***1. Introduction***

“We commit to evaluations that help us learn, understand and support the transformational and systemic changes needed in our countries and the world, as agreed upon in the 2030 Agenda for Sustainable Development. A sustainable balance between the social, economic and environmental domains is crucial in light of the existential threats of the climate crisis, mass extinction of species, growing local and global inequity, and ultimately unsustainable use of the resources of the planet,” (IDEAS, 2019, pp. 1).

Change is inevitable, but progress is not. Today, each of us is a globally connected witness to a time of profound and inter-related changes in human society and global ecological systems. The challenges facing humanity represent existential threats to the future of all societies (Pichler, Schaffartzik, Haberl & Gӧrg, 2017; Steffen et al., 2018), and require radical solutions if we are to be around in the coming centuries. The COVID-19 experience provides a clear lens on many of the frailties and failures of the planetary system (Andersen & Rockström, 2020; Lambert et al., 2020), and the need to reboot the current economic system (BMU & WRI, 2020). The call for transformation is a call for progress. The quote at the beginning of this article is taken from the Prague Declaration on Evaluation for Transformational Change, adopted at the Third International Conference of the International Development Evaluation Association (IDEAS, 2019). It reflects the growing attention to transformation in the evaluation of sustainable development following the adoption by the UN General Assembly of the landmark *Transforming our World: the 2030 Agenda for Sustainable Development* (henceforth “2030 Agenda”) and the Paris Agreement on Climate Change in 2015.

Some have argued (e.g., Sachs et al., 2019) that at least six major transformations are needed to achieve the Sustainable Development Goals (SDGs) identified in the UN 2030 Agenda. As the largest collaboration for collective welfare to date, adopted by and premised on the interdependent welfare of all UN Member States, the authors of the 2030 Agenda are explicit about its “supremely ambitious and transformational vision”, each calling for action by specific parts of government in collaboration with business and civil society. Both SDG 17.13 and 17.14 highlight that coherent, coordinated policy processes lie at the heart of the transformational process. This is particularly important in the design, implementation and evaluation of policies directed towards enhancing social-ecological sustainability and resilience (Sterner et al., 2019). In this complex and dynamic context, scientist and practitioner communities spanning natural, social, and human sciences need to work together, and with policymakers, to help inform the development of tractable solutions. Contributing in the most meaningful, effective way is the main challenge before the global community of evaluation specialists.

Capitalising on this agenda for change, in this article, we are keen to explore new developments in data science (Section 3), the potential of that data science frontier to aid sustainable development (Section 4), and for enabling and stimulating evaluation to make a transformational contribution to sustainable development, whilst at the same time being transformational for the field of evaluation (Section 5). Prefacing those sections, in Section 2, we provide a brief introduction describing the increasing alignment of complex systems thinking and sustainable development evaluation. In Section 6, the final substantive section, we discuss the feasibility and challenges of applying the developments outlined in Sections 4 and 5 in the current market-place of evaluation, and what that implies for the dual transformational potential of evaluation. We finish with some concluding insights.

But first, let us clarify some of the key concepts that are woven throughout this article. For our purposes, the use of *sustainable development* returns to its original conceptualization proposed by the Brundtland Report (WCED 1987, pp. 43): “(D)evelopment that meets the needs of the present without compromising the ability of future generations to meet their own needs.” We also include equitable social development and inclusion per the more expansive interpretation of sustainable development adopted 25 years later by the UN General Assembly (UN, 2012, pp. 1). We also acknowledge that *transformation* is an unfolding, open-ended process; therefore, we approach it heuristically rather than dogmatically and, given the international scope of sustainable development, we use transformation to refer to deep, rapid, and radical global systems change required to achieve the aspirations of the sustainable development goals (SDGs). This definition is often contrasted with incrementalism, reform, or transition to convey the magnitude of required change: “Unlike a ‘transition,’ which implies moving from one place or state to another, ‘transformation’ is more about completely reinventing shape or form – like the metamorphosis of a caterpillar to a butterfly,” (Waddock, Waddell, Goldstein, Linner, Schapke, & Vogel., 2020, pp. 4).

***2. The Rise of Complex Systems Thinking in Sustainable Development Evaluation***

Despite overtures towards more complex systems-informed thinking, policymaking (including associated measures such as projects and programmes) continues to operate largely in ignorance of complexity: insufficiently attuned to the reality of how complex systems function, and blinkered in its belief in a certain and controllable world (Geyer, 2012; Mueller, 2019). This is wholly at odds with the complex world we all live in, where outcomes from interventions may be radically different than those originally envisioned (Eppel & Rhodes, 2018). The persistence of this dominant worldview is problematical on two counts. First, its linear perspective on causality speaks to a particularly pervasive reductionist, bureaucratic and economistic accounting mentality (O’Brien, 2014), which fails to offer an integrative analysis of policies, interventions, and outcomes (Geyer, 2012). Second, it valorises the seeming neatness and certainty presented by numeric quantification over the ‘messiness’ of qualitative discourse: a perspective that continues to offer the ‘illusion of certainty and precision’, and peddle the falsehood that quantitative approaches are inherently superior to qualitative ones (Kovaciz, 2018). Instead, the more considered position would see these two approaches as equally valid and complementary investigative lenses both capable of revealing important and unique insights (Bamberger, Vaessen & Raimondo, 2016). The upshot, a kind of Freirean ‘false consciousness’ whereby only business-as-usual approaches can be imagined (Patton, 2018) – a reality that helps to explain the ubiquity of policy failures (Mueller, 2019).

Yet, integrating complexity concerns into policy processes and interventions continues to be highly challenging (Head, 2019; Nilsson & Eckerberg, 2009), which perhaps helps to explain that relative to other fields, evaluation has been slow to embrace systems and complexity thinking (Befani, Ramalingam, & Stern, 2015, pp. 5; Williams, 2015, pp. 9). However, the need for more ‘complexity appropriate’ policy and evaluation has been increasingly recognised (Quayle & Kelly, 2018) and, although experts may disagree on some of the finer details, they converge on the global scale and complexity of today’s sustainability challenges (Dodds & Bartram, 2016; Steffen et al., 2018; UN, 2013). In evaluation, the term ‘wicked problem’ describes complex problems that defy traditional analysis and solutions (Williams & van t’Hoff, 2014; Hopson & Cram, 2018). The ‘apparent’ intractability of these issues resides in their complexity; specifically, their multi-scale open dynamism, which means these social-technical-ecological systems behave in non-linear ways. Such behaviours enable properties like emergence, self-organization, adaptation, and uncertainty, magnifying the challenges of predicting system behaviours that are constantly changing, resisting change, or evolving in unexpected ways. Increasingly, ‘super wicked problem’ is used to refer to hyper-complex challenges, as mirrored by the SDGs, these problems are characterized by multiple interacting systems, scales of change (e.g. local, regional, and global), and intersecting interventions and actors (Levin, Cashore, Bernstein & Auld, 2012).

Sustainable development is especially fertile ground for super-wicked problems. Collectively, and individually, the 17 SDGs and their 169 targets and 232 indicators that comprise the 2030 Agenda respond to extraordinarily complex problems (Fu, Wang, Zhang, Hou & Li, 2019). Nested within this complexity is an assortment of actors, ranging from bilateral and multilateral aid organizations, philanthropic foundations and private donors, to civil society organizations, the national public sector, and local populations. In these circumstances, where multiple perspectives exist expressing varying priorities and agendas, and resource flows fluctuate in a global economy subject to recession and political change, combined with natural forces, the result is a high degree of uncertainty. For example, the convergence of the present pandemic, climate change and social unrest aptly illustrates the complex interactions of a super-wicked problem. COVID-19, a zoonotic disease, highlights how the steady encroachment on, and neglect of, the environment spills over to negatively impact the human system (Everard, Johnston, Santillo & Staddon, 2020).

In the policy studies and evaluation fields, there has been a concerted effort to address challenges related to non-linearity and uncertainty and; particularly, the difficulties these system dynamics pose for developing effective policy and the ability to forecast and attribute intervention-specific impacts and outcomes (Bicket et al., 2020; Byrne, 2013; Caffrey & Munro, 2017; Geyer & Cairney, 2015; Gates, 2016; Mowles, 2014; Walton, 2016). The adoption of the 2030 Agenda, alongside the Paris Agreement on Climate Change, has helped to further propel this uptake of complex systems thinking in the field of development evaluation (Bamberger, Vaessen, & Raimondo, 2016). The SDGs illustrate a more comprehensive, interconnected framework for global development than the narrower UN Millennium Development Goals (MDGs) adopted fifteen years earlier (**Figure 1**). While transformation is not defined explicitly in the 2030 Agenda, it encompasses the triple-bottom-line of economic, environmental, and social development (UNRISD, 2016). In other words, transformational change is systemic root-and-branch change of system structures, functions, and processes and fundamentally different from the way things are now (Schwandt, Ofir, Lucks, El-Saddik, & D’Errico, S., 2016, pp. 7).

[Insert Figure 1 Here]

Transformational development requires, amongst other things, a much-needed shift from siloed interventions and funding streams to more holistic, integrated approaches at all levels in the global system (AUTHOR, XXXX). Interventions in the arena of sustainable development have largely been dominated by single, clearly defined project and programmes provided by single agencies and funded by single donors. These interventions are typically treated as closed systems, with linear theories of change that overlook the broader context and the complex interactions and interdependencies in which they are unpacked. Narrow piecemeal approaches do not connect the dots required for more sustainable development, and risk overlooking important spill-over and side-effects, whether they are synergistic or crippling (Patton, 2019). Achieving complexity informed transformational change will require inter- and trans-disciplinary working across epistemic communities (AUTHOR, XXXY), open policy processes (Rutter, 2012; Talbot & Talbot, 2015) and methodological and data innovation (Ghaffarzadegan, Lyneis & Richardson, 2011; Hummelbrunner, 2015). It is to that last element we now turn.

***3. Data Science Technologies – A Primer***

We are currently living through a ‘Data Revolution’. The rapid development of new tools, techniques, and data types is growing exponentially. For instance, according to Statistia.com (June, 2020), the volume of data that humans are creating, consuming, capturing, and copying is set to reach 59 zettabytes in 2020, with the so-called ‘data sphere’ potentially reaching 175 zettabytes in just the next five years (that’s 175 followed by 21 zeros!). At the leading edge of this data revolution are the core trio of the data science toolbox: Big Data, Machine Learning (ML) and Artificial Intelligence (AI) (Giest, 2017; USAID, 2019; Vinuesa et al., 2020; York & Bamberger, 2020). Before discussing the transformative potential of these technologies in evaluation, let us first take a closer look at each and then how they relate to each other.

***Big Data***: Around since the 1990s, its ascendancy has grown with computing power and the vast volumes of data accumulated daily. Big Data can be neatly summarised by the four Vs, characterized by an increasing ***variety*** of data forms, in increasing ***volumes*** at rising ***velocities***, where there is often uncertainty regarding data quality, accuracy and completeness (***veracity***) in respect to its origins or source. Ultimately, Big Data is about larger, more complex datasets, frequently derived from new or novel sources that are generally beyond the capabilities of standard processing software to compute. In terms of data variety, Big Data mostly concerns unstructured and semi-structured data in the form of images, text, audio, and video, but also includes more traditional structured data (e.g. relational databases). The volume of this data is generally large but of low-density e.g. stemming from social media platforms, mobile phone applications, satellite imagery, and webpages. In many cases, such as social media data, the rate at which data is received (velocity) is in real or near real time.

***Machine Learning (ML)***: Initially pioneered by Arthur Samuel, who coined the term in 1959, ML is, at root, about enabling computers to learn to recognise patterns in data. The purpose of this pattern recognition is to make generalisations from existing data and predictions about new data. In ML, learning is based on training data, from which pattern seeing capability can be honed and subsequently applied, in order to produce models able to make predictions based on new unseen data. The training process allows ML ***algorithms*** (essentially implementable rules for problem solving or computation) to find patterns and relationships in existing data, which can then be used as the basis to develop specific rules to enable new predictions about novel data. ML models are generally based on ***classification*** or ***regression***. Classification concerns assigning instances of interest to specific categories based on previous learnings, whilst regression is about pattern ‘seeing’ and prediction more often than not based on scoring probabilities. Three sets of algorithms distinguish ML processes: ***supervised algorithms*** (where training data needs to be value labelled with values corresponding to the outcome variable of interest for it to correctly produce prediction rules); ***unsupervised algorithms*** (where no pre-labelling of values is required because their purpose is to find hidden patterns in data not to produce accurate predictions); and ***reinforcement learning*** (an algorithm learning mode that involves trial and error, in which an algorithm learns through interacting with its environment producing actions that lead to new discoveries). Importantly, reinforcement learning can produce both testable hypothesise and the optimisation of behaviour for a given situation or context.

***Artificial Intelligence (AI)***: Named in 1965 by John McCarthy to distinguish it from cybernetics, AI is generally regarded as an overarching term for both the science and technology of developing and creating intelligent systems. In many cases, AI is enabled by advanced ML. However, it goes further than ML in being able to enact, plan or do something within a real-world setting. It is generally regarded that AI is dependent upon technology that exhibits some or all of the following characteristics: ***perception*** (recognises different forms of input data such as visual, audio, textual and tactile); ***decision-making*** (can function as a diagnostic); ***prediction*** (capable of making forecasts for instance); ***pattern recognition and knowledge extraction***; ***interactive communication*** (can engage in a form of conversation like chatbots); and ***logical reasoning*** (able to develop theories from underlying premises).

***Digital Twins (DTs)***: In addition to the above trio, a new technology that is getting increasing attention for its transformative potential is Digital Twins (Arup, 2019; Deloitte Insights, 2020; Jones, Snider, Nassehi, Yon & Hicks, 2020). Originated by Michael Grieves in 2002, but coined by John Vickers in a 2010 NASA Roadmap Report (Piascik et al., 2010), DTs have started to become part of mainstream technology trends and more widespread in their application in corporate, business and industry sectors especially. As **Table 1** reflects, there is no definite consensus of what constitutes a Digital Twin, but rather a spectrum of views. That said, the central thread that holds these definitions together is the physical-to-virtual and virtual-to-physical connection, and the closed-loop of interaction and bi-directional informational exchange between the virtual and physical worlds (Jones, Snider, Nassehi, Yon & Hicks, 2020). Significantly, these definitions also emphasise that the physical components of DTs can concern people, processes, objects, places, and services thereby encompassing both systems and components, processes and behaviours, and assets and organisations (ARUP, 2019; Slingshot Simulations, 2020).

[Insert Table 1 Here]

Perhaps the most important transformational aspect about DTs is their potential for understanding and learning through a systems approach. A core facet to the development and utilisation of DTs is their proto-typing capacity. In other words, they have the possibility to reproduce things either in the real-world or have the potential to be in the real-world in the future (ARUP, 2019; Deloitte Insights, 2020). Digital Twins are therefore problem-solution focused, and their ‘value added’ rests on their ability to enable better planning for the future, monitoring the current state of systems, and problem discovery and correction. The advancement of DTs has been aided significantly by a number of other technological improvements, such as: simulator sophistication; new data sources; interoperability; visualisation; instrumentation and computing platforms, bringing together different data science technologies in the form of the Internet of Things (sensor data regarding the real-world), AI, data analytics, blockchain, edge and cloud computing, and 4G and 5G networks (ARUP, 2019; Deloitte Insights, 2020).

Digital Twins differ both in scope, scale, and function, and thus their degree of complexity and capabilities. ARUP (2019) identifies four key attributes the distinguish DTs: ***autonomy*** (i.e. level of capacity to act without human input); ***intelligence*** (i.e. performance capacity and ability to reproduce cognitive processes); ***learning*** (i.e. ability to use data to improve performance without prescriptive programming) and ***fidelity*** (i.e. the degree to which the system approximates the desired output). ARUP (2019) also identifies five levels of sophistication in terms of the degree to which these attributes are manifest from level one, which is the most limited functionality, to level five, which has high degrees of autonomy and reasoning. This makes DTs very flexible tools, capable to address what-if scenarios, enable virtual collaboration, process sensor data and simulate new conditions in almost real-time, and deliver output instructions that can manipulate the physical world (Deloitte Insights, 2020). DTs can therefore act as living models, adapting to change in real-time, bringing in elements of creativity and story-telling and offering a space to act as a virtual testbed such as in urban development policy (Slingshot Simulations, 2020), or personalised healthcare (Alber et al., 2019; Bruynseels, Santoni de Sio & Jeroen van den Hoven, 2018) and fitness management (Barricelli, Casiraghi, Gliozzo, Petrini & Valtolina, 2020).

Before we conclude this overview, it is important to understand that while the above discussion frames four data science technologies separately, they can only be properly understood in relationship with one another. **Figure 2** illustrates the interdependent relationships between a number of these technology “subsystems” within the data science system, highlighting the interconnections and data flows between them.

[Insert Figure 2 Here]

***4. Data Science and Sustainable Development: Frontiers of Transformational Change***

Accompanying this prodigious expansion in data technology and generation is a growing interest in the transformative potential of the data science frontier for sustainable development (Raftree & Bamberger, 2014; Yayboke, Nealer & Rice, 2017). For instance, as a milestone UN IEAG (2014, pp. 4) report on the data revolution and sustainable development states: “Data are the lifeblood of decision-making. Without data, we cannot know how many people are born and at what age they die; how many men, women and children still live in poverty; how many children need educating; how many doctors to train or schools to build; how public money is being spent and to what effect; whether greenhouse gas emissions are increasing or the fish stocks in the ocean are dangerously low; how many people are in what kinds of work, what companies are trading and whether economic activity is expanding”.

Given the dynamic landscape in which this discourse is evolving, different perspectives understandably emerge: some are aspirational and utopian, focusing on the benefits and a transformational data horizon, whilst other strands are more circumspect and cautious about digital technologies for universal good (e.g., Del Rio Castro et al., 2021; Delli Paoli et al., 2021; Goralski & Tan 2020; IATT 2019; Kong et al., 2020; Nara et al., 2021; Truby 2020; UN IEAG 2014; Vinuesa et al., 2020; 2030Vision, 2019). Ultimately, the achievement of the 2030 Agenda necessitates innovation and transformation on several fronts, including timely and accurate data collection, analysis, and use. As argued by Letouzé, Stock, De Chiara, Lizzi & Mazariegos (2019, pp. 17): “A growing number of innovative approaches, such as the Global Delivery Initiative and Global Learning for Adaptive Management Initiative, are promoting adaptive development as a way of Doing Development Differently and tackling complexity through more problem-driven and context-specific approaches. By leveraging agile methodology and modern technology, like machine learning and real-time data acquisition, to harness large data sets, these approaches may prove to be more effective in addressing the complexity of the SDGs and increasing development effectiveness”.

The remainder of this section will focus on the potential for data science technologies to help transform sustainable development to address today’s (and tomorrow’s) urgent, complex challenges. However, it is important to note that innovation in data science is not a magic bullet, which we will unpack in Section 6 of this article.

Artificial intelligence (AI) (inclusive of ML) has the potential to contribute substantially to both short and long-term gains in global productivity, environmental enhancement, and social justice (Truby, 2020). A recent assessment of the possible contribution of AI to sustainable development identified 134 targets across the 17 SDGs where AI could contribute to their achievement, but also 59 targets for which AI had the potential to be disrupt their achievement (Vinuesa et al., 2020, pp. 1). For societal outcomes, the study found that 82% of targets could benefit from AI-based data management and processing technologies, whilst 38% of the targets may be negatively impacted. Improvements were most likely in critical and basic services such as healthcare, food, and energy provision for transitions towards low carbon economies and smart city advancements. Negative impacts from the technologies were most likely in countries with different value systems, lower technical capacities, and fewer financial resources. The article also found that 70% of targets related to economic outcomes would likely benefit from the application of AI technologies, whilst at the same time it was possible that 33% of targets would be negatively affected. From an ecological perspective, the study indicated that AI had the potential to positively benefit 93%, of targets related to environmental SDGs (Vinuesa et al., 2020, pp. 2). Such research supports the perspective of the data revolution as “drivers for sustainable development” (Nara et al., 2021, pp. 103).

The AI potential is particularly notable at the intersection between the social, environmental, and economic dimensions of sustainable development, due to the dynamic complexity of these issues (2030Vision, 2019). For example, Hilbert & Mann (2019) examine 24 case studies to indicate how AI can support the fulfilment of nine SDGs, including smart agriculture (SDG 2), healthcare diagnostics (SDG3), virtual family planning for girls during pregnancy (SDG 5), and robotics for road repairs (SDG 11). Others have also noted AI applications for smallholder farm systems through the utilisation and development of smart agriculture. For instance, the FAO briefing paper *Digital Technologies in Agriculture and Rural Areas* (Trendov, Varas & Zeng, 2019) highlights how digital-based technologies could support rural farmers worldwide by improving their adaptability to changing weather conditions and pest and disease outbreaks. Related, AI can improve water security and quality challenges through smart water management and monitoring systems, with positive outcomes for water infrastructure, as well as health and sanitation through enhanced waterborne disease detection (Goralski & Tan, 2020).

In health settings, AI’s potential contributions include the diagnosing, monitoring, and modelling of disease outbreaks, pandemics, and chronic ailments, as well as the development and testing of new treatments and the use of chatbots in mental health and family planning services (USAID, 2019; 2030Vision, 2019). In the food sector, AI applications – beyond smart agriculture – include supply chain optimisation, financial management, and the combating of crop pests and diseases (2030Vision, 2019). Other noted areas of AI application include: fighting fraud and financial crimes; mobile phone e-commerce; improving urban planning; increasing the efficiency of organisational and supply chain logistics; transforming the circular economy; enhancing connectivity; innovating in renewable energy infrastructure design and operations; and combating the illegal wildlife trade (Goralski & Tan, 2020; Nikita et al., 2019; Vinuesa et al., 2020; Vision2030, 2019). In addition, AI can be used in disaster risk reduction and humanitarian situations; for example, modelling and forecasting food insecurity and famine events, and assessing social media to improve situational awareness of developing emergencies. AI also has a role in improving governance and democracy by, for instance, helping to detect and guard against tax evasion or highlight instances of violence against particular groups such as tracking violent acts against women reporting across different media channels (USAID, 2019).

Big Data has also been noted for its potential in complex development contexts to improve the understanding and examination of coupled environmental, social, and economic systems (Kong et al., 2020). For example, Big Data has enabled the development of additional SDG metrics via disaggregated geographical information (Delli Paoli et al., 2021); for instance, the use of Google Earth Engine data to assess the social and physical vulnerability of populations to flooding in Senegal (Schwarz et al., 2018). The capacity to expand the arsenal of analysable data for development outcomes has also been cited in relation to the emerging role of Big Data in urban sustainability, especially assessing human behaviours with regards to urban mobility, planning, public health and safety, energy utilisation, and environmental sustainability. It is argued that these aspects provide a human-centred vision of urban sustainability, whilst also delivering dynamic, timely and real-time information at a more fine-grained resolution (Kong et al., 2020). A neat example of this, is the use of mobile phone records and spatial data to assess the association between socio-economic and environmental conditions, and crime and safety vulnerabilities in neighbourhood districts of Bogota, Columbia (Nadai, Letouzé, González & Lepri, 2018).

Digital Twins (DTs)are a less proven technological innovation in sustainable development, but their use is already making headway improving the efficiency and effectiveness of system processes, procedures, and outputs across manufacturing, urban, energy, healthcare, transport, and supply chain sectors. DTs can enhance operational performance and productivity, resource and cost allocation, data and asset management, and human-environment interactions, primarily through enabling a systematic coupling of different data streams and active interventions that cut across technical, environmental, and social components (ARUP, 2019; Deloitte Insights, 2020). For example, as a recent report by Navigant Research (2019) *Leveraging Digital Twin Approach for Sustainable Manufacturing* outlines, DTs can improve the utilisation of energy and water resources across manufacturing processes leading to better operational decisions and a more circular approach to resource consumption.

Another promising area for DTs is urban development, regeneration, and smart cities (Nochta et al., 2019). DTs have been used to great effect in the development of Virtual Singapore – the virtual embodiment of the physical city. Through 3D semantic modelling, data can be related to the real world, with the simulation combining displays of land, transport, building and infrastructure attributes alongside real-time dynamics about demographics, climate, or traffic. This information can then be used to think about new ways of urban planning, such as how to increase urban mobility, where best to expand greening developments, and how to reduce the impacts of disasters (ARUP, 2019). One can imagine this being applied to areas of cities that are also highly under-developed, where urban planning process are inadequate, and many residents have poor access to basic services and infrastructure.

Importantly, human factors are central to DT functionality, and so these platforms provide opportunities for stakeholders to be involved in contributing to their development and refinement. Equally, there is also a larger role for stakeholder engagement through processes of visualisation and showcasing, policy testing at local and national government levels communication of ideas and decision-making around investments (Slingshot Simulation, 2020). Thus, the potential for DTs to enhance sustainable decision-making and governance, through more informed policymaking for example, and not just via improved sectoral outcomes within certain SDGs, is also significant (ARUP, 2019; Deloitte Insights, 2020).

***5. Technological Advances in Data Science for Development Evaluation***

These broadscale technological advances in data science present significant opportunities for the field of development evaluation. In particular, they represent a means to tap into a rich and vast array of data sources that can support programme management, facilitate understanding, and improve the analytical power and capacity to assess environments that are fast moving, complex, and where the volume of information produced is constantly increasing (Bruce, Gandhi & Vandelenotte, 2020). For example, in recent years, there has been an almost exponential increase in the use and distribution of mobile phones, satellite and remote sensing, and digital banking (York & Bamberger, 2020). This is a phenomenon not restricted to the Global North but also widespread across the Global South. In Africa, in 2018, there were approximately 346 million registered mobile money accounts compared to just 120 million bank accounts, propelled by the success of mobile phone-based money transfer services such as M-PESA and M-Shawri in Kenya (Infomineo, 2020). This growth is set to continue, with estimates suggesting that the global digital banking system market will generate a revenue of $1,702 million between 2019 and 2026 (Research Dive, 2020).

At the same time, a core challenge regarding the evaluation of sustainable development, is that this is development that is happening across multiple spatial and temporal scales, within and across nations. In turn, these multi-scalar issues amplify the interconnected and interdependent reality of the SDGs, and the need to achieve them on a joint basis. This is particularly so, because many are mutually reinforcing, rely on the achievement of several goals, or trade-off one another (Mainali et al., 2018; Nilsson et al., 2018; van Soest et al., 2019). The difficulties extend beyond managing the complex interactions between SDGs, to include the wider macro-environmental influences that impact national-level capabilities. For example, resource flows from development funders; the effectiveness of social and financial levers; historical and contemporary social, economic, and demographic disparities; the dynamics of multinational value chains; international conventions; and geopolitics (Ofir, Schwandt, D’Errico, El-Saddik & Lucks, 2016).

This means that there is a prolific amount of data and information that the development evaluation community need to handle and make sense of, some of which is compiled on platforms like the Open SDG Data Hub, but much that has not been collated and compiled in comprehensive databases and repositories as it comes through many different channels and streams. Not only does this clearly argue for joint evaluations (Carugi and Bryant, 2020) to ensure key learnings are captured across the SDGs (Patton, 2019); it also makes a strong argument for the evaluation community to increasingly adopt the tools, instruments and technologies of data science. Principally, because the strength of data science resides in being able to handle this type of complex data, and address the complex evaluation questions that sustainable development interventions generate (York & Bamberger, 2020).

The point is not to see these new resources as substituting for traditional evaluation data gathering and analytical approaches; but rather, as additional tools for evaluations to exploit and take evaluation to the next level. Certainly, the uptake and use of these technologies for evaluation is to a certain degree still nascent. For example, whilst there has been progress in text analytics and text mining in relation to sentiment analysis, less well developed and therefore of lower current mainstream utility and usability, are approaches such as automatic modelling, machine translation and abstract summarisation (Bruce, Gandhi & Vandelenotte, 2020). Nonetheless, advances and improvements in these areas are happening rapidly, so the time to harness these technologies for improving evaluation innovation and effectiveness is now, lest evaluation and the evaluation community gets left behind. Because, as York & Bamberger (2020) note, evaluators have been much slower compared to other practitioners to take-up and adopt the new tools afforded by the data revolution.

*5.1 Examples of Data Science Applications in Action*

To understand Big Data’s use in evaluation, it is useful to consider three generic categories: human-centred; administrative; and geospatial Big Data. Human-centred Big Data is associated with, for example, consumer-related information or behaviours derived from sensors and the Internet of Things, social media, or surveys. Administrative Big Data typically relates to programme-based data, such as monitoring data collected by service agencies or records from government and public databases and archives. Geospatial Big Data refers to data obtained from satellites, drones, or remote sensing. Across all these areas, one of the key advantages of Big Data is that it has the possibility of obtaining data coverage on an entire population, vastly improving resolution and granularity and, combined with AI algorithms and data mining techniques, has the capacity to enhance decision-making and the accuracy of predictions related to the efficacy of specific development interventions (Bertemann, Robinson, Bamberger, Higdon & Raftree, 2020; Letouzé, Areias & Jackson, 2016; York & Bamberger, 2020).

From a monitoring perspective, the consequences of Big Data are far reaching: for example, improving the identification of hard to reach groups where development impacts could have most value; providing data on a broader array of indicators across social, economic and environmental domains and the development of longitudinal datasets; monitoring traditionally difficult to research areas such as governance or human rights; and enabling the construction of integrated data platforms (Bertemann, Robinson, Bamberger, Higdon & Raftree, 2020). From the perspective of evaluation, Big Data has the potential to increase the sample size of the population of interest, and to do so at scale over long time periods at drastically reduced costs and with increased efficiencies. Furthermore, Big Data also increases the potential of sharing learning more rapidly, as it can be gathered, analysed, and visualised in near real-time, whilst also improving the ease with which to validate and ground-truth findings. At the same time, there is the opportunity to aggregate and combine data of different forms (i.e. quantitative, semi-quantitative and qualitative) in new ways. Broadening immensely the availability of tools, such as predictive analytics, that evaluators can employ. In this regard, the utilisation of Big Data has the prospect of enabling ‘complexity-responsive evaluation designs’, and promoting the application of systems-based thinking to examine and interrogate complex situations (Bertemann, Robinson, Bamberger, Higdon & Raftree, 2020; Letouzé, Areias & Jackson, 2016; York & Bamberger, 2020).

In the world of impact evaluation, as Rathinam et al. (2020) point out, Big Data provides a means to broaden the evidence base across sustainable development sectors by expanding the availability of data. Especially, through harnessing satellite imagery and mobile phone details, enabling access to areas where traditionally evaluations have stumbled due to data deficiencies. Yet, in their survey of 437 studies utilising Big Data to evaluate development outcomes, only 11% were rigorous impact evaluations: suggesting a continued low-level of use of Big Data science. Nonetheless, within the 48 impact evaluations identified, the use of Big Data covered a wide range of areas from environmental sustainability, livelihoods, economic and urban development, to energy, industry and infrastructure, and health and wellbeing. Some examples include: the use of social network data to assess collective citizenship in Tanzania; mobile data to evidence the pros and cons of mobile money platforms in Afghanistan; crowd sourcing to increase community engagement in childhood immunisation in Uganda; corporate credit card details to appraise the efficacy of financial education programmes in Mexico; large-scale public agency data to assess transport infrastructure investment in subway systems in Beijing; satellite data to evaluate high speed rail and urban expansion in China, and the effectiveness of protected areas and incentive-based environmental management interventions in the Tropics (Rathinam et al., 2020)

Machine Learning and AI have considerable contributions to make in obtaining basic measurements required for evaluation within a sustainable development context, especially in instances where there remain significant data gaps such as in the area of economic livelihoods. An important example of this is in poverty mapping and prediction. The work of Blumenstock, Cadamus & On (2015) in Rwanda, demonstrate how historical mobile phone data (and associated meta-data) can be used to infer peoples’ socio-economic status, helping to predict levels of poverty and wealth. At the same time, this research also demonstrates the ease of gathering data that is low cost, localised and timely: important considerations in resource constrained environments. Similarly, Jean, Burke, Xie, Davis, Lobell & Ernon (2016) working across Nigeria, Tanzania, Uganda, Rwanda and Malawi show that neural networks, using limited training and publicly available satellite data (i.e. Earth observation and geospatial data), can identify elements of images that can explain close to 75% of the variation in economic outcomes. More recently, the Asian Development Bank has reported, with great success due to improvements in data granularity, the use of similar AI and ML applications to map poverty in the Philippines and Thailand (ADB, 2020).

In relation to the potential of Digital Twins in evaluation, the work in this area is very new and emergent and thus at a nascent stage of development. Due to the history of DT development, it has been mainly used in engineering, aerospace, and infrastructure prototyping. However, the possibilities for open and real-time evaluations of small and large interventions is immediately apparent, especially at the ex-ante and implementation stages of an intervention. One example discussed by Oughton (2018) is the evaluation of telecommunication policies: a DT of a real-world telecommunications network was developed to allow for the exploration of possible future sates under differing policy scenarios for the roll-out of digital broadband under a market-based or subsidized-based policy model. This work found that the testing of different broadband deployment strategies within a virtual market context, and the subsequent evaluation of their effectiveness, provided much greater transparency of decision-making, and over the long-term could lead to new knowledge and innovation through testing different policy options. It’s highly plausible that as the utilisation of DTs within the policy space increases, DTs could be used to generate virtual counterfactuals for interventions against which those interventions can be evaluated (e.g. via running different scenarios to assess the degree the object of evaluation is relevant, fit for purpose, efficient, effective, having impact, or sustainable).

***6. Challenges of Data Science for Sustainable Development and Evaluation***

Whereas the data science revolution holds much promise across the environmental, economic, and social pillars of sustainable development, a growing part of the narrative is exploring the limitations and pitfalls of over-reliance on technological innovations in data acquisition and processing to solve large, multiscale, complex and multifactorial challenges. Related, the degree to which the data science revolution has been adopted and applied in the field of evaluation has been slow relative to other fields, such as applied research to affect economic, social and environmental policy-level transformation. Below we examine the underlying reasons, first focusing on challenges to the application of data science in development, and then looking at challenges to its uptake in evaluation.

*6.1 Key application challenges of data science in sustainable development*

Del Río Castro et al., (2021) point out that the evidence base regarding the “digital paradigm” to sustainable development is still noticeably lacking, largely due to a variety of issues that include: a poor understanding of the complexities and interconnections between SDGs; flawed indicators and assessment procedures; insufficient coordination which undermines implementation and is exacerbated by government bureaucracy; and a failure to fully engage with new data innovations and technological advances. Elsewhere, the UN IEAG (2014) identifies several key concerns with the uptake of data science including, the potential for abuse of privacy and human rights, as well as the stark data divisions and inequalities between groups of people, communities, sectors, and countries. Its report recommends different sectors and groups of actors from across public, private, and civil society domains come together to ensure improvements in data availability, accessibility, quality, and safeguarding. Four priority areas are identified: 1) **Principles and standards** largely relate to data rights, privacy and protection (particularly in relation to vulnerable and minority communities), transparency, openness and curation (especially in relation to data gathering and the useability of data gathered), and data quality, integrity, disaggregation and timeliness (particularly in relation to avoiding bias and exclusion); 2) **Technology innovation analysis** focuses on addressing research gaps, improving data sharing and enhancing SDG monitoring capabilities through leveraging a greater degree of data sources; 3) **Governance leadership** concerns establishing multiscale networks of key actors and governments to coordinate, manage and implement principles and standards; and 4) **Capacity resources** address the issue of attracting investment to build effective SDG monitoring systems, technical capacity and data literacy, and leveraging private sector involvement.

The prescience of these issues remains, reflecting current concerns. For example, Vinuesa et al., (2020) point out the need for AI to be supported by regulatory frameworks and oversight to avoid lapses in transparency and accountability and the evasion of ethical standards. Similarly, Truby (2020) calls for regulation through an agreed international governance framework to reduce the probability of AI undermining SDG progress, with particular supervision of Big Tech. A recent report by the UN Economic and Social Commission for Asia and the Pacific (UN ESCAP) and Google recommends improved private sector and government partnerships in the application of AI to government programmes, whilst at the same time cautioning that governments need strong regulatory frameworks to regulate these partnerships, widen public access, and ensure they are culturally appropriate (UN ESCAP, Google, 2019). This highlights the tension between ‘global efficiency and local diversity’ in terms of centralised versus decentralised solutions (Hilbert & Mann, 2019), whilst also echoing the need for improved safety to avoid so-called ‘AI catastrophes’ (Vinuesa et al., 2020). In addition, several authors note the need to improve e-infrastructure, connectivity, and related technologies to prevent the marginalisation and disenfranchisement of less resourced communities, especially those in the Global South (Broadband Commission, 2017; World Bank, 2016). Similarly, there is a strong emphasis on the need to address bias and inequality in AI designs and Big Data gathering and curation. Such biases in the underlying algorithms used by AI prorgammes can magnify human biases, resulting in large-scale impacts on minority and protected groups for example. Hence, concurrently, there are calls for greater investments to improve the fidelity and ethics of data mining, machine learning and AI technologies (Truby, 2020; Vinuesa et al., 2020; 2030Vision, 2019).

*6.2 Key uptake challenges of data science utilisation in evaluation*

The reasons for the digital paradigm’s lack of application within evaluation largely concerns issues cultural, ethical, and economic matters.

1. **Culture**: The different worlds and value systems occupied by data scientists and analysts versus those of evaluators ultimately impacts its uptake and use in evaluation. Key differences include the role and importance placed on theory and conceptual framings in data collection and evaluation design. Data scientists are not largely driven by theory, whilst evaluators tend to be more theoretically informed. Going forwards, it has been argued that decisions regarding evaluation designs will increasingly been made by data scientists instead of professional evaluators. At the same time, due to a lack of technical capacity building across the evaluation community, many evaluators remain unfamiliar with data science tools and techniques, especially machine learning, predictive analytics, and probability modelling. This is exacerbated by poor institutional connections between data science experts and evaluation teams. Unfortunately, the combination of reduced technical competency and disconnection become self-reinforcing (Bertemann, Robinson, Bamberger, Higdon & Raftree, 2020; York & Bamberger, 2020).

Furthermore, evaluators are generally more concerned with developing causal explanations for the successes and failures of programme interventions, whilst many ML and AI applications based on Big Data rely more on the sheer volume of data to produce statistically significant correlations, which evaluators consider less useful for policymakers. That said, the constant stream of real time information provided by Big Data sources can deliver operationally valuable insights. However, this aspect may also prove to be inhibitory, because although evaluations can be iterative to allow for ongoing learning, exemplified by developmental evaluation, such iterative modes of working may still find it difficult to accommodate the near real-time stream of data. In these instances, data will likely require sufficient time to be processed and analysed, and to be made sense of, in order for informed decisions and judgements regarding the evaluation process or approach to be considered and implemented. The temporal issue can also be problematic in the sense that Big Data information can extend well beyond the time horizon of a single project or programme evaluation. Moreover, some contend that data science applications are still not sufficiently developed to adequately assess complex systems, which is perhaps also a symptom of a lack widespread and adequate monitoring infrastructure. (Bertemann, Robinson, Bamberger, Higdon & Raftree, 2020; York & Bamberger, 2020).

1. **Ethics**: This prominent area of concerns include operational, organisational, and political challenges of data privacy, sensitivity, accessibility, representation, missing data, data manipulation, and elite capture. Some of these challenges stem from the fact that ML processes can cause over fitting, such that the models generated are biased towards the ‘trained data’ and do not produce accurate and generalisable predictions on new data. On the other hand, because AI and ML require large amounts of data, making them excel statistically, they can faulter spectacularly on individual cases. These issues have been noted to result in racial profiling and stereotyping, creeping surveillance, unfair outcomes for particular communities, and a lack of redress for mistakes as a result of model complexity (Bertemann, Robinson, Bamberger, Higdon & Raftree, 2020; Hagendorff, 2020; USAID, 2019; York & Bamberger, 2020).
2. **Economics**: Evaluation is embedded in the political economy, and therefore subject to the same market and power forces that shape its evaluand, such as sustainable development. In other words, the political economy of development shapes the industry of evaluation, and this includes the uptake and application of data science in evaluation. Some of the most fundamental practices of evaluators and evaluation commissioners are at odds with innovation in complex systems, such as the uptake of data science. While there may be rhetoric to complexity adaptive evaluation, for the most part evaluations are primarily commissioned and used for accountability purposes rather than adaptive learning, innovation, and improvement (Williams, 2015).

The emphasis on accountability favours commissioning methodologies and technologies for simple designs or framings of interventions, where measurement is more narrowly focused on a linear chain of desired results confined to a particular timeline dictated by funding cycles. As such, evaluators are accustomed and inclined to use experimental and quasi-experimental approaches that focus on the causal relationship between intended results by controlling other factors that could otherwise explain change. As such, evaluators are more accustomed to working with smaller population samples rather than large data sets that are steaming in real time.

In its study on the impact of the commissioning process on complexity-appropriate evaluation, the Centre for the Evaluation of Complexity Across the Nexus (CECAN) identify several barriers that inhibit complexity-appropriate methods (Cox, 2019). One important source of barriers relates to risk-adverse attitudes related to: experimentation amidst tight budgets and timelines; insufficient knowledge about and capacity in new methods and technologies to commission and manage complexity approaches; concerns as to whether the approach will deliver and needing to justify methods to key stakeholders; and the overall risk of assessing delivery of unknown evaluation methods. Such concerns reflect the market demand for accountability to the detriment of more innovative, technologies and data science.

***7. Moving Forward***

We have outlined some of the major challenges that limit the uptake and potential of data science in sustainable development and evaluation. However, the transformative potential and joint benefits of data science for development evaluation remain clear: improved modelling capabilities; better service delivery through enhanced targeting and granularity; broadly reduced levels of bias; efficiency gains via cost reductions and, just as important, the ability to have (near) real-time information on the system of interest (Bertemann, Robinson, Bamberger, Higdon & Raftree, 2020). The latter point is particularly important, because real-world development, especially super-wicked problems, are messy and do not fit well with predefined timeframes and funding cycles. The availability of real-time data can help shift evaluation away from narrow approaches focused on predetermined timeframes of purposive theories of change and; instead, adopt more real-time adaptive evaluation, monitoring ***as*** evaluation, that provides immediate feedback so that interventions can be nimbler and more responsive to change (AUTHOR, XXXZ). New developments, such as the complex evaluation framework, developed by the UK Government Department for Environment, Food and Rural Affairs in conjunction with CECAN, can aid in this process by providing commissioners of evaluation and evaluators with a practical structure for designing and navigating complex evaluations in complex policy contexts, and affording a basis for drawing-on data science technologies and applications as part of that process (CECAN Ltd, 2019).

At the same time, it is important to note that the direction of flow is not simply from data science to evaluation, it is also apparent that the evaluation field has important lessons to improve the utilisation of data science within evaluation. As Bamberger and York (2020), summarize, these lessons include the need to: (i) openly address data quality and validity, including construct validity; (ii) address social exclusion and sample bias; (iii) consult with a broad range of stakeholders to establish key evaluation questions of concern that data will need to address; (iv) reconsider the value of theory in data gathering, curation and analysis; (v) apply mixed method approaches combining traditional methods with new data science techniques; and (vi) recognise the importance of ground truthing what in many respects will be data streams not explicitly captured for evaluation purposes.

Returning to the premise of our article, we hope to have shown, firstly, that data science has a valuable transformational role to play in sustainable development evaluation, and that evaluation as a profession can contribute to that endeavour; and secondly, that evaluation is an innovative and vital field able to address complex pressing issues and questions that are central to the future of society in the Anthropocene.

***8. References***

2030Vision – Global Goals Technology Forum (2019) *AI & The Sustainable Development Goals: The state of play*. SustainAbility, London.

Alber, M., Tepole, A.B., Cannon, W.R., De, S., Dura Bernal, S., Garikipati, K., et al. (2019). Integrating machine learning and multiscale modelling – perspectives, challenges, and opportunities in the biological, biomedical, and behavioural sciences. *npj Digital Medicine*, 2:115

Andersen, I., & Rockström, J. (June 5, 2020) COVID-19 Is a Symptom of a Bigger Problem: Our Planet’s Ailing Health. *TIME Magazine*. Available at: <https://time.com/5848681/covid-19-world-environment-day/>

ARUP. (2019). *Digital Twin: Towards a meaningful framework*. ARUP, London.

Asian Development Bank (2020) *Mapping Poverty through Data Integration and Artificial Intelligence: A Special Supplement of the Key Indicators for Asia and the Pacific*. ISBN: 978-92-9262-313-5

Bamberger, M., Vaessen, J., & Raimondo, E. (2016). *Dealing with Complexity in Development Evaluation – A Practical Approach*. SAGE publications.

Barricelli, B.R., Casiraghi, E., Gliozzo, J., Petrini, A., & Valtolina, S. (2020). Human digital twin for fitness management. *IEEE Access*, 8, 26637-26664

Befani, B., Ramalingam, B., & Stern, E. (2015). Introduction – Towards Systemic Approaches to Evaluation and Impact. *IDS Bulletin* 46(1).

Bertermann, K., Robinson, A., Bamberger, M., Higdon, G.L., & Raftree, L. (2020). *Big data to data science: Moving from “what” to “how” in the MERL Tech Space*. MERL Tech.

Bicket, M., Christie, I., Gilbert, N., Penn, A., Hills, D., & Wilkinson, H. (2020). Complexity, and what it means for policy design, implementation, and evaluation. A CECAN Evaluation Policy and Practice Note for policy analyst and evaluators. Note No. 16. Centre for the Evaluation of Complexity Across the Nexus

Blumenstock, J., Cdamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone meta data. *Science*, 350(6264), 1073-1076

BMU and WRI. (2020). *Global dialogue on responding to the COVID-19 pandemic and economic crisis: Building back better aligned to the SDGs and the Paris Agreement*. Federal Ministry of the Environment, Nature Conservation and Nuclear Safety; and the World Resources Institute.

Broadband Commission for Sustainable Development. (2017). *The state of broadband: Broadband catalyzing sustainable development*. ITU and UNESCO.

Bruce, K., Gandhi, V.J., & Vandelenotte, J. (2020) *Emerging Technologies and Approaches in Monitoring, Evaluation, Research, and Learning for International Development Programs*. MERL Tech.

Bruynseels, K., Santoni de Sio, S., & van den Hoven, J. (2018). Digital twins in healthcare: Ethical implications for an emerging engineering paradigm. *Frontiers in Genetics*, 9:31

Byrne, D. (2013). Evaluating complex social interventions in a complex world. *Evaluation*, 19(3), 217-228

Caffrey, L., & Munro, E. (2017). A systems approach to policy evaluation. *Evaluation*, 23, 463–478.

Carugi, C., & Bryant, H. (2020). A joint evaluation with lessons for the Sustainable Development Goals Era: The joint GEF-UNDP evaluation of the small grants programme. *American Journal of Evaluation*, 41, 182-200

CECAN Ltd (2019) Complexity Evaluation Framework: Recognising complexity and key considerations for complexity-appropriate evaluation in the Department for Environment, Food and Rural Affairs, DEFRA. Commissioned by the Department for Environment, Food and Rural Affairs (Defra) Delivered by CECAN Ltd with Risk Solutions and the Tavistock Institute of Human Relations.

AUTHOR (XXXZ)

AUTHOR (XXXX)

 Gilpin-Jackson, Marguerite Welch, & Mitsunori Misawa. Palgrave MacMillan.

Cox, J. (2019). How does the commissioning process inhibit the uptake of complexity-appropriate evaluation? Centre for the Evaluation of Complexity Across the Nexus.

Del Río Castro, G., Fernández, MC.G., & Colsa, A.U. (2021). Unleashing the convergence and digitalization and sustainability towards pursuing the Sustainable Development Goals (SDGs): A holistic review. *Journal of Cleaner Production*, 280:122204

Delli Paoli, A., Addeo, F., Bevilacqua, C., Calabrò, F., & Della Spina, L. (2021) Big Data to Support Sustainable Development Goals (SDGs). *Smart Innovation, Systems and Technologies*, 178, 738-748

Deloitte Insights. (2020). *Tech Trends 2020*. Deloitte Development LLC.

Dodds, F., & Bartram, J. (Eds.) (2016) *The water, food, energy and climate Nexus: Challenges and an agenda for action*. Earthscan, Routledge. London

El Saddik, A. (2018). Digital twins: The convergence of multimedia technologies. *IEEE Multimedia*, 25(2), 87-92

Eppel, E.A., & Rhodes, M.L. (2018). Complexity theory and public management: a ‘becoming’ field. *Public Management Review*, 20, 949-959

Eurostat, (2019) Multi-purpose indicators within the EU SDG indicator set.gif. <https://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Multi-purpose_indicators_within_the_EU_SDG_indicator_set.gif&oldid=437034>

Everard, M., Johnston, P., Santillo, D., & Staddon, C. (2020). The role of ecosystems in mitigation and management of Covid-19 and other zoonoses. *Environmental Science & Policy*, 111, 7-17

Frierreira, B., Iten, M., & Silva, R.G. (2020). Monitoring sustainable development by means of earth observation data and machine learning: A review. *Environmental Sciences Europe*, 32: 120

Fu, B., Wang, S., Zhang, J., Hou, Z., & Li, J. (2019). Unravelling the complexity in achieving the 17 sustainable-development goals. *National Science Review*, 6(3), 386–388

Gartner (2020) Gartner Glossary: Digital Twin [online] <https://www.gartner.com/en/information-technology/glossary/digital-twin> (accessed October 2020)

Gates, E.F. (2016). Making sense of the emerging conversation in evaluation about systems thinking and complexity science. *Evaluation and Program Planning*, 59, 62–73.

Geyer R (2012) Can complexity move UK policy beyond ‘evidence–based policy making’ and the ‘audit culture’? Applying a ‘complexity cascade’ to education and health policy. *Political Studies*, 60, 20–43

Ghaffarzadegan, N., Lyneis, J., & Richardson, G.P. (2011). How small system dynamics models can help the public policy process. *System Dynamics Review*, 27(1), 22-44

Giest, S. (2017). Big data for policymaking: Fad or fastrack? *Policy Science* 50, 367-382

Goralski, M.A., & Tan, K.T. (2020). Artificial intelligence and sustainable development. *The International Journal of Management Education*, 18:100330

Greyer, R., & Cairney, P. (eds.) (2015) *Handbook on Complexity and Public Policy*. Edward Elgar Press.

Hagendorff, T. (2020). The Ethics of AI Ethics: An evaluation of guidelines. *Minds and Machines*, 30, 99-120

Head, B.W. (2019). Forty years of wicked problems literature: forging closer links to policy studies. *Policy and Society*, 38, 180-197

AUTHOR (XXXY)

Hopson, R., & Cram, F. (Eds.) (2018). *Tackling Wicked Problems in Complex Ecologies: The Role of Evaluation.* Stanford University Press

Hummelbrunner, R. (2015) Learning, Systems Concepts and Values in Evaluation: Proposal for an Exploratory Framework to Improve Coherence. *IDS Bulletin*, 46(1), 17-29

IATT (2019) (eds.). *Science-policy briefs on the impact of rapid technological change on the SDGs*. Contributions to the Technology Facilitation Mechanism by individual experts. New York: Interagency Task Team on Science, Technology and Innovation for the Sustainable Development Goals.

IBM (2020) What is a digital twin? [online]: <https://www.ibm.com/topics/what-is-a-digital-twin> (accessed October 2020)

IDEAS (2019). Prague Declaration on Evaluation for Transformational Change. Adopted by the International Development Evaluation Association (IDEAS), October 4, 2019.

Infomineo 2020 <https://infomineo.com/digital-banking-in-africa-origins-and-development-outlook/>

Jean, N., Burke, M., Xie, M., Davis, M.W., Lobell, D.B., & Ermon, S. (2016). Combining satellite imaging and machine learning to predict poverty. *Science*, 353(6301), 790-794

Jones, D., Snider, C., Nassechi, A., Yon, J., & Hicks, B. (2020). Characterising the Digital Twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology*, 29, 36-52

Kong, L., Liu, Z., & Wu, J. (2020). A systematic review of big data-based urban sustainability research: State-of-the-science and future directions*. Journal of Cleaner Production*, 273:123142

Kovacic, Z. (2018). Conceptualizing Numbers at the Science–Policy Interface. *Science, Technology, & Human Values*, 43, 1039-1065

Lambert, H., Gupte, J., Fletcher, H., Hammond, L., Lowe, N., Pelling, M. et al., (2020). Covid-19 as a global challenge: Towards an inclusive and sustainable future. *The Lancet Planetary Health*, 4, e312-e314

Letouzé, E., Areias, A. & Jackson, S. (2016). The evaluation of complex development interventions in the age of big data1. In Bamberger, M., Vaessen, J., & Raimondo, E. *Dealing with complexity in development evaluation* (pp. 221-250). Thousand Oaks, CA: SAGE Publications

Letouzé, F., Stock, M., Chiara, F. De., Lizzi, A., & Mazariegos, C. (2018). *Harnessing innovation data and technology to measure development effectiveness*. Occasional Paper Series No. 54. Southern Voice.

Levin, K., Cashore, B., Bernstein, S., & Auld, G. (2012). Overcoming the tragedy of super wicked problems: constraining our future selves to ameliorate global climate change. *Policy Sciences*, 45, 123–152

Mainali. B., Luukkanen, J., Silveira, S., & Kaivo-oja, J. (2018). Evaluating synergies and trade-offs among Sustainable Development Goals (SDGs): Explorative analyses of development paths in South Asia and Sub-Saharan Africa*. Sustainability,* 10: 815

Mowles, C. (2014). Complex, but not quite complex enough: The turn to the complexity sciences in evaluation scholarship. *Evaluation*, 20(2), 160-175

Muller, B. (2019, in press) Why public policies fail: Policymaking under complexity. *Economia* (<https://doi.org/10.1016/j.econ.2019.11.002>)

Nadai, M. De., Letouzé, E., González, M.C., & Lepris, B. (2018). Characterising and analysing urban dynamics in Bogota. AFD Research Paper Series, No. 2018-70

Nara, E.O.B., da Costa, M.B., Basierk, I.C., Schaefers, J.L., Benitez, G.B., Lima do Santos, L.M.A., & Benitez, L.B. (2021). Expected impact of industry 4.0 technologies on sustainable development: A study in the context of Brazil’s plastic industry. *Sustainable Production and Consumption*, 25, 102-122

Nilsson, M., & Eckerberg, K. (Eds.) (2009) *Environmental Policy Integration in Practice: Shaping Institutions for Learning*. Earthscan, London – Sterling, VA.

Nilsson, M., Chisholm, E., Griggs, D., Howden-Chapman, P., McCollum, D., Messerli, P., et al., (2018). Mapping interactions between the sustainable development goals: lessons learned and ways forward. *Sustainability Science*, 13, 1489–1503

O’Brien, D. (2014). Cultural value, measurement and policymaking. *Arts and Humanities in Higher Education*, 14(1), 79-94

Ofir, Z., Schwandt, T., D’Errico, S., El-Saddik, K., & Lucks, D. (2016). Five considerations for national evaluation agendas informed by the SDGs. IIED Briefing. International Institute of Environment and Development, London.

Oughton, E. (2018). Building a digital twin: Testing the effectiveness of telecommunication policies in a virtual world. TPRC 46: The 46th Research Conference on Communication Information and Internet Policy 2018.

Patton, M.Q. (2018). A historical perspective on the evolution of evaluative thinking. In Vo, A.T. and Archibald, T. (Eds.) Evaluative Thinking. *New Directions for Evaluation*, 158, 11–28

Patton, M.Q. (2019). Transformation to global sustainability: Implications for evaluation and evaluators. In G. Julnes (Ed.), Evaluating Sustainability: Evaluative Support for Managing Processes in the Public Interest. *New Directions for Evaluation*, 162, 103–117

Piascik, B., Vickers, J., Lowry, D., Scotti, S., Stewart, J., & Calomino, A. (2010) *Draft Materials, Structures, Mechanical Systems, and Manufacturing Roadmap.* Technology Area 12. National Aeronautics and Space Administration.

Pichler, M., Schaffartzik, A., Haberl, H., & Gӧrg, C. (2017). Drivers of society-nature relations in the Anthropocene and their implications for sustainability transformations. *Current Opinion in Environmental Sustainability*, 26–27, 32–36

Quayle, A., & Kelly, B. (2019). Building informal knowledge-sharing relationships between policy makers and academics: Insights from a PM&C engagement project. *Australian Journal of Public Administration*, 78, 311–318

Raftree, L., & Bamberger, M. (2014). *Emerging opportunities: Monitoring and evaluation in a tech-enabled world*. The Rockefeller Foundation Evaluation Office.

Rathinam, F, Khatua, S, Siddiqui, Z, Malik, M, Duggal, P, Watson, S, Vollenweider, X. 2020. Using Big Data for evaluating development outcomes: a systematic map [Online]. 3ie. Available at: https://gapmaps.3ieimpact.org/evidence-maps/big-data-systematic-map

Research Dive (2020) Coronavirus Pandemic Impact on Digital Banking System Market Report: https://www.researchdive.com/covid-19-insights/255/digital-banking-system-market

Rutter, J. (2012). *Evidence and evaluation in policymaking: A problem of supply and demand?* Institute for Government, London.

Sachs, J.D., Schmidt-Traub, G., Muzzicato, M., Messner, D., Nakicenovic, N., & Rockstrӧm, J. (2019). Six transformations to achieve the Sustainable Development Goals. *Nature Sustainability*, 2, 805-814

Schleich, B., Anwer, N., Matthieu, L., Wartzack, S. (2017). Shaping the digital twins for design and production engineering. *CIRP Annals – Manufacturing Technology*, 66(1), 10.1016/j.cirp.2017.04.040

Schwandt, T. A., Ofir, Z., Lucks, D., El-Saddik, K., & D’Errico, S. (2016) Evaluation: A Crucial Ingredient for SDG Success. An IIED Briefing. International Institute for Environment and Development, London.

Schwarz, B., Tellman, B., Sullivan, J., Kuhn, C., Mahtta, R., Pandey, B. et al. (2017). Socio-physical vulnerability to flooding in Senegal: An exploratory analysis with new data and Google Earth Engine. AFD Technical Report No. 2017-25.

Slingshot Simulation (2020) website: <https://www.slingshotsimulations.co.uk/digital-twins/digital-twin-series-coming-soon/> (accessed October 2020)

Steffen, W., Rockström, J., Richardson, K., Lenton, T.M., Folke, C., Liverman, D. et al. (2018). Trajectories of the Earth System in the Anthropocene. *Proceedings of the National Academy of Sciences*, 115, 8252-8259

Sterner, T., Barbier, E.B., Bateman, I., van den Bijgaart, I., Crépin, A-S., Edenhofer, O., et al. (2019). Policy design for the Anthropocene. *Nature Sustainability*, 2, 14-21

Talbot, C., & Talbot, C. (2015). Bridging the academic–policymaking gap: practice and policy issues. *Public Money & Management*, 35(3), 187-194

Trendov, N.M., Varas, S., & Zeng, M. (2019). *Digital technologies in agriculture and rural areas: Briefing Paper.* Food and Agricultural Organisation, Rome.

Truby, J. (2020). Governing artificial intelligence to benefit the UN Sustainable Development Goals. *Sustainable Development*, 28, 946-959

UN (2012) *The future we want: outcome document*. United Nations, New York. <https://sustainabledevelopment.un.org/rio20/futurewewant>.

UN ESCAP and Google. (2019). *Artificial intelligence in the delivery of public services*. United Nations Economics and Social Commission for Asia and the Pacific and Google. Bangkok, Thailand.

UN IEAG. (2014). *A world that counts: Mobilising the data revolution for sustainable development*. A report prepared at the request of the United Nations Secretary-General, by the Independent Expert Advisory Group on a data revolution for sustainable development.

UN. (2013). *World Economic and Social Survey 2013: Sustainable Development Challenges*. Department of Economic and Social Affairs. United Nations, New York

UNRISD. (2016). *Policy Innovations for Transformative Change: UNRISD Flagship Report 2016*. United Nations Research Institute for Social Development.

USAID. (2019). *Reflecting the past, shaping the future: Making AI work for international development*. United States Agency for International Development.

van Soest, H.L., van Vuuren, D.P., Hilaire, J., Minx, J.C., Krey, V., Popp, A., et al., (2019). Analysing interactions among Sustainable Development Goals with Integrated Assessment Models. *Global Transitions*, 1, 210-225

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domerscly, S., et al., (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, 11:233

Waddock, S., Waddell, S., Goldstein, B., Linner, B., Schapke, N., & Vogel., C. (2020). Transformation: How to spur Radical Change. In Scrutton, A. (Ed), *Our Future on Earth Report*, Future Earth. (pp.82-90).

Walton, M. (2016). Expert views on applying complexity theory in evaluation: Opportunities and barriers. *Evaluation*, 22, 410–423

WCED (1987). *Our Common Future*. World Commission on Environment and Development. Oxford University Press.

Williams, B. (2015) Prosaic or profound? The adoption of systems ideas by impact evaluation. *IDS Bulletin*, 46(1), 7-16

Williams, B., & Sjon van ’t Hof. (2016). *Wicked Solutions. A Systems Approach to Complex Problems* (2nd ed.).

World Bank. (2015). *World Development Report 2016: Digital Dividends Overview*. A World Bank Group Flagship Report. International Bank for Reconstruction and Development/The World Bank, Washington D.C.

Yayboke, E., Nealer, E., & Rice, C. (2017). *Harnessing the data revolution to achieve the Sustainable Development Goals: Enabling frogs to leap.* A report of the CSIS project on prosperity and development with the JICA Research Institute.

**List of Figures and Tables**

***Figure 1.*** *Illustrative mapping of the interactions between SDGs. Figure provided courtesy of Eurostat (2019)*



|  |
| --- |
| Table 1. Selected definitions of Digital Twins |
| Definitions | Source |
| A digital twin is a digital representation of a real-world entity or system. The implementation of a digital twin is an encapsulated software object or model that mirrors a unique physical object, process, organization, person, or other abstraction. Data from multiple digital twins can be aggregated for a composite view across a number of real-world entities, such as a power plant or a city, and their related processes. | Gartner (Research and Advisory company) (2020) |
| A digital twin is a virtual representation of a physical object or system across its lifecycle, using real-time data to enable understanding, learning, and reasoning. | IBM (2020) |
| A digital twin can be defined, fundamentally, as an evolving digital profile of the historical and current behaviour of a physical object or process that helps optimize business performance. The digital twin is based on massive, cumulative, real-time, real-world data measurements across an array of dimensions. | Deloitte Insights (2020) |
| A digital twin is the combination of a computational model of a real-world system, designed to monitor, control, and optimise its functionality. Through data and feedback, both simulated and real, a digital twin can develop capacity for autonomy and to learn and reason about its environment. | ARUP (2019) |
| A digital twin is a digital representation of an active unique product (real device, object, machine, service, or intangible asset) or unique product-service system (a system consisting of a product and a related service) that comprises its selected characteristics, properties, conditions, and behaviours by means of models, information, and data within a single or even across multiple life cycle phases. | CIRP Encyclopeadia of Production Engineering (2019) |
| A digital twin is a digital replica of a living or non-living physical entity. By bridging the physical and the virtual world, data is transmitted seamlessly allowing the virtual entity to exist simultaneously with the physical entity. | El Saddik (2018) |
| In synthesis, the vision of the digital twin describes the vision of a bi-directional relation between a physical artifact and the set of its virtual models. In this context, the virtual “twinning,” i.e., the establishment of such relations between physical parts and their virtual models, enables the efficient execution of product design, manufacturing, servicing, and various other activities throughout the product life cycle. | Schleich et al., (2017) |

***Figure 2.*** *Relationships between key components of Data Science. Image courtesy of Whatsthebigdata*