GIZ Guidance Tool for the Use of Alternative Data for Shock-Responsive Social Protection Systems

March 2023
ABOUT THIS TOOL

This tool was developed through close collaboration between Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) and Oxford Policy Management (OPM) by Emily Aiken and Tim Ohlenburg.
INTRODUCTION

SUMMARY BY DATA SOURCE
- Satellite imagery
- Mobile phone data
- Web and social media data
- Financial services data
- Digitised administrative data

SUMMARY BY DELIVERY CHAIN TASK
- Outreach
- Intake and registration
- Eligibility determination
- Provision of benefits
- Beneficiary management

KEY QUESTIONS THAT APPLY TO ALL DATA SOURCES
- Outreach, intake, and registration
- Eligibility determination
- Provision of benefits
- Beneficiary management: Communication and interaction
- Administration and digital governance

USEFUL (RE)SOURCES AND REFERENCES
INTRODUCTION
Novel data sources have become a topic of widespread interest in social protection. The deployment of satellites, phones, and social media, as well as the increasing use of non-social protection, administrative data sources during the COVID-19 pandemic, has brought the subject to the attention of policymakers around the globe. Interest in novel data sources is complementary to other trends in public administration, which include the digitalisation of public services, the growing acceptance and use of machine learning methods, and the shift to digital engagement in public service delivery. Due to the need for social distancing, the pandemic accelerated the shift towards digital engagement. Unable to conduct operations as before, several countries experimented with fully digital social protection programmes. Novel data sources played an important role in some of these and sparked interest in their application among policymakers elsewhere. Another factor is the convergence of social protection and humanitarian assistance embodied in the field of adaptive social protection, for which novel data sources can be especially useful. In view of these developments, the question arises: for which situations are novel data sources appropriate?

The purpose of this guidance tool is to provide a framework for an initial assessment of whether or not a certain data source may add value in a specific programme context and for which part of the delivery chain. In line with this objective, our main target audience is policymakers and practitioners from low- and middle-income countries who are interested in exploring novel data options for use in their context. The data sources we discuss are (i) satellite imagery, (ii) mobile phone data, (iii) web and social media data, (iv) financial data and (v) administrative data. Each type offers distinct opportunities and suffers from particular limitations that preclude certain applications. One aspect of this guidance tool is to summarise the data sources in terms of their basic features and potential applications, to provide an overview of the general landscape. The other aspect is to provide a framework for possible applications that are worth further investigation in a particular policy context, including the assessment of essential questions that potential programmes would need to address.

We have chosen the term ‘guidance tool’ for this report, as it is meant to be pragmatic and practical. However, there are at least three reasons why ‘guidance’ must remain fairly general in order to be useful to a diverse set of practitioners. Firstly, policymakers, including those from the same country, often have divergent requirements. Similarly, social protection systems differ widely among countries, as do the data ecosystems from which novel data sources can be drawn. Secondly, the diversity of sources provides a broad overview of the field, but it also requires the succinct treatment of each type in order to maintain the readability of the report. The third factor is
that there is limited experience with some novel data sources. This factor necessarily distinguishes this guidance tool from publications, such as the ISPA Tools¹, which provide more in-depth advice on established areas of practice. Their specificity is made possible by the extensive accumulated experience in the use of the topics they discuss.

Since the use of novel data sources, by definition, is relatively untested, the aim of this tool is to set the stage for further investigation. Once policymakers have identified suitable opportunities on this basis, they will need to chart their own course, as the experiences of peers can provide relevant insights, but no definite direction. In particular, there is limited experience with integrating novel data in existing social protection systems. During the COVID-19 pandemic, several high profile programmes that leveraged novel data – Togo’s Novissi 2.0, STEP-KIN in the Democratic Republic of the Congo (DRC), and the Social Relief of Distress Grant (SRD-350) in South Africa – were implemented in a fully digital format. Although shown to be viable in an emergency setting, they exposed challenges to do with access and inclusion, suggesting that there is an important place for traditional delivery channels in non-emergency social protection systems. Accordingly, we present potential novel data applications in the modular format of the social protection delivery chain, and recommend linkages that may allow the realisation of the potential of novel data while mitigating the associated limitations.

The structure of this report is as follows: Section 2 provides a summary of each type of data, based on a recent report on Novel Digital Data Sources for Social Protection: Opportunities and Challenges by the same authors (Aiken & Ohlenburg, 2023), presenting similar material in a slightly condensed format. Section 3 collates the limited existing evidence of novel data source applications along the social protection delivery chain (Lindert et al., 2020): outreach, intake and registration, eligibility determination, benefit provision, and beneficiary management are all considered. In addition to known applications, potential uses that are likely to be fruitful and appear realistic are also discussed. Section 4 rounds out the discussion by proposing cross-cutting questions for policymakers wishing to design and implement data-driven applications of the kind discussed in this report. Finally, Section 5 provides a number of suggestions for further reading that offer a more detailed treatment of key topics.

It is an exciting time for the social protection community. Rapid technological progress is transforming practices in this major field of public policy. The rapid pace of innovation during the COVID-19 pandemic was inspiring, as it revealed the potential for major change, as long as the right solution is chosen for the task at hand. However, harnessing innovations safely and effectively is no mean feat, given the lack of an evidence base to draw on. We hope that this tool will provide useful guidance for the first steps on a path towards integrating novel data sources responsibly and effectively in support of the vulnerable and disadvantaged.

¹ https://ispatools.org/all-tools/
SUMMARY BY DATA SOURCE
This section reviews the selection of data sources around which the guidance tool revolves. Satellite imagery, mobile phone data, and web and social media have emerged as sources recently and are beginning to have an impact on social protection systems. We also review financial services data and digitalised administrative data, as these are likely to be increasingly used in the digital delivery of social protection. Each data type is discussed in terms of the opportunities it offers, the data creation process and format that the data is generated in, the data sources and considerations necessary to access it, and, finally, salient issues of processing and capacity required to make use of it as part of a digital system. All data sources have inherent limitations and carry risks that need to be mitigated. To avoid repetition, these are addressed in the following section on implementation options along the delivery chain, as well as in the final section on key questions and issues to be considered in implementation. The material in this section is based on Aiken and Ohlenburg (2023), which provides a broader overview of novel data sources and goes into additional detail on some case studies.

**SATELLITE IMAGERY**

**Opportunities:** Satellite imagery is broadly applicable in social protection. It can inform a set of targeting methods, including geographic targeting, household-level targeting, and shock assessment for more adaptive and responsive social protection. It can also be layered with other targeting methods for hybrid targeting approaches that include a geographic component. Beyond the targeting method selected (and its eligibility criteria and qualifying conditions), satellite imagery can also more simply help to prioritise administrative capacity, outreach and registration efforts to areas that are particularly poor and vulnerable.

**Data creation & format:** The satellites of NASA and the European Space Agency are recording high-resolution images of the entire Earth’s surface every 10 to 14 days, collecting and storing unprecedented amounts of information about human behaviour and planetary conditions. Much of this imagery encodes high resolution information about population density, economic wellbeing, and environmental conditions, which can inform social protection policy – as long as organisations have the capacity to obtain, store, process, and learn from the data.

Satellite imagery is usually released as a set of multispectral bands. Applications using daytime satellite imagery typically involve the red, green, and blue bands, as well as sometimes near-infrared, infrared, and thermal bands. ‘Night-time lights’ imagery – images of the world at night that record the intensity of light emitted from each satellite tile – are a separate band, with publicly available datasets released
by the United States Geological Survey (USGS) and NASA alongside the Landsat day-time dataset.

**Sources & access:** There are two main sources of free and publicly available satellite imagery: (i) the Landsat dataset, from USGS and NASA, which covers the Earth’s surface at a 30 metre per pixel resolution with updates every two weeks, and (ii) the Sentinel dataset, from the European Union and the European Space Agency, which provides 10 metre per pixel resolution and refreshes every 10 days. There are also paid services that provide higher resolution and higher frequency updating of imagery: Planet and Digital Globe both provide up to sub-metre resolution satellite imagery, with updating on a daily to weekly basis. At present, the cost of private imagery sources ranges from 1 to 30 USD per kilometre.

While publicly available satellite imagery can meet some social protection needs, it is not as high resolution or as frequently updated as private imagery. However, access to private imagery may be prohibitively expensive for some programmes, with costs running into the millions of dollars for imaging a nation. Partnerships between private satellite data holders, governments, researchers, and multilaterals and humanitarian organisations may help ease cost burdens, particularly in cases where satellite-based estimates are publicly available or being used for related humanitarian tasks (e.g. Becker-Reshef *et al.*, 2020; Chi *et al.*, 2021; Rolf *et al.*, 2021) or other endeavours.

**Processing & capacity:** Processing satellite imagery requires substantial capacity investment. While some projects continue to examine satellite imagery by hand, many projects have turned to using machine learning methods to extract structured information. Neural networks are commonly used to draw inferences from satellite imagery, but have significant barriers to entry in terms of both human capacity and computational capacity. On the human side, these methods require substantial knowledge of computer science and machine learning to implement; on the computational side, they are costly in terms of memory and time, with most projects paying for cloud-based services to speed up computation. As one approach to easing this challenge, MOSAIKS, a set of pre-computed low-dimensional satellite ‘features’ that can be used in place of the images themselves for satellite-based applications (Rolf *et al.*, 2021), has a substantially lower barrier to entry than the standard satellite imagery pipeline, even if it does not have the spatiotemporal resolution of other satellite data sources.

---

**MOBILE PHONE DATA**

**Opportunities:** Mobile datasets have considerable potential for social protection. The focus here is on mobile phone metadata as a data source, as their primary function as a communication device can facilitate a variety of uses. First, mobile phones can serve as an avenue for outreach and registration, and mobile phone data can be integrated with outreach platforms for targeted outreach. In addition, mobile phone data can be used to measure mobility and economic wellbeing for targeting at the community and individual levels. Finally, as a high-frequency data source, mobile phone data has the potential to bring value to programme monitoring and evaluation.

**Data creation & format:** Mobile phone data can be broadly divided into two categories: the data traces recorded when subscribers use smartphones and the data traces recorded when subscribers use feature phones. As smartphone penetration continues to
increase globally, there is potential in the future to leverage data recorded by smartphones for social protection, such as fine-grained location traces recorded by mapping applications, data on application usage time and intensity, and footage from smartphone cameras. However, at present, while smartphone use is widespread in rich countries, the majority of subscribers in low- and middle-income countries still rely on feature phones (GSMA, 2021). As a result, most of the present potential for social protection initiatives to leverage mobile phone data centres on the metadata recorded in basic mobile transactions that are possible without a smartphone.

Data recorded by mobile network operators, called ‘call detail records’ (CDRs), consists of relational databases of information about each call or SMS transaction placed on a mobile network from any phone (smartphone or feature phone). A single record typically includes the phone number of the caller and recipient, the time and date of the transaction, and the mobile antennas through which the call was placed and received. CDRs for calls also typically include the duration of the interaction. Information on the location of mobile network towers can be merged with the antenna information recorded in the CDR to obtain information about the spatial movement of subscribers. Mobile network operators also record top-ups to the mobile balance of subscribers and other financially relevant transactions.

**Sources & access:** Mobile phone metadata is recorded by mobile network operators. The regulatory ecosystem of the mobile sector is unique to each country, and data access constraints vary from country to country and operator to operator. Country-specific data protection regulations provide a framework for operations that also need to comply with the stipulations of subscriber contracts. A major complication of working with mobile phone data is the presence of multiple network operators in most markets. It may be difficult to obtain information from all operators to harmonise data and avoid excluding subscribers. Unless a legal framework or strong institutions that promote cooperation are in place, programmes may be unable to agree on access with all operators, due to both commercial and competitive considerations. Data sharing agreements which include non-disclosure clauses and mandate safe storage typically need to be negotiated with each operator, and possibly be coordinated across operators.

**Processing & capacity:** CDR data requires a commercial-scale server or cloud processing infrastructure, due to the large data volume generated by networks with millions of users. In terms of human resources, the database structure of CDR facilitates processing and requires less specialist skills than satellite imagery, for example. The open source library CIDER² provides tools that facilitate CDR, processing, analysis and prediction, lowering the technical barriers. Lawson et al. (forthcoming) illustrates the software architecture of the CDR-based component of Novissi 2.0’s grant administration system, revealing a system of moderate complexity that can be implemented with a small team of data scientists, software developers, and database administrators.

---

² [https://github.com/Global-Policy-Lab/cider](https://github.com/Global-Policy-Lab/cider)
WEB AND SOCIAL MEDIA DATA

**Opportunities:** Web and social media platforms may be particularly influential for targeted outreach. Although few social protection programmes already use web-based infrastructure for outreach and enrolment, the potential for web-based interventions goes far beyond the status quo and has the potential to reach vulnerable populations on an unprecedented scale and with communications directed towards multiple and specific eligible groups. In terms of assessment, web and social media data have mainly been used by researchers to create high resolution poverty maps that can be used for geographic targeting, frequently in combination with other digital data sources. Some examples are Sheehan *et al.*’s (2019) use of geolocated Wikipedia articles to estimate village-level characteristics and Chi *et al.*’s (2021) use of proprietary aggregated data from Meta on device connectivity to generate gridded micro-estimates of poverty for all low- and middle-income countries. Web and social media data are also promising for dynamic programme monitoring, and possibly for beneficiary management communications and analyses. Finally, a fairly mature application of social media data is the tracking of mobility during a crisis, such as Meta’s work on disaster maps (Maas *et al.*, 2019), which tracks mobility following natural disasters using aggregated Facebook location data.

**Data creation & format:** Social media sites record enormous volumes of information, as all behaviour is logged and analysed, creating petabytes of daily data. Whereas markets have matured in high income countries, user numbers and engagement in low- and middle-income countries is still rising. This is supported by initiatives such as Meta’s Free Basics, which partners with mobile network operators to waive mobile data fees for access to Facebook and selected other sites in low-income countries, providing access to the Internet to the ‘next billion users’. As with mobile phones, the increasing ubiquity of data recorded in interactions with the online world opens up the possibility to use these data sources to inform social protection programmes. Web search and social media data is stored in extremely large, ‘web-scale’ non-relational databases. The largest companies have developed their own data management technologies, including proprietary file formats, query languages, and analytical tools that support operations, analysis, and business services such as targeted advertising.

**Sources & access:** Apart from crowd-sourced sites such as Wikipedia or OpenStreetMap, as well as aggregated search trends search engines such as Google and Bing, social media giants are the main players in this space. Meta (Facebook Instagram, WhatsApp), Google (YouTube), ByteDance (TikTok), and Twitter are some of the largest companies globally in terms of client numbers. They primarily monetise their user base by selling targeted advertising, but are careful to safeguard user information, both for reputational reasons and to protect the proprietary information that their business models are based on. Some aggregated information, such as mobility and device access data, is made available to researchers and the public, but targeted, user-level information can only be accessed indirectly via the purchase of communication and advertising services.

**Processing & capacity:** The computational requirements of large Internet companies are vast. The amount and frequency of information requires enormous cloud facilities, the maintenance of which is beyond the capacity of most database administrators. Similarly, the specialist analytical skills needed for web-scale distributed processing, the insight into proprietary data format and custom software tools, and the analytical skills for data science tasks such as network analysis at scale are limited to a relatively small pool of experts. Most governments would be unable to use the data if they had direct access to it. However, as the data is largely unavailable, the
investment in skills and resources needed to use it may not be worthwhile. A more realistic form of engagement would be the contracting of communications and advertising services, for which staff trained in digital marketing are more suitable. Unless frequent engagement is envisaged, such services may best be contracted, rather than cultivated in-house by social protection administrations. In addition, the creation of analytical and policy tools from aggregate data – e.g., poverty maps or population movement monitoring – is viable with fairly standard data science skills and tools, which may well be worth investing in.

**FINANCIAL SERVICES DATA**

**Opportunities:** Data on financial assets and liabilities, including bank balances, securities, and loans, can be used for affluence testing in assessment. The verification of financial status is already common practice in high-income countries with extensive banking systems, and where means testing would typically be used for assessments. For low- and middle-income countries, the disbursement of cash grants via payment platforms is the dominant use case of digital finance in social protection, but digital finance and social protection systems may evolve together to create additional touch points. Credit scores can serve as an indicator of economic wellbeing. Similarly, incoming and outgoing transactions are closely related to expenditure, consumption and income measures used to determine economic wellbeing, and can, thus, support means or proxy means tests (PMTs). In addition, automatic retrieval of financial data and validated personal information held by regulated financial businesses can support intake and registration. Finally, beneficiary management can benefit from the monitoring of benefit withdrawals, and financial data can prompt exit decisions.

**Data creation & format:** Financial service providers create data when they execute financial transactions. Unlike in the case of mobile phone data, where the ‘transaction’ itself – calls and messages – is not of interest and metadata is used to establish behaviour patterns, the content of financial transactions is the relevant information. Stocks and flows of currency units are recorded in transactional databases to keep accurate ledgers of customer finances. Periodic summaries, such as average balances and transfers, are the primary source of information. The nature of the transactions and their counterparts can provide additional insights, but such analysis is unlikely to be feasible in most countries, due to the opaque and protected nature of financial systems.

**Sources & access:** The financial system is decentralised and contains a large number of actors in a variety of categories. In most countries, commercial banks handle the bulk of payments and, thus, represent the most important sector from a data collection perspective. In some settings, microfinance institutions may hold relevant data, while in others – most famously Kenya – mobile money has become a major conduit for transactions. Here, mobile network operators play a similar role to banks, with agent networks of retail traders functioning as a branch network of sorts. Credit bureaus can perform the important function of aggregating financial information across the financial sector, and information provided by them, thus, offers a potential solution to the problem of collating data from a fragmented system. However, such private sector information may be biased and inaccurate, as discussed below.

Data access to financial services data is often restricted by data protection laws that are especially strict for financial transactions. Banking secrecy laws and financial regulatory restrictions can be major...
hurdles to access to financial data held by private sector actors. The difficulty of agreeing terms with a large number of private sector players, who may have little commercial incentive to engage, further complicates data access. In South Africa, a collaboration with the banking association allowed social protection administrators to overcome these hurdles, creating a privacy-preserving data exchange format that monitored financial flows without compromising banking secrecy.

**Processing & capacity:** Similar to CDR, data volumes for financial transactions are large due to the high number of transactions collected by major service providers. However, the storage and processing infrastructure required for social protection systems in most countries is likely to be moderate if the analysis is limited to data extracts of specific indicators at the person level. The human resources necessary for the handling and analysis of the standard relational database information is similar to that of other data types discussed here and likely to include a small number of data scientists, front- and back-end software engineers, and database administrators. The requisite skills are likely to be available for procurement in most countries, whether through direct hiring or outsourcing.

## DIGITISED ADMINISTRATIVE DATA

**Opportunities:** Social protection operations require extensive information, which is costly to collect and prone to errors, but can be retrieved cheaply and accurately from authoritative sources. Demographic data, such as civil registry information, is an example of highly relevant information for outreach and intake. Administrative data can also be used extensively in assessing applicants’ eligibility status, for instance, through automatic enrolment into categorically-targeted programmes (Barca & Hebbar, 2020), means testing via tax authority records, or PMTs in settings with limited tax compliance. For the latter, pre-filled data from administrative sources can contribute to the evaluation of economic well-being, including targeting out via exclusion criteria, which was widely deployed for COVID-19 cash transfer programmes (e.g. Sharpe & Barca, 2021).

The integration of public sector data in social protection systems also provides a possible template and useful insights for the use of the emerging data types discussed above.

**Data creation & format:** Administrative data is collected by public sector institutions in the regular course of their operation and includes both explicit records, such as those collected in public registries, and information gathered for operational reasons. Examples of administrative data types important for social protection systems include demographic and family relationship information from the civil registry, income and asset information from the tax authority, vehicle ownership from the vehicle register, land or building ownership from the land cadastre, education and health system usage from sectoral databases, and disability status from the disability registry.

Taking a broader view, administrative data can include information from third sector institutions, such as business association membership, which identifies business owners. The use of existing administrative data in social protection systems is closely tied to the concept of integrated databases (for a review, see Barca, 2017), in which variables from various sources are merged to create a combined database. In addition to confirming the identity of applicants or beneficiaries, national ID systems can play an essential role in the data integration process, providing a unique personal identifier that facilitates data merging and avoids duplicate entries.
**Sources & access:** As modern entities with digitalised operations, practically all government ministries, departments and agencies collect, store and process data in the course of their regular activities. The same applies to third sector and quasi-public institutions. Contingent on the data protection law, administrative data can typically be shared with other public bodies, if it supports their stipulated functions. Public IT systems designed with interoperability in mind can greatly support the uses described here. Application programme interfaces (APIs) and data service agreements that stipulate data formats, updates, maintenance, and access controls are key tools in promoting such interoperability. When accessing data not under the direct purview of a public body, local laws and the relationship with the public body will determine whether or not a contractual access framework, like that of private sector actors, is more appropriate.

**Processing & capacity:** Comparatively small data volumes and tabular database structures facilitate the processing of administrative data. Most data can be processed using existing IT infrastructure, but data transfer protocols between databases can be an issue. Without an established and safe transfer mechanism, data may be shared in an ad-hoc fashion that creates security risks, once more underlining the benefit of investment in interconnectivity. In terms of human capacity, IT skills such as front- and back-end developers, as well as database administrators, can be a bottleneck in leveraging the potential of existing data sources, as the public sector in low- and middle-income countries is often under-resourced in this regard. In cases where the internal supply of these skills is insufficient, procurement of additional capacity may well represent a productive investment in state capacity.
SUMMARY BY DATA SOURCE
SUMMARY BY DELIVERY CHAIN TASK
This section discusses applications of the data sources outlined previously. Firstly, it reviews existing applications that have already been implemented and that illustrate the practical potential of these sources. Secondly, it discusses potential applications that appear realistic and may be effective, but have not been implemented to the best of our knowledge at the time of writing. The organising framework is the social protection delivery chain (Lindert et al., 2020), which describes the essential operational building of most social protection programmes. The first link of the chain is outreach, which describes the process of informing potential beneficiaries about a programme, application modalities, and other information relevant at an early engagement stage. Intake and registration, during which applicant information is collected, is the second link. Eligibility determination follows, during which it is determined whether or not an applicant meets the criteria for eligibility and should become a programme beneficiary. The fourth link is the provision of benefits, such as a payment mechanism in the case of cash transfers. The final link is beneficiary management, which is mainly concerned with handling issues and grievances, maintaining accurate information, and handling the exit of beneficiaries from the programme.

OUTREACH

A number of social protection programmes already leverage digital technologies for outreach: blast SMS messages are sent to mobile phones to inform mobile subscribers of new social programmes; information about social programmes is available on the Internet and distributed via government social media channels; and radio advertisements provide an option for reaching households that are less digitally connected. Combining digital technologies with digital data, however, provides a new and largely unexplored option for running efficient, targeted outreach campaigns. For example, combining phone-based outreach with existing phone records can allow social programme designers to target specific populations with SMS messages containing information about programmes and eligibility criteria; likewise running programme advertisements on search engines and social media websites allows targeted advertising of programme information to specific demographic groups. While such approaches are particularly susceptible to digital exclusion – and, therefore, should in most cases, be combined with in-person outreach – the potential for fast, targeted, and wide-coverage outreach campaigns via digital data is vast.
How have digital data sources been used for programme outreach already?

Proactive and targeted outreach campaigns guided by digital data sources are not yet the norm for social protection programmes. However, three key examples deployed in the past three years demonstrate the ways in which digital data streams can guide an efficient and intelligent approach to outreach.

In Costa Rica, the NGO Prosperia has partnered with the national government to use satellite imagery to guide outreach and registration campaigns for the unified social registry (Aiken & Ohlenburg, 2023). Prior to the project with Prosperia, Costa Rica’s social registry had a coverage rate of approximately 70% of the population – but there was evidence that excluded households were poor and, if included in the social registry, would be eligible for social programmes. To help identify poor neighbourhoods with little social registry coverage, Prosperia used satellite-based poverty estimates at a neighbourhood level (produced with a custom machine learning pipeline), combined with existing estimated social registry coverage rates in each neighbourhood. They identified neighbourhoods with especially high poverty estimates and especially low social registry coverage and sent outreach workers to these priority neighbourhoods to register households en masse for the social registry and inform them of social programmes.

In Togo’s Novissi programme and the DRC’s STEP-KIN programme (both COVID-19 cash transfer programmes), mobile phone data was used to inform targeted outreach via SMS messages. The government of each country partnered with mobile network operators to obtain the approximate locations from which mobile subscribers placed calls (at the cell tower level, with each cell tower covering between a few hundred square metres and several square kilometres). ‘Home locations’, based on the location from which subscribers placed the most calls, were matched to estimates of neighbourhood-level poverty and vulnerability in rural Togo (for Novissi) and Kinshasa (for the DRC). In both programmes, SMS messages were sent to subscribers identified as living in poor or vulnerable geographies, encouraging them to register for cash transfers.
What other possibilities exist for digital data in outreach in the future?

Beyond informing potential beneficiaries of geographic programme expansion, digital data has the potential to guide targeted outreach at the individual or household level. As Internet penetration grows in low- and middle-income countries, web and social media data have the potential to be particularly powerful outreach tools. The existing advertising infrastructure built by search engines and social media companies could be leveraged through partnerships to ‘advertise’ social programmes and target such information based on data recorded by the hosting company (for example, data on geography, inferred poverty, gender, household size, and more). As in the example of Prosperia in Costa Rica, these digital data sources could be used to inform proactive outreach to individuals or areas that are likely to be poor, but are not covered in existing social registries.

What major pitfalls should be avoided?

The major risk with using digital data to inform programme outreach is digital exclusion: that individuals and households without access to digital services will fall through the cracks of digitally-informed outreach campaigns. The risk of digital exclusion is particularly high when leveraging a data source – like mobile phone data or web data – that in itself requires access to technology; for example, in Togo’s Novissi programme and the DRC’s STEP-KIN programme, individuals without mobile phones were unable to receive text messages informing them of the expanded cash transfer programmes.

Digital exclusion in digitally-informed outreach can be mitigated by combining digitally-informed outreach with in-person components. The Costa Rican example provides a blueprint for such an approach: while satellite imagery identified the general location to which social workers went to carry out outreach and registration campaigns, in-person presence from social workers reduced the possibility of very poor or ‘invisible’ households being missed.

More generally, digital exclusion is a major theme that cuts across data sources and delivery chain tasks. This is discussed in more detail in the first sub-section of section 4.
INTAKE AND REGISTRATION

Digital technologies enable powerful new intake and registration modalities: entirely remote registration is possible for digitally-connected individuals and households via text messages, interactive voice response (IVR), chatbots, and web forms. Such decentralised digital registration systems also have the potential to facilitate the dynamic updating of registration and associated data. Digital technology-enabled intake and registration is most powerful when integrated into larger pipelines integrating digital data, including targeted outreach and the near-instantaneous digital determination of eligibility.

How have digital data sources been used for intake and registration already?

A number of social protection programmes have integrated digital registration channels into larger social protection systems. Many of these social protection systems involve in-person components, either as an alternative to digital registration, or in other parts of the social protection delivery chain. Others have turned to a fully-digital model, where registration is digitised alongside outreach, eligibility determination, and the provision of benefits, for an entirely remote approach to social protection delivery. Across contexts, registration and payment are frequently the first components of social protection systems to be digitised, even when outreach, targeting, and beneficiary management still include in-person interaction.

During the COVID-19 pandemic, a number of programmes deployed SMS-based or Unstructured Supplementary Service Data (USSD)-based registration platforms on mobile phones. Such programmes allowed potential beneficiaries to dial a toll-free number to access a USSD platform, in which they entered key identifying information (such as a national ID number) and other information for their application. Prominent programmes incorporating USSD registration options include Togo’s Novissi programme, Pakistan’s Ehsaas programme, and the DRC’s STEP-KIN cash transfer programme.

Web-based registration platforms similarly allow potential beneficiaries to register for social protection benefits remotely using their mobile phone or home computer. Web-based registration options were made available in recent cash transfer programmes in South Africa, Brazil, and India — among many other low- and middle-income countries. Some countries also included customised apps that were easier to navigate from smartphones than websites.
What other possibilities exist for digital data in intake and registration in the future?

Beyond the fairly well-established USSD, SMS, and web-based registration options, new technologies may provide smoother and more data-driven registration pathways. Chatbots, for instance, have the potential to make digital registration experiences more straightforward, perhaps while answering questions about eligibility criteria and programme benefits along the way. Chatbots on WhatsApp and Facebook messenger were deployed as registration channels in South Africa’s Social Relief of Distress Grant, but a number of applicants reported difficulties navigating the chatbot’s registration flow (Senona et al., 2021). With advances in conversational AI (and the development of social protection-specific bots), chatbots may become a more feasible registration option in the future. Furthermore, as described in the previous section on outreach, digital data sources – and especially satellite imagery, as in the Prosperia example in Costa Rica – have the potential to inform targeted registration campaigns that can identify individuals or areas in a country that are likely to be poor, but that are not covered by existing social registries or social protection programmes.

What major pitfalls should be avoided?

In addition to general issues of digital exclusion, digital registration runs the risk of creating additional exclusion due to access to and ability to use digital technologies. For example, the primary registration channel for South Africa’s Social Relief of Distress Grant is via a website. However, many people living in poverty – especially in rural communities – have reported being unable to access the website, because they are not able to afford mobile data (Senona et al., 2021). Where possible, all phone- and web-based registration channels should be made available to registrants free of charge.

Literacy itself is a challenge for any text-based registration platform (such as USSD, SMS-based options, and traditional web forms). Particularly in countries with low literacy rates, IVR technologies may be a more accessible registration option for low-literacy populations. Providing multiple digital registration options – such as IVR, SMS-based, and Internet-based pathways – is likely to be accessible to the largest share of the population. While no IVR technologies have been deployed for social protection registration to date, they have been used in a number of other contexts, from automated surveys (Hortinela, 2019) to disseminating health knowledge (LeFevre et al., 2018) and citizen journalism (Mudliar, Donner & Thies, 2012). IVR holds promise for reducing the exclusionary nature of digital registration for social protection systems.
How have digital data sources been used for eligibility determination already?

Satellite imagery, mobile phone data, financial services data, and digitised administrative data have all been used to determine eligibility for social protection programmes.

Togo’s Novissi programme used satellite-based poverty estimates at the admin-3 level to identify rural regions of the country that would be eligible for cash transfers. The satellite-based poverty estimates used in Novissi came from the relative wealth index (RWI) (Chi et al., 2022), a publicly available source of micro-gridded poverty estimates for all low- and middle-income countries. The RWI also incorporates other digital data sources for predicting poverty, including cell phone tower coverage and proprietary indicators from Meta. The government of Nigeria is currently considering an expansion of its cash transfer programme based on the RWI (Smythe & Blumenstock, 2022). In Niger, satellite-based indicators of drought are being piloted as part of a new early warning system that will rapidly target cash transfers to drought-affected regions (Brunelin et al., 2022). The Niger example, in particular, highlights how the high temporal resolution of satellite imagery and other digital data sources can be leveraged for increasingly adaptive social protection measures, responding to environmental shocks in real time.

The most prominent example of using mobile phone data for targeting is, again, Togo’s Novissi programme, where cash transfers were provided to eligible mobile subscribers. Eligibility was determined using algorithmic predictions of poverty based on each subscriber’s history of phone use – including the number of calls they made, the diversity of their contact network, the general locations from which they placed calls, and how much mobile data and mobile money they used. Aiken et al. (2022) provides a detailed assessment of the accuracy...
of this targeting approach; in general, the approach is found to be more accurate than pure geographic targeting, but less accurate than survey-based targeting using a PMT. The DRC’s STEP-KIN programme took a similar approach, identifying subscribers eligible for cash transfers in Kinshasa on the basis of whether or not the cell towers they used were in poor areas and a simple eligibility criteria based on total phone bill amount. GiveDirectly has also piloted a similar phone-based approach to targeting in Malawi (GiveDirectly, 2022). Multiple types of financial services data have also been recently put to work for eligibility determination in large social protection programmes. In Colombia, financial data from DataCredito, a Latin American credit scoring agency, was combined with existing social registry data in the SISBEN targeting database (López Solano, 2020). Data shared by Experian included information on bank loans, telecommunications bills, and estimated total earnings. The financial information was combined with PMT indicators of household conditions and digitised administrative data from other branches of government (including health authorities and civil registries) to generate a single unified prediction of poverty. In South Africa, eligibility for the Social Relief of Distress Grant was determined via a means test based on the balance of a potential beneficiary’s bank account. If applicants met the means test criteria, eligibility was further restricted based on a number of heuristic criteria from other digitised administrative data sources.

While the above examples highlight that combined systems of eligibility determination frequently leverage digitised administrative data from a number of government agencies, it is worth noting that many other governments have implemented eligibility systems that rely on digitised administrative data alone, or in combination with in-person survey data. For example, Turkey’s Integrated Social Assistance System pulls data from civil registrations, a database of employment status, tax records, land and property ownership, vehicle registration, education, and health information. Pakistan’s Ehsaas programme uses data from vehicle registration, passport applications, employment, and travel records to ‘target out’ wealthy individuals from cash transfer eligibility. Similar ‘targeting out’ approaches have been taken in Chile, Namibia, Brazil, and elsewhere where high-quality government administrative records provide information on the poverty status of potential beneficiaries.
What other possibilities exist for digital data for eligibility determination in the future?

In the future, social protection programmes have the potential to take advantage of the real-time nature of digital data streams to make eligibility determination a more adaptive and shock-responsive process. While most of the programmes described above determined eligibility statically – with a one-time check of eligibility criteria – digital data sources may be able to track changing socioeconomic status over time. Such adaptivity could come in the form of frequent re-assessment of eligibility criteria, or re-assessment could be triggered by a change in or discontinuation of one of the digital data streams (Turkey’s Integrated Social Protection System provides an early example of this type of adaptivity, triggering eligibility re-assessment for all social protection programmes when a household’s information in any one of the administrative sources changes). Yet another opportunity for adaptivity is in responding to climate shocks and violent events: following the example of Niger, real-time digital data sources could be used to identify regions and households about to experience – or in the process of experiencing – an economic shock, and change their short-term eligibility status accordingly.

What major pitfalls should be avoided?

Beyond the cross-cutting concern of digital exclusion, three concerns are specific to eligibility determination: data access, privacy, and transparency. Any type of household or individual-level targeting will require access to data that is identifiable at the household or individual level. Many of the digital data streams described here – in particular, satellite imagery, web and social media data, and financial services data – are collected for non-social protection purposes, frequently by private companies. Data access in these scenarios may be difficult or expensive, and will require strong collaboration between implementing government agencies and private companies.

Moreover, many of these data sources may contain private information. For example, information about individual-level mobility patterns can be inferred from mobile phone data, and financial services data may contain sensitive information on the goods and services that an individual purchases. Data protection and data security protocols are essential for handling these types of data, and obtaining consent for the secondary use of data collected by private companies is required by data protection legislation in some countries. We discuss the issues of privacy and data protection in more detail in section 4.
A final concern for eligibility determination with digital data is that of transparency. Historically, PMTs and other traditional targeting approaches are not entirely transparent: while potential beneficiaries are aware they are being evaluated for social programme eligibility and of what data is being collected, they are typically not aware of the exact formula used to determine eligibility. With remotely-accessed digital data sources, on the other hand, potential beneficiaries may not even be aware of what data sources are being used to determine eligibility – or that they are being evaluated for social programme eligibility at all. This type of ‘black-box’ eligibility determination is likely to erode trust in the social protection system; as such, a degree of transparency on eligibility criteria may be necessary (or at least a notification or request for consent to access and use specific data fields) for long-term programme success. We discuss the question of transparency and explanations of eligibility criteria further in section 4.

PROVISION OF BENEFITS

Like intake and registration, the provision of programme benefits – especially in cash transfer programmes – has been transformed by digital technologies. Digital payment options – including bank transfers, mobile money, and pre-charged cards – allow programme implementers to distribute cash faster and more cheaply than delivering benefits in person (Aker et al., 2016). These types of technologies are particularly relevant during public health crises and in conflict-affected contexts, where it may not be possible to reach beneficiaries on the ground. The provision of benefits is also similar to intake and registration, in that digital delivery systems become more powerful when integrated into broader pipelines of digital data across the delivery chain. For example, the digital records produced by the provision of benefits via mobile money or bank accounts can be used to audit payment mechanisms and ensure that programme benefits are going to the intended beneficiary.
How have digital data sources been used in benefits provision already?

During the COVID-19 pandemic, around 80% of cash transfer programmes shifted to providing digital-only payments or incorporating a digital payment component (Gentilini et al., 2022). Mobile money payments were especially common in countries with low smartphone and financial services penetration; in countries with more financial services available to the poor, bank transfers or e-wallet payments allowed for rapid cash distribution.

Beyond being used as a payments endpoint, mobile money transfers and bank transfers have the advantage of recording the exact time and amount of government-to-person transfers. In Togo’s Novissi programme, for example, an external consulting firm was hired to audit the provision of benefits using mobile money logs. They were provided with records of all government-to-person mobile money transfers by the mobile network operators and used these records to confirm that all beneficiaries were receiving payments (and that no payments were made to non-beneficiaries). With cooperation from the banking sector, government-to-person bank transfers could be audited in a similar manner.

What other possibilities exist for digital data for benefits provision in the future?

Beyond simply auditing payments for correctness (that is, ensuring that payments are of the correct amount and going to the correct individuals), digital data streams could be used to track whether or not beneficiaries are able to access benefits, and if so, what they use benefits for. Mobile money provides a useful example: Are all beneficiaries able to ‘cash out’ the mobile money they receive from social protection programmes – and how long does it take? If money is used to purchase goods via vendors, what kind of goods? In the future, web and social media data may also provide clues as to how beneficiaries use cash provided by social protection programmes.
What major pitfalls should be avoided?

Like intake and registration, digital payments run the risk of excluding those without access to mobile phones and financial services, as well as those who face literacy barriers or other access challenges. Providing an alternative fallback payment option – such as in-person dispersal of payments at social programme offices or other pay points – will help alleviate this concern. Furthermore, extensive surveillance of the use of payments, as proposed above, risks invading the privacy of beneficiary. Programme designers should ensure that surveillance is justified by the utility of data traces for understanding whether or not payments are being successfully received, and they should collect consent for data analysis whenever possible.

BENEFICIARY MANAGEMENT

The real-time and up-to-date nature of many digital data streams makes them a natural fit for tasks in beneficiary management, including evaluating programme impacts and continuous re-assessment of eligibility criteria. However, there are limited examples in which digital data sources have been deployed for these tasks, outside of the research literature, so the potential of digital beneficiary management is still largely untested.
How have digital data sources been used in beneficiary management already?

The main uses of digital data for beneficiary management to date are (1) collecting post-programme outcomes using short phone-based surveys, and (2) using digital data streams to infer programme impacts.

Mobile phones have become a tool for the rapid gathering of opinions and feedback. This capability has been deployed for post-programme surveys about experiences with social protection programmes. For example, the STEP-KIN programme in the DRC used a short SMS-based survey to ask programme beneficiaries what major categories of goods they spent their STEP-KIN cash transfers on. They also used the survey to assess alignment of the programme’s eligibility criteria (based on mobile phone use and inferred home location) with the original programme goal (reaching the poorest individuals in Kinshasa). Similar rapid-fire SMS-based programme assessments have been deployed in a number of contexts. Longer-form phone surveys with a human enumerator are also a standard approach used to collect post-programme outcomes data from beneficiaries, and became particularly widely used when the COVID-19 pandemic rendered in-person surveys impossible in most countries.

Beyond using mobile phones as a tool for data gathering, digital data streams – including satellite imagery and mobile phone data – have been used to directly infer the outcomes of social protection programmes. The best example to date comes from Uganda (Huang et al., 2021), where information on building size and roofing material derived from satellite imagery was used to infer the treatment effects of a very large cash transfer programme run by GiveDirectly. An associated research study found that satellite-based treatment effects were similar in magnitude to treatment effects derived from standard in-person consumption surveys, albeit with wider confidence intervals.

Digital traces have also been used to understand programme impacts that are not directly related to the anti-poverty impacts of programmes. For example, the United Nations Development Programme (UNDP) partnered with GRANDATA – a Colombian mobility data aggregator – to use smartphone mobility traces to evaluate the impacts of COVID-19 cash transfers on mobility and adherence to social distancing measures (finding an on-average null effect on mobility from receiving a cash transfer). Similarly, researchers evaluated the impact of Togo’s Novissi cash transfers on phone sharing using CDRs provided by mobile network operators (finding that receiving a cash transfer increases both phone use and phone sharing).
What other possibilities exist for digital data for beneficiary management in the future?

Digital data hold enormous potential for the evaluation of programme impacts and the dynamic reassessment of eligibility. Existing programmes and research projects have only begun to explore the possibilities of using digital data for impact evaluation. If any of the data streams used for targeting – including satellite data, mobile phone data, web and social media data, and financial services data – can be used to track wellbeing in real time, they can also be used to evaluate programme impacts on the poverty level of beneficiaries. However, there is limited evidence to date that these types of data are able to capture dynamic measures of poverty (Yeh et al., 2020; Aiken et al., 2022). Digitised administrative data have the potential to track programme impacts on more heuristic measures of poverty – in just one example, if programme implementers have access to vehicle registrations in real-time, they can test if beneficiaries are more likely to purchase a vehicle post-programme than non-beneficiaries.

Similarly, if digital data are updated frequently enough, they can be used to trigger reassessments of eligibility criteria (as discussed in the earlier section). For example, if the estimate of a beneficiary’s poverty status derived from mobile phone data or web data changes, they could be identified as newly-eligible or newly-ineligible. Similarly, changes in any form of digitised administrative record could be used to trigger reassessment. A simple form of this dynamic reassessment was included in South Africa’s Social Relief of Distress Grant, in which eligibility was re-assessed on a monthly basis via (1) a means test based on the bank account balances of beneficiaries, and (2) heuristic criteria from a number of digitised administrative records. These criteria included not qualifying to receive benefits from the national Unemployment Insurance Fund, not receiving a stipend from the Student Financial Aid Scheme, not resident in a government-subsidised institution (including prisons), not receive any other government support, and receiving no income (Department of Social Development, 2021).
What major pitfalls should be avoided?

Using digital data for beneficiary management necessitates access to such data at a high temporal frequency. Many existing data sharing agreements between social protection agencies and data providers (such as other government agencies, private companies, and researchers) rely on one-time or fixed-term data sharing agreements (Aiken & Ohlenburg, 2023). Real-time or near-real-time access to data requires closer cooperation with data providers and different kinds of data sharing practices. Such real-time data access also raises the issue of data vendor lock-in: if the data provider is unable to provide up-to-date records, it may compromise the functioning of the social protection programme.

In addition to data access issues, the accuracy of digital and data-driven approaches to beneficiary management is still largely untested. As these interventions have only been introduced in the past few years, there are few examples in which they have been deployed outside of the research literature, and there is little evidence from which to draw conclusions about their efficacy.
KEY QUESTIONS THAT APPLY TO ALL DATA SOURCES
This section takes the form of a ‘Frequently Asked Questions’ list that covers large-scale policy, regulation, and ethical questions. The questions, organised according to the stages of the social protection delivery chain, apply to programmes reliant on any of the data sources discussed above.

**OUTREACH, INTAKE, AND REGISTRATION**

1. How will access barriers, including illiteracy, disability and age, mediate outreach and registration?

   Access barriers may limit or preclude a programme’s ability to engage with potential beneficiaries. Important barriers include disability, age, literacy, and gender. The former two may restrict potential beneficiaries’ interaction with outreach and registration channels through physical limitations. Similarly, both literacy and digital literacy mediate the effectiveness of outreach and registration channels. Many digital outreach and registration channels – including the Internet, mobile messaging, and SMS-based options – rely on text only to communicate programme information and facilitate registration. For individuals with low literacy, difficulties with reading and understanding text may prohibit programme registration. Voice-based approaches, including outreach via call centres and IVR technologies are text-free digital alternatives to text-heavy Internet and phone-based options.

   In considering which combination of outreach and registration channels to deploy, it is essential to consider how vulnerable subgroups of the population will experience these channels. When it comes to literacy, women and the elderly are particularly likely to experience difficulties accessing and using digital and text-based registration platforms. Complementing these platforms with in-person (or at least voice-based) outreach can help ensure that vulnerable groups are not disproportionately excluded from programme registration. Further, racial minorities, immigrants, and other minority language communities may have difficulty accessing outreach and registration materials if they are only available in majority languages. Ensuring that whichever outreach and registration channel is selected is available in as many languages as possible will reduce the eventual exclusion of these groups from programme benefits. Issues of literacy and digital literacy in other parts of the social protection delivery chain are also covered further in questions 10 and 11.
2. What support networks will be available for applicants who struggle with registration?

To avoid the exclusion of particularly vulnerable groups, programme designers can include some form of support for beneficiaries who struggle to access digital registration channels. Here we cover four options for providing such support, or leveraging existing networks to provide support in registration:

**Leverage existing social protection services:** Many social protection agencies will have an existing network of offices and case workers who provide on-the-ground support to communities. A key benefit of digitising social protection delivery is that digitisation can allow governments to shrink the on-the-ground footprint, generating cost savings and increasing efficiency. However, even if on-the-ground presence is reduced, maintaining a presence – particularly in communities with low levels of literacy and digital literacy – can ensure that vulnerable individuals are not inadvertently excluded from programme registration. For example, Togo’s Novissi programme deployed a team of field workers to travel from village to village assisting potential beneficiaries with SMS-based registration channels. Moreover, providing training on digital registration services to case workers serving rural communities will ensure that they are equipped to help these communities with digitised systems.

**Partner to leverage other existing in-person networks:** If in-person social protection services are not available (for example, in conflict or crisis settings, or if costs are prohibitively high), it may be possible to work with NGOs and for-profit organisations to leverage existing in-person networks for support with receiving payments. For example, for South Africa’s Social Relief of Distress Grant, case workers from the South African Social Security Administration were not trained to help people access grants, but case workers from the Black Sash, a longstanding human rights NGO, were able to help a large number of beneficiaries register for the programme and access payments via their in-person support offices (Matthews et al., 2020). As an NGO with the aim of helping unemployed individuals access social benefits and lobbying for government-run universal basic income schemes, the Black Sash was a natural fit for aiding applicants who experienced difficulty registering (although they did not partner directly with the government). For-profit actors could be trained to provide similar support: for example, when cash transfers are delivered via mobile money, mobile money cash out agents could provide critical assistance to those struggling with programme registration (as well as other parts of the social protection delivery chain, including accessing funds).
Create digital support channels: To complement in-person support networks – or in contexts where any form of in-person support is not feasible – digital support channels may provide some assistance for beneficiaries struggling with digital registration. For example, a number of emergency cash transfer programmes launched during the COVID-19 pandemic turned to toll-free call centres to provide assistance to those struggling with programme registration and payments (Gentilini et al., 2022). However, results on the efficacy of such call centres are mixed: for example, qualitative results suggest that long wait times and difficulty with connectivity may have compromised the efficacy of call centres (e.g. Mathews et al., 2020). To maximise call helpline effectiveness, call centres should employ agents who speak as many of the relevant languages as possible, and ensure that call centres are well staffed to avoid long wait times, or that a specific call centre team is trained on the particularities of assisting potential beneficiaries with digital registration. Chat-based services may also be relevant in high-literacy contexts, along with frequently-asked-questions lists (distributed digitally or non-digitally) that cover solutions to common registration problems, or digital forums facilitating community support for registration (see following section).

Leverage the community of potential beneficiaries: Growing evidence suggests that within-community support may fill gaps in formal support networks for accessing benefits from digital social protection programmes. Qualitative evidence from a digital cash transfer programme set up during the COVID-19 pandemic in South Africa, for example, suggests that when little or no formal in-person support for registration is provided to beneficiaries, beneficiaries frequently relied on other beneficiaries or other members of their community at large for support in registration and receiving payments (Mathews et al., 2020). The implication is that in contexts where little in-person support is provided, the labour and costs of support are shifted to community members themselves. On the other hand, if governments and other programme implementers could find ways to formally leverage community support networks, they may provide a low-cost option for helping beneficiaries overcome literacy and digital literacy challenges. For example, by providing training to community members on how to assist with registration and cash-out and/or formally compensating community members who assist others with accessing benefits, implementers could create large teams with the skills and incentive to help others access programme benefits at a relatively low cost.
To reduce fraud in social programme registration, it is essential to find a method for deduplicating potential beneficiaries. In this section, we discuss the advantages of using a foundational national identification system, where available. In the following two sections, we discuss alternatives to national IDs for contexts where they are not available or where coverage is low.

Many social protection programmes require potential beneficiaries to provide a national ID number upon registration. A national ID number alone can ensure that an individual or household cannot register for a social protection programme more than once. Where additional data is available, national ID numbers can also be linked to demographic information about potential beneficiaries, which can help ensure that target groups are being included in programme benefits. These data, as well as other administrative information linked to national ID numbers, may also be helpful for eligibility determination (see question 5).

A number of countries have implemented foundational unique ID systems linked to social protection programmes – typically with the goal of near-universal ID coverage. For example, as of 2022, 99.9% of adults in India and 92.7% of all residents have been issued an Aadhaar identification number, which can be used to register for a number of social protection programmes (Sharma, 2022). The Aadhaar database also records biometric data – including fingerprints and iris scans – for rapid and accurate verification and deduplication. Other countries, including Argentina, Indonesia, Peru, Uruguay, and Vietnam, have implemented similar foundational unique IDs – often with biometric data – which can be integrated with social protection programme registration (Burt, 2019; World Bank, 2019).

Where foundational unique IDs are not available, other forms of functional national IDs may be sufficient to deduplicate applicants in social protection programmes (ID4D, 2021). Birth certificates and civil registration systems are one form of functional national ID that can be repurposed for social protection identification, although the coverage of birth registration in many low-income countries is prohibitively low (ID4D, 2021). Other functional IDs are not created with the purpose of providing identification for social protection systems or other government-run programmes, but with sufficient coverage, they may suffice for deduplication where foundational IDs are not available. For example, Togo used a national voter ID (with an estimated coverage of 86% of adults) for deduplication in a digital COVID-19 cash transfer programme (Aiken et al., 2022). Similarly, taxpayer IDs, social security numbers, and driver’s licence numbers may all suffice as sources of identification for social programmes, if their coverage is high enough (ID4D, 2021), and if IDs can be matched to other data sources used for the programme (especially targeting data sources). In contexts with low coverage, a combination of identification systems or alternative approaches may be necessary (see questions 4 and 5).
4. What is the coverage of the identification database, overall and by relevant groups?

Naturally, when unique IDs are required for social registration programmes, it is essential that individuals, households, and communities are not excluded on the basis of not possessing an ID. It is, therefore, important to check for the coverage of an identification database before linking it to a social protection programme. In addition to validating that coverage in the overall population is high, it is worth checking specifically whether or not coverage is also high among vulnerable groups (such as women, the elderly, immigrants, people with disabilities, and racial and political minorities). Is there a risk that a vulnerable group will be disproportionately excluded by the identification source selected? As just one example, in Afghanistan – where women have to be accompanied by a man to apply for a national ID card – 87% of those lacking IDs are women (Livani & Haddock, 2020). In such a context, using alternative or complementary sources of identification data will help maximise programme coverage and accessibility.

5. If there is no government-issued ID with substantial coverage, is there a suitable alternative approach?

In some countries, the coverage of national IDs – whether foundational or functional – is prohibitively low, whether in the overall population or among specific vulnerable groups. One option in these settings is to accept multiple forms of ID. While this strategy complicates programme administration and reduces the ability of an ID to directly deduplicate applicants, it does ensure wider coverage of social protection programmes in contexts where no individual ID system has high penetration. When even combining ID options does not yield wide coverage of the target population, an alternative is to deduplicate based on in-person registration. Caseworkers – especially those who are members of the community they serve – can ensure that individuals or households only register once for a social protection programme. Co-designing a national ID with a social protection programme has the advantage of potentially incentivising both government investment in the programme and registration among potential beneficiaries. Caseworker deduplication, however, has the downside of opening the door to fraud in the registration process.
Another alternative for settings with low ID coverage is to set up biometrics for programme registration independently of a national ID. For example, the NGO GiveDirectly has worked with third-party consulting firms that use facial recognition software to deduplicate applicants in their cash grant programmes in Uganda, using pictures of recipients taken by field workers. The NGO has also piloted the use of palm vein readers for more accurate biometric identification (GiveDirectly, 2016). While costly to set up – both in terms of equipment costs and on-the-ground staffing – independent biometric identification can fill the gap in contexts with little ID coverage. There are also concerns around the accuracy of facial recognition systems, as well as related surveillance implications; these concerns have not been fully explored in the social protection context (Introna & Wood, 2004).

In programmes for which registration occurs on mobile phones, deduplicating using the mobile phone number is another option – albeit one that is subject to fraud and manipulation. For example, the STEP-KIN programme in Kinshasa was set up entirely remotely during the COVID-19 pandemic, with little availability of national IDs. Registration for the programme occurred on smartphones and feature phones via a USSD platform, and each phone number (i.e. each SIM card) could register only once for the programme. To reduce the potential for individuals registering on multiple phones and receiving multiple transfers, phone-based registration and deduplication was paired with relatively stringent targeting criteria based on inferred home location and cost of phone bills (Mukherjee et al., 2023). However, except in extreme circumstances, deduplicating on phone numbers should be avoided, as this deduplication method has the potential to be regressive (allowing richer individuals with multiple SIMs or mobile phones to register more times than poorer individuals using single or shared devices), and to incentivise manipulation (including buying phones or SIM cards to register for social protection benefits).
6. What is the target population of the programme?

The clear specification of the target population is essential for effective programme design. It informs programmatic aspects, such as needs assessment, and the simulation of programme impacts as well as operational aspects, including the selection of appropriate communication, interaction, and payment channels. Common criteria include demographics, economic welfare, disability, and vulnerability. Errors of exclusion may be of greater concern in a crisis response setting than errors of inclusion, and a targeting out approach that provides benefits to all applicants, except those who are screened out, can be suitable here. In this case, the target population is defined negatively, e.g. by affluence testing which excludes the upper quantile of income distribution.

At the legislative stage, the target population may be defined in broad terms that require translation into measurable characteristics for policy implementation. For example, instead of referencing 'the economically vulnerable', a clear definition could be the 'households in the lowest two deciles of the consumption distribution'. Although data limitations will typically constrain precise identification, the mapping from policy intention to specific characteristics is one of the major responsibilities at the technical policy design stage. In the context of programmes using novel data sources or delivery mechanisms, an iterative process that takes data availability and policy simulations into account may be necessary to specify the target population in a suitable way.
Are you able to simulate the achievement of targeting objectives?

Simulating the effects of targeting policies on the basis of data and theory can inform programme design to yield predictable and reliable outcomes. Such evidence-based policymaking can be especially important in an adaptive social protection setting where time is of the essence and immediate impact, rather than gradual adjustment, is called for. Where possible, policymakers should create policy simulations that allow for the systematic evaluation of design parameters such as the definition and size of the target population, the nature and quantity of benefits to be provided, and the disbursement modalities.

Data sources for simulation may include surveys, such as socio-economic, labour force or other population sample surveys, registries and programme databases, and administrative data. Each of these data sources has the potential to provide ‘ground truth’ information on individual-level poverty. When matched with digital data sources, these combined datasets can be used to simulate expected targeting accuracy (for at least a subset of the population). For example, satellite imagery can be linked to ground-truth survey data by matching the GPS coordinates of households collected in surveys to satellite tiles. Mobile phone data and financial data can be linked – with informed consent from surveyed individuals – via unique identifiers (for mobile phone data, typically the individual’s phone number; for financial data, typically bank account or credit card identifiers).

In many cases, assumptions will need to be made to combine different sources in order to create the data foundations for relevant policy simulations, and it will often be preferable to use imperfectly combined sources than to set parameters by intuition. The estimation of metrics, including poverty impacts as well as inclusion and exclusion errors, and the fiscal costs of different options are likely to be the key results in question. Although policy choices will be influenced by political considerations, such technical inputs can steer programme design in an efficient and effective direction.
8. Will you be able to access the data, and is it legal to do so?

The existence of relevant data is necessary, but not sufficient, to implement a given targeting approach, unless the data is accessible for programme purposes. Factors that may preclude data access are excessive cost, commercial sensitivity of the data which deters sharing by the proprietor, data protection, legal obstacles, and procurement challenges. Different types of data have different accessibility profiles. In addition, different types of data also have different availability profiles. Survey data is generally accessible for policy purposes, including tasks such as PMT design and impact simulations, as it is partly collected for this purpose. Many administrative data sources are available for public programmes with a basis in law, as various parts of the government are legally the same entity and can, thus, share information. Satellite data is constrained by technical availability and cost. Images without cloud cover may not be available for given sites at needed times and at the desired resolution. Where it is, and especially where it would need to be collected specifically for this purpose, costs may be prohibitive.

Phone and financial data may or may not be shared by mobile network operators, depending on the country’s legal framework, the willingness of operators to cooperate, and competitive or data protection concerns. Both types can be highly sensitive, as they reveal personal behaviours and circumstances, hence, data protection issues also loom large. The use of data generated in a different context requires consent from data subjects, but for those in need of social protection, genuine consent will often be indistinguishable from enforced consent. A complicating factor is that no single source exists for such data, as multiple financial and phone companies split the market among themselves. Access may, thus, need to be secured from several players, who may distrust each other and not be willing to contribute to a sector-wide initiative. An early exploration of these issues is prudent, including consultations with lawyers, procurement specialists, and other public bodies that may influence commercial firms’ behaviour.

After accessing data, it must be ensured that the necessary infrastructure and capacity are available to store and analyse the targeting data sources. Issues of infrastructure and capacity are discussed in more detail in questions 21 and 24.
9. How will alignment with initial targeting and outcome objectives be evaluated over the life cycle of the programme?

A well-defined target population and target outcome specify the programme’s desired outcome, which can be monitored. Simulations provide an approximate quantitative measure of expected impacts on the target population in terms of specific measures. The measurement of actual programme outcomes during deployment can either confirm these expectations or show deviations that expose flawed assumptions. In particular, an investigation of reasons for downside deviations can enable remedial action to realise original programme intentions. A clear evaluation strategy is especially important in an adaptive social protection setting, where beneficiaries’ needs have to be met quickly, and where regular programme evaluation cycles may not yield sufficiently timely feedback. In addition, the experimental nature of the innovative designs discussed here calls for closer monitoring than would be needed for established methods. An assessment strategy that provides timely feedback, such as a phone-based post-distribution survey, can provide actionable insights. In addition, the systematic monitoring of operational metrics in comparison with a roll-out plan can reveal issues and support improvements. The monitoring strategy should be considered a key operational aspect that is included in programme design, resourced adequately and rolled out early.
How will access barriers mediate access to digital payment channels?

As with registration in a digital social protection programme, a number of different types of access barriers can mediate an individual or household’s ability to efficiently receive programme payments, particularly when those payments are provided digitally (via mobile money, e-wallet, or bank transfers). Disability, age, gender, and literacy are intersectional issues that can limit access to digital payment channels. The following discussion of literacy subsumes the variety of access barriers that beneficiaries may experience.

First, payment notifications for these digital services are typically provided in text (via SMS, email, an application, or another messaging service). Text-based notifications are challenging for those with low language literacy; moreover, many services will only provide text-based notifications in certain languages, effectively excluding those who do not speak one of the available languages (a particular concern in contexts with a diversity of languages and in countries where certain languages are unwritten). This concern applies less to the provision of prepaid cards or in-hand cash, the provision of which typically involves person-to-person interaction. For mobile money, digital wallet, and bank transfers, the language literacy concern can be ameliorated by providing payment notifications and instructions for accessing funds in every language spoken by beneficiaries, providing voice-based instructions where possible, and ensuring sufficient support for beneficiaries who struggle with receiving payments (see questions 1 and 2).

Beyond language literacy, digital literacy may also mediate access to payment channels for intended beneficiaries. Digital literacy poses a challenge for mobile money, digital wallet, and bank transfers, which all require competency with phones or other digital devices. Prepaid cards pose a different set of digital literacy challenges, including beneficiaries’ comfort with ATM machines and PIN numbers (Ornellas & Zastrau, 2020). Moreover, even those with basic digital skills may lack confidence in those skills, turning instead to in-person support networks for help with accessing benefits (Matthews et al., 2020). As with language literacy, digital literacy is naturally less of a barrier in programmes that deliver cash face-to-face. For digitally delivered payments, it is critical to consider how support networks can be created – and how existing support networks might be leveraged – to ensure that even beneficiaries with low levels of language and digital literacy can access benefits (see questions 1 and 2 above). In the context of digital payments – and particularly when payments are provided by a private sector company – it is important to ensure that beneficiaries are protected from financial products sold in an unregulated manner.4

4 For example, when South Africa’s Social Security Administration worked with private company Net1 to provide bank transfers to social grant beneficiaries, Net1 targeted beneficiaries for loans and deductions from cash transfers without their consent (Torkelson, 2020).
11. What support networks will be available for applicants who struggle with digital payment channels?

Support networks implemented to ensure that those struggling with digital registration channels are able to register for social protection programmes can also serve as support networks for accessing benefits. Options for creating such support networks – and leveraging existing ones – are covered in question 4 above.

12. To what degree will programme applicants be informed of why they are or are not targeted for benefits?

There is no consensus on whether or not social programme applicants should receive an exact explanation as to why they are or are not eligible to receive benefits, particularly in programmes that use digital data and/or predictive algorithms as part of the targeting system (Banerjee et al., 2020; Budlender, 2014; Grosh et al., 2022; Kidd & Wylde, 2011). Here we cover the pros and cons of providing such explanations.

Providing explanations of targeting decisions helps endow programme applicants with agency to appeal decisions that they believe are incorrect. Providing explanations can also curb confusion and frustration among programme applicants, as decisions provided without explanation are more likely to be viewed as random or illegitimate. For example, in South Africa’s Social Relief of Distress Grant, implemented from 2020 to the present, the combination of opaque targeting criteria and cryptic explanations for decisions has led to the belief among applicants that selection is arbitrary, as well as accusations of corruption in programme administration (Senona et al., 2021). The targeting criteria for the grant are complex (involving a number of administrative and financial data sources), and eligibility for the programme was re-assessed on a monthly basis, so programme applicants were frequently confused as to why their eligibility status changed frequently. Moreover, while explanations for eligibility decisions were available via some registration and beneficiary outreach channels, they were frequently difficult to access digitally, especially for those with access barriers or low digital literacy.
On the other hand, in some contexts, providing exact explanations for eligibility decisions may incentivise manipulation of the decision rule. Particularly in the context of predictive algorithms like PMTs or machine learning algorithms using digital data sources, providing the exact formula used to determine eligibility provides an opportunity for applicants to manipulate their observable characteristics to meet eligibility standards (Banerjee et al., 2020). In the context of black-box predictive algorithms, which do not provide easy-to-interpret decisions, providing exact explanations for decisions may also be logistically impossible, although innovations from the explainability in machine learning literature allow for approximate explanations (Ribeiro, Singh & Guestrin, 2016).

A middle-ground option is to provide explanations on which data are used in determining eligibility, but not the exact weight placed on each data source. Such an approach would allow potential beneficiaries to understand the overall thrust of a targeting strategy, although they might not understand exactly why they or others were or were not selected for a given social protection programme. This approach is what is typically implemented in programmes that use PMTs for targeting: potential beneficiaries are aware of what information is used to determine eligibility – because the questions are asked of them directly – but do not know the exact PMT formula (e.g. Banerjee et al., 2020; Castañeda & Fernandez 2005). This middle-ground approach respects the autonomy of potential beneficiaries and allows for the contestation of programme decisions without incentivising manipulation by revealing the exact formula used to determine eligibility (also see question 15).
13. **How will programme applicants be informed of what data is used for targeting?**

Regardless of whether or not a programme releases information on how individual decisions are made, it is essential that the underlying data sources that are used in the decision process are revealed to applicants. Transparency on data usage is consistent with Europe’s General Data Protection Regulation (GDPR) and other similar data protection frameworks. Moreover, transparency on data usage facilitates agency among programme applicants and allows civil society to hold governments accountable for lawful data use and data protection.

Information on data use should be made available to programme applicants in an accessible format. In contexts with low digital penetration (see questions 1, 2) information on data use should be made available in non-digital formats, such as through community leaders or radio announcements.

14. **Will informed consent be collected for data used for targeting?**

Where possible, consent for data use should be collected from programme applicants prior to data analysis, using either an opt-in or opt-out framework. In an opt-in framework, households or individuals are asked for informed consent prior to accessing the data used for programme targeting; such data is only accessed for those who consent. In an opt-out framework, households or individuals are informed that data will be accessed and given a period to inform programme implementers if they would like to opt out of data access. However, meaningful consent is challenging to collect in the context of social protection programmes, particularly due to power imbalances between programme implementers and potential beneficiaries. One way to make consent more meaningful is to ensure that there are other, non-data-intensive paths to benefits for applicants who do not consent to data usage.
15. Which programme staff will have access to information on why applicants are, or are not, targeted for benefits?

When individual applicants do not receive explanations for targeting decisions, providing such explanations to direct services staff affiliated with the programme can help address issues of frustration among programme applicants and allegations of random targeting (or worse, corruption). In some cases with simple eligibility criteria, making the underlying data used for targeting available to direct services staff could allow staff to clarify targeting decisions for applicants. In other cases, when a black-box predictive algorithm is used to determine eligibility decisions, a level of intermediate explanation provided to caseworkers could facilitate explainability without fully revealing the decision rule. The social enterprise Prosperia is currently piloting such intermediate explanations in a PMT targeting approach in partnership with the government of Costa Rica (Carillo et al., 2021).

It is critical that when targeting data is available to caseworkers, it is available in a read-only format, so that direct services staff cannot directly edit underlying data (which could facilitate collusion with programme applicants). By making the data available in a read-only format, caseworkers can ‘sanity-check’ decisions for applicants they are familiar with, and help applicants appeal targeting decisions through official channels.

16. What grievance redressal channels will be available to applicants?

Grievance redressal mechanisms are essential to social protection programmes, and digital programmes are no different. Applicants should be given opportunities to appeal targeting decisions that they believe are incorrect, lodge complaints about issues with the payment of benefits, and contact programme administrators in regard to other issues with the programme. Grievance redressal channels should be consistent with the programme’s operating environment. Most importantly, in areas with low digital penetration, non-digital channels for grievance redressal should be made available (see questions 1 and 2 for literacy and access issues).

While grievance redressal channels are a standard part of most social protection programmes, many digital social protection programmes launched during the COVID-19 pandemic did not include such options (Aiken & Ohlenburg, 2023). It is common for programmes that use digital data for targeting (as well as programmes that use traditional survey-style data) to neglect the provision of targeting data and explanations to applicants (see questions 13 and 15), thereby eliminating the agency of programme applicants to appeal decisions. It is essential that post-COVID digital social protection systems reintroduce explanations and the option to appeal targeting decisions.
17. How will appeals be adjudicated?

Targeting decisions that are appealed should always be assessed for validity with respect to the targeting criteria, and may be assessed for validity with respect to the original targeting objective (if the targeting criteria is only a proxy for the original objective). Firstly, the decision should be audited for correctness with respect to the targeting criteria. Was all underlying data used for targeting correct, or could an issue of data entry have generated a false exclusion? Checks on underlying data are particularly crucial for programmes using simple decision rules based on digitised administrative data or a ‘targeting out’ approach.

For more complicated decision criteria – for example, when a predictive algorithm is being deployed (particularly one that relies on digital data sources, rather than data gathered in an in-person home visit), decisions that are appealed might also be checked for ‘external validity’ using an alternative targeting method (Kidd & Wylde, 2011). For example, if a decision based entirely on digital data (such as satellite imagery or mobile phone data) is appealed, an in-person visit to verify household poverty using a PMT-like method could provide an external validation check. Where resources are available for such secondary evaluations, errors of exclusion based on digital data sources could be substantially reduced.
18. Can you recruit a high-level sponsor to champion the programme and data use?

The digital shock-responsive programmes discussed here differ in their implementation from standard programmes, as novel data sources and digital channels require unusual activities such as untested procurement processes, cooperation with private sector organisations, or new engagement channels. Legal, organisational, financial and regulatory hurdles may arise that existing decision and control processes are not prepared for. As a timely response is essential to mitigate the shock, programme designers and administrators need to overcome hurdles quickly for successful implementation.

A senior decision maker who takes on formal or informal responsibility for the digital aspects of a social protection programme can perform key functions in hastening implementation, particularly resource mobilisation, convening power, tie breaking in case of intractable disputes, and the direction of required actions. A senior rank in government, such as a cabinet post, or leadership of a high-profile institution can provide the requisite formal and informal authority. High profile programmes that were implemented quickly under challenging circumstances, such as Togo’s Novissi and South Africa’s Social Relief of Distress Grant, benefited from such ‘sponsorship’ (in Togo’s case, from the presidency and the Minister of Digital Economy; in South Africa’s case, from the presidency and the Minister of Social Development). Although civil servants would not usually have the authority to appoint a sponsor, they may catalyse a process that leads to the appointment of one by compiling a dossier that outlines key issues that may require high-level inputs and proposing an institution that could solve these. If this analysis suggests that the programme could likely be managed more productively by a different institution then its re-assignment should be considered, taking data protection and governance issues into account.
19. Are mechanisms in place to coordinate the main stakeholders?

As a new digital shock-responsive programme or a new aspect of an existing one is created, unforeseen challenges are likely to arise. Apart from unexpected issues, a complicating factor is a potentially wider range of partners than for traditional programmes that are managed in-house. The institution in charge of the programme needs to work hand-in-hand with the various stakeholders to solve problems collaboratively as they arise. To be able to move quickly, smooth communication and coordination modalities should be established early on. With such coordination mechanisms in place, a high-level sponsor who can adjudicate disputes should then only be required in exceptional circumstances.

Key players likely include the ministry or agency in charge, data providers, digital channel partners and other commercial partners, the finance ministry, and possibly civil society organisations and others that support beneficiaries. Not all stakeholders need to interact with all others; coordination can be organised in thematic or sector groups, such as administrative/legal/financial, data and IT, outreach/communication/payment channels, and beneficiary management. Where private sector partners are concerned, the involvement of a relevant industry association may be helpful.

20. Does the envisioned programme and data use have a legal basis?

A legal basis for operation is essential for two reasons. One is that public bodies require one to obtain and disburse budget allocations, whether directly for benefits or for programme overheads. The other reason is that the collection and processing of personal information – a practical requirement for digital social protection programmes – requires a legal basis in many settings that have implemented GDPR-style data protection laws. Funds may come from external sources, without a basis in national law, but in such a setting the consent of applicants and beneficiaries for data processing may need to be obtained. For public programmes, data processing is generally provided for by a public interest legal basis, but note that this basis may only be valid if there is no alternative targeting approach that yields better results. Programme designers are advised to consider and simulate alternatives to ensure that the legal basis is given. In some cases, it is possible to amend and adjust existing programme structures for crisis responses, as was done in the case of South Africa’s Social Relief of Distress Grant, which was governed under South Africa’s Disaster Management Act. Such a setting
obviates the need for new legislation and can speed up the intervention. If new laws are required, then it may be advisable to develop draft texts with law officers ahead of time. Such preparation should provide for third-party service and data provision as required, resulting in legal provisions that may differ meaningfully from those of existing programmes. A sunset clause may be suitable for situation-specific legal provisions.

21. Are the necessary skills for programme development and deployment available?

Unlike programmes with a significant physical footprint, purely digital programmes can be executed with minimal staffing. On the other hand, a concentration of responsibilities puts a premium on the skills of the few specialists who are involved. IT skills, both in terms of software development and management, take centre stage with digital programmes. The maintenance of the programme will require front and back-end developers to maintain the data and interfaces with applicants/beneficiaries, data providers, and communication and payment channels. Those with data-driven targeting mechanisms additionally require expertise in computational statistics. For example, the processing of satellite imagery, mobile phone data, and financial services data will typically require expertise in both data science and development economics. In some legal settings, a data protection officer may be stipulated by law. Where third-party inputs are procured via tenders or direct purchases, particular expertise for the structuring and evaluation of bids, as well as the management of the suppliers, will be needed.

Some skills may not require full-time staff throughout the programme lifespan. For instance, statistical modelling of targeting methods may be labour-intensive at the design stage, but then only require monitoring for continued fit and occasional updates. Such temporary inputs can be provided on a consultancy basis, by temporary staff, or via outsourcing. For other staff, it depends on whether or not the project structure is meant to be maintained beyond the shock response it was created for. The establishment of the requisite internal capacity may take some time, and its central importance means that it should be one of the first issues tackled in programme planning.

In addition to technical skills for data-intensive tasks, core social protection expertise will remain an important skill, even in social protection programmes that are entirely or partially digital, particularly in conceiving and designing programme objectives. Developing digital social protection solutions will require close collaboration between social protection experts, IT staff, data analysts, and those with expertise in data protection.
22. Are the necessary skills for programme development and deployment available?

Although it may appear that digitally implemented programmes are economical to administer, an analysis of three pandemic-type programmes (Okamura, Tesliuc & Ohlenburg, forthcoming) suggests that overall costs were similar to those of regular programmes. Although a physical footprint with attendant costs may not be necessary, the fees payable for the use of digital channels may more than outweigh such savings.

During the design and implementation of a crisis-time programme, much progress is needed quickly. Sufficient resourcing is crucial for access to the requisite skill set and infrastructure set-up. The programme is then likely to require significant resources for infrastructure maintenance and payment to communication and payment partners. Beneficiary management is the third aspect, and one that is commonly under-resourced until significant issues are publicised. Appropriate, early provision for on-going costs, including beneficiary management, should be planned in order to foster smooth functioning and a high level of service quality.

23. Where necessary, can a procurement process ensure the availability and acquisition of third party inputs?

Data and IT services are examples of essential inputs to digital shock-responsive programmes that may require acquisition through a procurement process. Market supply and industry structure are key factors that determine whether a simple process is likely to lead to successful acquisition. A competitive market with interchangeable suppliers, similar to what may be available for IT services such as cloud computing or database administration, is likely to allow straightforward purchasing. Where a competitive market is not available, a more complicated process with longer lead in time and higher costs may be necessary. In such cases, there may also be a significant risk of vendor lock-in, where the data supplier becomes irreplaceable and, thereby, acquires the ability to extract excess profits or other benefits. Careful contracting and potentially resorting to regulation and legislation may be necessary to manage this risk.
Oligopolistic markets, in which few competitors control most of the market, pose particular challenges. Unless a trusting relationship exists with the government, firms may want to avoid cooperation to protect commercially sensitive data. This is likely to be the case for mobile network operators and perhaps also for financial services. Both data types are also highly sensitive, as they reveal personal behaviour and circumstances, hence, data protection issues also loom large. A complicating factor is that no single representative source exists for such data, as multiple financial and phone companies split the market among themselves. Access may, thus, need to be secured from several players, who may distrust each other and not be willing to contribute to a sector-wide initiative. An early exploration of these issues is prudent, including consultations with lawyers, procurement specialists, and other public bodies that may influence the behaviour of commercial firms.

24. Where will the data be stored and processed?

The computer systems on which digital programmes run are equivalent to the physical locations through which traditional social protection programmes operate. The fundamental importance of the IT system for the operation of a digitally-based programme elevates it from a technical detail to a design aspect worthy of early consideration. The main deployment options are a server or cloud set-up, and either in-house operation or outsourcing to a suitable IT contractor.

Especially in an adaptive social protection setting and given its ready availability, the scalability and fast deployment of cloud resources may be an advantage over a server-based solution that might require procurement. Similarly, the outsourcing of a system’s operation may allow faster and more flexible deployment than internal management, which implies hiring new staff. Programme designers can prepare by exploring the availability of cloud providers and outsourced service providers, and by determining relevant IT and procurement policies that may constrain options. In addition, data sovereignty may be a formal or informal constraint on the public use of foreign-based computational resources.
25. Which personally identifiable information is in the data?

Social protection programmes usually rely on the collection of sensitive personal information. Family structure, economic conditions, health status and contact details are examples of the sensitive information that many social protection programmes require for targeting and administration. The more sensitive the information, the more care needs to be taken in its collection, handling and storage. The first step towards the responsible handling of such data is to assess the extent of the issue, by documenting items of sensitive personal information. At a minimum, the documentation should include a listing of sensitive items, the rationale for processing and storing them, details of the data’s intended use, and the major risks associated with the disclosure or manipulation of each data item. The GDPR provides a useful analytical framework, with its data protection impact assessments.

26. How will sensitive information be protected?

Data security, which protects data from unauthorised third-party access, is a fundamental responsibility of administrators of social protection systems. The information contained in programme or social registries can be misused in a variety of ways, including identity theft, misappropriation of benefits, or the creation of ghost recipients. Approaches that rely on novel data or that combine administrative data sources may pose a heightened risk compared to regular programmes. That is especially true if they contain commercially sensitive items such as financial data, including mobile money transactions, or administrative variables that could facilitate identity theft.

Detailed prescriptions of data security measures cannot be provided here, as the appropriate systems and actions depend on various factors, including the sensitivity of the data, the local data security landscape, and the state of the art in cybersecurity; hence, specialist, context-specific advice needs to be sought. This should include organisational aspects, such as access rights and data sharing protocols, which are as important as software and hardware security measures. In cases where third parties have access, as is often the case for IT contractors, data security measures need to be enforced at the same level. If third-party data is used, such as mobile phone or financial services data, the safest approach may be to process it remotely and limit the transfer of information to the minimally required aspects. Special consideration must be given to known weaknesses in data security measures, to avoid the repetition of prior instances of data misuse. For more details on data protection in digital and non-digital social protection programmes, refer to Wagner and Ferro (2020).
27. Which groups will have read and write privileges for the data?

This guidance document considers aspects of both the planning of social protection programmes and their implementation. During the planning stage, for instance, in the simulation of programme impacts, personally identifiable information may not be required. Once relevant datasets with such information have been merged, personal identifiers should not be made visible to stakeholders conducting simulations. In contrast, the implementation of social protection programmes does require personally identifiable information, with sensitive variables typically at centre stage. Different items are relevant in different parts of the delivery chain (Lindert et al., 2020). For example, contact information and household composition may be important for outreach, whereas variables related to economic wellbeing may be used for eligibility determination. Different actors perform design and implementation tasks, and each should only get access to the information required for their respective responsibilities.

Misuse of sensitive information is one side of data risk, data manipulation is the other. Examples include changes that falsely alter eligibility or increase benefit amounts unduly. As with read access, write access – the ability to alter data – needs to be restricted according to functional necessity. Only actors who have a role in collecting or correcting data entries should be able to write to the database, and only for those variables that they are responsible for. This group is likely to be smaller than that requiring read access, and any changes should be logged, creating the possibility to generate an audit trail where necessary. Read and write access should be managed as part of the overall data policy, which also includes a risk assessment and retention policy.

28. How long will the data be stored for?

The storage of sensitive personal data poses a risk to the individuals who are the subject of the database. Although some of the information may lose relevance as it becomes outdated, other items maintain their detrimental potential. Deletion of personal data is an important aspect of risk mitigation, and one that can be managed systematically. A data retention policy that determines the length of time each data item can be stored may complement the listing of sensitive information described under question 25. In addition, protocols for regular review and for deletion are pillars of a data retention policy.
USEFUL (RE)SOURCES AND REFERENCES
Below are some particular sources and resources that readers may wish to consult for more detailed treatment of particular areas. Subsequently, a list of references is provided.

• Aiken, E. and Ohlenburg, T., *Novel digital data sources for social protection: opportunities and challenges*, GIZ, 2023


• Digital Convergence Initiative [https://spdci.org](https://spdci.org)

• Prosperia.ai [https://www.prosperia.ai/](https://www.prosperia.ai/)


ADDITIONAL REFERENCES


• Barca, V., *Integrating data and information management for social protection: Social registries and integrated beneficiary registries*, GIZ, 2017


• Budlender, D., *Conditional cash transfers: Learning from the literature*, UN Women, 2014


• Chi, G., Fang, H., Chatterjee, S. and Blumenstock, J. E., Microestimates of wealth for all low-and middle-income countries, *Proceedings of the National Academy of Sciences*, 119(3), e2113658119, 2022

• Department of Social Development, *The rapid assessment of the implementation and utilisation*
of the special COVID-19 SRD Grant, Department of Social Development Republic of South Africa, 2021

- Gentilini, U., Cash transfers in pandemic times: Evidence, practices, and implications from the largest scale-up in history, World Bank, Washington, DC, 2022
- GiveDirectly, Biometrics in Uganda, Blog post, GiveDirectly, 6 January 2016
- GiveDirectly, Canva partnership tackling extreme poverty in Malawi one year on, [Online], GiveDirectly, 9 November 2022
- GSMA, The state of mobile Internet connectivity 2021, Global System for Mobile Communications Association (GSMA), 2021
- Hortinela, C., Make the most impact with interactive voice response (IVR) in developing countries, [Online], EngageSpark, 2 October 2019
- Huang, L. Y., Hsiang, S. M. and Gonzalez-Navarro, M., Using satellite imagery and deep learning to evaluate the impact of anti-poverty programs (No. w29105), National Bureau of Economic Research, 2021
- LeFevre, A. E., Dane, P., Copley, C. J., Pienaar, C., Parsons, A. N., Engelhard, M., ... and Mohan, D., Unpacking the performance of a mobile health information messaging program for mothers (MomConnect) in South Africa: Evidence on program reach and messaging exposure, BMJ Global Health, 3, 2018
- Livani, T. and Haddock, S., Women’s access to identification cards can accelerate development in
Afghanistan, [Online], World Bank End Poverty in South Asia Blog, 2020

• López Solano, J., Experimenting with poverty: The SISBEN and data analytics projects in Colombia, Fondacion Karisma, 2020 https://www.researchgate.net/publication/351699844_Experimenting_with_poverty_The_SISBEN_and_data_analytics_projects_in_Colombia


• Matthews, T., Gronewald, C. and Moolman, B., It's a lifeline, but it's not enough: The COVID-19 social relief of distress grant, basic income support, and social protection in South Africa, Black Sash, 2020


• Ornellas, A. and Zastrau, E., Researching and engaging the decommissioning of SASSA paypoints, Black Sash, 2020


• Rolf, E., Proctor, J., Carleton, T., Bolliger, I., Shankar, V., Ishihara, M., ... and Hsiang, S., A generalizable and accessible approach to machine learning with global satellite imagery, Nature Communications, 12(1), 4392, 2021

• Senona, E., Torkelson, E. and Zembe-Mbakile, W., Social protection in a time of COVID: Lessons for basic income support, Black Sash, 2021

• Sharma, A., 99.9% of adults in India have Aadhar number and ‘use it at least one a month,’ says UIDAI, [Online], First Post, 23 July 2022

• Sharpe, Joanne, and Barca, Valentina, Dynamic social protection: The role of on demand updating, Social Protection Approaches to COVID-19 Expert Advice Service (SPACE), London, 2021
• Sheehan, E., Meng, C., Tan, M., Uzkent, B., Jean, N., Burke, M., ... and Ermon, S., Predicting economic development using geolocated Wikipedia articles, in: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2698–2706, July 2019

• Smythe, I. S. and Blumenstock, J. E., Geographic microtargeting of social assistance with high-resolution poverty maps, *Proceedings of the National Academy of Sciences*, 119(32), e2120025119, 2022


• World Bank, *Global ID coverage, barriers, and use by the numbers: An in-depth look at the 2017 ID4D-Findex Survey*, World Bank, 2019

• Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., ... and Burke, M., Using publicly available satellite imagery and deep learning to understand economic well-being in Africa, *Nature Communications*, 11(1), 2583, 2020
Published by:
Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

Registered offices:
Bonn and Eschborn

Address:
Friedrich-Ebert-Allee 32 + 36               Dag-Hammerskjöld-Weg 1-5
53113 Bonn / Germany                       65760 Eschborn / Germany
T +49 228 44 60 - 0                         T +49 61 96 79 - 0
F +49 228 44 60 - 17 66                     F +49 61 96 79 - 11 15

E info@giz.de
I www.giz.de

Programme/project description:
GIZ Guidance Tool for the Use of Alternative Data for Shock-Responsive Social Protection Systems

Author:
© 2023 Emily Aiken and Tim Ohlenburg

The views expressed in this publication are those of the author and do not necessarily reflect the views and policies of the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH.

Coordination:
Susanne Lein, Dominique Leska-See

Editing:
Susan Sellars

Layout / Illustrations:
EYES-OPEN K15 GmbH, Berlin

URL links:
Responsibility for the content of external websites linked in this publication always lies with their respective publishers. GIZ expressly dissociates itself from such content.

Photo credits:
p.15: SeventyFour @Shutterstock; p.55: Yuri A @Shutterstock

Responsible for the content:
Ralf Radermacher (GIZ)
Bonn, June 2023