the .Stat community as benefits, but also noted that in situations where mapping or centralised development is necessary, funding must be available in order to reach the priority list. The stakeholder also cautioned that in consideration of an interoperable system, funding for development is not a one-time cost, and ongoing maintenance and evolution plans are essential in the long term. The stakeholder further noted that custom-built tools may work well initially but are more difficult to maintain than community-based tools.

SDMX and .Stat Use Case: Australian Bureau of Statistics

A country-level example of an orchestrator is the Australian Bureau of Statistics (ABS)[16], which states on its website its purpose: "The ABS purpose is to inform Australia's important decisions by partnering and innovating to deliver relevant, trusted, objective data, statistics and insights."

ABS is mandated with a statutory role to collect national statistics relating to the population and business. This includes the national census, price indexes, gross domestic product and other information of this type. Currently, ABS collects and publishes its own data through surveys and interviews of businesses and households but is seeking to establish a more robust role across sectors through establishing partnerships across government.

ABS is currently in the process of a major transformation effort in order to modernise its processes and infrastructure, align further towards a user-focused approach to data collection, improve collaboration with partners and move further towards becoming a national hub for data collection. The effort includes an AUD 257 million investment.

16 https://www.abs.gov.au/

Members of ABS cited benefits of using .Stat to be the user interface, metadata and structures and the fact that APIs allow access to external users who can consume information directly without bureaucratic processes. The use of a standard allows member agencies to streamline reporting of the indicators they collect to UNESCO and other international bodies with whom they have agreements, and ABS was able to develop its own tools to fit unique and shared .Stat community needs for its use in-country. Additionally, stakeholders from ABS noted that it is not just a dissemination tool but enables them to manage data as well. The agency manages important issues of *privacy* by ensuring that confidentiality is achieved before the data ever reaches ABS, in accordance with restrictions on personal identification in Australian law.

ABS is able to control the flow of data to internal and external stakeholders through ensuring that published data is meant for public consumption while retaining internal data. Stakeholders noted that previous versions of .Stat had some challenges with regard to determining levels of access but that this is being addressed in new releases. As part of the setup of its involvement with .Stat, ABS retains ownership and control of its own data, which it had historically hosted locally. However, ABS is currently setting up a "private cloud" to host some aspects of its data. Like the UIS, ABS also manages the dissemination of microdata and other specialised needs such as those of researchers through various channels outside of the API.

ABS has a small team of two to three people who manage the database, supported periodically by three to four other individuals in different units in ABS. Teams of two to three also work with data owners to model and ensure compliance with standards and data structures. Two full-time developers built the system to allow for extraction of data and loading onto .Stat, and additional contract staff includes business analysts and support for the architecture of the system. Other personnel include one to two individuals concerned with firewalls, safety and security, which is a concern of the organisation. ABI staff noted that the organisation has a good track record of protecting its data but that remaining on the alert is of high importance.

ABS stakeholders noted that the use of .Stat is part of the data management process embedded in the General Statistics Business Process Model (GSBPM) and suggested mapping of all components in order to ensure robust systems of data and interoperability. They noted that decisions need to be based on the requirements of clients and that data collection has to meet clients' needs. In other words, the beneficiaries and their needs have to be determined and the entire process considered through a process of working backwards from those needs. One of the major challenges which they cited was when the needs of beneficiaries did not align.

Lessons and relevance:

A number of important lessons which should be considered in the development of the PSET digital ecosystem are highlighted in the example of the SDMX standard and .Stat community.

- A system orchestrator or a designated body is necessary to manage information and dissemination. The orchestrator should be exempt from political mandates but operate within policy frameworks and mandates.
- Release dates for data should be scheduled according to a work plan, with additional requests managed by an API and/or at a charge.
- The relationships between organisations using the standard are as important as the standard itself, and a community provides opportunities to leverage disparate strengths for shared benefits.
- Ongoing support and training of the standard and associated tools for community members are of utmost importance to maintain relevant and usable instruments.

- Development of standards should work backwards from the needs of participating organisations, with particular consideration to shared needs and benefits. Relationship management is often necessary when needs do not align.
- Designing, building and maintaining the standard is an ongoing process which requires a committed, long-term budget.
- Ownership and storage of as well as access to data at different levels need to be considered carefully in the design of a standard.
- Standards and good operating procedures can increase organisational efficiency.

The Ed-Fi Alliance[17]

The Michael and Susan Dell Foundation (MSDF) has done a range of important work with data standards, including establishing the Ed-Fi Alliance (Ed-Fi), an MSDF subsidiary in the United States. Ed-Fi was established to create a common language for education data that is relevant to instruction and student outcomes and to create a way to move data from disparate source systems to a single source of information that can be leveraged for integration, analysis, and reporting in order to empower educators with a complete view of every student.

The Ed-Fi Data Standard, therefore, serves as the foundation for enabling interoperability among secure education data systems. The Ed-Fi Alliance imperatives are:

- Unlocking data residing in multiple systems;
- Unifying data to provide a comprehensive view of each student;
- Providing flexibility to meet accountability and reporting needs;
- Creating data-rich insights for educators, administrators, decision-makers;
- Fuelling student-centred learning with current, comprehensive data;
- Fostering a thriving practitioner community and vendor ecosystem;
- Leveraging and sharing field implementation experience, expertise and solutions.

The purpose:

Ed-Fi was created as an answer to the following reported issues of data users:

- Inability of systems to deliver data that is timely, relevant and actionable;
- Data collection model imposes a significant burden on organisations and institutions;
- Lack of state-wide or country-wide standards for data systems;
- Difficulty of integrating data across data sources owing to lack of unique identifiers for learners;
- Cumbersome and inefficient reporting and analytics;
- Inability to easily access longitudinal data.

17 https://www.ed-fi.org/

The Ed-Fi Data Standard serves as the foundation for enabling interoperability among secure education data systems

How it works:

The Ed-Fi Alliance Open Community is characterised by the following:

- Ed-Fi technology is developed in collaboration with a community of educators and education and technology professionals.
- A community-sourced approach allows innovations in a single district or state to benefit students and teachers across the nation.
- Ed-Fi leverages common standards, technologies, tools and practices across a network of practitioners.
- Ed-Fi scales and directly supports schools, classrooms, and individual teachers and students.
- Community engagement is a key ongoing priority of the Alliance (regular communications, events, working groups, etc.).

The Ed-Fi Alliance also publishes free and open tools that integrate and streamline K-12 data systems.

Lessons and relevance:

When considering how to extend the alliance, the following has been learnt:

- Conceptually, extending the Ed-Fi Data Standard is easy: you simply add an entity or element, specify its properties and specify where the extension fits in the data model.
- Practically, however, extensions are challenging.

The MSDF has been careful to record lessons from case study research, and these are summarised below.

Figure 18: Three key lessons from case study research (Moffatt, date unknown)

A **Widespread adoption is often catalysed by major step-change adoption** or endorsements, and can require **years of sustained effort**

B **A clear simple use case** that solves real pain points, and **ease of adopting/updating a standard are critical**

C Standard bodies are **firmly established in their community**

MSDF also reports that in creating and delivering data-driven education solutions of this kind, the hardest and most expensive aspect has been gathering and aggregating disparate data from disparate sources.

"Middleware": OpenFn (Open Function)[18]

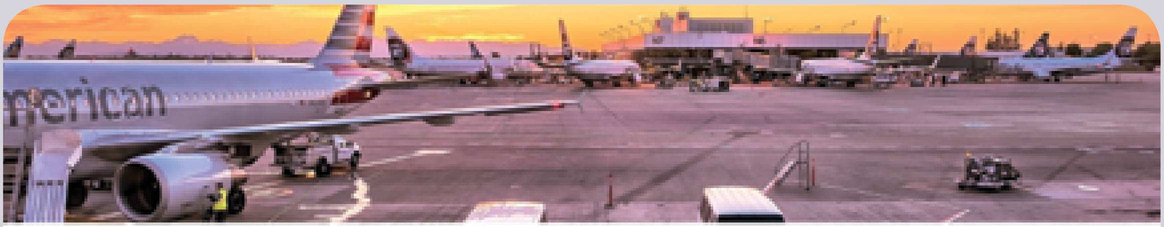
OpenFn is an associate of Vera Solutions[19], a technology company founded in 2010 which builds data systems primarily for non-profit organisations and services 300 clients globally. Vera Solutions' mission, as stated on its website, is "to amplify the impact of the social sector using cloud and mobile technology".

The purpose:

The development of OpenFn was undertaken in response to an observed need for intermediary software which could manage interactions between different tools and methods of data capture, essentially automating the process of reconciling information between systems. The aggregate results could then be pulled into and displayed in new applications or uploaded to shared databases for further processing. The stakeholder interviewed noted that with Vera Solutions' tools, OpenFn custom-builds architecture for clients arising from the clear need for a solution which can automate the flow of data between different devices and sources of data to facilitate the process of building or using interoperable systems.

A use case example given on the OpenFn website is of a commercial airline (OpenFn, date unknown):

Figure 19: OpenFn use case example (OpenFn, date unknown)



Helping an airport soar: Synchronizing asset data between the tarmac and HQ

An international airport, USA

The Problem

An international American airport that serves 18 million passengers and 180,000 aircraft operations annually manages an extensive network of assets, including heavy machinery, across a large area with intermittent network coverage. To comply with FAA requirements, the airport conducts regular walk-downs in which field staff rigorously document each asset and flag and photograph required repairs. The airport uses IBM Maximo EAM to manage assets and schedule repairs at HQ and CommCare to capture asset and work order data offline on the tarmac.

They relied on a manual process to transfer data between the two tools, wasting time and money and risking errors.

The Solution

The airport uses OpenFn to integrate the two tools. This integrated asset management system saves time, money, and prevents errors by reducing manual, paper-based data transfer between systems. It also enables field and HQ staff to each use the tool that suits them best. On the tarmac, field staff use an intuitive mobile tool to receive new maintenance orders from HQ, document assets, and report and photograph required maintenance. Updates to assets and required maintenance are automatically pushed up to HQ staff who can then use a leading industry tool to store asset data, schedule maintenance, and perform performance analysis.

18 https://www.openfn.org/
19 https://www.verasolutions.org/

How it works:

OpenFn engages in rule-based work, so routines need to be set up and programmed into OpenFn by individual users. OpenFn then responds based on timers or events to convert, cleans and loads data to a database. Data validity checks can be built in through field validation or validation between fields. The software also logs activities and the history and sends alerts if normal parameters are breached so that events or transactions can be replayed as needed for troubleshooting or amendments.

The platform can be run by an administrator who is trained, although technical expertise is necessary in the case of errors or routine maintenance, which must be planned for; the stakeholder interviewed recommended a consultant on retainer for such tasks and for initial set-up. Data on the platform as well as data created by the platform is owned by the platform user and can be hosted remotely in the cloud or on local servers. OpenFn creates a record of the data in order to check information and replay transactions. However, security is a concern for some organisations, and OpenFn can be set up to manage the interoperability work and remove the copy of the data as it flows through. Movement of data would likely have to be managed through security resources in cases where data is not hosted locally.

Lessons and relevance:

OpenFn, in itself, is not an interoperability strategy. However, middleware like OpenFn can potentially facilitate reconciliation of data sources as a part of such a strategy. This may be of particular relevance in environments where participating organisations and stakeholders may have low levels of data management and technical skills.

The initiatives outlined demonstrate different methods of achieving data interoperability, while organisational cases provide concrete examples of standard adoption and utilisation

Semantic interoperability

Interview participants highlighted that semantic interoperability is an often-underestimated component of interoperability. In linguistics, semantics is the term used to denote the study of meaning, a complex undertaking with various discrete and sometimes competing theories (see for example the description of computational semantics in Wilks & Charniak, 1976 and truth-conditional semantics in Davidson & Harman, 1972). Computer science uses semantics to refer to the outcome of particular operations, as opposed to their form. For example, the Python operation `x += 1` and the Pascal operation `x := x + 1` both achieve the same objective of adding 1 to x and storing the result in the variable x. So the syntax, or structure of the sentence or argument may be different, but the semantics are not.

Interoperability relies on a degree of both syntactic, but especially, semantic congruence. In order to create usable interoperable systems, common semantics must be used to describe aspects of the concerned industry or industries. Some of the tensions in this include regional or industry variations in the use of terminology: these variations can be addressed through the development or use of standard nomenclatures such as the Systematized Nomenclature of Medicine[20].

Beyond ensuring consistency of meaning, semantic architecture can also define relations between items, and proper development of semantic architecture enables machines to read metadata, resulting in, for example, searchable databases. Consistency of tags and nomenclatures is an essential part of ensuring that data collected can be leveraged meaningfully by user interfaces. The Credential Engine[21], one of the three case studies presented further on in this report, is an example of an initiative in higher education which is especially concerned with semantic interoperability.

Advantages of interoperability for mobility of student data

One critical key to creating optimal leverage of education platforms has to do with the availability and usability of learner data across a system and even systems. In Europe, there are a number of initiatives which aim to ensure data repositories of learner profiles and progression records that are accessible across government systems. A growing trend is the internationalisation of education and the concomitant requirement for interoperable datasets that allow for the required mobility of learning. Key platforms and organisations which are working towards this end include:

- CHESICC[22], the Chinese Credentials Verification Authority;
- Emrex[23], a Nordic initiative that allows students from Nordic countries to access and share educational credits earned in such a way that these individuals are empowered to control their own student data and exchange it across borders for various purposes; and
- My eQuals[24], a platform in Australia and New Zealand which allows universities to certify and upload digital versions of academic transcripts and graduation documents and which gives students, graduates, employers and third parties, anytime, anywhere, access to certified degrees and transcripts.

All of these platforms are focused not only on the idea of the learner owning his or her own data but also on ensuring the mobility of this data. Both Emrex and My eQuals are growing and expanding their footprints beyond their immediate geography, a sign of the growing commitment to this philosophy internationally.

- In addition, significant work is being undertaken in both national and regional contexts to ensure mobility and its underlying principles: credentials must be mapped for equivalency across contexts (including institutional, national and regional contexts) and systems developed which enable smooth transfer between institutions and, ultimately, labour markets. To this end, there are a number of international organisations, associations and networks which seek to enable or promote the aggregation or interoperability of particularly international datasets to support higher education

20 http://www.snomed.org/
21 https://credentialengine.org/
22 https://www.chsi.com.cn/en/
23 http://www.emrex.eu
24 https://www.myequals.edu.au

students and institution networks and mobility. These include: **The European Association for International Education**, "the acknowledged European centre for expertise, networking and resources in the internationalisation of higher education"[25] and which aims to serve the international higher education community.

- **The European Higher Education Area** (EHEA)[26]**,** an international collaboration of 48 countries which agree to and adopt reforms on higher education on the basis of common key values, including the free movement of students and staff: through this process, countries, institutions and stakeholders of the European area continuously adapt their higher education systems, making them more compatible and strengthening their quality assurance mechanisms.
- **Nuffic**, the Dutch organisation for internationalisation in education, from primary and secondary education to vocational and higher education and research[27].
- **The Credential Engine**, a non-profit which works to promote transparency in credentialing[28].
- **The East African Community**, an intergovernmental organisation of six member states which has formed a Common Higher Education Area[29].
- **The Groningen Declaration Network**, an international network focused on learner ownership of data and digital student data portability[30].

The last three formed case studies for this paper and were chosen as they represent both a geographic spread and various approaches to address challenges in student mobility and ownership.

> One critical key to creating optimal leverage of education platforms has to do with the availability and usability of learner data across a system and even systems

25 https://www.eaie.org/about-eaie.html
26 http://www.ehea.info
27 https://www.nuffic.nl/en/subjects/who-are-we/
28 https://credentialengine.org/
29 https://www.eac.int/
30 https://www.groningendeclaration.org/

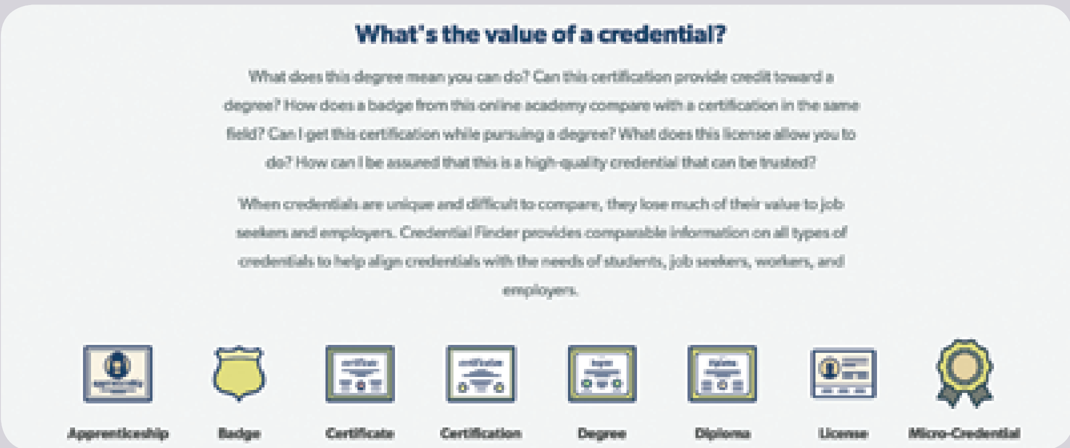
The Credential Engine

The Credential Engine is a non-profit based in the United States that seeks to improve transparency in the more than 330,000 available credentials in the United States, in part through collecting credential information using a common language and mapping the relationships between credentials. Transparency in credentialing allows degree-seekers to make informed choices, enables equivalency degree determination, assists employers to evaluate the relevance of a degree for the skills desired in the workplace, and also provides an avenue for a feedback loop between industry and academia. The Credential Engine is a collaborative attempt to standardise meaning across institutions and the broader community in terms of higher and ongoing education and create transparency in the types of credentials offered and their links to marketplaces. The Credential Engine further works to allow stakeholders to compare credential data in real-time as well as to curate credential data in new and more meaningful ways.

The purpose:

Even before the rise of platform-based educational opportunities, the extent and variety of credentials offered by and across institutions created a challenging landscape for degree and credential-seekers, exacerbated by disparate terminologies used at the institution, area or country level to describe identical credentials on the one hand, and identical terminologies used to describe disparate credentials on the other. For example, a credential labelled "data science" could refer to any number of time commitments and/or could refer to a set of learning outcomes quite discrete from another institution's interpretation of "data science". Therefore, determining the value of a credential is a more complex task than simply ascertaining the title of the credential and the venue through which it is delivered.

Figure 20: Credential Engine problem statement (Credential Engine, date unknown)



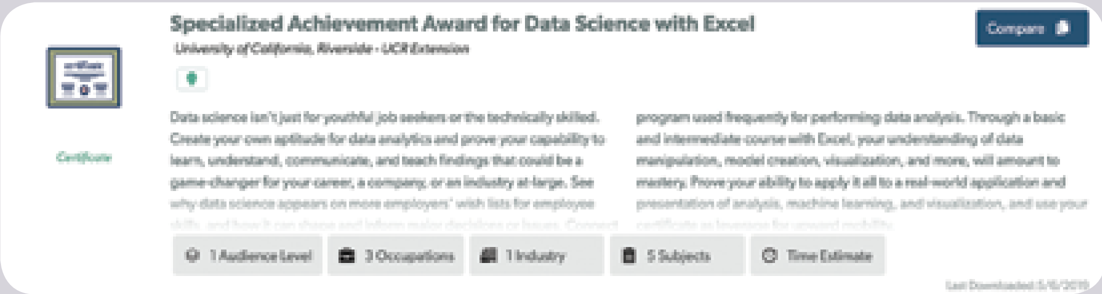
How it works:

The Credential Engine platform includes resources for both agencies and organisations which offer credentials (the Credential Publisher) as well as resources for those seeking credentials (the Credential Finder).

The Credential Finder has a search filter which is applied across the 6,000+ credentials and learning opportunities currently registered with the Credential Engine. An applied search will produce a series of results with a credential title, the offering organisation and a description, with tabs which provide additional information on the audience level (beginner, advanced, etc.), relevant occupations, industries, learning outcomes/subjects included and the time estimated for completion.

Credential seekers can select and compare credentials by adding them to a "compare" list, which then provides a side-by-side comparison including the description, audience level (beginner, advanced, etc.), relevant industries and occupations, estimated cost and estimated time to completion.

Figure 21: Sample result from the Credential Engine Finder[31] (Credential Finder, date unknown)



The Credential Engine also provides a framework for future development which will see the linking of career-seekers to relevant credentials based on interests, career plans and current aptitudes, and resources for employers to determine the credentials best suited to their employment needs.

The Credential Engine has leveraged a series of advisory committees for Higher Education, Business, Certification and Licensure, Quality Assurance and Technical Matters which provide oversight to different aspects related to the Credential Engine. There is a process of application to join advisory committees. In addition, smaller committees, which are open to interested members of the community, are leveraged for specific development procedures. In this way, the creation of the credential maps and comparative frameworks is undertaken in a participatory manner.

Lessons and relevance:

In addition to the construction of advisory committees, demonstrating one type of multi-stakeholder partnerships which could benefit the PSET digital ecosystem, the undertakings of the Credential Engine highlight the importance of ensuring shared terminology at the organisational, cross-organisational and systemic levels.

Accordingly, one consideration for building an interoperable data system rests on different types of semantic interoperability, one of which is the ability to transparently navigate through a complex field of PSET opportunities.

As more national and international opportunities for education (for example, distance education, micro-credentials and other forms of credentialing such as those offered by the FutureSkills Portal) become available to citizens and become more commonplace, the importance of initiatives such as the Credential Engine increases for both students and the labour market. The Credential Engine provides an example of a platform which seeks to both increase transparency in higher education and enable tighter collaboration between education and the labour market.

The necessity of such an undertaking is also relevant in the context of broader undertakings, both in Africa and internationally, related to the mobility of students and credentials.

31 https://credentialfinder.org/

East African Community and the Common Higher Education Area

The East African Community (EAC), which has six member states, has formed a Common Higher Education Area, a geographic area that brings together a set of geographically proximate sovereign states with a mutual interest in and common goals for the development of higher education.

The purpose:

The EAC's aim has been for the region to have "comparable, compatible, coherent and harmonised systems of higher education" (EAC Common Higher Education Area, date unknown, b). The EAC Partner States wish to share approaches to quality, criteria, standards and learning outcomes in order that they can promote student and labour mobility in the EAC. "The Treaty for the Establishment of EAC clearly stipulates the commitment of the Partner States to harmonise their education and training systems in order to provide for, among others, concerted efforts in the development of human resources, and mobility of people, labour, and services" (EAC Common Higher Education Area , date unknown, a).

Mobility is fundamental to the current transformation of higher education in Africa, the scope and scale of which is changing rapidly. The number of students who leave the continent to pursue higher education remains low, yet mobility within the continent, and within sub-regions in Africa, is growing rapidly. Beyond student mobility, the demand for higher education is pushing providers of PSET to cross borders and set up operations in neighbouring countries.

How it works:

The approach has borrowed from the Bologna Process, the series of ministerial meetings and agreements between European countries that ensured comparability in the standards and quality of higher-education qualifications. To date, the following has occurred:

- The harmonising of parts of the curricula of the education systems of the six EAC partner states with the development of a framework for the PSET sector currently underway;
- The identification of centres of excellence in the different member states and the facilitating of mobility among these partner states;
- Kenya has started to use a single education management information system (EMIS) number for a learner throughout the education system, and now the EAC seeks to ensure interoperable data standards in order to connect data sets on a regional platform.

Further, the initiative seeks to address the lack of data on mobility in higher education by setting up a data collection system for mobility data across the six member countries of the EAC and strengthening the capacities of member countries to collect, analyse and accordingly use evidence on mobility for evidence-based interventions in higher education. The system has the following planned outcomes:

- Strengthened capacity to collect and manage data on the mobility of students, programmes and providers in the six EAC Partner States;
- Strengthened capacity to produce evidence; and
- Strengthened capacity to utilise data on mobility for evidence-based policy interventions.

The programme will include the design of an interoperable Data Collection System for mobility in the region.

Lessons and relevance:

The EAC Common Higher Education Area provides a model for pursuing complex outcomes through stakeholder engagement and partnerships. It demonstrates the possible positive outcomes that can be achieved through political will as well as the strong potential for regional cooperation. Additionally, as a regional collaborator with a specific interest in post-secondary education, the Common Higher Education Area forms a potential future partner for the PSET digital ecosystem.

Groningen Declaration Network[32]

The Groningen Declaration Network (GDN), formally established in 2012, serves as an example of an international network focused on a non-binding agreement: the Groningen Declaration on Digital Student Data Depositories Worldwide. The declaration seeks to establish a "global area of convergence on digital student data depositories" (GDN, 2012, 1), with an emphasis on sharing best practice and moving towards worldwide exchange of digital student data. Focus areas include semantic interoperability, mobility of student data and promoting acceptance of digital student data in place of paper documents.

The purpose:

The GDN seeks to enhance student data portability and enable citizens world-wide "to consult and share their authentic educational data with whomever they want, whenever they want, wherever they are" (GDN, date unknown).

How it works:

Targeted participants of the network are centralised student data administration systems, with invitations also extended to other organisations with a direct connection to or significant association with the institutional owners of student data. These organisations include ministries of education, database managers, qualification authorities, quality assurance bodies associated with credentials and/or education, individual universities and research institutions (GDN, date unknown).

The philosophy that guides the approach of the network is that the learner is the central point of the data system and the rightful owner of his or her own data, with that data being a virtual learning currency that a learner can choose to utilise or not, with consent being the key foundational principle of the approach (GDN, date unknown).

Lessons and relevance:

In addition to supporting key principles such as individual data ownership and ethical use of data, the network has links with a range of international platforms with similar intentions regarding the portability of data and can provide endorsement and also, importantly, introduction to potential partners of this kind.

32 https://www.groningendeclaration.org/signatories/

Challenges in interoperability

In considering interoperability challenges, it is useful to leverage the aspects of interoperability as drafted by HiMSS: foundational, structural, semantic and operational.

In terms of foundational infrastructure for interoperability, the first consideration is connectivity. Interoperability can be built on local area networks (LANs) or wide area networks (WANs). LANs are a collection of networked devices generally within the same building, while WANs are a collection of LANs which can span a geographical location or multi-site entity. LANs typically share an information source such as a server and do not technically require internet connectivity to run; however, without connectivity, resources are limited to whatever is available within the network, and options such as cloud storage are not available. This can be a boon to security as, without external connectivity, access to files and information is strictly limited; however, the amount and size of data and files which will need to be stored in the network must also be considered, and this is not a suitable solution for systems which operate across locations or even different buildings. WANs require internet access of some type to function across LANs in various locations.

Internet connectivity is still a challenge for 3.9 billion people, largely in developing countries (McKinley, 2018). Beyond that, the cost of connectivity can also provide a challenge in some areas. The Alliance for Affordable Internet[33] provides statistics on broadband price trends and comparisons on the cost of internet relative to the average national income for 59 countries. Their analysis shows that in 2017, data costs in these countries ranged from 0.33% to 32.9%, and all ten countries with the most expensive data were in Africa[34]. On the individual level, inability to access internet connectivity can result in poorer access to resources and knowledge, inability to access new trade and industry tools, and even fewer job opportunities (McKinley, 2018). On the national and policy levels, connectivity needs to be considered in terms of the network and system to be established, with recognition of the fact that if not all components or stakeholders can access the system reliably and continuously, those that cannot will be disadvantaged.

In most cases, data will be externally hosted, and this requires another decision on the part of the user community. For-profit providers are available, but the aims, objectives and shareholders of a private provider may not closely align with a project's needs and goals; there is an implicit risk in private providers regarding sustainability and potential ethical concerns (such as privacy, ownership and use of data) which must be managed. Preferable options are dedicated project sites or neutral organisations (ODINE, 2016).

Access to data is another foundational challenge which must be mitigated. Organisations or other contributors to data ecosystems must be willing as well as able to share their data. Fields in which open data[35] is perceived as most valuable include education, environment, agriculture and healthcare (ODINE, 2017). However, interoperability in healthcare, in particular, has struggled as a result of stakeholders who are unwilling to share data due to fears of exposing themselves to liability and/or abdicating a competitive edge.

Structural challenges can relate to the data standard employed. To begin with, all parties have to agree that the information being shared across the system has value (ODINE, 2017). Additionally, the cost of converting existing data to a data standard must be weighed; if the cost of Extract, Transform and Load (ETL) outweighs the benefits of data sharing, the undertaking may be inadvisable. Cost-benefit analysis can be particularly difficult in situations where the return on investment is indirect, delayed or of a non-monetary nature (Cameron, 2011), strong considerations in the field of education.

In addition, a data standard crafted in the realm of a theoretical understanding of the desired outputs and outcomes may underestimate the practical realities of the data sets included. Datasets may be incomplete, biased or flawed in other ways, and provision and capacity must be available to deal with these discrepancies on a case by case basis (ODINE, 2016).

Semantic interoperability can pose its own significant challenges because the meaning of information is context-dependent and changes depending on its field, use, collection method or even organisation, and therefore semantic interoperability relies on rules which can govern multiple contexts. For this reason, semantic

33 https://a4ai.org/
34 The analysis is based on the lowest available cost for a 1G mobile broadband contract in each country.
35 Open data is discussed elsewhere in this report.

frameworks, or shared repositories of data models, are a key component of interoperability which, beyond dealing with shared vocabulary and formatting, also concern the ways in which information is linked using data elements (Shukla et al., 2010).

Additionally, operational challenges must be taken into consideration from the beginning of the project's conceptualisation. In many cases, *the biggest challenges in interoperability are based on people and adoption rather than technology*. In addition to data being interoperable, stakeholders on both the production and consumption sides of an industry must adopt the tools necessary to share data and be connected to allow for optimal flow and use of data (ODINE, 2017).

Logistical challenges are another concern which must be addressed. For example, Kurtz (2018) reported on an Education Week Research Center and Consortium for School Networking survey undertaken in the United States regarding interoperability in education systems, which found that the top challenges in improving software interoperability in school districts were budget, followed by a lack of widely agreed-upon technical standards. The survey also found that officials with concerns about staff level expertise were more likely to be in districts with higher poverty rates, and leaders in rural districts were less likely to feel that interoperability would help them achieve goals such as streamlining state and federal reporting.

ARTIFICIAL INTELLIGENCE

The third major opportunity which has opened as a result of advancing technology relates to the creation and use of artificial intelligence (AI). This section outlines the potential roles of AI in education, which centre on both the integration of concepts into the education system and, more importantly for this paper, the potential uses of AI in education to improve efficiencies and outcomes. While the focus of this review is on the creation of interoperable data systems, one of the benefits of such a system which should not be overlooked is the existence of a sufficient pool of data for both the creation and application of education-related AI. Therefore, this section provides a brief overview of what AI is, its current influence, the relevance of open data for AI and potential use cases of AI in education.

> The better the data available, the more accurately and quickly AI will be able to return the desired outputs

Much of the Fourth Industrial Revolution hinges on AI, robotics or some combination of the two. Given the premise that connectivity and interoperability between devices, or the ability of devices to share and use shared data, is achieved, the marriage of robotics and AI often results in what is referred to as the Internet of Things, or interlocking systems of devices capable of achieving outcomes or performing tasks without sustained human contribution or direction. For example, based on information from soil sensors, an AI software programme could determine irrigation needs, and an irrigation system could be cued by the AI to turn on for specific durations of time. The example includes both AI, in the form of a software programme, and robotics, in the form of both soil sensors and the irrigation system itself.

Robotics is a field which is often linked with AI and constitutes the most tangible point at which AI interacts with the physical world. Robotics is a branch of engineering which deals with the design, construction, operation and application of robots or machines which are programmable to complete a task or series of tasks autonomously or semi-autonomously. Robotics has a broad application in almost every field, and many of these uses do not require AI as the robot is programmed to complete a single or set of tasks. However, the introduction of AI into robotics has expanded the range and capability of robots, which can use sensory input mechanisms such as cameras or touch sensors to gather information for algorithmic processing, giving robots the ability to simulate a human decision-making process and act accordingly.

The Indian National Strategy for Artificial Intelligence Discussion Paper, published in June 2018 by the National Institution for Transforming India (NITI Ayog) (2018,7) describes AI as "the ability of machines to perform cognitive tasks like thinking, perceiving, learning, problem solving and decision making", and argues that thanks to "advances in data collection, processing and computation power", the uses of AI have far outstripped its original conceptualisation as a tool which could "mimic human intelligence". Now that AI is being increasingly put into practice, the full range and implications of machines which can mimic human intelligence, even imperfectly, are being made clear.

AI is augmented by technical advancements in data collection and storage as well as computer vision and audio processing which can be used by processing and inference engines to understand, analyse and react. Reactions can include searching, sorting and storing data, a linguistic response, making a dinner reservation or even physical movement.

Most recent AI relies on *machine learning*, or "the ability to learn without being explicitly programmed" (Samuel, 1959, cited in McClelland, 2017[36]). Machine learning involves the use of algorithms to segment and process data, learn from the data and return a meaningful result. Results can be descriptive, meaning they explain what has happened; predictive, meaning they explain the probable result; or prescriptive, meaning they return a suggested course of action.

Use cases of disruptive AI technology are as widespread as sectors themselves. In agriculture, AI is improving crop yields through real-time advisories, advance detection of pest attacks and market predictions. In retail, AI technology has been utilised to provide personalised, preference-based advertising as well as customer demand anticipation, inventory and delivery management. In manufacturing, the AI sector is robust in fields like engineering, supply chain management, production, maintenance, quality assurance and in-plant logistics and warehousing, and advanced robotics is decreasing the demand for low-skill labour. In the energy sector, AI is being developed to improve system modelling, predictive maintenance and forecasting to decrease unpredictability. AI can be used for traffic control, and for driving. In finance, AI is improving personalising customer interactions, providing low-cost customer assistance through chatbots, decreasing administrative costs by taking over rule-based back-office operations, determining eligibility and risk profiles of customers and producing credit scores by using bank history and social media data; and AI is improving wealth management through algorithmic trading and automated transactions. In education, AI holds the potential to customise learning, reduce administrative costs, offer more meaningful assessments and trace and predict learner outcomes.

Through applications such as these, AI is expected to increase rates of profitability by an average of 38% by 2035, leading to an economic boost of USD 14 trillion across 16 industries in 12 economies by 2035 (Purdy & Daugherty, 2017). Economic growth is expected to be driven by innovations and productivity-boosting technology (Frontier Economics, 2018) such as open data resources, data standards, and open cloud-based AI resources such as TensorFlow[37], an example in which AI democratises access to innovative and productivity-boosting technology (Access Partnership, 2018). However, the invention, diffusion and effective use of new technology are in turn likely to be influenced by other factors, including economic conditions, institutions and social conditions (Frontier Economics, 2018) as well as concrete investments in both digital and physical connectivity infrastructure (Lin & Rosenblatt, 2012). In other words, capitalising on the potential of AI in a sector and even across an economy requires a substantial investment in creating an enabling environment in order to support both the creation of AI and its appropriate use, an undertaking which requires attention to human capital, infrastructure and access to open data.

If policy, infrastructure and other conditions are sufficiently enabling for AI, it can serve as a mechanism for increased productivity as well as a key enabler behind innovative systems shift. For example, one of the most disruptive innovations of the 21st century is largely driven by AI, namely, platforms. As previously discussed, platforms are business models which connect customers to services more efficiently than brick-and-mortar stores – at times without even owning the products, as exemplified by two of the most highly valued platform companies, Airbnb and Uber (Parker et al., 2016). In South Africa, the health provider Discovery can be pointed out as a pioneer in the use of data and analytics to drive decision-making and customer offerings, tactics also adopted by supermarket chains.

Other selected use cases of AI include:

- In the health sector, improved patient tracking across institutions as well as improved diagnosis;
- In the finance sector, improved risk profiling and institutional efficiencies through automation of clerical tasks;

36 Note that while this quote is often attributed to Samuel, 1959, it is actually paraphrased from the article, "Some Studies in Machine Learning Using the Game of Checkers".

37 https://www.tensorflow.org/

- In manufacturing, preventative maintenance of infrastructure as well as improved efficiencies through automation of routine jobs and tasks;
- In agriculture, adjusted crop yields through forecasting of market requirements and automated responses to weather conditions;
- In transportation, the development of automated vehicles; and
- In marketing, advertising targeted to customer preferences.

The creation of AI which responds to situational contexts such as the use cases described above often requires not only the technical expertise to create the AI but also inputs from the sector in which the AI will be used. In most cases, large amounts of data are required to train the AI system. The better the data available, the more accurately and quickly AI will be able to return the desired outputs. Given the importance of data in the creation of viable AI-based products and the efficacy of AI in improving both efficiencies and outcomes, it is reasonable to expect an increasing emphasis on the creation of open data repositories by governments worldwide.

Open data

Open data is freely available to be used without restriction, and the creation of open data repositories is a key element of current policy discussions the world over, as both the prevalence and benefits of AI as an industry in itself and to improve the efficiencies of other industries becomes clearer. Governments, universities, non-governmental organisations, donors, entrepreneurs and corporates are all potential sources as well as potential users of open data.

In the education sphere, open data as a public resource with universal access, participation and transparency can help stakeholders understand the education landscape and has the potential to be a key resource for system improvement. Open data can be leveraged for the creation of educational resources as well as potential AI applications and is important to researchers and policymakers (both as evidence and in the design of effective education policies), as well as to parents and learners themselves, where it can provide evidence and guidance to inform choices on educational programmes and institutions. The EU's Eurydice report (European Commission/EACEA/Eurydice, 2017) emphasises that policy-makers need better access to data in order to address policy issues. The *UNESCO Handbook on Education Policy Analysis and Programming* (UNESCO, 2013) asserts that for policies to be implementable and effective, they must be based on reliable data. In short, open data can offer perspectives and a range of tools to further an understanding of educational problems and to support the development of solutions.

Once access is achieved, privacy and capacity building are crucial issues to address regarding open data. Educational data for research and policy purposes must ensure privacy for teachers and learners alike. With the availability of educational data, governments must prevent the identification of individuals and collectives and, in addition, also ensure the ethical use of this data by public and private stakeholders.

> Open data can be leveraged for the creation of educational resources as well as potential AI applications

Further, education systems must ensure the capacity is developed to both create and use data: "Data literacy – defined here as the ability to identify, retrieve, evaluate and use information to both ask and answer meaningful questions – is an important civic skill that forms the foundation of an innovative knowledge economy and increasingly data-driven society" (McAuley, Rahemtulla, Goulding & Souch, 2014, 89). A very real and concrete risk appears in situations where adequate capacity in data creation and use are not wide-spread, and this risk must be handled through both leveraging available external capacity and engaging in meaningful capacity development with multiple stakeholder audiences.

Challenges in AI implementation

Human capital poses the most serious concern in realising the potential of AI. The total global talent pool of PhD level AI researchers is estimated by Element AI to be 22,000 world-wide. If this pool is expanded to include lower levels of education, it may reach 300,000 (Element AI, 2018; Khan, 2018). Therefore, the labour market for AI-related talent is incredibly competitive, and the first 18 AI policies developed (Dutton, Barron, & Boskovic, 2018) largely focus on efforts to attract and develop national talent. Examples include financial incentives, the development of AI research centres, hubs and industrial parks equipped for the development of AI, and incentives or funding for the development of new courses or training schemes at universities. (See for example the Pan-Canadian AI Strategy, 2017; China's New Generation of Artificial Intelligence Development Plan, 2017; France's Strategy for AI, 2018; Japan's Artificial Intelligence Technology Strategy, 2017; and South Korea's Artificial Intelligence R&D Strategy, 2018. See also footnote 41 on page 53). At the same time, the shortage of AI talent is resulting in a majority of practitioners choosing employment at top firms rather than universities as both the pay and the working conditions are preferable in the industry environment (Boland, 2018; Sample, 2017). However, the repercussions are clear: the talent which would be necessary to rapidly skill AI professionals is difficult to draw to academia (Benali, 2018). Joint appointments are one solution to this challenge.

Beyond concerns with acquiring the necessary talent to create AI, there are also challenges pertaining to the use of AI. The report *Artificial Intelligence and Life in 2030: One hundred year study on artificial intelligence* by the Stanford University Study Panel (Stanford University, 2016) considered eight domains to be the most salient for AI-based disruption: transportation; service robots; healthcare; education; low-resource communities; public safety and security; employment and workplace; and entertainment. However, AI has influenced each of these domains in a different way, and the challenges that have resulted are different across the domains. For instance, challenges include: the difficulty of smoothly interacting with human experts (healthcare and education); gaining public trust (low-resource communities and public safety and security); overcoming fears of marginalising humans (causing unemployment); and the social and societal risk of diminishing interpersonal interactions (creation of a virtual world) (Stanford University, 2016).

In addition to these considerations, cross-cutting issues which affect AI policies and use include ethics, bias and safety (Brookfield Institute, 2018).

In the context of AI, ethics refers to debates about whether and when machines should be entrusted to make decisions and the extent to which the implicit values guiding those decisions reflect the values of both the relevant society and broader moral and behavioural norms. An example of a challenge of this nature is the use of autonomous military drones for combat: although the technology is available, it is unclear whether their decision-making in combat situations is an adequate moral substitute for a human operator. Another example would be how autonomous vehicles decide between human lives in the event of a collision (Brookfield Institute, 2018, 9).

> Human capital poses the most serious concern in realising the potential of AI

Bias can occur and even be amplified by AI when training data used to teach systems is skewed, incomplete, non-representative or contains its own forms of bias (Brookfield Institute, 2018, 9). Researchers have found that facial recognition AI, for example, displays both gender bias (Simonite, 2017) and racial bias (Tucker, 2017) due to factors such as the datasets used to train the AI and the technological capacity of AI. Predictive policing tools have also raised concerns about the amplification of biases in existing datasets on which algorithms are based (Lartey, 2016).

Safety refers to the ability of AI systems to operate without posing a risk or causing harm to humans (Brookfield Institute, 2018, 11). While the premise might seem simple, in reality, this is a complex issue which spans multiple areas. Some aspects of harm are easier to differentiate than others, and considerations about safety also often require specific legal frameworks such as liability, which can become increasingly complex as development chains for technology, AI and robotics span multiple companies. In 2018, a Tesla car operating on autopilot was involved in a fatal crash when the car sped up and crashed into a damaged road barrier.

Potential liability lies with numerous aspects of the technology which can be faulted and potentially with different developers, with the human driver who failed to retake control of the car, and with the government agencies responsible for road repair, all of which may have played a role in the sensors failing to register the barrier (Levin, 2018).

Beyond legal considerations, the technological advancements which enable AI also enable new forms of cyberterrorism which need to be considered and countered by relevant agencies, while at the same time AI advancements are underpinning advanced security systems such as AI-enabled biometrics and new forms of malware detection and protection such as Darktrace[38]. However, some critics caution that an over-reliance on AI in cybersecurity raises potential challenges such as the possibility of supervised learning systems being easily defeated by something as simple as relabelling code. And deep learning systems still have a "black box" factor in which the process of decision-making is not explicit. In the realm of cybersecurity, neither of these situations is ideal, and failures on the part of cybersecurity systems can lead to harm that varies from an individual's Ashley Madison account unexpectedly going public (Hosie, 2017; Panda Security, 2017) to ransomware demands (O'Flaherty, 2018) and even to economic instability (Pisani, 2018).

Broader social conversations about AI have to do with a larger conversation about benefit, and particularly equality of benefit. The *Indian National Strategy for Artificial Intelligence Discussion Paper* (NITI Ayog, 2018) discusses at length the necessity of a focus on equity and use of AI to optimise social impact and achieve development goals. In accordance, the purpose of the national strategy is stated as economic growth, social development and inclusive growth, with an underlying principle that "India's approach to implementation of AI has to be guided by optimisation of social goods, rather than maximisation of topline growth" (NITI Ayog, 2018, 5).

One of the main challenges in achieving this goal is that a majority of investment in AI is driven by the private sector (CB Insights, 2018b), as only nine fully-funded government AI strategies have been developed to date (Dutton et al., 2018). By far the largest drivers of AI adoption are compelling business use cases such as improved efficiency, accuracy, forecasting and decision-making leading to direct financial profitability; and in industries with far-reaching social consequences with delayed, unclear or non-monetary returns such as education and health, the "externalities from adoption of AI far outweigh the economic returns realised by private initiatives, and hence the role of government becomes pivotal in ensuring large scale AI interventions" (NITI Ayog, 2018, 22).

In addition, if economic growth is perceived as the sustainable growth of society, it is more difficult to anticipate the outcome of increased use of AI, and determining this will need specific measures to be put in place. It is possible that AI can drive both economic and social progress and help countries achieve national objectives like inclusive growth and development, but to achieve this, the technology must be developed in a way that is human-centred or focused on more than financial gain (Access Partnership, 2018). The conditions under which this is likely require both private and public investment and concerted effort on the part of government to ensure human-centred development.

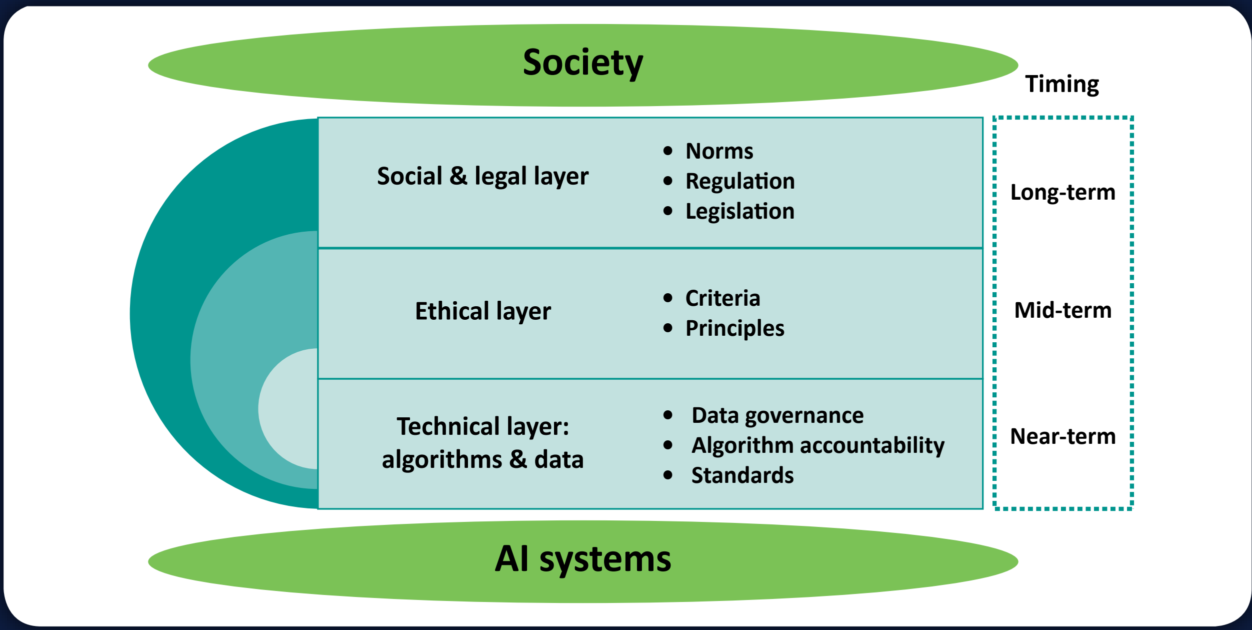
Policy responses to AI

Governments must and are moving swiftly in order to address both the rising prevalence and potential of AI and/or robotics and conceptual and concrete challenges which arise from their use. Government responses to AI include: direct policies which specifically govern AI-based technologies, such as driverless cars; indirect policies which apply to broader fields that aspects of AI can be grouped within, such as intellectual property laws; and relevant policies which govern areas of interest in which AI does or is likely to play a significant role, such as education, urban planning, welfare and health (Brundage & Bryson, 2016). Areas of interest include education, quality of life, economic prosperity, national security, healthcare, energy, transport, city planning, robotics, mobility, information security, environment, agriculture and life sciences (DTPS, 2018; Dutton et al., 2018; NITI Ayog, 2018: 22).

However, governance of AI is currently complicated by a number of factors. These include the fact that governments typically have limited expertise in AI and AI-

38 Darktrace Enterprise Immune System is a cybersecurity system based on unsupervised learning, or an algorithm which generates outputs without explicit search variables (Darktrace, date unknown). https://www.darktrace.com/en/products/enterprise/

Figure 22: Gasser and Almeida's Proposed Layered Model for AI Governance (adapted from Gasser & Almeida, 2017, 5)



related technologies; a lack of a broad social consensus on definitions, applications, strategies and tools for AI use, much less governance; contextual differences within and across nations; and the speed at which the field is evolving (Gasser & Almeida, 2017; Horvitz, 2017; National Science & Technology Council Committee on Technology, 2016).

Noting that AI governance is a complex undertaking and that AI governance models must provide for interoperability between both frameworks and approaches and apply to various contexts and geographies, Gasser and Almeida (2017) developed a layered model for AI governance, shown in Figure 22. The model considers the governance of the technical aspects of the AI itself to be the most immediate concern, with the addition of ethical, social and legal layers in the medium- to long- term.

It is proposed that technical aspects applying to the collection, use and management of data by AI algorithms should address responsibility, explainability, accuracy, auditability and fairness – the set of principles developed as part of a Dagstuhl Seminar[39] on Data Responsibility (Abiteboul et al., 2016).

The ethical layer should address "high-level ethical concerns" such as human rights principles (Gasser & Almeida, 2017, 6). The general principles for AI and autonomous systems (AS) of the Institute of Electrical and Electronics Engineers (IEEE) have been suggested as an alternative. Winfield and Halverson (2017) explain the importance of the IEEE principles, noting that they are built upon Isaac Asimov's "three laws of robotics" and the Engineering and Physical Sciences Research Council's *Principles of Robotics.* The IEEE principles outline a set of five high-level ethical questions and concerns:

1. How can we ensure that AI/AS do not infringe human rights?
2. Traditional metrics of prosperity do not take into account the full effect of AI/AS technologies on human well-being.
3. How can we assure that designers, manufacturers, owners and operators of AI/AS are responsible and accountable?
4. How can we ensure that AI/AS is/are transparent?
5. How can we extend the benefits and minimise the risks of AI/AS technology being misused?[40]

39 Academic seminars on computer science organised by the Daghstul Computer Science Research Centre in Germany. The seminars are conducted as retreats and are designed to "promote personal interaction and open discussion among researchers of international standing from academia and industry" (see https://www.dagstuhl.de/en/program/dagstuhl-seminars/)

40 The full standards document can be accessed at: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_general_principles.pdf

Finally, the social and legal layer is seen as the means to address the process of devising an oversight or regulatory body for AI (Gasser & Almeida, 2017).

In practice, immediate policy responses have started from the social and legal layer, with a number of policies setting up new ministries such as the Ministry of Artificial Intelligence in the United Arab Emirates or the Office of AI and the AI Council in the United Kingdom. In other instances, the responsibilities of existing ministries have been adapted to include AI (Dutton et al., 2018). While Gasser and Almeida's framework implies this as the most distant objective, the formation of such bodies is one method of mitigating a key challenge identified by Gasser and Almeida and others (see Benali, 2018): the fact that AI expertise is rare among the population and rarer in elected representatives. Locating AI within specific sub-bodies clarifies the responsibility for personal upskilling and identifying the necessary expertise, a logical step for large governing bodies. In addition to the setting up of new ministries and commissions, common strategies to address expertise in government include partnering with for-profit and educational institutions and the establishment of common centres, hubs and/or platforms to gather expertise[41] .

Public resources have been committed in a number of countries, including Japan, China, the United Kingdom and France, at the national and sub-national levels, as local governments also recognise the potential for growth in the technology sector (Dutton et al., 2018). There has been a significant increase in allocation of public funding for AI initiatives, including research and development, industrial and investment funds, and investments in related digital infrastructure (Purdy & Daugherty, 2017).

Importantly, policy should not be static, particularly in the context of an ongoing and developing industrial revolution. The speed at which innovations are created, markets are disrupted and social and economic changes are brought about requires agile and timely responses. Marume (2016) states that basic requirements of policy are adaptability and regular reviewing to ensure relevance and applicability. Šinko (2016) opines that there is a special place in issues of policy change that is occupied by external events (shocks), or what Šinko terms as "independent variables". These factors can include macro-economic circumstances (e.g. democratic transitions, economic and financial crises), new systemic government coalitions and natural disasters, amongst other things.

It may be argued that technology is also an independent variable, as seen in the developing countries which are on the receiving end of technological advancements and have to play catch up with the global North through policy and other measures to regulate technology. Marume (2016) states that policy is influenced by technological developments, population increases and urbanisation of the population, by crises, natural disasters, war and depression, international relations, economic and industrial development, policies of political parties, and research and investigations, inter alia.

Concluding comments on AI

While the challenges in utilising AI for the education space are far from insignificant and are hardly static, use cases for AI in education are emerging around the world and are tackling common problems. Some examples include:

- Predictive infrastructure maintenance to increase efficiencies;
- The use of big data repositories to target educational interventions;
- The use of AI-enabled gamification, personalised tutoring and chatbots to support individualised learning pathways;
- The use of algorithms to support career and education guidance initiatives; and
- The use of AI to map labour markets and related credentials.

The potential of AI-based solutions to support both system efficiencies and efficacy is an outcome of any construction of an interoperable data system dealing with large quantities of data, and related questions about development, access, privacy and use should be carefully considered.

41 See Dutton (2018). An overview of National AI strategies. Medium. Retrieved from: https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd for links to country AI strategies.

CONCLUSIONS AND RECOMMENDATIONS

This paper set out to determine enablers and inhibitors of interoperable data management and usage, give evidence of best practices and learnings in terms of approaches to designing, developing and maintaining complex data systems, and explore opportunities for partnership with existing systems or platforms.

INTEROPERABLE DATA MANAGEMENT AND USAGE

Interoperable data management is far more complex than the simple application of a data standard. In order for a system to be truly interoperable, especially one which has to do with PSET in a developing context, a number of critical factors need to be addressed.

First, a review of the technical requirements, including hardware and software access as well as human capital, should be undertaken with the relevant bodies and institutions within the sector in order to determine the logistical feasibility of an interoperable system. This is inextricably linked to further considerations in terms of approach: namely, the extent to which organisations can achieve foundational interoperability will assist in determining whether an adoption or crafted standard approach is more feasible. In situations with low foundational interoperability, an adoption approach can begin with a smaller pool and broaden as access increases.

Second, the selection of a data standard should carefully consider the needs of the organisations and the sector. Shared needs should be assessed, and use cases should be determined; only once this has been achieved can the question of whether to join an existing standard or create a bespoke solution be reasonably explored. The benefits of a bespoke solution include a crafted response to particular mandates; however, such an approach has an increased reliance on internal capacity and/or funding and/or the establishment of a new community of practice which may be in direct competition with existing communities. Joining a standard gives access to a community and its combined resources, and standards have different degrees of flexibility. Some standards and communities, such as SDMX, enable and share bespoke components added on to core components, providing a degree of flexibility as well as support which can reduce costs and build capacity.

Third, especially in the PSET sector, semantic interoperability becomes a key focus – the extent to which terminology is shared across institutions and credentials has severe implications for both skills development and labour market entry. The Credential Engine is an example of an initiative with a focus on semantics and skills and is in the process of creating a complex mapping between credentialing institutions, credential-seekers and employers: the Credential Engine platform provides opportunities to link between these beneficiaries in new and more meaningful ways.

Fourth, a focus on data creation is insufficient; for interoperability to be meaningful, organisational interoperability must feature. This speaks to both organisational capacity to deliver data, timeliness of aggregate releases and the ability of beneficiaries to make meaningful use of data. Data Driven Districts and .Stat are examples of initiatives which take aggregate data and reinterpret it for meaningful dissemination; key lessons learned from these initiatives include that a focus on the technical or narrow functionalities of interoperability is insufficient: for interoperability to be meaningful, significant development is necessary at the organisational level.

If properly achieved, the integration of a data standard and system interoperability can lead to improved system efficiencies, particularly with regard to reduced redundancy and discrepancy across the system. Other benefits are the improved use of data and improved decision-making by institutions such as government and policy-makers as well as by individuals, thus also achieving the broad functionalities of the Data Commons Framework (Goldstein et al., 2018).

Key recommendation 1: A focus on technical or narrow interoperability is insufficient and broader semantic frameworks and organisational interoperability must be considered. The purpose of an interoperable system is much broader than the aggregation of data. Therefore, from the onset of the development or implementation of such a system, factors such as the creation and use of data must also be considered, and the methods and means through which organisations and individuals will engage in the system should be defined.

BEST PRACTICE IN DEVELOPING AND MAINTAINING COMPLEX DATA SYSTEMS

The GSBPM, as discussed previously, provides an in-depth overview of the steps inherent in the creation of a strong statistics-based product such as an interoperable data system and consists of eight primary phases: **specifying needs, design, build, collect, process, analyse, disseminate and evaluate,** all of which contain sub-processes to be considered. Relevant findings in terms of some of these steps are presented below. In particular, recommendations focus on the establishment of a system, from specifying the needs to building the system and the analysis and dissemination of data.

Specifying needs and consultative processes

In terms of specifying needs, strong consultative procedures are recommended. Before the determination to even create an interoperable system for data management and use, shared objectives and the benefits of such a system should be defined. The cost against the perceived benefits should be weighed, with consideration for non-monetary and delayed returns. Costs related to upskilling institutions should also be considered. A business case, sometimes referred to as a situational analysis and project plan in the education sector, is a useful component of this stage which can clarify existing practices against proposed solutions to reach shared objectives. In addition, a governmental mandate can enable and provide leverage for discussions about system implementation. Therefore, it is important to secure leadership from government and engage on multiple levels, and the ideal advocacy strategy will include both top-down and bottom-up engagements.

Key recommendation 2: Invest substantial time in initial preparation. Initial preparation includes not only stakeholder engagement but also marketing and advocacy, and investments should be made in both initial and ongoing advocacy for and marketing of the benefits of the ecosystem or platform. Understanding the purpose and intended outcome of the ecosystem is vital at this stage.

Designing and building through agile development and collaboration

As discussed above, one key decision which must be undertaken is whether to rely on an existing standard and platform or develop a bespoke solution. One of the advantages of a bespoke solution is that it can provide a more “perfect fit” solution to organisational or sectoral needs, faster, through the engagement of experts and a more expert-driven approach. However, this can result in later challenges with buy-in and requires intensive upskilling efforts and generally continuous expensive redesigns if usage of the system expands to include more stakeholders, particularly if there are interoperability challenges between the bespoke system and those in wider use in the country, region or internationally.

Predeveloped standards and tools, on the other hand, may not be perfectly suited to organisational needs, and it is unlikely an off-the-shelf solution will offer all the required features relevant to local stakeholders and nationally determined outcomes. An *open suite with a component architecture* can provide the middle ground, in which components can be adopted, or not, based on the needs of individual organisations, and bespoke products can be developed as new components which are then available for others.

Key recommendation 3: Explore and leverage available systems. The benefits of joining an existing community, in most cases, will far outweigh the benefits of creating a bespoke system. However, this should be done subsequent to robust internal consultation in which the needs of users and the intended outcomes of the system are determined to ensure the needs of beneficiaries are met. It is notable that even in cases where specific objectives may not be reached, it is still likely to be more effective to leverage available education standards which are collaborative in nature and allow for individual API development on top of a standard or core offering. An open suite with a component architecture is thus recommended.

The design of an interoperable data management system should focus on identified user needs and particularly typical use cases, and the technology, from the interactive platforms and APIs to the data collected, should be built to enable and deliver what users want. The best platforms are built with an interaction-first, not a technology-first mindset; therefore, technology should be built only after understanding the interaction that needs to be facilitated for the participants. In particular, a minimum viable product (MVP) should take into account the most common or desired use cases.

Platforms can range in function and ability from the dissemination of data to the mapping of credentials to skills planning: the development of a data ecosystem for the PSET sector must, therefore, consider the intended outcomes of the data system and especially the needs of users in relation to these outcomes.

There is a wide range of education-related standards and national and international organisations devoted to the development of standards, and many operate on a community or open-source basis. Before creating a bespoke solution, carefully consider the relevance, costs and benefits of available options.

Key recommendation 4: Work backwards from the needs of users to design the system and its components, and allow ample opportunities for innovation to come from and in collaboration with users. In designing and building the system, agile development should be a primary focus and the system should be built in iterative cycles. A responsive system is adaptive and innovative and allows for connection, integration and collaboration. It is recommended that a customer discovery and validation process be undertaken in the first instance, where assumptions and hypotheses are tested.

Key recommendation 5: Continuously innovate through innovative cycles of development. The strongest systems will allow for innovation to come from a broad community and will encourage and reward the sharing and open source nature of tools developed by community members. At the same time, it is imperative to not only rely on community members but to ensure that ongoing maintenance and redesign are adequately budgeted for and/or funded on an ongoing basis.

Together with design, attention must be paid to collection instruments, methodologies and timelines as well as dissemination methodologies and timelines. Metadata descriptions are an additional component which must be considered, and means must be determined to deal with semantic interoperability between component parts of the system. In addition, development must take into account future additions and period data influx and must be built to accommodate these changes without disruption. The strongest systems will allow for individually developed APIs so that users with unique needs can leverage the data for unanticipated purposes (within the limits of ethics). Tiered access systems, with different specifications for different users, are appropriate for the education system. Building the system should not be viewed as an event but as an ongoing component (and cost) of the system.

Key recommendation 6: Plan for future updates, additions and adjustments to data. It may seem obvious, but it is of paramount importance to consider the ongoing collection of data required by education systems, a consideration which may be overlooked by technical experts. Systems should have a sound and agreed-upon semantic architecture, and whatever tools/APIs are associated with the project should have the functionality to manage updated or replaced datasets or databases without a redesign. Leveraging international and/or national standards can significantly reduce the initial build time and cost.

The collection and processing of data is a concern with complex systems, and quality-assurance procedures must be robust. Most systems utilise a central processing organ (such as a Bureau of Statistics) to manage the central collection, cleaning, validation and so forth of data. Significant capacitation efforts are often necessary regarding the submission of data and data standards in use; intermediary software can sometimes assist with automating the adherence of standards to systems.

Key recommendation 7: Define clear parameters for ownership and use of data, platforms and data standards. Concerns about ethical access and use of data must be resolved. Consideration must, therefore, be given to ownership and access to data as well as data portability. It is advisable that individuals are positioned as the primary owners of their data, and systems must work to ensure the portability and beneficial use of such data.

In addition to data, ownership of platforms and data standards developed as well as assignment of associated roles and responsibilities and levels of access must be determined.

Open source data and shared software are an advantage many international systems seek and can be leveraged for purposes which range from evaluation to innovation, particularly with the current trends in AI and its development. However, human-centred system development is paramount; therefore, even above the needs of the organisations involved in the system, the needs of individuals, particularly students, must be taken into account; and further, systems and processes must adhere to national legislation regarding privacy and related concerns.

Therefore, questions concerning the ethical collection and management of an individual's data as well as the ways in which individuals in the system will directly benefit from the collection and use of their data must be taken into account. The GDN is an example of an international network which has undertaken deep consideration of these issues, and members have committed themselves to a philosophy in which the student is the central point of the data system and the owner of her/his own data, a virtual learning currency which the student may use or not and that may only be shared with the student's consent.

Analysis, dissemination and use of data

Analysis of data can be undertaken as part of the functionality of platforms and/or APIs; for most users, the output is more relevant than the process of analysis. Dissemination can take place on multiple fronts, with the primary being large-scale, system-wide releases managed by a central agency, and the other being the ability of discrete users to pull relevant information through an API. Mass dissemination should be regularly scheduled and timelines adhered to; it makes sense for these to be aligned to national and/or regional requirements in terms of reporting frameworks.

Key recommendation 8: Schedule data releases, encourage the use of APIs and charge for custom requests. Discrete requests from a central management agency can be accommodated, but these should require a financial commitment to prevent abuse and further encourage use of the available APIs and platform functions.

Platforms such as Data Driven Districts or .Stat produce the results of an analysis, updated and disseminated at standardised intervals to relevant beneficiaries who have tiered access to the results. In this way, platforms with robust, user-focused interfaces and user-friendly APIs can make data more accessible and assist with the flow of data between and to beneficiaries. However, this does not eliminate the need for skills capacitation in both the system and the broader underpinnings of data interpretation and the use of data for decision-making. In short, capacitation is necessary not only to ensure robust and aligned data submission/collection, but also in terms of the utilisation of data.

Key recommendation 9: Plan for investments in capacitation for all beneficiary groups. In the PSET system, this includes government departments, PSET institutions (including higher education institutions, qualifications authorities, sector authorities, etc.), labour market and work-integrated learning representatives, PSET staff and students.

In the reviewed standards there is a central organisation (generally the unit or body responsible for statistical processes) which ensures compliance to the standard and handles ongoing maintenance, data processing and regular releases of aggregate data as well as special requests for data. Coding, standards as well as data management processes form an integral part of the core capacity of these organisations, although additional support is generally brought in for specific development tasks.

Key recommendation 10: Consider from the onset who will hold a system developed long-term. This body may be different from a governance structure and must be sufficiently capacitated and funded. It is more efficient if the unit can leverage additional resources and expertise for specific development tasks. Notably, a central managing organisation does not have to assume ownership of the data or the outputs of the data; in some cases, the organisation is responsible for only the development process.

The discussion on dissemination refers to one aspect of the overall strategic approach to the initiative: Who will control and who will manage the data collection, analysis, dissemination, evaluation and use? If a central agency is leveraged, at which points do data enter the ambit of such an agency, and at which point does the agency relinquish responsibility?

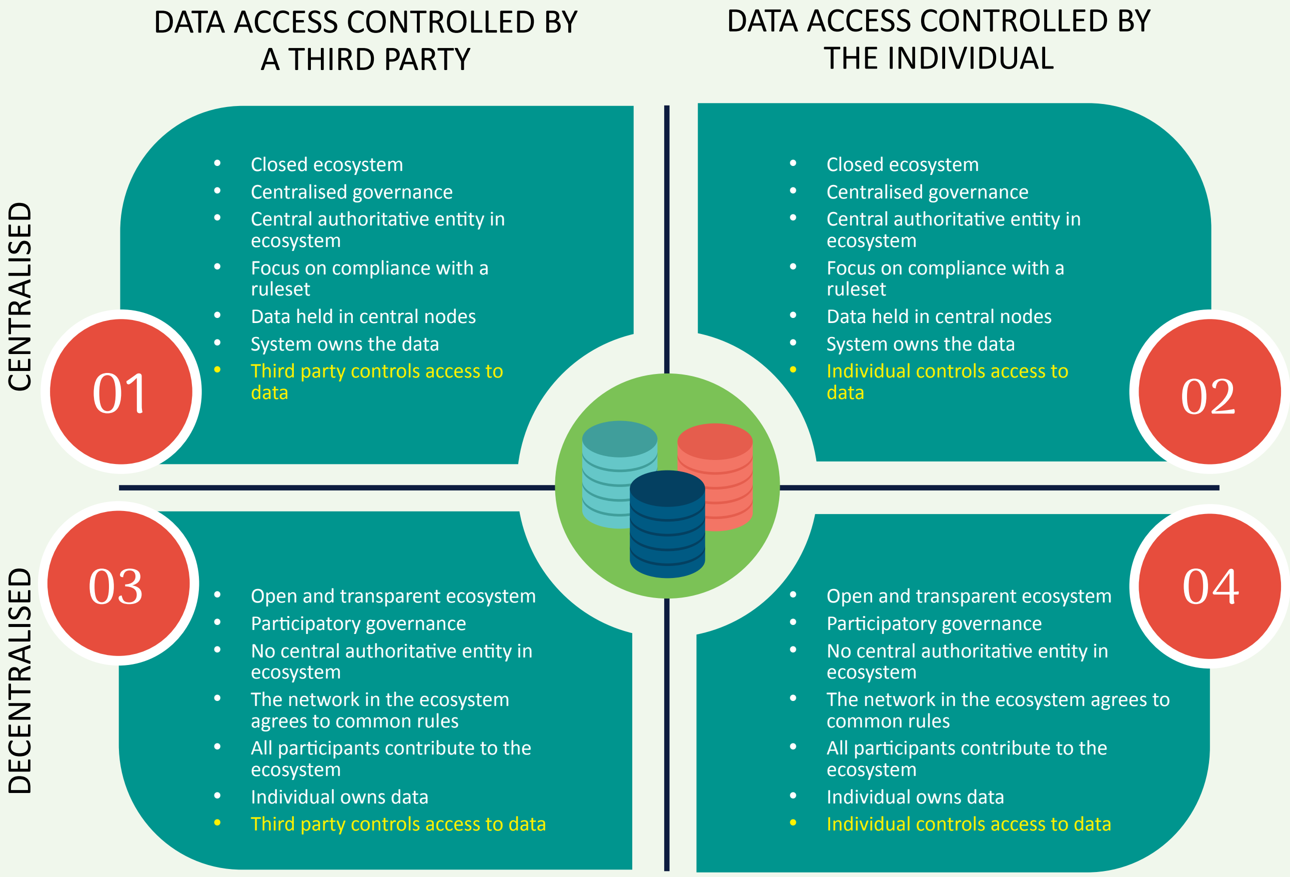
In order to address these questions, it is useful to consider the features of centralised and decentralised data ecosystems. A centralised data ecosystem is closed and has centralised governance. There is an authoritative entity in the ecosystem which ensures compliance to a ruleset. Data is held in central nodes, and the system (or an entity within the system) owns the data and possesses full control of managing the system. This approach is often easier than a decentralised approach in that capacity can be largely centralised, and the ecosystem is, therefore, more efficient to construct and run. However, transparency becomes a primary concern and, depending on the extent to which data usage is restricted, a centralised data ecosystem may limit the uptake and use of data and thereby eliminate many of the benefits of an interoperable system. In addition, while a central agency can act as a quality assurance body, significant capacity is required within the central entity to collect, process and disseminate data, particularly if usage is restricted in such a way that the entity is required to process data on demand from stakeholders.

The possibility exists for a system to operate without a central managing agency, with governance structures determined by the individuals within the organisation and maintenance occurring through organic revision led by direct users: aspects of the Credential Engine and the .Stat community function in this manner. This sort of decentralised data ecosystem is open and transparent and has a participatory governance approach. There is no central authoritative entity in such an ecosystem. Instead, the network in the ecosystem agrees to common rules and governance structures, and all participants contribute to the ecosystem. In a decentralised data ecosystem, the individual owns her/his own data. This approach ensures a distribution of power, access and control away from a central entity and hence increased levels of transparency and trust. However, notably, high levels of capacity are necessary throughout to meaningfully leverage a highly autonomous system.

The approach to be taken in a national data ecosystem depends on its context, relative weighting of advantages and disadvantages and underlying philosophy regarding centralisation and who controls access to the data. The intersection of these two variables and a description of their concurrent features are provided in the Figure below.

- Option 1 describes a closed and centralised ecosystem owning the data and controlling access to it, much like many data systems run by, for example, governments today.
- Option 2 sees a closed and centralised ecosystem owning the data, but in this scenario, individuals have authority over their own data and, at their own discretion, can grant others permission to access it.
- Option 3 sees a decentralised model at work, with all that that entails, including the individual owning her/his own data, but the system is still run by an intermediary that controls access to the data.
- The final option is a radically decentralised ecosystem that allows individuals to both own their data and control access to it.

Figure 23: Intersections between ownership and type of system



Key recommendation 11: **Develop an approach suited to context.** Each system must consider its positionality in terms of both philosophical underpinnings and legal and policy frameworks regarding privacy and, where in effect, use of personal data. In addition, consideration must be given to system-wide capacity and distribution of capacity, both in terms of infrastructure and human resources. The conversation must be framed in terms of the beneficiaries of the system and the intended outcomes of the system as determined by a thorough consultative process and systemic review of the four levels of interoperability.

OPPORTUNITIES FOR PARTNERSHIPS

SDG 4, Quality Education, aims to "ensure inclusive and equitable quality education and promote lifelong learning opportunities for all" (UN, 2015, 14). The provision of tertiary education is deemed vital for lifelong learning, and mobility in tertiary education is seen as an asset and opportunity that should be enhanced to develop students' competencies and global competitiveness. Given the growing trend of the internationalisation of education, interoperable datasets are needed that allow for not only the required mobility of learning but also the possibility of life-long learning journeys made up of both formal and informal learning events.

In order to maximise the potential of higher education systems, South Africa should seek to become part of a community related to data systems and management. Doing so will enable greater cross-pollination of ideas as well as grant the ability of South Africa to both contribute and draw from the international community. Within the Southern African Development Community (SADC) in particular, the opportunity exists for South Africa to become a leader in establishing such a community.

By extension, a platform should not be developed in isolation but should be devised in order to link into various systems, including both national and international reporting frameworks. Systems designed will be most useful if they are, at the very least, interoperable with the systems of structures which regularly require data reports. In addition, leveraging work which has already been completed and is in use can reduce both initial and ongoing costs and enable engagements with a community for capacitation.

As previously mentioned, a number of initiatives and a wide range of standards are in use in individual countries, regions and internationally. Many of these are concerned with challenges which exist across contexts, and partnerships have been formed which allow beneficiaries to both input, and benefit from, expertise across the collaboration.

Of particular note are:

- The SDMX and .Stat community, which provides a standard and platform which can be leveraged with bespoke components;
- The Credential Engine work, which could be leveraged to improve links between further education and training and the labour market;
- OpenFN, which could be leveraged to decrease the capacity demands of individual organisations in adherence to a standard; and
- The EAC, which could provide a template for similar initiatives or be expanded to additional regions or areas across the African continent.
- In addition, local collaboration with relevant government departments such as the SABS as well as all components of the PSET system is of paramount importance.

Key Recommendation 12: Join a community. There are a number of communities which provide opportunities for collaboration and growth as well as further capacitation and shared development. Local and user capacity should also be leveraged to add value to an interoperable system and enable increased levels of innovation.

CLOSING NOTE

South Africa has the opportunity to ensure that its citizens and institutions have access to integrated and broad-ranging data, which could be utilised for both effective decision-making and enhancing the accessibility, articulation and relevance of skills training to the world of work. Key to realising this vision is the political will and openness to move away from discrete data collection and housing towards the development and provision of an integrated, open and interoperable digital ecosystem for the PSET sector.

Such a bold innovation would also promote increased levels of individual agency in the broader international labour market. If individuals are positioned as the primary owners of their data, and the wider ecosystem works to ensure the portability and beneficial use of such data, life-long learning journeys made up of both formal and informal learning events will be possible, as will the mobility of data and credentials in a wider, international setting. In this way, rates of employment could be increased, and long-term economic benefits realised.

Now is the time for a brave decision to be made, to maximise the potential of 4IR technology in ensuring the success of skills development in South Africa and beyond.

References

1-Page Ltd. (2016). *New Pew study shows the future looks bright…And full of retraining*. Retrieved from: https://www.marketscreener.com/1-PAGE-LTD-18206408/news/1-Page-New-Pew-Study-Shows-the-Future-Looks-Bright-hellip-And-Full-of-Retraining-23562719/

Abiteboul, S., Miklau, G., Stoyanovich, J. & Weikum, G., eds. (2016). Data, responsibly. Dagstuhl Seminar 16291. *Dagstuhl Reports*, 6(7): 42–71. Retrieved from: http://drops.dagstuhl.de/opus/volltexte/2016/6764/pdf/dagrep_v006_i007_p042_s16291.pdf

Access Partnership. (2018). *Artificial intelligence for Africa: An opportunity for growth, development, and democratisation*. Retrieved from: https://www.accesspartnership.com/artificial-intelligence-for-africa-an-opportunity-for-growth-development-and-democratisation/

Alliance for Affordable Internet. (2017+). *Price of broadband data (1GB mobile prepaid) as % of GNI per capita, 2017*. Retrieved from: https://a4ai.org/mobile-broadband-pricing-data/

Andreessen, M. (2011). Why software is eating the world. *Wall Street Journal*, 20(2011): C2.

Benali, I. (2018). The million dollar graduates: making sense of AI talent migration. *SingularityNET,* 14 November 2018. Retrieved from: https://blog.singularitynet.io/the-million-dollar-graduates-making-sense-of-ai-talent-migration-142997d5e018

Benko, C. & Anderson, M. (2010). *The corporate lattice: Achieving high performance in the changing world of work*. Boston: Harvard Business Review Press.

Bloomberg, J. (2018). Digitization, digitalization, and digital transformation: Confuse them at your peril. Forbes, 29 April 2018. Retrieved from: https://www.forbes.com/sites/jasonbloomberg/2018/04/29/digitization-digitalization-and-digital-transformation-confuse-them-at-your-peril/#9030c4d2f2c7

Boland, H. (2018). Britain faces an AI brain drain as tech giants raid top universities. *The Telegraph*, 3 September 2018. Retrieved from: https://www.telegraph.co.uk/technology/2018/09/02/britain-faces-artificial-intelligence-brain-drain/

Brookfield Institute. (2018). *AI + public policy: Understanding the shift*. Toronto: Brookfield Institute, Ryerson University. Retrieved from: http://brookfieldinstitute.ca/wp-content/uploads/AI_BackgroundMaterials_ONLINE-1-1.pdf

Brown, W. (2017). *Inequality, poverty, and jobs: South Africa's relationship to the Fourth Industrial Revolution: And their relationship to South Africa's Electronic Communications Amendment Bill 2017.* Retrieved from: https://www.ellipsis.co.za/wp-content/uploads/2018/03/Mr.-Walter-Brown-ECA_Bill_2017.pdf

Brundage, M. & Bryson, J. (2016). *Smart policies for artificial intelligence*. Retrieved from: https://arxiv.org/ftp/arxiv/papers/1608/1608.08196.pdf

Cameron, J. (2011). Social cost-benefit analysis – principles. In Cameron, J., Hunter, P., Jagals, P. & Pond, K., eds. *Valuing water, valuing livelihoods*. London: World Health Organization and IWA Publishing, pp. 199–206. Retrieved from: https://apps.who.int/iris/bitstream/handle/10665/44635/9781843393108_eng.df;jsessionid=065AE9087727B6A6B3B9FB57613E65BD?sequence=1

CB Insights. (2018a). *AI industry series: Top healthcare trends to watch*. New York, NY: CB Insights. Retrieved from: https://www.cbinsights.com/research/report/ai-trends-healthcare/

CB Insights. (2018b). *What's next in AI? Artificial intelligence trends 2019*. New York, NY: CB Insights. Retrieved from: https://www.cbinsights.com/research/report/ai-trends-2019/

CB Insights. (2019). *China is starting to edge out the US in AI investment*. Retrieved from: https://www.cbinsights.com/research/china-artificial-intelligence-investment-startups-tech/

Credential Engine. (Date unknown). *What's the value of a credential?* Retrieved from: https://credentialfinder.org/home/about

Dadios, E.P. et al. (2018). *Preparing the Philippines for the Fourth Industrial Revolution: A scoping study.* Philippine Institute for Development Studies. Discussion paper series no. 2018-11. Quezon City, Philippines: PIDS. Retrieved from: https://pidswebs.pids.gov.ph/CDN/PUBLICATIONS/pidsdps1811.pdf

Darktrace. (Date unknown). *Darktrace Enterprise*. Retrieved from: https://www.darktrace.com/en/resources/ds-enterprise.pdf

Davidson, D. & Harman, G. (1972). Editorial introduction. In Davidson, *Semantics of natural language.* 2nd ed. Dordrecht: D. Reidel.

Deloitte. (2016). *Industry 4.0. Is Africa ready for digital transformation?* Retrieved from: https://www2.deloitte.com/content/dam/Deloitte/za/Documents/manufacturing/za-Africa-industry-4.0-report-April14.pdf

Department of Telecommunications and Postal Services (DTPS). (2016). *National integrated ICT policy white paper.* Pretoria: DTPS. Retrieved from: https://www.dtps.gov.za/images/phocagallery/Popular_Topic_Pictures/National_Integrated_ICT_Policy_White.pdf

Department of Telecommunications and Postal Services (DTPS). (2018). Invitation to nominate candidates for the Presidential Commission on Fourth Industrial Revolution. *Government Gazette* 42078, 4 December 2018, 4-21. Retrieved from: https://www.gov.za/sites/default/files/gcis_document/201812/42078gen764.pdf

Donovan J. & Benko, C. (2016). AT&T's talent overhaul. *Harvard Business Review*, 68–73.

Dutton, T., Barron, B. & Boskovic, G. (2018). *Building an AI world: Report on national and regional AI strategies.* Canada: CIFAR.

East African Community (EAC) Common Higher Education Area. (Date unknown, a). *Student mobility policy.* Retrieved from: http://iucea.org/eahea1/student-mobility-policy/

East African Community (EAC) Common Higher Education Area. (Date unknown, b). *What is a Common Higher Education Area?* Retrieved from: http://iucea.org/eahea1/east-african-community-common-higher-education-area/

Element AI. (2018). *Global AI talent report 2018.* Retrieved from: https://jfgagne.ai/talent/

Environmental Protection Agency (EPA). (Date unknown). *Learn about data standards*. Retrieved from: https://www.epa.gov/data-standards/learn-about-data-standards

European Commission. (2018). *Policy: Artificial intelligence*. Retrieved from: https://ec.europa.eu/digital-single-market/en/artificial-intelligence

European Commission/EACEA/Eurydice. (2017). *Support mechanisms for evidence-based policy making in education.* Eurydice report. Luxembourg: Publications Office of the European Union. Retrieved from: https://eige.europa.eu/resources/206_EN_Evidence_based_policy_making.pdf

EY, FICCI & NASSCOM. (2017). *Future of jobs in India: A 2022 perspective*. Retrieved from: https://cdn.ey.com/echannel/in/en/services/tax/future-jobs-india/ey-future-of-jobs-infographic.pdf

Frontier Economics. (2018). *The impact of artificial intelligence on work: An evidence review prepared for the Royal Society and the British Academy.* Europe & Australia: Frontier Economics Network. Retrieved from: https://royalsociety.org/~/media/policy/projects/ai-and-work/frontier-review-the-impact-of-AI-on-work.pdf

Gasser, U. & Almeida, V. (2017). A layered model for AI governance. *IEEE Internet Computing,* 21(6): 58–62.

Goldstein, E., Gasser, U. & Budish, R. (2018). *Data Commons Version 1.0: A framework to build toward AI for good: A roadmap for data from the 2018 AI for Good Summit*. Cambridge, MA: Berkman Klein Center for Internet and Society, Harvard University. Retrieved from: Data https://cyber.harvard.edu/story/2018-06/data-commons-version-10

Gonzalez, C. (2018). *China's plan to become a robotic powerhouse.* Machine Design. Retrieved from: https://www.machinedesign.com/motion-control/china-s-plan-become-robotic-powerhouse

Groningen Declaration Network (GDN). (2012). *Groningen declaration on digital student data depositories worldwide.* The Hague: Ministerie van Onderwijs, Cultuur en Wetenschap. Retrieved from: https://www.groningendeclaration.org/wp-content/uploads/2018/02/groningendeclaration_final_final-1.pdf

Groningen Declaration Network (GDN). (Date unknown). Our goal. Retrieved from https://www.groningendeclaration.org/

Healthcare Information and Management Systems Society (HiMSS). (Date unknown). *What is interoperability?* HiMSS. Retrieved from https://www.himss.org/library/interoperability-standards/what-is-interoperability

Herweijer, C. et al. (2017). *Enabling a sustainable Fourth Industrial Revolution: How G20 countries can create the conditions for emerging technologies to benefit people and the planet.* PwC. Retrieved from: https://www.pwc.co.uk/sustainability-climate-change/assets/enabling-a-sustainable-fourth-industrial-revolution.pdf

Heubusch, K. (2006). Interoperability: What it means, why it matters. *Journal of AHIMA,* 77(1): 26–30.

Horvitz, E. (2017). AI, people, and society. *Science*, 357(6346): 7.

Hosie, R. (2017). Ashley Madison hacking: What happened when married man was exposed. *The Independent,* 16 January 2017. Retrieved from: https://www.independent.co.uk/life-style/love-sex/ashley-madison-hacking-accounts-married-man-exposes-cheating-website-infidelity-rick-thomas-a7529356.html

International Federation of Robotics (IFR). (2017). *The impact of robots on productivity, employment and jobs: A positioning paper by the International Federation of Robotics April 2017*. Retrieved from: https://ifr.org/img/office/IFR_The_Impact_of_Robots_on_Employment.pdf

Janson, E. (Date unknown). Ed-Fi Technology – 101: Ed-Fi Data Standard. In*: Ed-Fi Bootcamp 2017.* Presentation. Ed-Fi. Retrieved from: https://www.ed-fi.org/assets/2018/09/2017-Ed-Fi-Bootcamp-Introduction-101-Ed-Fi-Technology.pdf

JET Education Services (JET) & merSETA. (2019). *Digital interoperability: The new frontier for academic integrity in the post-school education and training system*. Presentation prepared by Keevy, J., Rajab, R., Nomvete, S. & Manda, M. Johannesburg: JET and merSETA.

Khan, J. (2018). *Just how shallow is the artificial intelligence talent pool?* Bloomberg. Retrieved from: https://www.bloomberg.com/news/articles/2018-02-07/just-how-shallow-is-the-artificial-intelligence-talent-pool

Kituyi, M. (2018). The 'fourth industrial revolution' can power sustainable development, if we get it right. Geneva: United Nations Conference on Trade and Development. Retrieved from: https://unctad.org/en/pages/newsdetails.aspx?OriginalVersionID=1746

Kurtz, H. (2018). Lack of clear standards, and money, trips up district tech leaders on interoperability. *Education Week,* 30 October 2018. Retrieved from: https://www.edweek.org/ew/articles/technology/2018/10/31/lack-of-clear-standards-and-money-trips.html

Lartey, J. (2016). Predictive policing practices labelled as 'flawed' by civil rights coalition. *The Guardian*, 31 August 2016. Retrieved from: https://www.theguardian.com/us-news/2016/aug/31/predictive-policing-civil-rights-coalition-aclu

Lee et al. (2018). How to respond to the fourth industrial revolution, or the second information technology revolution? Dynamic new combinations between technology, market, and society through Open Innovation. *Journal of Open Innovation: Technology, Market, and Complexity,* 4(3): 21. Retrieved from: https://www.mdpi.com/2199-8531/4/3/21

Levin, S. (2018). Tesla fatal crash: 'autopilot' mode sped up car before driver killed, report finds. *The Guardian,* 7 June 2018. Retrieved from: https://www.theguardian.com/technology/2018/jun/07/tesla-fatal-crash-silicon-valley-autopilot-mode-report

Liao, Y., Loures, E. R., Deschamps, F., Brezinski, G., & Venâncio, A. (2018). The impact of the fourth industrial revolution: a cross-country/region comparison. *Production*, 28. Retrieved from: http://www.scielo.br/pdf/prod/v28/0103-6513-prod-28-e20180061.pdf

Lin, J. & Rosenblatt, D., eds. (2012). *Shifting patterns of economic growth and rethinking development.* Washington, DC: World Bank.

Maharajh, R. (2018). Africa and the Fourth Industrial Revolution: The need for "creative destruction" beyond technological change. *Perspectives,* 4 December 2018. Retrieved from: http://za.boell.org/2018/12/04/africa-and-fourth-industrial-revolution-need-creative-destruction-beyond-technological

Marr, B. (2016). What everyone must know about industry 4.0. *Forbes,* 20 June 2016. Retrieved from: https://www.forbes.com/sites/bernardmarr/2016/06/20/what-everyone-must-know-about-industry-4-0/#1b7ffe36795f

Marume, S.B.M. (2016). Public policy and factors influencing public policy. *International Journal of Engineering Science Invention*, 5 (6): 06–14.

McAfee, A. & Brynjolfsson, E. (2017). *Machine, platform, crowd: Harnessing our digital future.* New York: WW Norton & Company, Inc.

McAuley, D., Rahemtulla, H., Goulding, J., & Souch, C. (2011). How Open Data, data literacy and Linked Data will revolutionise higher education. In Coiffait, L. ed. (2011). *Blue Skies: New Thinking about the Future of Higher Education: a Collection of Short Articles by Leading Commentators.* Pearson Centre for Policy and Learning, pp. 88–93. Retrieved from https://www.academia.edu/1945211/The_Open_Data_Revolution_and_Data_Literacy_in_Higher_Education

McClelland, C. (2017). *The difference between artificial intelligence, machine learning, and deep learning.* A Medium Corporation. Retrieved from: https://medium.com/iotforall/the-difference-between-artificial-intelligence-machine-learning-and-deep-learning-3aa67bff5991

McKinley, S. (2018). Before we solve the world's problems, we need to connect it to the internet. *Quartz,* 20 March 2018. Retrieved from: https://qz.com/1233010/before-we-solve-the-worlds-problems-we-need-to-connect-it-to-the-internet/

Miah, M. & Omar, A. (2012). Technology advancement in developing countries during digital age. *International Journal of Science and Applied Information Technology*, 1(1).

Michael and Susan Dell Foundation (MSDF). (Date unknown). *Data driven districts*. Retrieved from https://www.msdf.org/initiatives/data-driven-districts/

Moffatt, C. (Date unknown). Education data standards. MSDF. Presentation supplied.

Morales, L. & Orrell, T. (2018). *Data interoperability: A practitioner's guide to linking up data in the development sector.* Collaborative on SDG Data Interoperability. Retrieved from: https://unstats.un.org/wiki/display/InteropGuide/Home

National Institution for Transforming India (NITI Aayog). (2018). *National strategy for artificial intelligence. Discussion paper.* Delhi: NITI. Retrieved from https://www.niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf

National Science & Technology Council Committee on Technology. (2016). *Preparing for the future of artificial intelligence.* Washington, DC: Executive Office of the President of the United States. Retrieved from: https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

National Science & Technology Forum (NSTF). (2018). *Discussion Forum: The 4th Industrial Revolution.* Pretoria: NSTF. Retrieved from: http://www.nstf.org.za/wp-content/uploads/2018/08/184IRConcept.pdf

Ndemo, B. & Weiss, T. (2017). Making sense of Africa's emerging digital transformation and its many futures. *Africa Journal of Management*, 3(3–4): 328–347.

Ng'ambi, D., Brown, C., Bozalek, V., Gachago, D. & Wood, D. (2016). Technology enhanced teaching and learning in South African higher education – A rearview of a 20 year journey. *British Journal of Educational Technology,* 47(5): 843–858.

O'Flaherty, K. (2018). *How to survive a ransomware attack – and not get hit again*. Forbes. Retrieved from: https://www.forbes.com/sites/kateoflahertyuk/2018/08/17/how-to-survive-a-ransomware-attack-and-not-get-hit-again/#1779d256cd36

Open Data Incubator for Europe (ODINE). (2016). *Best practices for data standards*. Retrieved from: https://opendataincubator.eu/files/2017/07/2016-08-01-data-standards.pdf

ODINE. (2017). Creating a European ecosystem for data innovation: Reflections on ODINE. *The Guardian,* 20 July 2017. Retrieved from: https://opendataincubator.eu/creating-a-european-ecosystem-for-data-innovation-reflections-on-odine/

Open Data Watch. (2018). *The data value chain: Moving from production to impact, Executive summary.* Retrieved from: https://opendatawatch.com/reference/the-data-value-chain-executive-summary/

OpenFN. (Date unknown). *Helping an airport soar: Synchronizing asset data between the tarmac and HQ: An international airport, USA*. Retrieved from: https://www.openfn.org/success-stories

Organisation for Economic Co-operation and Development (OECD). (2014). *.Stat and SIS-CC Strategy: 2014–19.* Paris: OECD. Retrieved from: https://siscc.org/wp-content/uploads/2018/10/dotStat-and-SIS-CC-Strategy-2014-19.pdf

Panda Security. (2017). *A dating site and corporate cyber-security lessons to be learned.* Panda Mediacenter. Retrieved from: https://www.pandasecurity.com/mediacenter/security/lessons-ashley-madison-data-breach/

Parker, G.G., Van Alstyne, M.W. & Choudary, S.P. (2016). *Platform revolution: how networked markets are transforming the economy and how to make them work for you.* New York, NY: WW Norton & Company.

Pew Research Center. (2016). *The State of American Jobs: How the shifting economic landscape is reshaping work and society and affecting the way people think about the skills and training they need to get ahead.* Retrieved from: https://www.markle.org/sites/default/files/State-of-American-Jobs.pdf

Pisani, B. (2018*). Cybersecurity, trade tensions rank as top threats to markets in 2019, survey finds*. CNBC. Retrieved from: https://www.cnbc.com/2018/12/11/cybersecurity-trade-war-rank-as-top-threats-to-2019-markets-survey.html

Puentedura, R.R. (2010). *The SAMR model: Background and exemplars*. Retrieved from: http://www.hippasus.com/rrpweblog/archives/2012/08/23/SAMR_BackgroundExemplars.pdf

Purdy, M. & Daugherty, P. (2017) *How AI boosts industry profits and innovation*. Accenture. Retrieved from: www.accenture.com/aiboostsprofits

Puskar, J. (2018). Edtech investment is growing – but the majority of investment isn't for the classroom. *EdtechTimes,* 7 March 2018. Retrieved from: https://edtechtimes.com/2018/03/07/record-funding-in-edtech-reveal-growing-disparity/

Rakyan, S. (2017). *Attracting, retaining and engaging your digital workforce in an era of digital transformation.* Willis Towers, Watson. Retrieved from: https://www.towerswatson.com/en/Insights/Newsletters/Asia-Pacific/points-of-view/2017/attract-retain-and-engage-your-digital-workforce

Sample, I. (2017). We can't compete: Why universities are losing their best AI scientists. *The Guardian,* 1 November 2018. Retrieved from: https://www.theguardian.com/science/2017/nov/01/cant-compete-universities-losing-best-ai-scientists

Schäfer, M. (2018). The fourth industrial revolution: How the EU can lead it. *European View 2018,* 17(1): 5–12.

Schwab, K. (2017). *The fourth industrial revolution*. New York, NY: Crown Business.

Shukla, A., Harris, S. & Davies, J. (2010). *Semantic interoperability in practice.* Proceedings of the 43rd Hawaii International Conference on System Sciences. Oxford University Computing Laboratory. Retrieved from: https://www.computer.org/csdl/proceedings/hicss/2010/3869/00/04-01-05.pdf

Simonite, T. (2017). Machines taught by photos learn a sexist view of women. *WIRED*, 21 August 2017. Retrieved from: https://www.wired.com/story/machines-taught-by-photos-learn-a-sexist-view-of-women/

Šinko, M. (2016). Policy change – review of classification, measurement and factors. *Teorija In Praksa let*., 1 (53): 228–247.

Stanford University. (2016). *Artificial intelligence and life in 2030: one hundred year study on artificial intelligence; report of the 2015 study panel*. Stanford, CA: Stanford University. Retrieved from: https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai_100_report_0831fnl.pdf

Statistical Information System Collaboration Community (SIS-CC). (2018). *Building national data backbones: Empowering countries through capacity development and technology*. Workshop Highlights Report from the 2018 SIS-CC Workshop. Paris, France. Retrieved from: https://siscc.org/wp-content/uploads/2018/10/sis-cc-workshop-2018-highlights-report.pdf

Stratus. (2017). *Asia: get future-ready for the fourth industrial revolution*. Stratus. Retrieved from: https://www.stratus.com/assets/apac/Asia-Forth-Industrial-Revolution.pdf

Tucker, I. (2017). 'A white mask worked better': why algorithms are not colour blind. *The Guardian*, 28 May 2017. Retrieved from: https://www.theguardian.com/technology/2017/may/28/joy-buolamwini-when-algorithms-are-racist-facialrecognition-bias

United Kingdom. (2018). *AI sector deal*. Retrieved from: https://www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal

United Nations (UN). (2015). *Transforming our world: the 2030 Agenda for Sustainable Development*. Geneva: UN. Retrieved from: https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E

United Nations Economic Commission for Europe (UNECE). (2019). *Generic Statistical Business Process Model. GSBPM. (Version 5.1, January 2019).* Geneva: UNECE. Retrieved from: https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2013). UNESCO *Handbook on education policy analysis and programming, Volume 1: Education policy analysis*. Bangkok: UNESCO Bangkok, Asia and Pacific Regional Bureau for Education. Retrieved from: https://unesdoc.unesco.org/ark:/48223/pf0000221189

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2016). *The 4th Industrial Revolution must be a development revolution.* UNESCO at Davos. https://en.unesco.org/news/4th-industrial-revolution-must-be-development-revolution-unesco-davos

Vrountas, T. (2018). *How demand-side platforms enable wider reach & efficient ad buying.* Instapage. Retrieved from: https://instapage.com/blog/demand-side-platform

Wakefield, S. (2017). *Education system digital strategy: Transforming education for the digital age*. Presented at Tertiary ICT Conference, Wellington, 6–8 September 2017. Retrieved from: https://www.tertiaryictconference.co.nz/uploads/1/8/1/3/18139115/stuart_wakefield.pdf

Walsh, K. (2019). How AI is transforming agriculture. *Forbes,* 5 July 2019. Retrieved from: https://www.forbes.com/sites/cognitiveworld/2019/07/05/how-ai-is-transforming-agriculture/#3b116e1e4ad1

Wang, B. (2018). China increasing industrial robots by ten times to 1.8 million by 2025. *Nextbigfuture*, 22 September 2018. Retrieved from: https://www.nextbigfuture.com/2018/09/china-increasing-industrial-robots-by-ten-times-to-1-8-million-by-2025.html

Wilks, Y. & Charniak, E. (1976). *Computational semantics: An introduction to artificial intelligence and natural language understanding.* Amsterdam: North-Holland Press.

Winfield, A. & Halverson, M. (2017). Artificial intelligence and autonomous systems: Why principles matter. *IEEE Future Directions*. IIE. Retrieved from: http://sites.ieee.org/futuredirections/tech-policy-ethics/september-2017/artificial-intelligence-and-autonomous-systems-why-principles-matter/

World Bank Singapore Hub/Global Infrastructure Connectivity Alliance. (2018). Why connectivity matters. Washington, DC: World Bank Group. Retrieved from: https://www.gica.global/resources/discussion-paper-why-connectivity-matters

World Economic Forum. (2017). *Harnessing the Fourth Industrial Revolution for Sustainable Emerging Cities.* World Economic Forum Retrieved from: http://www3.weforum.org/docs/WEF_Harnessing_the_4IR_for_Sustainable_Emerging_Cities.pdf

Wortmann, F. & Flüchter, K. (2015). Internet of things: Technology and value added. *Business & Information Systems Engineering,* 57(3): 221–224.

Zarkadakis, G. (2018). Do platforms work?. *Aeon,* 28 May 2018. https://aeon.co/essays/workers-of-the-world-unite-on-distributed-digital-platforms

Interoperable data ecosystems

An international review to inform a South African innovation

Interoperability, data and ecosystems are three concepts that, to date, have not been used together enough internationally, and even less so in South Africa.

We now stand at the cusp of a new era in which we have an over-supply of data, while our ability to harvest the data simply does not keep up with the new approaches that artificial intelligence in a second machine age places at our disposal. Many countries, including South Africa, have a range of datasets that cover many critical aspects of their education and training systems – some of these may be well-established but perhaps archaic in design and based on outdated software, while others may be well-designed, containing pristine data that is not available anywhere else in the national system.

The problem with this situation is that this lack of interoperability leads to a weak national data ecosystem, made up of only a few willing partners cooperating across datasets, often only because they are legally compelled to do so. As a result, the ability of the national system to link a latent workforce (the supply side) with existing and new opportunities (the demand side) is severely constricted, if functional at all.

In all of this, and more so in developing countries like South Africa, government is the slow and steady gatekeeper, and innovative people and companies will not wait for government to catch up, as we can see in the increase of the many platforms that link supply and demand across the world. While public–private partnerships that harness these new technologies have huge potential to support emerging economies, if left uncoordinated, they often result in increased inequality – something that South Africa can ill afford.

This international review attempts to explore some of these debates as part of a larger South African initiative that aims to ultimately give South African citizens the ability to make informed labour market decisions that lead to employment. International readers will also find value in the analysis as the move towards interoperable data ecosystems is certainly a global narrative that can only be ignored at your own peril.